A neural network approach to predicting outcomes in heart failure using cardiopulmonary exercise testing

Jonathan Myers a,⁎, Cesar Roberto de Souza b, Audrey Borghi-Silva b, Marco Guazzi c, Paul Chase d, Daniel Bensimhond d, Mary Ann Peberdy e, Euan Ashley f, Erin West f, Lawrence P. Cahalin g, Daniel Forman h, Ross Arena i

a Division of Cardiology, VA Palo Alto Health Care System/Stanford University, United States
b Department of Physical Therapy, Federal University Sao Carlos, Brazil
c Division of Cardiology, University of Milano, Milano, Italy
d Department of Internal Medicine, Virginia Commonwealth University, Richmond, VA, United States
e Cardiovascular Medicine, Stanford University, Palo Alto, CA, United States
f Department of Physical Therapy, Leonard M. Miller School of Medicine, University of Miami, Miami, FL, United States
h Division of Cardiovascular Medicine, Brigham and Women’s Hospital, Boston, MA, United States
i Department of Physical Therapy, College of Applied Health Sciences, University of Illinois Chicago, Chicago, IL, United States

A R T I C L E   I N F O

Article history:
Received 25 August 2013
Received in revised form 26 October 2013
Accepted 14 December 2013
Available online xxxx

Keywords:
Heart failure
Cardiopulmonary exercise testing
Mortality
Oxygen uptake

A B S T R A C T

Objectives: To determine the utility of an artificial neural network (ANN) in predicting cardiovascular (CV) death in patients with heart failure (HF).

Method: ANN is used to weight inputs in multiple layers of mathematical connections in order to predict outcomes from multiple risk markers. This approach has not been applied in the context of cardiopulmonary exercise testing (CPX) to predict risk in patients with HF.

Methods: 2635 patients with HF underwent CPX and were followed for a mean of 29 ± 30 months. The sample was divided randomly into ANN training and testing sets to predict CV mortality. Peak VO2, VE/VCO2, rate recovery, oxygen uptake efficiency slope, and end-tidal CO2 pressure were included in the model. The predictive accuracy of the ANN was compared to logistic regression (LR) and a Cox proportional hazards (PH) score. A multi-layer feed-forward ANN was used and was tested with a single hidden layer containing a varying number of hidden neurons.

Results: There were 291 CV deaths during the follow-up. An abnormal VE/VCO2 slope was the strongest predictor of CV mortality using conventional PH analysis (hazard ratio 3.04; 95% CI 2.2–4.2, p < 0.001). After training, the ANN was more accurate in predicting CV mortality compared to LR and PH; ROC areas for the ANN, LR, and PH models were 0.72, 0.70, and 0.69, respectively. Age and BMI-adjusted odds ratios were 4.2, 2.6, and 2.9, for ANN, LR, and PH, respectively.

Conclusion: An ANN model slightly improves upon conventional methods for estimating CV mortality risk using established CPX responses.

1. Introduction

Over the last two decades, the cardiopulmonary exercise test (CPX) has become an important procedure for quantifying the degree of exercise intolerance in patients with chronic heart failure (HF). Many recent studies have demonstrated that CPX responses powerfully stratify risk for mortality, hospitalization, and other adverse outcomes in patients with HF [1–3]. While the focus of the CPX has historically centered on peak VO2 [4,5], more recently ventilatory abnormalities, heart rate recovery (HRR), and other responses have been demonstrated to provide clinically significant and independent information for estimating prognosis in these patients [1–3]. In particular, the recognition that indices of ventilatory inefficiency, alone or in combination with peak VO2, powerfully predict outcomes in HF has been a major advancement for the clinical utility of this procedure [1–3,6]. Other approaches designed to optimize risk stratification have included the application of algorithms that combine peak VO2 with clinical and hemodynamic data [7–9], the combination of several CPX indices [2,10–13] or the application of composite multivariate scores [7,10]. All of these methods have been shown to more accurately estimate risk for adverse events in patients with HF when compared to peak VO2 alone [2,7–13].

⁎ Corresponding author at: VA Palo Alto Health Care System, Cardiology Division (111C), 3801 Miranda Avenue, Palo Alto, CA 94304, United States. Tel.: +1 650 493 5000 x4661; fax: +1 650 852 5473.
E-mail address: fDrj9939aeol.com (J. Myers).

