HEART
and Education in Heart

DEVELOPMENT OF A COMPOSITE MODEL DERIVED FROM CARDIOPULMONARY EXERCISE TESTS TO PREDICT MORTALITY RISK IN PATIENTS WITH MILD-TO-MODERATE HEART FAILURE

Journal:	*Heart*
Manuscript ID:	heartjnl-2013-304614.R2
Article Type:	Original research article
Date Submitted by the Author:	22-Jan-2014
Complete List of Authors:	Ingle, Lee; University of Hull, Sport, Health & Exercise Science Rigby, Alan; Hull, Sloan, Rebecca Carroll, Sean Goode, Kevin Cleland, John Clark, Andrew
Keywords:	HEART FAILURE, EXERCISE ECG < ELECTROCARDIOGRAPHY < IMAGING AND DIAGNOSTICS, RISK STRATIFICATION < CORONARY ARTERY DISEASE

DEVELOPMENT OF A COMPOSITE MODEL DERIVED FROM CARDIOPULMONARY EXERCISE TESTS TO PREDICT MORTALITY RISK IN PATIENTS WITH MILD-TO-MODERATE HEART FAILURE

Lee Ingle[1*] PhD, Alan Rigby[2] PhD, Rebecca Sloan[1] PhD, Sean Carroll[1] PhD, Kevin Goode[2] PhD, John G Cleland[2] MD, Andrew L Clark[2] MD

*Address for correspondence

[1] 104 Don Building
Department of Sport, Health & Exercise Science
University of Hull
Cottingham Road
Kingston-upon-Hull
HU7 6RX
United Kingdom

Tel: 01482 463141
Fax: 01482 463855
Email: L.Ingle@hull.ac.uk

[2]Department of Cardiology
Castle Hill Hospital
Kingston-upon-Hull
HU16 5JQ
United Kingdom

Running title: Composite CPET risk model in CHF

Conflicts of Interest: None declared

Total Word Count: 2,395 (excluding abstract, tables, figures, references).

Abstract:

Objective: Cardiopulmonary exercise testing (CPET) is used to predict outcome in patients with mild-to-moderate heart failure (HF). Single CPET-derived variables are often used, but we wanted to see if a composite score achieved better predictive power.

Methods: Retrospective analysis of patient records at the Department of Cardiology, Castle Hill Hospital, Kingston-upon-Hull. 387 patients [median (25^{th}-75^{th} percentile)] [age 65 (56-72) years; 79% males; LVEF 34 (31-37) %] were included. Patients underwent a symptom-limited, maximal CPET on a treadmill. During a median follow up of 8.6 ± 2.1 years in survivors, 107 patients died. Survival models were built and validated using a hybrid approach between the bootstrap and Cox regression. Nine CPET-derived variables were included. Z-score defined each variable's predictive strength. Model coefficients were converted to a risk score.

Results: Four CPET-related variables were independent predictors of all-cause mortality in the survival model: the presence of exertional oscillatory ventilation (EOV), increasing slope of the relation between ventilation and carbon dioxide production (VE/VCO_2 slope), decreasing oxygen uptake efficiency slope (OUES), and an increase in the lowest ventilatory equivalent for carbon dioxide ($VEqCO_2$ nadir). Individual predictors of mortality ranged from 0.60 to 0.71 using Harrell's C-statistic, but the optimal combination of EOV + VE/VCO_2 slope + OUES + $VEqCO_2$ nadir reached 0.75. The Hull CPET risk score had a significantly higher area under the curve (0.78) when compared to the Heart Failure Survival Score (AUC=0.70; P<0.001).

Conclusions: A composite risk score using variables from CPET out-performs the traditional single variable approach in predicting outcome in patients with mild-to-moderate HF.

Keywords: CPET score, prognosis, risk, CHF, exercise, EOV, OUES

What is already known about this subject?

There is a substantial literature describing the performance of different variables derived from cardiopulmonary exercise testing (CPET) as predictors of prognosis. Optimal cut-points delineating higher risk patients are often identified using binary statistical approaches. However, the traditional approach often focuses on the top performing variable(s) whilst discounting the additive or cumulative effect of a combination of different predictor variables. Composite risk scores, which combine the level of risk across a number of variables, have become increasingly prevalent especially in studies including patients with heart failure (HF).

What does this study add?

None of the composite approaches to risk scores derived by CPET have combined both conventional variables, such as peak VO_2 and VE/VCO_2 slope, and more recently described variables, such as the presence of exertional oscillatory ventilation, $VEqCO_2$ nadir, peak circulatory power and oxygen uptake efficiency slope. We have developed a composite risk model derived from CPET-related variables (traditional and contemporary) to predict mortality in patients with mild-to-moderate HF. Our model derived from CPET-only variables outperformed the Heart Failure Survival Score.

How might this impact on clinical practice?

Clinicians should consider using composite risk scores from CPET studies to improve risk prediction in patients with mild-to-moderate heart failure.

