Cardiopulmonary exercise testing for prognosis in chronic heart failure: continuous and independent prognostic value from VE/VCO₂ slope and peak VO₂

D. P. Francis¹, W. Shamim¹, L. Ceri Davies¹, M. F. Piepoli¹, P. Ponikowski¹,2, S. D. Anker¹,3 and A. J. S. Coats¹

¹National Heart & Lung Institute, Imperial College of Science, Technology and Medicine, London, U.K.; ²Department of Cardiology, Military Hospital, Wroclaw, Poland; ³Franz-Volhard-Klinik at Max Delbrück Centrum, Charité, Campus Berlin-Buch, Berlin, Germany

Background Chronic heart failure carries a poor prognosis. Cardiopulmonary exercise testing is useful in predicting survival. We set out to establish the prognostic value of peak VO₂ and VE/VCO₂ slope across a range of threshold values.

Method and Results Three hundred and three consecutive patients with stable chronic heart failure underwent cardiopulmonary exercise testing between 1992 and 1996. Their age was 59 ± 11 years (mean ± SD), peak VO₂ 17.8 ± 6.6 ml kg⁻¹ min⁻¹, VE/VCO₂ slope 37 ± 12. At the end of follow-up in January 1999, 91 patients had died (after a median of 7 months, interquartile range 3–16 months). The median follow-up for the survivors was 47 months (interquartile range 37–57 months). The areas under the receiver-operating characteristic curves for predicting mortality at 2 years were 0.77 for both peak VO₂ and VE/VCO₂ slope. With peak VO₂ and VE/VCO₂ slope viewed as continuous variables in the Cox proportional-hazards model, they were both highly significant prognostic indicators, both in univariate analysis and bivariate analysis (\(P<0.001\) for VE/VCO₂ slope, \(P<0.003\) for peak VO₂).

Conclusions Lower peak VO₂ implies poorer prognosis across a range of values from 10 to 20 ml kg⁻¹ min⁻¹, without a unique threshold. Gradations of elevation of the VE/VCO₂ slope also carry prognostic information over a wide range (30–55). The two parameters are comparable in terms of prognostic power, and contribute complementary prognostic information.

(Eur Heart J 2000; 21: 154–161) © 2000 The European Society of Cardiology

Key Words: Cardiopulmonary exercise testing, prognosis, chronic heart failure.

Introduction Chronic heart failure is characterized by poor prognosis and an increasing prevalence in the developed world. Allocation of scarce resources, both for medical therapy and for surgical intervention, is based upon clinical judgement of prognosis. Accordingly, there is an ongoing quest to increase our understanding of the predictive value of clinical data on which such judgements are based.

Cardiopulmonary exercise testing with metabolic monitoring is established as a mainstay of the objective assessment of functional ability in chronic heart failure. The key measurement is peak oxygen uptake (peak VO₂), whose usefulness as a prognostic marker is widely accepted[1]. Oxygen uptake rises during incremental exercise, and peak VO₂ represents the highest rate of oxygen uptake achieved. A reduction in its level may result from a variety of factors[2], including limitation in cardiac output, poor peripheral blood flow[3,4], impaired skeletal muscle metabolism[5], or early termination of the test because of cardiac-related or other symptoms. The remarkable prognostic power[6] of peak VO₂ may result from its dependence on a clinically relevant combination of these numerous mechanisms of impairment of functional capacity.

The prognostic power of peak VO₂ has previously been addressed using thresholds, a variety of competing values of which now appear in the literature[6,7-10]. However, since peak VO₂ is a continuous variable, it
might carry prognostic information over a range of values. We used two major lines of approach to study this. Firstly, we applied a variety of cut-off values and assessed the prognostic specificity and sensitivity of each, forming a receiver-operating characteristic curve. Secondly, we analysed its prognostic value as a continuous variable using the Cox proportional-hazards model.

