Incremental Role of New York Heart Association Class and Cardiopulmonary Exercise Test Indices for Prognostication in Heart Failure: A Cohort Study

Pedro Henrique de Borba Engster, André Zimerman, Thomas Schaan, Marina S. Borges, Gabriel Souza, Giovanni Donelli Costa, Luis Eduardo Rohde, Anderson Donelli da Silveira

Universidade Federal do Rio Grande do Sul – FAMED, Porto Alegre, RS – Brazil
Hospital de Clínicas de Porto Alegre, Porto Alegre, RS – Brazil
Harvard Medical School, Boston, Massachusetts – USA
Hospital Moinhos de Vento – Cardiologia, Porto Alegre, RS – Brazil

Abstract

Background: The accuracy of the New York Heart Association (NYHA) classification to assess prognosis may be limited compared with objective cardiopulmonary exercise test (CPET) parameters in heart failure (HF).

Objective: To investigate the prognostic value of the NYHA classification in addition to Weber class.

Methods: Adult outpatients with HF undergoing CPET in a Brazilian tertiary care center were included. The physician-assigned NYHA class and the CPET-derived Weber class were stratified into “favorable” (NYHA I or II; Weber A or B) or “adverse” (NYHA III or IV; Weber C or D). Patients with one favorable class and one adverse class were defined as “discordant.” The primary endpoint was time to all-cause mortality. A 2-sided p value < 0.05 was considered statistically significant.

Results: A total of 834 patients were included. Median age was 57 years; 42% (351) were female, and median left ventricular ejection fraction was 32%. Among patients with concordant NYHA and Weber classes, those with adverse NYHA and Weber classes had significantly higher all-cause mortality compared to those with favorable classes (hazard ratio [HR]: 5.65; 95% confidence interval [CI]: 3.38 to 9.42). Among patients with discordant classes, there was no significant difference in all-cause mortality (HR: 1.38; 95% CI: 0.82 to 2.34). In the multivariable model, increments in NYHA class (HR: 1.55 per class increase; 95% CI: 1.26 to 1.92) and reductions in peak VO₂ (HR: 1.47 per 3 ml/kg/min decrease; 95% CI: 1.28 to 1.70) significantly predicted mortality.

Conclusions: Physician-assigned NYHA class and objective CPET measures provide complementary prognostic information for patients with HF.

Keywords: Heart Failure; Prognosis; Exercise Test.

Introduction

Heart failure (HF) is one of the leading causes of morbidity and mortality worldwide, affecting over 64 million people. One of the cornerstones of HF management is the definition of a patient’s New York Heart Association (NYHA) classification, proposed in 1921 to measure functional impairment. This subjective measurement has been widely used as an inclusion criterion for clinical protocols. Patients considered asymptomatic at ordinary physical activity (namely, NYHA class I) have been systematically excluded from HF trials. Consequently, clinical guidelines frequently use an NYHA class cutoff to determine eligibility for treatments such as mineralocorticoid receptor antagonists, sodium-glucose cotransporter-2 inhibitors, and cardiac resynchronization therapy. The NYHA classification is an established, powerful predictor of HF prognosis at a group level. Recent studies, however, have questioned the reproducibility of the NYHA classification and its ability to discriminate the prognosis of patients with HF at the individual level. These limitations have encouraged efforts to attain more accurate and reproducible parameters of functional capacity in patients with HF, ranging from structured questionnaires to objective measurements of functional capacity, such as cardiopulmonary exercise testing (CPET). CPET is a non-invasive method to analyze cardiopulmonary fitness and establish functional status. Currently, CPET is used to assess HF severity, monitor disease progression, and determine eligibility for heart transplantation.
For patients with HF, both the physician-assigned NYHA class and the objective CPET parameters have been shown to be independent prognostic factors. However, as two surrogates of functional capacity, one subjective and one objective, their combined ability for prognostication is less clear. For example, for patients who have undergone a CPET, it is plausible that the NYHA classification lacks additional prognostic value. In the current analysis, we investigated the interplay of CPET-derived indices and NYHA class to refine prognostic assessment in patients with HF, particularly when CPET and NYHA class depicted conflicting results.