Introduction:

Chronic heart failure (HF) is very common and has a poor prognosis despite medical therapy.(1) The patient's performance during a cardio-pulmonary exercise test (CPET) is used to assess prognosis and to help select patients who might benefit from heart transplantation. CPET broadly provides two types of prognostic variable: a measure of exercise capacity, such as peak oxygen consumption (VO_2) or peak circulatory power (markers of oxygen delivery and extraction), and a measure of the ventilatory response to exercise, such as the slope of the relation between ventilation and carbon dioxide production (V_E/V_{CO_2} slope), reflecting the abnormal stimulus to ventilation in HF.(2)

There is a large literature describing the performance of different variables derived from CPET as predictors of prognosis. Optimal cut-points delineating higher risk patients are often identified using binary statistical approaches. However, the traditional approach often focuses on the top performing variable(s) whilst discounting the additive or cumulative effect of a combination of different predictor variables.(3) Composite risk scores, which combine the level of risk across a number of variables, have become more commonplace.(3-5) None of the composite approaches has combined both conventional variables, such as peak VO_2 and VE/VCO_2 slope, and more recently described variables, such as the presence of exertional oscillatory ventilation, peak circulatory power and oxygen uptake efficiency slope. We aimed to develop a composite risk model derived from CPET-related variables to predict mortality in patients with mild-to-moderate HF.

Methods:

The Hull and East Riding Ethics Committee approved the study, and all patients provided informed consent. We recruited consecutive patients referred to a community heart failure clinic with symptoms of breathlessness (NYHA functional class II-III) who had left ventricular systolic dysfunction on investigation. Clinical information obtained included past medical history and drug and smoking history. Clinical examination included measurement of body mass index (BMI), heart rate, rhythm, and blood pressure. Patients were excluded if they were unable to exercise because of non-cardiac limitations (such as osteoarthritis), or had significant respiratory disease (defined as a predicted $FEV_1/FVC < 70\%$).

Heart failure was defined as the presence of current symptoms of HF, or a history of symptoms controlled by ongoing therapy, and impaired left ventricular systolic function. Left ventricular function was assessed by estimation on a scale of normal, mild, mild-to-moderate, moderate, moderate-to-severe, and severe impairment using echocardiography. Left ventricular ejection fraction (LVEF) was calculated using the Simpson's formula from measurements of end-diastolic and end-systolic volumes on apical 2D views where possible, and LVSD was diagnosed if LVEF was ≤45%. When LVEF could not be calculated, LVSD was diagnosed if there was at least "mild-to-moderate" impairment.

Patients underwent a symptom-limited, maximal CPET on a treadmill using the Bruce protocol modified by the addition of a Stage 0 (2.74 $km \cdot h^{-1}$ and 0% gradient) at the onset of exercise. Metabolic gas exchange was measured with an Oxycon Delta metabolic cart (VIASYS Healthcare Inc., Philadelphia, PA). Prior to start of CPET, patients rested in a seated

position and resting end-tidal CO_2 (resting $PETCO_2$) was averaged over the first two minutes.(6)

To develop the composite risk score, we focused on variables that have previously been identified as independent predictors of mortality in patients with HF:

- Exercise oscillatory ventilation (EOV) during CPET using the criteria described by Leite and colleagues.(7) The criteria included: (a) at least three oscillatory fluctuations in minute ventilation during warm-up and exercise; (b) regular oscillations, as defined by a standard deviation of 3 consecutive cycles (time between 2 consecutive nadirs) within 20% of the average; (c) a minimal average ventilation amplitude of at least 5 litres, defined as peak VE of one oscillation minus the average of two adjacent nadirs.(8)
- Peak oxygen uptake (peak VO_2) was calculated as the average VO_2 for the final 30s of exercise.
- Peak circulatory power (PCP) was defined as the product of peak VO_2 and peak systolic arterial pressure at peak exercise intensity.(9)
- The oxygen uptake efficiency slope (OUES) was derived from the linear relationship between oxygen consumption (VO_2) and $\log_{10}$ ventilation (VE).(10)
- The ventilatory anaerobic threshold (AT) was calculated by the V-slope method.(11)
- The gradient of the relationship between VE and VCO_2 (VE/VCO_2 slope) was calculated by linear regression analysis using data acquired from the whole test.
- The lowest point of the relation between ventilation and carbon dioxide production ($VEqCO_2$ nadir) was identified by plotting the consecutive 30-second readings of the data.(12, 13)

- The peak respiratory exchange ratio (pRER) was calculated as the mean VCO_2/VO_2 ratio for the final 30s of exercise.
- Heart rate reserve (HRR) was calculated as the difference between resting heart rate and peak heart rate achieved during the CPET.

Statistical Methodology

Continuous data are summarised by the median (25^{th} to 75th percentile ranges); categorical data by percentages. The primary outcome measure was all-cause mortality. Models were built and validated using a hybrid approach between the bootstrap and Cox regression.(14) We have used the method in previous work.(15) Bootstrapping is sampling data with replacement meaning that an individual may be selected more than once per bootstrapped sample.(16) There should be no more than 200 bootstraps per sample,(17) and the method is not without its critics.(18) However, Sauerbrei (14) contended that with respect to bootstrap sampling, key variables will be included in most replications, and the inclusion frequency may be used as a criterion for the importance of a variable. To test the adequacy of the model we generated 10 test samples using 10-fold cross-validation. The data was split into 10 equal sub-samples (S1-S10) at random. From this we generated 10 test samples leaving out 10% of the data at a time. Thus, test sample 1 was generated leaving out sub-sample S1 and so on. Royston (19) recently developed a new measure of explained variance for use with censored survival data which was applied to the test samples.