A separate abnormality is also observed during the cardiopulmonary exercise test in chronic heart failure patients. The rate of increase in ventilation per unit increase in carbon dioxide production (VE/VCO₂ slope) is greater in patients with chronic heart failure than in normals[11]. Patients with slopes above the upper limits of normal have a poorer exercise tolerance and a significantly worse prognosis[12]. However, the ability of this parameter to predict outcome over a range of cut-offs has not been established. We therefore set out to determine the sensitivity and specificity of different VE/VCO₂ cut-offs for predicting mortality, and to plot the receiver-operating characteristic curve. The final question is whether these two variables yield identical pieces of prognostic information, or whether they are to some extent independent. If the latter is the case, it may be important to consider both figures when interpreting the result of a cardiopulmonary exercise test.

Methods

Subjects

Three hundred and three outpatients with chronic heart failure at the Royal Brompton Hospital underwent cardiopulmonary exercise testing between January 1992 and December 1996. The diagnosis of chronic heart failure was based on a history of dyspnoea and symptomatic exercise intolerance with signs of pulmonary congestion or peripheral oedema, and echocardiographic or radionucleide ventriculographic evidence of impaired ventricular function. Patients with chronic lung disease, primary valvular disease, neuromuscular disease, or myocardial infarction within the previous 3 months were not included. Survival was determined from outpatient records, the registry maintained by the Office of National Statistics (ONS) and, where necessary, telephone calls to the general practitioners or to the patients’ residences. Follow up was complete for at least 2 years in all patients.

Cardiopulmonary exercise testing

Exercise testing was performed on a treadmill using a modified Bruce protocol[13], in which there was a stage zero of 1 mph at 5% gradient. Ventilation, oxygen uptake, and carbon dioxide production were monitored continuously using a respiratory mass spectrometer (Amis 2000, Innovision, Odense, Denmark). Patients exercised to the limit of their symptoms. The VE/VCO₂ slope, which relates the rate of increase in ventilation per unit increase in carbon dioxide production, was obtained by linear regression analysis[11].

Statistical analysis

Statistical calculations were performed using the Statview 4.5 package (Abacus Concepts, Berkeley, CA, U.S.A.). Numerical values are presented as mean ± standard deviation. Patients were categorized into quartiles by peak VO₂ and by VE/VCO₂ slope. Survival curves for patient groups were calculated using the Kaplan–Meier method. The prognostic value of exercise parameters considered as continuous variables were determined using the Cox proportional-hazards linear regression model. Finally, a non-parametric analysis was performed: for each potential combination of peak VO₂ and VE/VCO₂ slope, the mortality of the 30 patients whose results most closely resembled that combination was determined and plotted.

Results

The overall patient group had a mean peak VO₂ of 17.8 (± 6.6) ml kg⁻¹ min⁻¹, and a mean VE/VCO₂ slope of 37 (± 12). The mean age was 59 ± 11 years. Two hundred and sixty-seven patients were men, 26 were women. Fourteen percent of the patients were in NYHA class I, 36% in class II, 39% in class III and 11% in IV. Aetiology of chronic heart failure was considered to be ischaemic in 178 patients. Radionucleide ventriculography (n=204) showed an average left ventricular ejection fraction of 25 ± 11%. Echocardiographic quantification of left ventricular dimensions (n=173) showed a mean end-diastolic dimension of 7.1 ± 1.0 cm, and a mean end-systolic dimension of 6.0 ± 1.1 cm. At the end of the follow-up period in January 1999, 91 patients had died (median time to death 7 months, interquartile range 3–16 months). Seventy-three of these deaths occurred within the first 24 months of testing. The median follow-up duration for the survivors was 47 months (interquartile range 37–57 months).

A notional threshold in peak VO₂ of 14 ml kg⁻¹ min⁻¹ yielded a 57% sensitivity and a 80% specificity for prediction of mortality at 24 months. Application of lower thresholds yielded greater specificity at the expense of lower sensitivity, and higher thresholds gave higher sensitivity although lower specificity, as shown in Table 1. A notional threshold in VE/VCO₂ slope of 40 gave a sensitivity of 58% and specificity of 78% for predicting mortality at 24 months. Changes in this threshold also resulted in a trade-off between sensitivity and specificity, as shown in Table 2.

