Heart Failure Prognosis: Comorbidities Matter

Raya E. Kheirbek, MD,1,2 Farrokh Alemi, PhD,1,4 and Ross Fletcher, MD1,3

Abstract

Background: Prior risk prediction models have included a selective group of broad comorbidities in scoring prognosis of heart failure (HF) patients.

Objective: We examined whether scoring a comprehensive set of comorbidities separately, could improve the performance and accuracy of predicting HF prognosis.

Methods: This is a cross-validated, longitudinal, retrospective, observational study. Data were collected on 602,050 unique HF patients, who received care through the Veterans Administration (VA) between October 1, 2006 and September 31, 2011. The dependent variable was mortality in six months. The independent variables were all International Classification of Disease (ICD) comorbidities, without grouping into broad disease categories.

Results: The area under the receiver-operating curve (AROC) for the MM index was 0.784 (95% confidence interval [CI]: 0.781–0.786). The MM index was significantly (alpha < 0.05) more accurate than the Quan variant of the Charlson Index (AROC = 0.656), the comorbidity categories within the Care Assessment of Need (CAN) Index (AROC = 0.677), the von Walraven variant of the Elixhauser Index (AROC = 0.639), chronological age (AROC = 0.649), or ejection fraction (EF) (AROC = 0.533).

Conclusion: Inclusion of additional comorbidities improves the accuracy of HF prognostic indices. Future studies are needed to drive HF prognostic indices with the MM index as a component.

Introduction

In the last four decades, several teams of investigators have developed prognostic indices for heart failure (HF) patients (Table 1).1–10 These indices vary considerably in how they score patients’ comorbidities. Some indices have not used comorbidities at all and focused on heart-specific pathophysiology—e.g., ejection fraction (EF), the New York Heart Association Functional Classification of HF, heart rate, etc. These indices report a lower predictive accuracy, as measured by the area under the receiver-operating curve (AROC). For example, the 2006 revised Seattle HF model did not rely on comorbidities and had lower AROC than the original version that did include selective comorbidities.11

Furthermore, other investigators have attempted to include comorbidities in their models but have relied on broad categories of selective diseases. This approach was first developed by Charlson and colleagues,12 and some indices have included the exact Charlson Index (Lee et al.)5 The Charlson Index classifies patients’ diagnoses into 17 disease categories (e.g., moderate to severe liver diseases, any malignancy except skin cancer) and scores each category from one to six points. Unfortunately, it performs poorly in predicting HF patients; it has an AROC of 0.66 in predicting one year mortality.13 Quan and colleagues updated the scoring of the Charlson Index to reflect improvements in management of diseases.14 Elixhauser developed a broader set of diagnostic categories and von Walraven and colleagues transformed this set into a prognostic indicator.15 The relative accuracy of these improved indices in predicting HF prognosis is not known.

The most comprehensive approach for including comorbidities in HF indices can be found in the Care Assessment of Need (CAN) Index. This index includes 23 broad categories of diseases, the pairwise interaction among these broad categories, and the Deyo variant of the Charlson Index.16 It is the only index in the HF prognostication literature that examines interactions among pairs of comorbidities.

To understand how HF prognostic indices score comorbidities, it is illustrative to examine the scoring of the

1Office of Chief of Staff, Veterans Affairs Medical Center, Washington, DC.
2School of Medicine & Health Sciences, George Washington University, Washington, DC.
3School of Medicine, Georgetown University, Washington, DC.
4Department of Health Administration and Policy, George Mason University, Fairfax, Virginia.
Accepted February 24, 2015.
Table 1. Heart Failure Prognostic Indices