Nine CPET-derived variables were included: EOV, OUES, $VEqCO_2$ nadir, VE/VCO_2 slope, resting $PETCO_2$, peak VO_2, HRR, AT, and PCP. Optimal binary cut-points were calculated from receiver operating characteristic (ROC) curves. The cut-offs were determined by maximising

the sum of sensitivity and specificity (equivalent to the Youden index). ROC curves were drawn using methods of Hanley and McNeil.(20)

Goodness-of-fit was explored by Cox-Snell residuals. We fitted a multivariable Cox regression model on the nine variables. We expressed each variable's predictive strength by its Z-score (model coefficient/standard error). Model coefficients were converted to a risk score by multiplying by 10 (for ease of rounding) and summing over all patients.(21) The proportionality of hazards (PH) assumption was based on the global PH test.(22)

Prognostic comparisons between models were made by Harrell's concordance c-statistic, a summary measure of model accuracy (23) similar to the area under the curve of a receiver-operator characteristics curve. The validity of the c-statistic is not dependent on parametric distribution. A Kaplan-Meier curve was plotted;(24,25) risk groups were compared by the log-rank test. An arbitrary level of 5% statistical significance (two-tailed) was assumed. The Stata statistical computer package was used to analyse the data (Stata Statistical Software, v10, StataCorp, College Station, Texas).

External validation

We validated our CPET risk score against the more established Heart Failure Survival Score (HFSS).(26) The HFSS uses seven variables that independently predict prognosis: peak oxygen uptake, resting heart rate (HR), mean blood pressure (mBP), LVEF, serum sodium, presence/absence of ischaemic heart disease, presence/absence of an intraventricular defect (IVCD). The presence of IVCD was defined as QRS interval ≥120 msec due to left or right bundle branch block or ventricular-paced rhythm (irrespective of CRT).(27)

Dichotomous variables were coded as 1 = present and 0 = absent. The HFSS was calculated in each patient by the following formula:

HFSS = [(0.0216 x resting HR) + (– 0.0255 x mBP) + (– 0.0464 x LVEF) + (– 0.047 x serum sodium) + (– 0.0546 x peak VO_2) + (0.608 x presence or absence of IVCD) + (0.6931 x presence or absence of ischemic heart disease)].

Results:

From 423 patients undergoing CPET, we identified 387 patients [median (25^{th}-75^{th} percentile)] (age 65 (56-72) years; 79% males; LVEF 34 (31-37) %; peak VO_2 21 (17-25) $mL \cdot kg^{-1} \cdot min^{-1}$; VE/VCO_2 slope 32 (29-38) with a complete data set (Table 1). Of these, 70% were taking ACE-inhibitors, 72% beta-blockers, and 61% loop diuretics. One hundred and seven patients (28%) died during follow-up. The median follow up in survivors was 8.6 ± 2.1 years.

Table 2 shows the inclusion frequencies of the potential variables (*P*=0.05) from 100 bootstrapped models. The PH assumption was not violated for any model. Four variables appeared in more than one-third of all models (EOV 98%, VE/VCO_2 slope 46%, $VEqCO_2$ nadir 41%, and OUES 34%). The optimal binary cut-points used in our study were as follows: EOV (risk score= 10; higher risk threshold = EOV present); VE/VCO_2 slope (risk score= 9; higher risk threshold >34); $VEqCO_2$ nadir (risk score = 5; higher risk threshold >33); and OUES (risk score=5; higher risk threshold <1.8 $L \cdot min^{-1}$). The maximum possible risk score is 29. None of our models were over-fitted.(28,29) Prognostic strength ranged from 0.60 to 0.71 (based on Harrell's c-statistic) for individual predictors of mortality. However, prognostic strength increased when using a combination of different independent predictors; the optimal combination was EOV + VE/VCO_2 slope + OUES + $VEqCO_2$ nadir (Harrell's c-statistic = 0.75).

Table 3 shows the distribution of our risk score for all patients and includes observed and expected deaths (under the assumption of no relationship between survival and risk score). For a risk score of 29 points, there was an excess of deaths in a ratio of 3.8:1. Table 4 shows the consistency of the optimal model (EOV + OUES + $VEqCO_2$ nadir + VE/VCO_2 slope) when 10 random samples of 90% of the distribution were re-tested. The number of events in the test samples ranged from 90-104. Explained variance for each random model was similar (range 43%-53%) demonstrating good consistency in the overall model. Table 5 shows changes in explained variance for CPET variables when presented as single variables, or in two-way or three-way combinations. Explained variance in single CPET variables ranged from 12% - 26%; variance increased when CPET variables were considered in two-way combinations (30%-37%); and was highest when variables were presented in three-way combinations (36%-43%). For example, EOV + $VEqCO_2$ nadir + OUES demonstrated an explained variance of 43% (29-55%). This analysis provides further supporting evidence for our 4-variable risk score. Figure 1 shows a Kaplan-Meier curve: as the risk score increases, so does the probability of death (log-rank test=109.2, df=3, *P*<0.001).