Receiver operating characteristic curves

The interplay between sensitivity and specificity with changes in threshold can be represented by plotting one
Table 1 Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of thresholds in peak VO₂ for detecting mortality at 24 months

<table>
<thead>
<tr>
<th>Threshold (ml.·kg⁻¹·min⁻¹)</th>
<th>Peak VO₂</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10</td>
<td>18%</td>
<td>96%</td>
<td>59%</td>
<td>79%</td>
<td></td>
</tr>
<tr>
<td>&lt;12</td>
<td>40%</td>
<td>90%</td>
<td>57%</td>
<td>83%</td>
<td></td>
</tr>
<tr>
<td>&lt;14</td>
<td>57%</td>
<td>80%</td>
<td>48%</td>
<td>86%</td>
<td></td>
</tr>
<tr>
<td>&lt;16</td>
<td>80%</td>
<td>64%</td>
<td>42%</td>
<td>91%</td>
<td></td>
</tr>
<tr>
<td>&lt;18</td>
<td>87%</td>
<td>48%</td>
<td>35%</td>
<td>92%</td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>93%</td>
<td>37%</td>
<td>32%</td>
<td>94%</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of thresholds in VE/VCO₂ slope for detecting mortality at 24 months

<table>
<thead>
<tr>
<th>Threshold (ml.·kg⁻¹·min⁻¹)</th>
<th>VE/VCO₂ slope</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;30</td>
<td>93%</td>
<td>41%</td>
<td>33%</td>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>&gt;35</td>
<td>78%</td>
<td>60%</td>
<td>38%</td>
<td>90%</td>
<td></td>
</tr>
<tr>
<td>&gt;40</td>
<td>58%</td>
<td>78%</td>
<td>46%</td>
<td>85%</td>
<td></td>
</tr>
<tr>
<td>&gt;45</td>
<td>47%</td>
<td>88%</td>
<td>55%</td>
<td>84%</td>
<td></td>
</tr>
<tr>
<td>&gt;50</td>
<td>32%</td>
<td>93%</td>
<td>60%</td>
<td>81%</td>
<td></td>
</tr>
<tr>
<td>&gt;55</td>
<td>20%</td>
<td>96%</td>
<td>65%</td>
<td>79%</td>
<td></td>
</tr>
</tbody>
</table>

Continuous models

When peak VO₂ and VE/VCO₂ slope were viewed as continuous variables in the Cox proportional-hazards prognostic model, they were found to be highly significant prognostic indicators, both in univariate analysis (P<0.001 for each) and bivariate analysis (P<0.001 for VE/VCO₂ slope, P<0.003 for peak VO₂). To avoid the assumption that mortality hazard varies linearly with peak VO₂ and VE/VCO₂ slope, a non-parametric risk surface was plotted (Fig. 5). This confirms the independent prognostic value of the two variables.

Multivariate analysis

Using individual univariate Cox proportional-hazards survival regressions, we identified univariate predictors of mortality. In descending order of statistical significance (as shown in Table 4) they are: VE/VCO₂ slope, peak VO₂, NYHA class (dichotomized into I/II and III/IV), age, left ventricular ejection fraction and ischaemic aetiology. We then performed multivariate analysis using the forward–backward method, in which only VE/VCO₂ slope, age, left ventricular ejection fraction and NYHA class were retained in the model, as shown in Table 4.

Discussion

This study shows that reduced peak VO₂ predicts mortality across a wide range of thresholds and that enhancement of the VE/VCO₂ slope is also prognostically useful across a broad range of values. It further shows that they have equivalent power to predict death at 2 years, and that they provide complementary prognostic information.

In chronic heart failure cardiopulmonary exercise testing is widely accepted as an objective means of defining functional status[14,15]. Low peak VO₂ values are known to identify patients at higher mortality risk. Several threshold values in peak VO₂ have been put forward. Initially a threshold of 14 ml.·kg⁻¹·min⁻¹ was proposed by one study[16] of a cohort of patients whose mean peak VO₂ was 14·7 ml.·kg⁻¹·min⁻¹ and another[17] whose mean peak VO₂ was 13·7 ml.·kg⁻¹·min⁻¹. This figure was incorporated into major treatment guidelines[11] for chronic heart failure.