Definitions and endpoints

The NYHA classification is a subjective, physician-defined measure of a patient’s physical limitation, ranging from no limitation at ordinary physical activity (class I) to symptomatic at rest (class IV). The Weber class is derived from the maximum oxygen consumption during exercise (peak VO\(_2\)) measured during CPET and is categorized into class A (peak VO\(_2\) > 20 ml/kg/min), B (16 to 20 ml/kg/min), C (10 to 16 ml/kg/min), and D (< 10 ml/kg/min). In this study, we stratified NYHA and Weber classes into “favorable” (NYHA I or II; Weber A or B) or “adverse” (NYHA III or IV; Weber C or D). Subjects with one favorable class and one adverse class (i.e., NYHA I or II with Weber C or D, or NYHA III or IV with Weber A or B) were classified as “discordant.” The primary endpoint of this study was all-cause mortality. Vital status was prospectively evaluated using electronic health records and telephone calls. As part of a sensitivity analysis, we also stratified patients into favorable and adverse classifications regarding minute ventilation/carbon dioxide production (VE/VCO\(_2\)) slope and percent-predicted peak VO\(_2\) (ppVO\(_2\)). Favorable VE/VCO\(_2\) slope was defined as VE/VCO\(_2\) \(\leq\) 36, and adverse VE/VCO\(_2\) slope was defined as VE/VCO\(_2\) > 36. Favorable ppVO\(_2\) was defined as ppVO\(_2\) \(\geq\) 50%, and adverse ppVO\(_2\) was defined as ppVO\(_2\) < 50%.

Methods

Patients and study design

This cohort study included consecutive patients with HF who underwent a CPET in a tertiary care hospital in Brazil between January 2008 and November 2020. The first CPET of each patient was included in this analysis. NYHA class was determined immediately before CPET or in the previous outpatient visit. Eligible patients were 16 years or older with documented HF, diagnosed by clinical, laboratory, and echocardiographic criteria. Subjects had to be clinically stable prior to CPET and using optimal medical therapy. There were no left ventricular ejection fraction (LVEF) eligibility criteria, i.e., patients with reduced, mildly reduced, and preserved LVEF were eligible for enrollment. Patients who were unable to perform a CPET were excluded. This study was approved by the local research ethics board, and all participants provided written informed consent for participation.
Cardiopulmonary exercise testing

CPET methodology has been previously reported by our institution, and it follows previously validated recommendations. CPET was conducted by experienced and trained cardiologists using standardized institutional protocols. In brief, CPET was performed on a treadmill (General Electric T-2100, GE Healthcare, USA) with breath-by-breath gas analysis (MetaLyzer 3B, Cortex, Leipzig, Germany or Quark CPET, COSMED, Rome, Italy). Symptom-limited maximal exercise testing with an individualized ramp protocol was used to yield fatigue-limited exercise duration of 8 to 12 minutes. Peak VO₂ was determined by the highest measure of a 20-second rolling average of breath-by-breath values. VE/VO₂ slope was determined by a linear regression model using data from the entire duration of the test. Oxygen uptake efficiency slope (OUES) was derived from a similar model, and ppVO₂ estimations used Wasserman and Hansen’s algorithm, considered the preferred equation for patients with HF.

Statistical analysis

Continuous variables are displayed as median (25th and 75th percentiles), as a Shapiro-Wilk test indicated that all continuous baseline variables significantly differed from a normal distribution. Categorical variables are displayed as absolute numbers and percentages. Kruskal-Wallis tests were used to compare continuous values, and chi-square tests were used to compare proportions. No post hoc tests were used. For the main analysis of time to all-cause death, the two groups of subjects with discordant NYHA and Weber classes (i.e., favorable NYHA and adverse Weber class, and adverse NYHA and favorable Weber class) were compared using a Cox proportional hazards model. Time to all-cause death was used to produce Kaplan-Meier estimates and analyzed with log-rank statistics. Furthermore, to visually examine the association between peak VO₂, NYHA class, and mortality, we developed a multivariable Cox model to compute the predicted 5-year mortality rates according to peak VO₂ and NYHA class, adjusted for age and sex at baseline. All analyses were performed using R v4.0.2 (R Foundation for Statistical Computing, R Core Team, 2023). A 2-sided p value < 0.05 was considered statistically significant. The dataset used for this manuscript is not openly available, but we encourage colleagues to contact the corresponding author if they are interested in collaborating.

Results

Patient characteristics

The clinical characteristics of the 834 patients included are described in Table 1. Median age was 57.1 years; 42% (351) were female, and median LVEF was 32.0%. Median follow-up time was 3.1 years (interquartile range: 1.6 to 5.1). Overall, patients were well distributed between NYHA classes I, II, and III, with only 3% classified as NYHA IV. Patients in milder HF classes were more likely to be male, to have preserved (versus reduced) LVEF, and to be using angiotensin-converting enzyme inhibitors or angiotensin receptor blockers.