We explored the goodness-of-fit of our CPET risk score by plotting Cox-Snell residuals against the 45^{o} line, and found that the residual plot was closely related (Figure 2). The HR for our risk score was 1.10 (95% CI=1.08-1.10; AUC: 0.78; *P*<0.001). The Hull CPET risk score had a significantly higher area under the curve (0.78) when compared to the Heart Failure Survival Score (AUC=0.70; *P*<0.001; Figure 3).

Discussion:

We have shown that a composite risk score using variables from CPET out-performs the traditional single variable approach in predicting outcome in patients with mild-to-moderate HF. We validated our model against the Heart Failure Survival Score (26) and found that variables collected solely from CPET significantly outperformed the well established HFSS. The HFSS includes only one CPET-related variable, peak oxygen uptake. The other six variables are derived from a combination of standard investigative methods including echocardiography, electrocardiography, blood pressure monitoring, blood biochemistry, and patient history (aetiology of disease). Recently, Goda and colleagues (27) showed that the HFSS outperformed peak oxygen uptake alone for stratifying risk in CHF patients in the presence of implantable cardioverter-defibrillators and/or cardiac resynchronisation therapy. Risk stratification models should investigate the efficacy of combinations of CPET-related variables in the era of device therapy.

Historically, most exercise-related composite risk scores have been developed in patients with coronary artery disease prior to the widespread adoption of CPET.(5) Perhaps the most accepted integrated exercise risk score is the Duke Treadmill Score (DTS) which has both prognostic (30) and diagnostic (31) predictive power. The DTS combines exercise time (using a Bruce protocol) with ECG abnormalities (ST segment depression) and symptoms of angina. It was originally described in patients with coronary artery disease, though successful validation in other sub-groups has been reported.(32) Few composite risk score have been developed which have specifically included CPET variables for risk stratification among HF patients.

Risk models using both non-invasive and invasive data (with and without catheterisation data) in combination with peak VO_2 was developed in 268 patients with advanced HF.(33) The models were prospectively validated on 199 similar patients, and the non-invasive model performed well in both samples, and model performance did not improve with the addition of invasive catheterisation-related data. The authors concluded that the selection of candidates for cardiac transplantation may be improved by using a non-invasive risk-stratification model which included peak VO_2. Myers and colleagues (3) recruited 710 patients with HF (80% male; 56 ± 13 years; LVEF 33 ± 13 %) from four different institutions in Italy and the USA. CPET-derived variables included in the risk score included peak VO_2, VE/VCO_2 slope, OUES, resting $PETCO_2$, heart rate recovery and chronotropic index. The VE/VCO_2 slope (optimal cut point ≥34) was the strongest predictor of risk and attributed a relative weight of 7 points. A cumulative CPET score >15 points was associated with an annual mortality of 27% and a relative risk of 7.6. The authors recently published a validation study (34) of their original work in a larger sample size using a different statistical approach. Our study extends these findings by including other independent predictors of mortality in our risk algorithm such as ventilatory variables, EOV and $VEqCO_2$ nadir, and circulatory-related variables including heart rate reserve and peak circulatory power.

Guazzi and co-workers (5) used peak VO_2, VE/VCO_2 slope, EOV, to develop a prognostic risk score in 695 patients with heart failure. EOV was the strongest single predictor of cardiac mortality. The greatest contribution to the risk score was provided by EOV, followed by VE/VCO_2 slope, whereas peak VO_2 added minimal prognostic value. However, one of the major limitations of the study was that it only included 3 variables and did not include other potentially important predictive variables such as OUES, $VEqCO_2$ nadir, and peak circulatory

power. Recently, Italian clinicians published data from a multi-centre study (34) designed to build a new risk score for patients with systolic HF, integrating CPET measures with established clinical, laboratory and echocardiographic risk factors in order to identify patients at risk of cardiovascular death and urgent heart transplant. The MECKI score combined % predicted peak VO_2, VE/VCO_2 slope, LVEF, haemoglobin, sodium, and modification of diet in renal disease. A ROC analysis of the MECKI score for predicting CV death and heart transplant was 0.804 at year 1 which decreased to 0.760 by year 4. The MECKI score appears to further advance holistic risk models such as the Heart Failure Survival Score (26) and the HF-Action Predictive Risk Score Model (35) by the inclusion of a well established CPET-related risk algorithm.

Limitations

In our study, most patients were men (79%) and further work is needed to validate the risk model in women. We only included all-cause mortality as our primary clinical end-point; however, this method has advantages over a cardiovascular mortality end-point.(36) We accept that dichotomising continuous variables can be problematic; however, we wanted to develop a risk score which was simple and easy to use. Using continuous data would mean that every variable would have a different score and would require the operator to calculate a score per variable and then do a summation. We felt a more pragmatic approach was required. We acknowledge that there are many other variables and biomarkers which are important predictors of mortality in patients with HF, however, our focus for this risk score was solely cardiometabolic variables collected from CPET.