Published by Elsevier Ireland Ltd.

0016-755X/– see front matter. Published by Elsevier Ireland Ltd.
http://dx.doi.org/10.1016/j.ijcard.2013.12.031

Please cite this article as: Myers J, et al, A neural network approach to predicting outcomes in heart failure using cardiopulmonary exercise testing, Int J Cardiol (2013), http://dx.doi.org/10.1016/j.ijcard.2013.12.031
A great deal of effort in recent years has been directed toward the optimal application of CPX variables for estimating risk of adverse outcomes in patients with HF. Integrating several clinical and CPX responses into a single estimate of overall risk is advantageous [2,7–13], but is an obvious challenge for clinicians who likely do not have the time to perform complex, multivariate calculations. In addition, conventional statistical methods for assessing survival, such as proportional hazards analysis, are limited because they assume that a risk marker is associated with an outcome in a linear fashion (e.g. relative risks between subjects are constant over time), which is rarely the case in epidemiologic studies. In recent years, artificial neural networks (ANNs), a form of artificial intelligence which processes large amounts of information and considers complex non-linear relationships that inevitably exist in epidemiologic data, have been applied in various domains of medicine [14–16]. ANNs process information by means of networks of interconnected processing elements, similar to neurons. ANNs can detect non-linear relationships between dependent and independent variables and uncover all possible interactions between predictors, and have the capacity to produce more accurate estimations of risk than conventional methods [14–17]. Although ANNs have been successfully used to optimize diagnostic and prognostic performance of numerous tests in medicine [16,18], to our knowledge this technique has not been applied in the context of the CPX for the estimation of risk among patients with HF. In the current study, we hypothesized that by training and testing an ANN using multiple CPX variables based on a validated score [10,19], prediction of risk for adverse events in patients with HF would be improved relative to conventional survival analyses.

2. Methods

The study was performed as part of a HF consortium; multicenter, retrospective analysis including HF patients from the exercise laboratories at the VA Palo Alto Health Care System and Stanford University, Palo Alto, California, San Paolo Hospital, Milan, Italy, Virginia Commonwealth University, Richmond, Virginia, Brigham and Women’s Hospital, Boston, Massachusetts, and the Leuflaer Cardiovascular Research Foundation, Greensboro, North Carolina. A total of 2635 patients with chronic HF, tested between 1993 and 2010, were included. The sample included 1974 males and 651 females, with a mean age of 56 ± 14 years. Inclusion criteria consisted of a diagnosis of HF [20] and evidence of left ventricular systolic (ejection fraction [EF] < 40%) and/or with HF preserved EF by two-dimensional echocardiography obtained within one month of exercise testing. HF with preserved systolic function was considered to be present if the ejection fraction was normal (45%) and the subject had a history of decompensated heart failure. Subjects received routine follow-up care at the five institutions included in the study. All subjects were stable and receiving optimal medical therapy at the time of testing. The subjects completed a written informed consent and institutional review board approval was obtained at each institution.

2.1. CPX procedure and data collection

Symptom-limited CPX was performed on all patients using treadmill or cycle ergometer ramping protocols [21]. A cycle ergometer was used at Brigham and Women’s Hospital and in the European center, while a treadmill was used at the other centers. We previously observed that optimal peak VO2 and VE/VO2 slope thresholds for estimating validity were similar irrespective of mode of exercise in patients with HF [22]. Ventilatory expired gas analysis was performed using a metabolic cart at all five centers (Medgraphics CPX-D or ULTIMA PXP, Minneapolis, MN, Orca Diagnostics, Santa Barbara, CA, Parvo Medics TrueOne 2400, Sandy, UT or CareFusion Oxycon Pro, San Diego, CA). Before each test, the equipment was calibrated in standard fashion using reference gases. A standard 12-lead electrocardiogram was obtained at rest, each minute during exercise, and for at least 5 minutes during the recovery phase; blood pressure was measured using a standard left arm sphygmomanometer. Minute ventilation (VE, BTPS), oxygen uptake (VO2, STPD), carbon dioxide production (VCO2, STPD) and other CPX variables were acquired breath-by-breath and averaged over 10- or 15-second intervals. VE and VO2 responses throughout exercise were used to calculate the VE/VO2 slope via least squares linear regression (y = mx + b, m = slope). Previous work by our group and others has shown this method of calculating the VE/VO2 slope to be optimal for estimating prognosis [23,24]. The OUES was calculated using ([VO2]y (1/min) = m ([log]VE) + b, where m = OUES) [25]. Heart rate recovery was defined as [maximal heart rate minus heart rate at 1 min in recovery] [11]. Resting end-tidal CO2 pressure (PetCO2) was derived from the average of a two minute sitting resting period prior to the test [26].