Studies conducted in patients with milder heart failure have identified correspondingly higher thresholds: one[18] whose mean peak VO₂ was 17·0 ml.·kg⁻¹·min⁻¹ and another[19] whose mean peak VO₂ was 17·6 ml.·kg⁻¹·min⁻¹ pointed to a cutoff at 17 ml.·kg⁻¹·min⁻¹. Conversely, at the severe end of the spectrum, a study[20] whose mean peak VO₂ was 11·0 ml.·kg⁻¹·min⁻¹ identified 10 ml.·kg⁻¹·min⁻¹ as a useful prognostic threshold, as did another analysis[21] of a cohort whose peak VO₂ averaged

against the other, using all possible thresholds, to produce the receiver operating characteristic curve. Receiver-operating characteristic curves for peak VO₂ (thick line) and VE/VCO₂ (thin line) are shown in Fig. 1. The areas under the receiver-operating characteristic curves were 0·77 for peak VO₂ and also 0·77 for VE/VCO₂.

Prognostic implications of categories of peak VO₂ and VE/VCO₂ slope

Patients were subdivided into quartiles according to peak VO₂ (P1:<13·0 ml.·kg⁻¹·min⁻¹, P2:13·0–16·5 ml.·kg⁻¹·min⁻¹, P3:16·6–21·6 ml.·kg⁻¹·min⁻¹, P4: >21·6 ml.·kg⁻¹·min⁻¹). Their respective Kaplan–Meier survival curves were significantly different (Fig. 2, P<0.0001). Mortalities at 24 months were 48%, 32%, 12%, and 4% respectively. Patients were separately subdivided in quartiles according to VE/VCO₂ slope (V1:<27·7, V2:27·7–34·5, V3:34·6–42·1, V4: >42·1), whose Kaplan–Meier Survival curves were again distinct from each other (Fig. 3, P<0.0001). Mortalities at 24 months were 3%, 17%, 26%, and 49% respectively. With the patients separated into groups according to both peak VO₂ and VE/VCO₂ slope, a gradient of increase in 24-month mortality was observed both in the direction of decreasing peak VO₂ and in that of increasing VE/VCO₂ slope (Table 3 and Fig. 4).
Figure 1  Receiver-operating characteristic (ROC) curves representing the ability of peak VO₂ (thick line) and VE/VCO₂ slope (fine line) to predict mortality at 24 months.

Figure 2  Kaplan–Meier survival curves of the patients grouped into quartiles by peak VO₂.
A recent study [13] showed that while patients with peak VO$_2$ below 10 ml . kg$^{-1}$ min$^{-1}$ had a poor prognosis, and those above 18 ml . kg$^{-1}$ min$^{-1}$ had a good prognosis, those who achieved between 10 and 18 ml . kg$^{-1}$ min$^{-1}$ were a uniform group with an intermediate outcome.

Disputes over the choice of threshold are tacitly based on the assumption that there is a unique threshold which identifies patients with poor prognosis with particular clarity. However, since peak VO$_2$ is a continuous variable, it is conceivable that its prognostic significance is similarly continuous, with successively lower values.

Figure 3  Kaplan–Meier survival curves of the patients grouped into quartiles by VE/VCO$_2$.

Figure 4  Survival of patients grouped according to quartiles of peak VO$_2$ and VE/VCO$_2$ slope.
indicating progressively poorer outcomes. In such a situation, it is useful to assess the prognostic value of several potential thresholds and examine the sensitivity and specificity of each (Table 1). The trade-off between sensitivity and specificity can be seen clearly when they are plotted against each other in the receiver-operating characteristic curve (Fig. 1). A high peak VO\textsubscript{2} threshold (18 ml kg\textsuperscript{-1} min\textsuperscript{-1}) correctly identifies the vast majority of patients who subsequently died within 2 years (high sensitivity, 87%), but incorrectly classifies a majority of the survivors (low specificity, 48%). A lower threshold (10 ml kg\textsuperscript{-1} min\textsuperscript{-1}) solves the latter problem by correctly identifying as ‘low risk’ the vast majority of those who survived more than 2 years (high specificity, 96%) but at the cost of failing to identify most of those who subsequently died (low sensitivity, 18%). Intermediate thresholds offer a compromise between these extremes.