<table>
<thead>
<tr>
<th>Index name</th>
<th>Study year</th>
<th>Number of subjects in the study</th>
<th>Number of HF-specific predictors</th>
<th>Number of comorbid predictors</th>
<th>Timeframe</th>
<th>AROC in cross-validated dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Failure Survival Score</td>
<td>1997</td>
<td>199</td>
<td>80</td>
<td>0</td>
<td>1 year</td>
<td>0.76</td>
</tr>
<tr>
<td>Zugck Two-Variable Model</td>
<td>2001</td>
<td>208</td>
<td>2</td>
<td>0</td>
<td>1 year</td>
<td>0.72</td>
</tr>
<tr>
<td>Charlson Index and its variations</td>
<td>2001</td>
<td>56,735</td>
<td>2</td>
<td>15–28</td>
<td>1 year</td>
<td>0.63–0.75</td>
</tr>
<tr>
<td>Bouvy Model</td>
<td>2003</td>
<td>152</td>
<td>20</td>
<td>1</td>
<td>1.5 years</td>
<td>0.85</td>
</tr>
<tr>
<td>Heart Failure Risk Scoring System</td>
<td>2003</td>
<td>4031</td>
<td>20</td>
<td>5</td>
<td>30 days</td>
<td>0.79</td>
</tr>
<tr>
<td>Digitalis Investigation</td>
<td>2004</td>
<td>2145</td>
<td>7</td>
<td>0</td>
<td>1 year</td>
<td>0.76</td>
</tr>
<tr>
<td>Acute De-Compensated Registry</td>
<td>2005</td>
<td>65,275</td>
<td>25</td>
<td>14</td>
<td>Hosp. stay</td>
<td>0.69</td>
</tr>
<tr>
<td>Seattle Heart Failure</td>
<td>2006</td>
<td>11,067</td>
<td>27</td>
<td>4a</td>
<td>1 year</td>
<td>0.73–0.76</td>
</tr>
<tr>
<td>Get with the Guidelines Heart Failure</td>
<td>2010</td>
<td>39,783</td>
<td>4</td>
<td>3</td>
<td>In hospital</td>
<td>0.75</td>
</tr>
<tr>
<td>CAN</td>
<td>2012</td>
<td>198,460</td>
<td>137</td>
<td>23</td>
<td>1 year</td>
<td>0.76</td>
</tr>
</tbody>
</table>

*Includes only HF-specific diagnoses and no laboratory tests. **Includes no diagnoses but laboratory values, age, and gender.

AROC, area under the receiver operating curve; Care Assessment of Need.

Metastatic Cancer category, used in several indices. There are 141 cancers that fit this category. All are scored in the same fashion. Obviously, all metastatic cancers are not equally fatal; and scoring them as if they are could, at least theoretically, reduce accuracy of predictions. In addition to the problem with use of nonhomogeneous broad categories, there are also concerns that selective inclusion of diseases could lead to missing important diseases. For example, metastatic cancers may be included but fatal nonmetastatic cancers, such as brain cancer, may be left out. We hypothesize that an approach that relies on neither broad diagnostic categories nor selective inclusion of diseases would provide more accurate predictions of HF patient’s prognosis.

To test the above hypothesis, we developed the multimorbidity (MM) index which scores all diagnoses separately. The MM index is comprehensive and does not rely on broad disease categories. This paper compares the predictive accuracy of the MM index to the comorbidity component of CAN, the Charlson Index, and the von Walraven version of the Elixhauser Index. These indices are compared in predicting all-cause mortality within six months after discharge from the hospital.

Methods

This is a cross-validated, twofold, longitudinal, retrospective study. Data were collected on 602,050 unique HF cases, who received care through the Veterans Administration (VA) between October 1, 2006 and September 31, 2011. The definition of HF was based on the ICD-9 codes developed by the Healthcare Cost & Utilization Project of the Agency for Healthcare Research and Quality (AHRQ). Note that more than 50% of HF patients were hospitalized for noncardiac comorbidities, e.g., cancer. We divided the data into a training dataset and a validation dataset. The training dataset used 80% of data; the validation dataset used 20% of the data. The dependent variable was defined as all-cause mortality within six months after discharge from the hospital. Dates of death were verified from the Veterans Affairs Beneficiary Identification and Record Locator System File. The VA system contains data reported to the Social Security Administration and those dates have been shown to be 95% complete.\textsuperscript{17,18} All independent variables were calculated from data available prior to or during the index admission. The independent variables were:

- **MM index**: There were 8687 unique diagnoses in the training set. We calculated the likelihood ratio associated with each unique diagnosis within the training set and used these likelihood ratios to score patients in the validation set. The overall prognosis was scored as:

\[
\text{MM score} = \prod L_{\text{Diagnosis}}
\]

where \( L_{\text{Diagnosis}} \) indicates the likelihood ratio associated with the diagnostic code in the training set. It is calculated as the prevalence of the diagnosis among patients who die divided by the prevalence of the diagnosis among patients who lived.