Table 1 – Baseline characteristics

<table>
<thead>
<tr>
<th>Characteristics*</th>
<th>NYHA I (N=246)</th>
<th>NYHA II (N=362)</th>
<th>NYHA III (N=197)</th>
<th>NYHA IV (N=29)</th>
<th>Overall (N=834)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>57.1 (48.0-64.0)</td>
<td>56.8 (48.9-63.4)</td>
<td>58.2 (49.8-65.7)</td>
<td>56.1 (49.0-62.6)</td>
<td>57.1 (49.0-64.1)</td>
<td>0.346</td>
</tr>
<tr>
<td>Female sex</td>
<td>89 (36.2%)</td>
<td>151 (41.7%)</td>
<td>98 (49.7%)</td>
<td>13 (44.8%)</td>
<td>351 (42.1%)</td>
<td>0.039</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>26.1 (23.9-29.4)</td>
<td>28.2 (24.4-32.6)</td>
<td>27.7 (23.9-31.8)</td>
<td>26.6 (24.7-29.4)</td>
<td>27.4 (24.1-31.6)</td>
<td>0.004</td>
</tr>
<tr>
<td>Hypertension</td>
<td>117 (47.6%)</td>
<td>200 (55.2%)</td>
<td>108 (54.8%)</td>
<td>12 (41.4%)</td>
<td>437 (52.4%)</td>
<td>0.15</td>
</tr>
<tr>
<td>Diabetes</td>
<td>65 (26.4%)</td>
<td>120 (33.1%)</td>
<td>72 (36.5%)</td>
<td>11 (37.9%)</td>
<td>268 (32.1%)</td>
<td>0.11</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>43 (17.5%)</td>
<td>71 (19.6%)</td>
<td>50 (25.4%)</td>
<td>8 (27.6%)</td>
<td>172 (20.6%)</td>
<td>0.15</td>
</tr>
<tr>
<td>LVEF, %</td>
<td>34.0 (25.0-45.3)</td>
<td>32.0 (25.0-45.0)</td>
<td>30.0 (23.0-38.0)</td>
<td>28.0 (20.0-53.0)</td>
<td>32.0 (25.0-43.0)</td>
<td>0.003</td>
</tr>
<tr>
<td>LVEF &lt; 40%</td>
<td>150 (61.0%)</td>
<td>239 (66.0%)</td>
<td>149 (75.6%)</td>
<td>20 (69.0%)</td>
<td>558 (66.9%)</td>
<td>0.009</td>
</tr>
<tr>
<td>LVEF 40.0% to 49.9%</td>
<td>42 (17.1%)</td>
<td>53 (14.6%)</td>
<td>19 (9.6%)</td>
<td>1 (3.4%)</td>
<td>115 (13.8%)</td>
<td>0.27</td>
</tr>
<tr>
<td>LVEF ≥ 50%</td>
<td>52 (21.1%)</td>
<td>63 (17.4%)</td>
<td>24 (12.2%)</td>
<td>8 (27.6%)</td>
<td>147 (17.6%)</td>
<td>0.11</td>
</tr>
<tr>
<td>Ischemic cardiomyopathy</td>
<td>59 (24.0%)</td>
<td>111 (30.7%)</td>
<td>72 (36.5%)</td>
<td>8 (27.6%)</td>
<td>250 (30.0%)</td>
<td>0.038</td>
</tr>
<tr>
<td>Beta blocker use</td>
<td>228 (92.7%)</td>
<td>345 (95.3%)</td>
<td>182 (92.4%)</td>
<td>26 (89.7%)</td>
<td>781 (93.6%)</td>
<td>0.34</td>
</tr>
<tr>
<td>ACEI or ARB use</td>
<td>214 (87.0%)</td>
<td>304 (84.0%)</td>
<td>157 (79.7%)</td>
<td>16 (55.2%)</td>
<td>691 (82.9%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Spironolactone use</td>
<td>137 (55.7%)</td>
<td>229 (63.3%)</td>
<td>125 (63.5%)</td>
<td>14 (40.3%)</td>
<td>505 (60.6%)</td>
<td>0.11</td>
</tr>
</tbody>
</table>