2.2. Endpoints

The endpoint was cardiac-related mortality. Subjects were followed for major cardiac-related events for three years after their exercise test using the Social Security Death Index and/or hospital and outpatient medical chart review. Follow-up was performed by the HF program at each respective institution, providing a high likelihood that all major events were captured. Individuals conducting the CPX were not involved in decisions regarding cause of death or heart transplant/IVAD implantation.

2.3. Statistical analysis

NCSS (Kaysville, UT), software was used to perform unpaired t-tests to compare continuous variables, and chi-square tests were used to compare categorical variables between those who experienced a cardiac event and those who did not. Receiver operating characteristic (ROC) curve analysis was used to define optimal threshold values for each CPX response. Optimal thresholds for each of the CPX variables were as follows: VE/VO2 slope (≥ 34) abnormal HRR (≤ 6 beats at 1 min), OUES (≤ 1.4), PetCO2 (≥ 33 mm Hg), and VO2 (≤ 14 ml·kg⁻¹·min⁻¹).

Three sequential approaches were used to evaluate prediction of cardiovascular events. All survival analyses were adjusted for age, BMI and gender. Initially, Cox proportional hazards analysis was used to determine age-adjusted hazard ratios for the 5 CPX variables included in the model, each expressed dichotomously using the threshold value. For comparison with ANN and logistic regression, a composite CPX score was used as described previously [10,19], in which each variable was assigned a weight according to the hazard ratios, and summed to calculate the composite score. Stepwise logistic regression was applied using each of the CPX variables, and odds ratios were calculated along with their 95% confidence intervals. Receiver operating characteristic (ROC) curves were constructed to compare the predictive value of each method for predicting cardiovascular mortality. Z-tests were used to compare the areas under the ROC curves.

ANN analyses were performed using Accord.NET Framework (Sao Carlos, Brazil). The ANN employed is commonly viewed as a mathematical model that parallels neurotransmission in biology, in which computer technology is used to model complex relationships to find patterns in data. An ANN consists of a highly interconnected network of processing units (analogous to neurons in biology) including inputs (predictor variables), hidden layers (data processing elements), and outputs (outcomes). The ANN is initially presented with input variables for which the outcome (in the current study, cardiovascular death) is known. At each input node, the inputs are weighted numerically; these weights are randomly derived initially before learning takes place. The weighted sum of the input data is determined and presented to each node in the hidden layer. A transfer function then takes the weighted sum and calculates a value that numerically scales the strength of each node’s output. The output of each succeeding layer then becomes the input for the next layer. A major distinction between an ANN and conventional methods is its ability to “learn”. Learning occurs when an output is calculated; it is compared with the known output from the training set, and differences between the two are propagated through the ANN and the weights are adjusted in a repetitive manner.

The ANN used a feed-forward multilayer learning strategy with a back-propagation training method. It was created using age, BMI, ejection fraction and the 5 CPX variables included in the proportional hazards model described above. Of the 2635 patients in the sample, 550 randomly selected tests were set aside for testing and were not included in the training set. A feed-forward neural network with a sigmoid activation function was used. The feed-forward ANN is a multi-layer network in which no loops occur in the network path. The Levenberg-Marquardt algorithm was used during the learning phase of the ANN to optimize fitting [27]. The Bayesian regularization method was employed to constrain the fitting and enforce capacity control through regularization [28]. The ANN was tested with the random 550 tests using a single hidden layer containing a varying number of hidden neurons.

3. Results

During a mean follow-up of 2.4 ± 2.5 years, 291 subjects died from cardiovascular causes. Clinical characteristics and exercise test responses between survivors and those who experienced cardiovascular death are shown in Table 1. Subjects who died were older, had a lower ejection fraction and a higher New York Heart Association class. Among CPX variables, peak VO2 (18.7 ± 8.5 vs. 14.1 ± 5.5 ml·kg⁻¹·min⁻¹), peak heart rate, HRR, OUES, and PetCO2 were higher among those with no events. Conversely, resting heart rate, the VE/VO2 slope and the CPX weighted summed score were lower among survivors vs. those who died from cardiac causes.