Conclusion

We have found that a composite risk score using four CPET-derived variables (EOV + VE/VCO_2 slope + OUES + $VEqCO_2$ nadir) out-performs the traditional single variable approach in predicting outcome in patients with mild-to-moderate HF.

Competing Interests

The authors have no competing interests to declare.

Contributors: LI led on the drafting and revision of the paper – he is guarantor. RS was responsible for cleaning of data, calculation of variables, drafting of manuscript. SC was responsible for calculation of variables and drafting of manuscript. AS and KG were responsible for data analysis. JG and ALC provided critical input and revised the initial drafts of the manuscript.

Authors Declaration:

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in HEART editions and any other BMJPGL products to exploit all subsidiary rights.

List of Figures

Figure 1. Kaplan-Meier curve showing risk distribution of the risk score (maximum score = 29)

Figure 2. Hull CPET survival score: Cox-Snell residual plot

Figure 3. Comparison between the Hull CPET survival score (AUC=0.78, 95%CI=0.72-0.82), and the HFSS (AUC=0.70; 95%CI=0.64-0.75; *P*<0.001).

Table 1. Baseline clinical characteristics in 387 patients with HF [median (25^{th}-75^{th} percentile)]

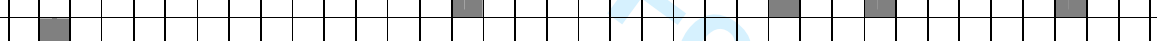
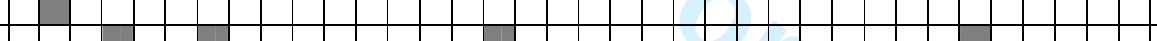
Variables	All patients
Age (years)	65 (56-72)
Males (%)	79
BMI ($kg \cdot m^{-2}$)	27 (24-30)
LVEF (%)	34 (31-37)
Loop diuretic (%)	61
ACE-I (%)	70
Beta-blocker (%)	72
Resting HR (bpm)	75 (64-85)
FEV_1 (% predicted)	90 (76-104)
Sodium ($mmol \cdot L^{-1}$)	139 (137-141)
Creatinine ($\mu mol \cdot L^{-1}$)	104 (86-125)
Haemoglobin (g/dL^{-1})	13.9 (13.0-15.0)
Systolic BP (rest, mmHg)	137 (119-152)
Diastolic BP (rest, mmHg)	84 (73-94)
DCM (%)	11
Hypertension (%)	6
IHD (%)	83
HR (peak, bpm)	133 (116-155)
EOV (%)	34
OUES ($L \cdot min^{-1}$)	1.9 (1.5-2.5)

$VEqCO_2$ (nadir)	31.7 (28.0-36.4)
VE/VCO_2 slope	32.1 (28.5-37.8)
Resting $PETCO_2$ (mmHg)	33.3 (30.2-35.7)
Peak VO_2 ($mL \cdot kg^{-1} \cdot min^{-1}$)	21.0 (16.6-25.0)
AT ($mL \cdot kg^{-1} \cdot min^{-1}$)	14.2 (11.4-18.1)
HRR ($beats \cdot min^{-1}$)	57 (42-77)
PCP ($mmHg \cdot mL^{-1} \cdot min^{-1}$)	3498 (2559-4662)
Peak RER	1.05 (1.00-1.11)
Exercise time (seconds)	537 (383 -743)
Diastolic BP (peak, mmHg)	90 (78-140)
Systolic BP (peak, mmHg)	171 (150-200)
QRS duration ($m \cdot sec^{-1}$)	106 (88-128)
HFSS	9.2 (8.6-9.8)

DCM: dilated cardiomyopathy; IHD: ischaemic heart disease; EOV: Exertional oscillatory ventilation; OUES: oxygen uptake efficiency slope; Resting $PETCO_2$: end-tidal CO_2 at rest; AT: anaerobic threshold; HRR: heart rate recovery; PCP: peak circulatory power; peak RER: peak respiratory exchange ratio; HR: heart rate; BP: blood pressure; HFSS: heart failure survival score.

Table 2. An illustration of bootstrapped Cox models for each individual and combined CPET variable(s). An example of models 65 to 100 is presented below.

Each column represents one model with the shaded part a variable within the model found to be a predictor of outcome. The frequency with which a variable was found from all 100 bootstrapped models is reported as is Harrell's c-statistic which shows the prognostic strength for each individual and combined variable(s).

Variable	Bootstrapped models	Frequency	C-statistic
EOV		98	0.60
OUES		34	0.70
$VEqCO_2$ nadir		41	0.71
VE/VCO_2 slope		46	0.69
Resting $PETCO_2$		9	0.61
Peak VO_2		8	0.67
AT		5	0.63
HRR		10	0.64
PCP		18	0.70
EOV + OUES		-	**0.73**
EOV + $VEqCO_2$ nadir + VE/VCO_2 slope		-	**0.74**
EOV + OUES + $VEqCO_2$ nadir + VE/VCO_2 slope		-	**0.75**

EOV: Exertional oscillatory ventilation; OUES: oxygen uptake efficiency slope; $PETCO_2$: end-tidal CO_2; AT: anaerobic threshold; HRR: heart rate recovery; PCP: peak circulatory power; peak RER: peak respiratory exchange ratio.