In a minority of cases, when a patient presented with a diagnosis not seen in at least 99 cases in the training set, the likelihood ratio associated with a broader diagnostic category was used to score the patient. A typical International Classification of Disease version 9 diagnosis (ICD9) is a five-digit number consisting of three initial numbers, a period, followed by two additional digits. The first three digits represent a category of disease. Each additional digit after the period represents further refinements. The study sequentially matched the patient’s diagnoses to five-digit, four-digit, or three-digit ICD9 codes until a match was found. In the validation dataset, 98.35% of diagnoses were matched to the five-digit codes, 0.67% to the four-digit codes, and 0.98% to the three-digit codes. The likelihood ratios used in the MM index were provided online (see openonlinecourses.com/LR.xlsx) so that other investigators can test the index in different datasets.

- **Age**: We used age to provide a benchmark for the performance of comorbidity indices. Patients’ age on admission was calculated as the difference of year of birth and date of admission. Patients younger than 35 were grouped together; patients older than 100 were also grouped together.
Ejection Fraction (EF) was used to provide benchmarks for the performance of the comorbidity indices. EF score was available on 263,995 unique patients, 83,479 of whom were patients in the current study. Accuracy of EF was examined in both the larger and the smaller data. (There was no difference in accuracy levels achieved based on the size of the sample; therefore we report the accuracy of EF using the larger sample).

**Variants of the Charlson index:** The Charlson Index (17 broad categories of diseases) was calculated using the enhancements described by Hude et al. (2005) and Quan et al.14

**Variant of Elixhauser list:** We used the von Walraven et al. index developed from Elixhauser’s list of 30 broad categories of comorbidities.15

**The CAN comorbidity index:** This index examines 5429 diagnoses and classifies these diagnoses into 23 broad categories as well as the 17 categories within the Charlson Index. The comorbidity component of the CAN Index was calculated as:

\[
\text{CAN comorbidity score} = 0.104 \times \text{Deyo-Charlson Index} + 0.197 \times \text{MI} + 0.015 \times \text{CABG} - 0.387 \times \text{MI} + 0.030 \times \text{Unstable angina pectoris} + 0.351 \times \text{Respiratory failure} + 0.132 \times \text{Valvular heart diseases} - 0.232 \times \text{Respiratory failure & Valvular heart disease} + 0.144 \times \text{Renal failure} + 0.295 \times \text{Chronic obstructive pulmonary diseases (COPD)} - 0.136 \times \text{Renal failure & COPD} + 0.007 \times \text{Hospitalized in prior year} + 0.009 \times \text{Hospitalized & Valvular heart diseases} + 0.246 \times \text{Stroke} - 0.088 \times \text{Hospitalized & Stroke} - 0.190 \times \text{Renal failure & Stroke} + 0.155 \times \text{Function diseases} - 0.177 \times \text{COPD & Dementia} + 0.112 \times \text{Pneumonia} + 0.110 \times \text{Diabetes} + 0.136 \times \text{Malnutrition} + 0.088 \times \text{Peripheral vascular diseases} + 0.924 \times \text{Metastatic cancer} - 0.027 \times \text{Trauma} + 0.020 \times \text{Hospitalized & Trauma} + 0.124 \times \text{Psychotic disorders} + 0.336 \times \text{Liver diseases} + 0.027 \times \text{Trauma} + 0.020 \times \text{Hospitalized & Trauma} + 0.124 \times \text{Psychotic disorders} + 0.336 \times \text{Liver diseases} + 0.027 \times \text{Trauma} + 0.020 \times \text{Hospitalized & Trauma} + 0.124 \times \text{Psychotic disorders} + 0.336 \times \text{Liver diseases} + 0.027 \times \text{Trauma} + 0.020 \times \text{Hospitalized & Trauma} + 0.124 \times \text{Psychotic disorders} + 0.336 \times \text{Liver diseases}.
\]