Age-adjusted CPX predictors of cardiovascular mortality are presented in Table 2. Each component of the CPX score was significantly associated with cardiovascular mortality. The VE/VO2 slope generated the highest hazard ratio and an ROC area similar to that of peak VO2 (0.72), while an abnormal OUES had an ROC area of 0.75 (all p < 0.01). Test performance characteristics of the CPX variables from logistic regression analysis are presented in Table 3. All the CPX variables except the OUES were significant predictors of cardiovascular mortality. The highest odds ratio for predicting mortality was generated by an abnormal VE/VO2 slope (OR 2.83, 95% CI 2.4–3.3, p < 0.001). A comparison...
of the performance characteristics for predicting mortality using ANN (from the testing set), logistic regression, and proportional hazards analysis is presented in Table 4. The ANN had a slightly higher ROC area, sensitivity, and odds ratio compared to logistic regression and proportional hazards analysis (ROC areas 0.72, 0.70, and 0.69, respectively, Fig. 1). Differences in ROC areas between the three methods were not significant. Fig. 2 illustrates ROC areas produced by the number of hidden neurons; performance of the ANN peaked at 5 hidden neurons, after which a reduction in the ROC occurred due to overfitting.

### 4. Discussion

The appeal of ANNs is evident from their escalating use in various fields over the last two decades [14,16]. The advantages of ANNs include their ability to explore multiple hidden patterns to find non-linear interactions between independent variables, their freedom from statistical assumptions, and their ability to “learn” [17,18,29–31]. In cardiovascular medicine, these techniques have been applied to improve prediction of mortality risk from resting ECGs [32], to predict acute coronary syndrome or myocardial infarction [33,34], to associate conventional risk factors with 25-year mortality [35], and other applications [36]. To our knowledge, ANNs have not been applied in the context of CPX to predict outcomes in patients with HF. In the current study, we determined the predictive accuracy of established CPX variables using an ANN, and compared it to more conventional methods to estimate risk for cardiovascular mortality.

In training the ANN, we employed 5 easily obtained CPX variables that have been demonstrated to be strong predictors of mortality risk in patients with HF [1,2,10]; we recently reported that a multivariate CPX score [19] and proportional hazards analysis were compared to logistic regression and ANN. Proportional hazards analysis is the most commonly used method for estimating survival, with extensive applications in medicine over many decades. Logistic regression is a statistical fitting model that is highly effective in creating estimates for the likelihood that an event will occur. It provides coefficients that have intuitive clinical interpretations [37]. Biomedical knowledge is a well-established method used to model medical problems and provide effective in creating estimates for the likelihood that an event will occur. It provides coefficients that have intuitive clinical interpretations [37].

### Table 1

Clinical characteristics and exercise test responses between survivors and non-survivors (mean ± SD).