Table 3. Distribution of the risk score

Risk Score	Frequency	Dead Observed (O)	Dead Expected (E)	O/E ratio
0	90	7	29.34	0.23
5	16	1	5.58	0.17
9	47	6	13.77	0.43
10	58	12	17.66	0.67
14	13	3	3.69	0.81
15	7	1	2.27	0.44
19	103	39	24.24	1.60
20	3	2	0.77	2.59
24	8	6	1.73	3.16
29	42	30	7.96	3.76
Total	387	107	107.00	

Chi-squared = 111.9, df=9, *P*<0.01; Expected cell frequencies based on no association between survival and the risk score.

The risk score is cumulative based on the four most prominent CPET variables we have identified:

If, EOV is present: risk score= 10;
VE/VCO_2 slope > 34: risk score= 9;
$VEqCO_2$ nadir >33: risk score = 5;
OUES <1.8 $L \cdot min^{-1}$ = risk score=5;
The maximum possible risk score is 29.

Footnote: At each distinct failure-time the contribution to the test statistic comes as a sum of the difference between the observed/expected deaths in each of the 10 groups. The expected number of deaths is obtained under a null hypothesis of no differences between the survival experiences of the 10 groups. Pearson's Chi-squared test is calculated as $O\text{-}E^2/E$.(37)

Table 4. Optimal model testing (EOV + OUES + $VEqCO_2$ nadir + VE/VCO_2 slope) based on random samples of 90% of the distribution

Random sample	**Events**	Estimated explained relative risk **(%) (95% CI)**
1	91	49 (41-64)
2	96	46 (33-62)
3	92	43 (28-58)
4	90	49 (33-65)
5	94	47 (36-61)
6	98	50 (35-66)
7	94	53 (60-66)
8	102	44 (36-60)
9	104	46 (32-59)
10	102	47 (33-65)

Table 5. Model testing showing changes in estimated explained relative risk in single, two-, and three-way combinations of CPET variables

Variable	Estimated explained relative risk **(%)** **(95% CI)**
Single variables:	
EOV	12 (5-22)
$VEqCO_2$ nadir	26 (16-40)
OUES	22 (12-47)
VE/VCO_2 slope	25 (14-36)
Two-way combinations:	
EOV + $VEqCO_2$ nadir	37 (24-52)
EOV + OUES	33(19-45)
EOV + VE/VCO_2 slope	35 (20-48)
$VEqCO_2$ nadir + OUES	33 (22-50)
$VEqCO_2$ nadir + VE/VCO_2 slope	34 (22,50)
OUES + VE/VCO_2 slope	30 (17-46)
Three-way combinations:	
EOV + $VEqCO_2$ nadir + OUES	43 (29-55)
EOV + $VEqCO_2$ nadir + VE/VCO_2 slope	43 (31-58)
EOV + OUES + VE/VCO_2 slope	39 (28-53)
$VEqCO_2$ nadir + OUES + VE/VCO_2 slope	36 (22-50)

References:

1. Mosterd A, Hoes AW. Clinical epidemiology of heart failure. Heart 2007;**93**: 1137-46.

2. Ingle L, Sloan R, Carroll S, et al. Abnormalities of the ventilatory equivalent for carbon dioxide in patients with chronic heart failure. Pulm Med 2012; 589164.

3. Myers J, Arena R, Dewey F, et al. A cardiopulmonary exercise testing score for predicting outcomes in patients with heart failure. Am Heart J 2008; **156**:1177-83.

4. Ashley E, Myers J, Froelicher V. Exercise testing scores as an example of better decisions through science. Med Sci Sports Ex 2002; **34**:1391-8.

5. Guazzi M, Boracchi P, Arena R, et al. Development of a cardiopulmonary exercise prognostic score for optimizing risk stratification in heart failure: the (P)e(R)i(O)dic (B)reathing during (E)xercise (PROBE) study. J Card Fail 2010; **16**:799-805.

6. Arena R, Myers J, Abella J, et al. The partial pressure of resting end-tidal carbon dioxide predicts major cardiac events in patients with systolic heart failure. Am Heart J 2008; **156**:982-8.

7. Leite JJ, Mansur AJ, de Freitas HF, et al. Periodic breathing during incremental exercise predicts mortality in patients with chronic heart failure evaluated for cardiac transplantation. J Am Coll Cardiol 2003;**41**:2175-81.

8. Zurek M, Corra U, Piepoli MF, et al. Exercise training reverses exertional oscillatory ventilation in heart failure patients. Eur Resp J 2012; **40**(5): 1238-44.

9. Cohen-Solal A, Tabet JY, Logeart D, et al. A non-invasively determined surrogate of cardiac power ('circulatory power') at peak exercise is a powerful prognostic factor in chronic heart failure. Eur Heart J 2002; **23**:806-14.

10. Baba R, Tsuyuki K, Kimura Y, et al. Oxygen uptake efficiency slope as a useful measure of cardiorespiratory functional reserve in adult cardiac patients. Eur J Appl Physiol Occup Physiol 1999; **80**:397-401.