Although sensitivity and specificity are important qualities of a clinical test, they do not directly tell the clinician about the likely outcome for a particular patient. 

Table 3 Two-year mortality of patients grouped according to quartiles of peak VO\textsubscript{2} and VE/VCO\textsubscript{2} slope. In each cell the number of patients who died and the total number of patients is given.

<table>
<thead>
<tr>
<th>Peak VO\textsubscript{2} (ml kg\textsuperscript{-1} min\textsuperscript{-1})</th>
<th>13&lt;0</th>
<th>13–16</th>
<th>16–21</th>
<th>&gt;21</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;13·0</td>
<td>0/13</td>
<td>0/24</td>
<td>1/36</td>
<td></td>
</tr>
<tr>
<td>13&lt;0</td>
<td>1/13</td>
<td>0/24</td>
<td>1/36</td>
<td></td>
</tr>
<tr>
<td>13&lt;0</td>
<td>0/24</td>
<td>1/36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13&lt;0</td>
<td>1/36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4 Results of Cox survival regression analyses conducted by univariate and multivariate (forward-backward) methods.

<table>
<thead>
<tr>
<th>Univariate analyses</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>VE/VCO\textsubscript{2}</td>
<td>70·9</td>
</tr>
<tr>
<td>Peak VO\textsubscript{2}</td>
<td>40·9</td>
</tr>
<tr>
<td>NYHA</td>
<td>34·0</td>
</tr>
<tr>
<td>Age</td>
<td>16·8</td>
</tr>
<tr>
<td>LVEF</td>
<td>15·4</td>
</tr>
<tr>
<td>Ischaemic aetiology</td>
<td>8·4</td>
</tr>
</tbody>
</table>

LVEF=left ventricular ejection fraction. NYHA=New York Heart Association function class, dichotomized into I/II vs III/IV.
patient: for answers to this type of question, the positive and negative predictive values can be consulted (Table 1). In our study, amongst the group of patients with a peak VO₂ < 10 ml kg⁻¹ min⁻¹, 59% died within 2 years (positive predictive value 59%), whereas amongst the remainder of the patients only 21% died (negative predictive value 79%). A higher threshold, such as 18 ml kg⁻¹ min⁻¹, yields a stronger negative predictive value (92%) but a weaker positive predictive value (35%).

The widely discordant thresholds previously identified may result from the observed tendency of the threshold from a study to be close to the average peak VO₂ of that study. One explanation is that there may truly be a global threshold in peak VO₂ below which prognosis is poor: local factors, such as systematic bias in the estimation of peak VO₂ arising from equipment differences, may result in this global physiological threshold manifesting at different measured peak VO₂ values in different hospitals. A second, and perhaps more plausible, explanation is a quirk of statistics. A proposed peak VO₂ cut-off that is far from the central value of a particular cohort will yield groups of markedly unequal sizes, which would reduce the statistical power to detect a mortality difference between them. In contrast, approximate bisection of the patient group using a near-median peak VO₂ value is more likely to yield a statistically significant result. It should therefore not be surprising that different workers, whose patient cohorts differed in their severity of heart failure, obtained different thresholds. We believe that this is indirect evidence of a genuine spectrum of risk, increasing with decreasing peak VO₂ (Figs 1 and 2). Arguments over selection of an optimal threshold may be sterile unless this is taken into account. Sensitivity-vs-specificity analyses offer a way of comparing the findings of different institutions by looking beyond the tendency to identify the middle of one's own patient group as the optimal diagnostic threshold.

Aside from reduction in exercise capacity, chronic heart failure is also characterized by an increase in the ventilatory response to exercise (the VE/VCO₂ slope). A variety of pathophysiological abnormalities, known to be present in chronic heart failure, have been linked to this phenomenon, including increased anatomical dead space, ventilation-perfusion mismatch, abnormal pulmonary vascular haemodynamics and disordered ventilatory reflex control. The extent of elevation of the VE/VCO₂ slope is more marked in patient groups with more severe functional limitation, and those above the 95th percentile of the normals are known to have a poorer prognosis.