<table>
<thead>
<tr>
<th>Variables</th>
<th>All Subjects (n = 2635)</th>
<th>Survivors (n = 2344)</th>
<th>Cardiovascular death (n = 291)</th>
<th>p-Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>55.7 ± 14.2</td>
<td>55.1 ± 14.3</td>
<td>60.9 ± 13</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>28.52 ± 6.01</td>
<td>28.6 ± 6.1</td>
<td>27.8 ± 5.6</td>
<td>0.024</td>
</tr>
<tr>
<td>Ejection fraction (%)</td>
<td>35.46 ± 15.77</td>
<td>36.03 ± 15.9</td>
<td>31.1 ± 14.4</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>NYHA class</td>
<td>2.42 ± 0.84</td>
<td>2.38 ± 0.84</td>
<td>2.69 ± 0.75</td>
<td>0.001</td>
</tr>
<tr>
<td>Medications, n(%)</td>
<td>1611 (61.1)</td>
<td>1450 (61.8)</td>
<td>161 (55.3)</td>
<td>0.04</td>
</tr>
<tr>
<td>Beta blocker</td>
<td>1469 (55.7)</td>
<td>1301 (55.4)</td>
<td>168 (57.7)</td>
<td>0.47</td>
</tr>
<tr>
<td>Diuretics</td>
<td>1248 (47.4)</td>
<td>1104 (47.1)</td>
<td>144 (49.5)</td>
<td>0.004</td>
</tr>
<tr>
<td>Exercise test responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resting heart rate (b/min)</td>
<td>75 ± 14</td>
<td>75 ± 14</td>
<td>79 ± 15</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Peak heart rate (b/min)</td>
<td>129 ± 26</td>
<td>130 ± 27</td>
<td>119 ± 22</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Peak SBP (mm Hg)</td>
<td>18.15 ± 8.32</td>
<td>18.70 ± 8.5</td>
<td>14.1 ± 5.5</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Peak VO₂ (ml·kg⁻¹·min⁻¹)</td>
<td>33.35 ± 9.23</td>
<td>32.70 ± 8.8</td>
<td>38.3 ± 11</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>VE/VCO₂ slope</td>
<td>19.29 ± 13.2</td>
<td>20.26 ± 13.2</td>
<td>13.06 ± 11.04</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>HRR (beats)</td>
<td>20.1 ± 0.88</td>
<td>2.08 ± 0.89</td>
<td>1.58 ± 0.64</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>OUES</td>
<td>159.4 ± 461</td>
<td>33.73 ± 487</td>
<td>32.43 ± 4.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Weighted summed score</td>
<td>5.59 ± 4.41</td>
<td>5.31 ± 4.27</td>
<td>7.89 ± 4.84</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

NYHA = New York Heart Association; HRR = heart rate recovery 1-minute post exercise; OUES = oxygen uptake efficiency slope; PetCO₂ = end-tidal CO₂ pressure.

* p-Value reflects comparison between survivors and non-survivors.

### Table 2

Test performance characteristics for predicting cardiovascular mortality from CPX variables included in the neural network analysis.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Hazard ratio</th>
<th>95% CI</th>
<th>ROC area</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Wald p value</th>
<th>Likelihood ratio p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak VO₂</td>
<td>1.72</td>
<td>1.3–2.3</td>
<td>0.72</td>
<td>76</td>
<td>55</td>
<td>&lt;0.0001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>VE/VCO₂ slope</td>
<td>3.04</td>
<td>2.2–4.2</td>
<td>0.72</td>
<td>72</td>
<td>62</td>
<td>&lt;0.0001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HRR</td>
<td>2.17</td>
<td>1.4–3.3</td>
<td>0.68</td>
<td>73</td>
<td>61</td>
<td>0.0001</td>
<td>0.001</td>
</tr>
<tr>
<td>OUES</td>
<td>1.59</td>
<td>1.1–2.3</td>
<td>0.75</td>
<td>75</td>
<td>56</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>PetCO₂</td>
<td>1.34</td>
<td>0.99–1.8</td>
<td>0.63</td>
<td>71</td>
<td>47</td>
<td>0.06</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Hazard ratios from proportional hazards analysis, adjusted for age, BMI, and ejection fraction; HRR = heart rate recovery at 1-minute; OUES = oxygen uptake efficiency slope; PetCO₂ = end-tidal CO₂ pressure.

### Table 3

Test performance characteristics from logistic regression analysis.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Odds ratio</th>
<th>95% CI</th>
<th>Wald test p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak VO₂</td>
<td>1.72</td>
<td>1.3–1.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>VE/VCO₂ slope</td>
<td>2.83</td>
<td>2.4–3.3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HRR</td>
<td>2.30</td>
<td>1.7–3.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>OUES</td>
<td>1.12</td>
<td>0.71–1.1</td>
<td>0.29</td>
</tr>
<tr>
<td>PetCO₂</td>
<td>1.20</td>
<td>0.71–0.99</td>
<td>0.04</td>
</tr>
</tbody>
</table>

### Table 4

Performance characteristics for predicting mortality for ANN, logistic regression, and proportional hazards analysis using a CPX score.