* Continuous data are displayed as median (Q1-Q3); categorical data are displayed as N (%). Sodium-glucose cotransporter-2 inhibitors were not used. ACEI angiotensin converting enzyme inhibitor; ARB: angiotensin receptor blocker; LVEF: left ventricular ejection fraction; NYHA: New York Heart Association. Missing measurements accounted for 1.7% or less of each variable.
Cardiopulmonary exercise test characteristics

Table 2 illustrates the distribution of CPET parameters by baseline NYHA class. None of the continuous variables were normally distributed. Patients were well distributed across Weber classes, with approximately one third of patients in classes A, B, and C, while only 2% were in class D. Peak VO₂ was significantly lower in patients with higher NYHA class, and median values ranged from 19.1 (NYHA class I) to 13.6 (NYHA class IV) ml/kg/min. ppVO₂ varied from 66.8% (NYHA I) to 48.1% (NYHA IV). Median VE/VCO₂ slope ranged from 36.7 (NYHA I) to 49.8 (NYHA IV), and median OUES ranged from 1.40 (NYHA I) to 1.07 (NYHA IV). Across all variables, there was a statistically significant association between unfavorable CPET parameters and higher NYHA class (p < 0.001).

Prognostic value of NYHA and Weber classes

A total of 64% (535) patients had concordant NYHA and Weber classes (i.e., NYHA I or II with Weber A or B, or NYHA III or IV with Weber C or D). Among those with concordant classes, patients with both adverse classifications had significantly higher mortality (Figure 1). Of the 299 patients with discordant classifications, 208 classifications had significantly higher mortality (Figure 1).

Table 2 – CPET parameters by NYHA class

![Table 2 - CPET parameters by NYHA class](image)

*Continuous data are displayed as median (Q1-Q3); categorical data are displayed as N (%). NYHA: New York Heart Association; OUES: oxygen uptake efficiency slope; VE/VCO₂: minute ventilation/carbon dioxide output slope; VO₂: oxygen consumption. Missing measurements accounted for 0.8% or less of each variable.
Discussion

In a large cohort of patients with HF undergoing CPET, both NYHA classification and peak VO\textsubscript{2} were independent predictors of all-cause mortality. Patients who exhibited higher NYHA classes consistently presented worse results in the CPET. Furthermore, patients with a favorable NYHA class and an adverse Weber class had an intermediate risk of all-cause mortality that was not significantly different from patients with an adverse NYHA class and a favorable Weber class.

Previous studies have analyzed the prognostic importance of the NYHA classification. Muntwyler et al. showed NYHA class to be an independent prognostic factor in multivariable analysis, with 1-year mortality ranging from 7.1% in patients with NYHA II to 28.0% in those with NYHA IV.\textsuperscript{6} NYHA classification also remained a powerful predictor of mortality for at least 10 years.\textsuperscript{7} Several other studies, however, have suggested that NYHA classification might be an unreliable marker of prognosis on an individual level. Caraballo et al. showed significant heterogeneity of mortality risk in NYHA II and III patients across studies, suggesting that the prognostic implication of the NYHA classification is largely dependent on the baseline risk of the patient being assessed.\textsuperscript{13} More recently, Blacher et al. showed significant overlap in several metrics between NYHA I and II patients, suggesting that the NYHA class, by itself, may be an insufficient discriminator of individual patients with mild HF.\textsuperscript{11} The question of whether changes in NYHA class over time can predict prognosis has been studied as well. Greene et al. showed that improvements in NYHA class did not lead to better outcomes in patients with HF, while improvement in Kansas City Cardiomyopathy Questionnaire Overall Summary Score was correlated with improved prognosis.\textsuperscript{23} Rohde et al. suggested that changes in NYHA class over time may have limited predictive value, particularly in mild HF.\textsuperscript{12}

CPET has been increasingly proposed as a way to improve prognostic assessment in patients with HF by offering objective and reproducible metrics.\textsuperscript{10,15} CPET has been used as a tool to aid in cardiac transplantation decision-making for over three decades,\textsuperscript{18} given its reliability in distinguishing risk among patients with severe HF. There have also been calls for the inclusion of CPET as part of the enrollment and endpoint criteria in HF trials as early as 1988.\textsuperscript{15} Many CPET metrics have been shown to have prognostic implications, including peak VO\textsubscript{2}, OUES, VE/VCO\textsubscript{2} slope, resting end-tidal CO\textsubscript{2} pressure, and exercise oscillatory ventilation.\textsuperscript{18,19,26-28}