In our study, the VE/VCO₂ slope manifested considerable prognostic value. Patients with a VE/VCO₂ slope exceeding 55 had a 2-year mortality of 65% (positive predictive value), while those below this threshold had a 2-year mortality of only 21% (negative predictive value 79%). Alternatively, a much higher negative predictive value (96%, indicating 2-year mortality of only 4%) can be obtained by using a lower threshold such as 30, although the positive predictive value is correspondingly weaker (33%). The Kaplan–Meier survival curves for patients categorized by VE/VCO₂ slope (Fig. 3) show divergence between the quartile groups no less clearly than do the curves for the patients categorized by peak VO₂ quartile (Fig. 2). The receiver-operating characteristic curve for VE/VCO₂ slope to predict 2-year survival is similar in shape and area to that for peak VO₂, implying equivalent prognostic power. The final result of our study was that peak VO₂ and VE/VCO₂ slope can be combined to provide a risk surface (Fig. 5) for evaluating prognosis in chronic heart failure.

There are several possible reasons why progressively higher VE/VCO₂ slopes might predict higher mortality across a range of peak VO₂ values. Firstly, steepening of the VE/VCO₂ slope is associated with reduced peak cardiac output and elevated pulmonary vascular resistance. Secondly, it may indicate enhanced ventilatory reflex sensitivity, which may be prognostically significant in a manner analogous to impairment of baroreflex sensitivity and depression of heart rate variability. A third consideration is that the VE/VCO₂ slope is computed using far more data than is peak VO₂: it is accordingly less susceptible to vagaries of chronic heart failure such as irregular breathing and early subjective fatigue which may sometimes interfere with determination of peak VO₂.

Regardless of mechanism, the prognostic value of quantifying the enhancement of the VE/VCO₂ slope is evident in this study. Patients need suffer no invasive instrumentation, nor does the laboratory need additional equipment. Its calculation is straightforward, and the result simple to report alongside the peak VO₂. Because VE/VCO₂ slope enhancement has similar prognostic power to peak VO₂, and they convey independent predictive information, we propose that it should be a routine component of the analysis of the cardiopulmonary exercise test in chronic heart failure.

When multiple variables are considered, NYHA class, age and ejection fraction may convey sufficient information to eclipse the statistical significance of peak VO₂. In contrast, the VE/VCO₂ slope remains a powerfully significant prognosticator even in the presence of these other variables, reflecting its probing of aspects of physiology other than symptoms and ventricular function.

**Limitations**

This study considered the actual peak oxygen uptake and VE/VCO₂ achieved by each patient. No patient was excluded for reasons of fatigue prior to the demonstration of anaerobic metabolism. We believe this reflects actual clinical practice, since there is a need to assess all patients with chronic heart failure.

**Conclusions**

Reduced peak VO₂ implies poor prognosis. Assessment of sensitivity and specificity, and of positive and negative
predictive values, reveals that the prognostic value of peak VO$_2$ extends over a broad spectrum of thresholds from 10 to 20 ml . kg$^{-1}$ min$^{-1}$. Steepness of the VE/ VCO$_2$ slope, over a range from 30 to 55, is as good at predicting outcome as peak VO$_2$. Finally, peak VO$_2$ and VE/VCO$_2$ slope make complementary contributions to the prognostic evaluation of patients with chronic heart failure. The information obtained from VE/VCO$_2$ slope is far more independent of symptoms, age and ventricular function than is the information obtained from peak VO$_2$: it should always be considered in multivariate models.

Dr Francis is supported by the Estate of the late Mabel Grace Parker (CRC 9716) and by the British Heart Foundation (FS 98005). Dr Davies is supported by the Robert Luft Fellowship. Dr Anker is supported by a postgraduate research fellowship of the Max Delbrück Centrum, Charité, Berlin-Buch, Germany. Prof Coats is supported by the Viscount Royston Trust.

References