11. Beaver WL, Wasserman K, Whipp BJ. A new method for detecting anaerobic threshold by gas exchange. J Appl Physiol 1986;**60**:2020-7.

12. Sun XG, Hansen JE, Garatachea N, et al. Ventilatory efficiency during exercise in healthy subjects. Am J Resp Crit Care Med 2002;**166**:1443-8.

13. Ingle L, Sloan R, Carroll S, et al. Prognostic significance of different measures of the ventilation-carbon dioxide relation in patients with suspected heart failure. Eur J Heart Fail 2011;**13**:537-42.

14. Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. Appl Statistics 1999;**48**:313-29.

15. Ingle L, Rigby AS, Carroll S, et al. Prognostic value of the 6 min walk test and self-perceived symptom severity in older patients with chronic heart failure. Eur Heart J 2007;**28**:560-8.

16. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Sci 1986;**1**:54-77.

17. Boos D, Stefanski L. Efron's bootstrap. Significance 2010;**7**:186-8.

18. Young G, Daniels H. Bootstrap bias. Biometrika 1990;**77**;179-185.

19. Royston P. Explained variation for survival models. Stata J 2006; **6**:83-96.

20. Hanley JA, McNeil BJ. The meaning and use of the area under the receiver operating characteristic (ROC) curve. Radiology 1992; **143**:29-36.

21. Clayton TC, Lubsen J, Pocock SJ, et al. Risk score for predicting death, myocardial infarction, and stroke in patients with stable angina, based on a large randomised trial cohort of patients. BMJ 2005;**331**(7521):869.

22. Schoenfeld D. Partial residual estimation for the proportional hazards regression. Biometrika 1982;**69**:239-41.

23. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;**15**:361-87.

24. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet 2002;**359**(9318):1686-9.

25. Pocock SJ, McCormack V, Gueyffier F, et al. A score for predicting risk of death from cardiovascular disease in adults with raised blood pressure, based on individual patient data from randomised controlled trials. BMJ 2001;**323**(7304):75-81.

26. Aaronson KD, Schwartz JS, Chen TM, Wong KL, Goin JE, Mancini DM. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. Circulation 1997; **17**:2660-7.

27. Goda A, Lund LH, Mancini D. The Heart Failure Survival Score outperforms the peak oxygen consumption for heart transplantation selection in the era of device therapy. J Heart Lung Transplant 2011; **30**:315-25.

28. Concato J, Peduzzi P, Holford TR, et al. Importance of events per independent variable in proportional hazards analysis .1. Background, goals, and general strategy. J Clin Epidemiol 1995;**48**:1495-501.

29. Peduzzi P, Concato J, Feinstein AR, et al. Importance of events per independent variable in proportional hazards regression analysis .2. Accuracy and precision of regression estimates. J Clin Epidemiol 1995;**48**:1503-10.

30. Mark DB, Shaw L, Harrell FE, Jr., et al. Prognostic value of a treadmill exercise score in outpatients with suspected coronary artery disease. New Engl J Med 1991;**325**(12):849-53.

31. Shaw LJ, Peterson ED, Shaw LK, et al. Use of a prognostic treadmill score in identifying diagnostic coronary disease subgroups. Circulation 1998;**98**:1622-30.

32. Kwok JM, Miller TD, Christian TF, et al. Prognostic value of a treadmill exercise score in symptomatic patients with nonspecific ST-T abnormalities on resting ECG. JAMA 1999;**282**:1047-53.

33. Myers J, Oliveira R, Dewey F, et al. Validation of a cardiopulmonary exercise test score in heart failure, Circ Heart Fail 2013; **6**: 211-218.

34. Agostoni P, Corra U, Cattodori G, et al. Metabolic exercise test data combined with cardiac and kidney indexes, the MECKI score: A multi-parametric approach to heart failure diagnosis, Int J Cardiol 2012; epub ahead of print.

35. O'Connor CM, Whellan DJ, Wojdyla D, et al. Factors related to morbidity and mortality in patients with chronic heart failure with systolic dysfunction: the HF-ACTION predictive risk score model. Circ Heart Fail 2011;**5**:63–71.

36. Lauer MS, Blackstone EH, Young JB, et al. Cause of death in clinical research: time for a reassessment? J Am Coll Cardiol 1999;**34**:618-20.