Most prognostic studies have focused either on NYHA classification or CPET, but rarely on both. This leaves clinicians unsure of how to interpret the information derived from simultaneous NYHA and CPET assessments. This is especially important when they are faced with conflicting information, such as a patient categorized in an advanced NYHA class with CPET showing favorable Weber classification (class A or B). Our study aimed to combine these assessments to refine the prognostic evaluation in patients with HF. In this analysis of a large cohort of patients with HF undergoing CPET, both NYHA classification and peak VO\textsubscript{2} were predictive of all-cause mortality after adjusting for age and sex. Across all NYHA classes, decreases in peak VO\textsubscript{2} were associated with increased mortality; likewise, across the spectrum of peak VO\textsubscript{2}, increments in NYHA classification were also linked to increased mortality. The notable exception was the lack of a significant difference between NYHA classes I and II, in conformity with prior studies.\textsuperscript{9,11,12} This finding is critical because patients classified as NYHA I have been excluded from HF clinical trials based on the assumption that they constitute a uniformly low-risk group, and NYHA I patients are thus ineligible for several life-prolonging therapies that are well established for patients with HF in NYHA class II and above.\textsuperscript{14} Furthermore, we sought to analyze the prognostic value of both classifications when patients had CPET results that apparently conflicted with their physician-assigned NYHA class. We found...
no significant difference in all-cause mortality between patients with discordant classes (NYHA I or II with Weber C or D versus NYHA III or IV with Weber A or B), suggesting that both NYHA and CPET metrics are complementary prognostic variables. Patients with discordant classes displayed an intermediate prognosis compared to those with concordant favorable (NYHA I or II and Weber A or B) or concordant adverse (NYHA III or IV and Weber C or D) classifications. Our findings were consistent in sensitivity analyses using VE/VCO₂ slope and ppVO₂ instead of peak VO₂ to determine conflicting classes. Our results are in conformity with a previous study by Ritt et al. that demonstrated an association between NYHA and Weber classes, albeit with low concordance between them.²⁹

Our study had limitations that merit consideration. First, although ordinarily performed within weeks, a period in which the functional status of a patient with HF is not expected to shift, the exact timing between NYHA determination and the CPET was not recorded. Second, this study was retrospective and included a one-time NYHA assessment, and results cannot be extrapolated to NYHA class variation over time. Third, for the longitudinal analysis, it is unclear how CPET findings were used to guide therapeutic decisions. Furthermore, we did not study the impact of repeated CPETs in this population. Finally, the study time frame spans over a decade, and clinical practice might have shifted over that time.

**Conclusion**

NYHA classification and CPET parameters provide complementary prognostic information that is more accurate than using either alone. CPET may be a valuable tool to discriminate risk in patients with HF across all NYHA classes, particularly for those in NYHA classes I and II.

**Author Contributions**

Conception and design of the research: Engster PHB, Zimerman A, Schaan T, Rohde LE, Silveira AD; Acquisition of data: Engster PHB, Zimerman A, Schaan T, Borges MS, Souza G, Costa GD, Rohde LE, Silveira AD; Analysis and interpretation of the data: Engster PHB, Zimerman A, Schaan T, Borges MS, Souza G, Costa GD, Rohde LE, Silveira AD; Statistical analysis: Engster PHB, Zimerman A, Borges MS, Souza G, Rohde LE, Silveira AD; Obtaining financing: Engster PHB, Zimerman A, Silveira AD; Writing of the manuscript: Engster PHB, Zimerman A, Schaan T, Borges MS, Souza G, Costa GD, Rohde LE, Silveira AD; Critical revision of the manuscript for important intellectual content: Engster PHB, Zimerman A, Rohde LE, Silveira AD.

**Potential conflict of interest**

No potential conflict of interest relevant to this article was reported.

**Sources of funding**

There were no external funding sources for this study.

**Study association**

This study is not associated with any thesis or dissertation work.

**Ethics approval and consent to participate**

This study was approved by the Ethics Committee of the HCPA under the protocol number 2014-0162. All the procedures in this study were in accordance with the 1975 Helsinki Declaration, updated in 2013. Informed consent was obtained from all participants included in the study.

**References**


2. White PD, Myers MM. The Classification of Cardiac Diagnosis. JAMA. 1921;77(8): 1414-5.


*Supplemental Materials
For Supplementary Figure, please click here.