<table>
<thead>
<tr>
<th>Model</th>
<th>ROC Area</th>
<th>ROC, SE</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>0.72</td>
<td>0.032</td>
<td>0.79</td>
<td>0.63</td>
<td>4.2</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.70</td>
<td>0.033</td>
<td>0.71</td>
<td>0.65</td>
<td>2.6</td>
</tr>
<tr>
<td>Proportional Hazards</td>
<td>0.69</td>
<td>0.033</td>
<td>0.74</td>
<td>0.53</td>
<td>2.9</td>
</tr>
</tbody>
</table>

ANN = artificial neural network.
techniques, ANNs could complement conventional methods by providing a measure of validation in a particular data set [30,38–40]. In addition, the use of ANNs may be justified in the many instances in which the assumptions required by proportional hazards analysis are not met or cannot be verified, a common problem in epidemiological research [41].

Our initial analysis involved the more conventional logistic regression and proportional hazards models. Unlike the ANN, these permit the determination of the most predictive explanatory variables. As we [2,10–12] and others [1,13,24] have observed previously, the VE/VCO₂ slope was the strongest predictor of risk, exhibiting the best performance characteristics in the test set (Tables 2–4). This was followed by training and testing the ANN in an independent subset of patients. The ANN, using 3 clinical variables (age, ejection fraction and BMI) along with the 5 CPX variables, exhibited a slightly higher predictive accuracy when compared to proportional hazards and logistic regression. The superior performance of the ANN was evidenced by a higher ROC area, a more optimal balance of sensitivity and specificity, and a higher odds ratio versus the other methods (Table 4). While this suggests a potential role for ANNs in estimating risk based on the CPX, incorporating the ANN model would require structured reporting software that is currently unavailable in metabolic systems. Such programs would permit the clinician to apply the probability calculations generated by the software and potentially aid in clinical decision making.

There have been a number of comparisons between ANN and either logistic regression or proportional hazards analysis for predicting outcomes in various chronic conditions [30,35,38–40,42–44]. Some of these studies have demonstrated better risk estimation from ANNs, but these differences have generally been small; others have shown no differences between ANN and conventional linear-based methods. Although the performance characteristics of the ANN in the current study were not strikingly different from conventional methods, each technique provides particular advantages. ANNs represent a more elegant approach, although inputs and outputs are subtle and non-intuitive. While this better reflects the complexity of clinical dynamics, it may seem disconcerting to clinicians who prefer the relatively more quantifiable and concrete relationships provided by traditional methods. However, logistic regression and proportional hazards are limited in that they are dependent upon assumptions that can distort relationships and result in biases that provide inaccurate estimates of risk. Their appeal is related to the fact that they provide direct estimates of variables that predict outcomes based on coefficients and corresponding odds or hazard ratios which are intuitively attractive to clinicians. While additional studies are needed to explore the utility of ANNs in estimating prognosis from CPX responses, an optimal approach might be the combination of linear and non-linear statistics since each has elements that account for limitations of the other [30,39,40,42,43].

The current results confirm the many studies performed over the last two decades demonstrating the value of CPX responses for estimating mortality risk in patients with HF. Importantly, the VE/VCO₂ slope was the most powerful predictor of risk (Tables 2 and 3), which is in agreement with numerous studies over the last decade [1,2,23]. Our results also provide further validation of a multivariate CPX score, which has been shown to provide significant net reclassification improvement compared with individual CPX and clinical variables [19]. Notably, the ANN provided superior risk stratification relative to these well-established predictors in our sample. It should also be noted that the test sample, chosen at random from our larger data set, had a comparatively low mortality rate relative to other studies from our group [2,8,10–12] and others [2,5,13,24]. Since predictive value is strongly related to prevalence, this may explain the comparatively modest prognostic characteristics of the 3 approaches (Table 4). While it would be of value to replicate these findings in an independent sample of patients with HF, the results nevertheless provide a valid comparison between ANN, proportional hazards analysis, and logistic regression.

### 4.1. Limitations

Our main interest in the current study was limited to assessing the performance of an ANN using CPX data; there are many other variables and biomarkers that have been used to define risk in HF, and superior prediction of cardiovascular mortality or other events might be achieved by including additional clinical variables. We only included 5 CPX variables that were part of a validated score [10]; there are other CPX responses that predict risk, particularly oscillatory breathing [1,2], which were not included. In addition, the sample was 75% male, and the results may not be as applicable to women.

### 4.2. Conclusion

These findings suggest that an ANN employing established CPX responses improves modestly upon conventional survival analyses for estimating CV mortality risk in patients with HF.
References