37. Bland JM, Altman DG. Statistics Notes: The logrank test. British Medical Journal 2004; **328**:1073-1074.

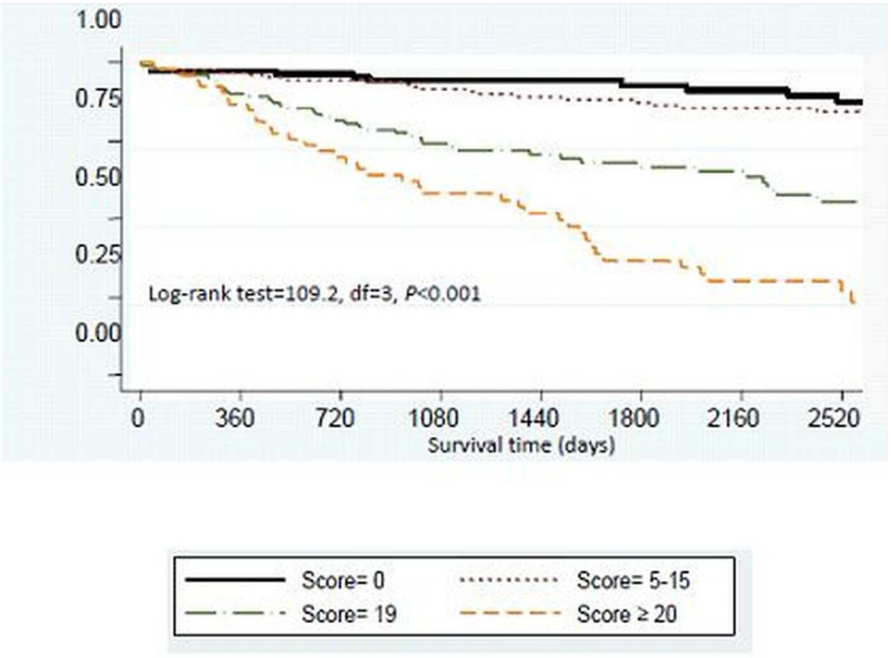


Figure 1.

Figure 1
125x91mm (300 x 300 DPI)

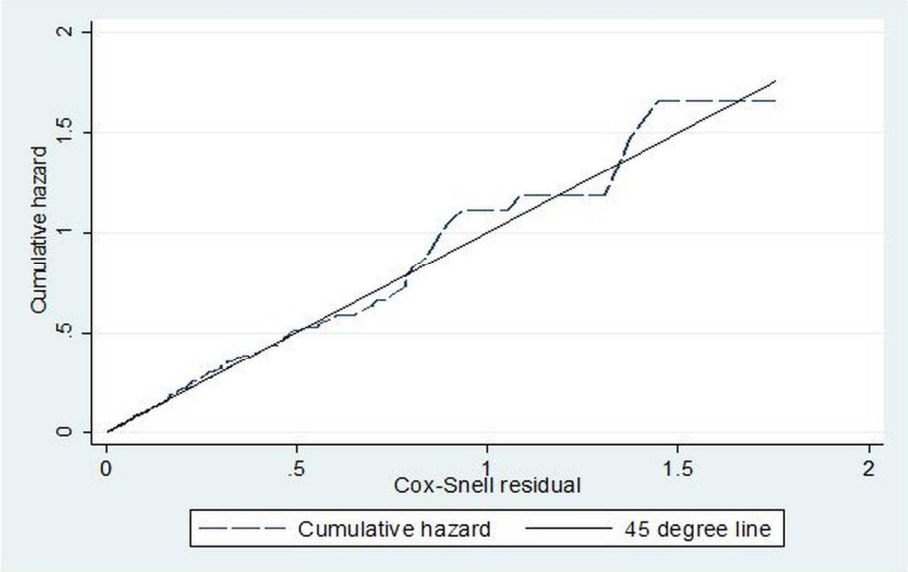


Figure 2.

Figure 2
177x117mm (300 x 300 DPI)

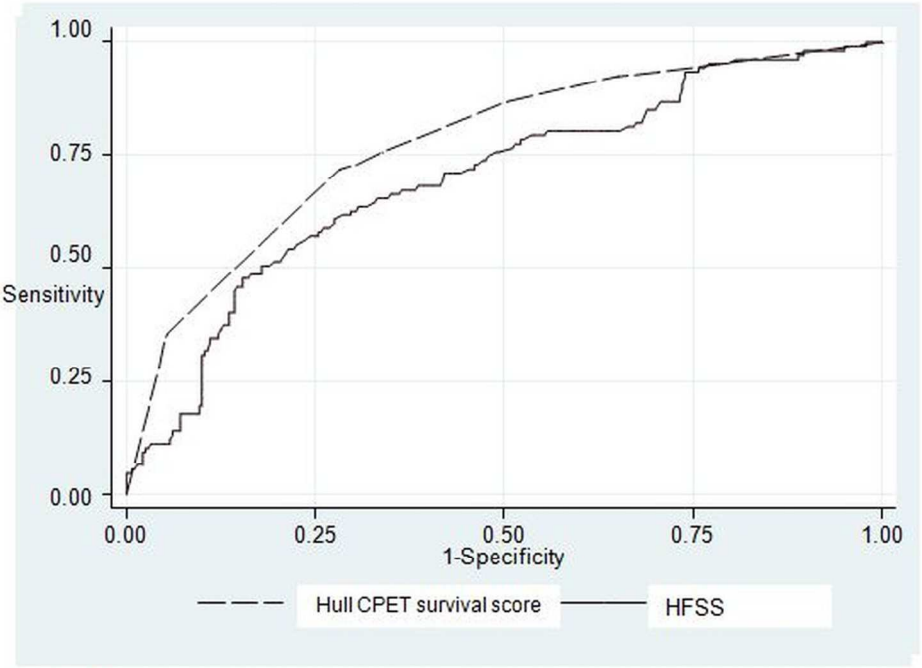


Figure 3.

Figure 3
156x119mm (300 x 300 DPI)