The accuracy of predictions was calculated using AROC, or c-statistic.19 We used the pROC function in R to calculate 95% confidence interval (CI) for the AROC. This function calculated AROC using 2000 stratified bootstrap replicates.20

This work was approved by the Washington DC VA Medical Center (protocol 01565).

**Results**

The patients examined were veterans with HF, 74% white, 98% male, 56% married, 20% divorced, and 15% widowed (see Table 2). The training and validation data were selected randomly. Because the sample size was large, small differences were statistically significant. If we require a minimum effect size at 1%, these data show that the training and validation samples were similar.

The homogeneity of broad disease categories can be examined by contrasting the odds ratio (OR) and the likelihood ratio of mortality for each diagnosis within these categories (Table 3). Listed categories contained between 2 and 29 different diagnoses. Table 3 also lists the range of OR

**Table 2. Demographics of Patients with Heart Failure**

<table>
<thead>
<tr>
<th>Category</th>
<th>Training data</th>
<th>Validation data</th>
<th>Effect size</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of patients</td>
<td>479,254</td>
<td>107,056</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average age (standard deviation)</td>
<td>77 (10.5)</td>
<td>78 (10.38)</td>
<td>1 year</td>
<td>–28.44&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Percentage male</td>
<td>97.87%</td>
<td>97.85%</td>
<td>0.02%</td>
<td>0.378</td>
</tr>
<tr>
<td>Percentage married</td>
<td>55.84%</td>
<td>56.72%</td>
<td>–0.88%</td>
<td>–5.233&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Percentage divorced</td>
<td>20.27%</td>
<td>19.21%</td>
<td>1.06%</td>
<td>7.926&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Percentage widowed</td>
<td>14.62%</td>
<td>14.98%</td>
<td>–0.36%</td>
<td>–2.963&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Percentage never married</td>
<td>6.30%</td>
<td>6.24%</td>
<td>0.06%</td>
<td>0.754</td>
</tr>
<tr>
<td>Percentage white</td>
<td>73.95%</td>
<td>74.03%</td>
<td>–0.08%</td>
<td>–0.507</td>
</tr>
<tr>
<td>Percentage African American</td>
<td>13.28%</td>
<td>13.07%</td>
<td>0.21%</td>
<td>1.849&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Percentage Asian</td>
<td>1.10%</td>
<td>1.12%</td>
<td>–0.01%</td>
<td>–0.406</td>
</tr>
<tr>
<td>Percentage Native American</td>
<td>0.44%</td>
<td>0.44%</td>
<td>0.00%</td>
<td>0.005</td>
</tr>
</tbody>
</table>

<sup>a</sup>Statistically significant at alpha levels <0.05.

**Table 3. Selected Comorbidities and Impact on Six-month Mortality**

<table>
<thead>
<tr>
<th>Category (number of diagnoses in category)</th>
<th>Admissions (percent of total)</th>
<th>Likelihood ratio (95% confidence interval)</th>
<th>Odds ratio for diagnoses in the category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancers of prostate (2)</td>
<td>29,458 (3.97%)</td>
<td>1.46 (1.45–1.53)</td>
<td>1.03</td>
</tr>
<tr>
<td>Secondary malignancies (28)</td>
<td>19,231 (2.59%)</td>
<td>8.98 (9.56–10.15)</td>
<td>3.24</td>
</tr>
<tr>
<td>Lung cancer &amp; cancer of bronchus (7)</td>
<td>16,807 (2.27%)</td>
<td>4.05 (4.12–4.39)</td>
<td>1.39</td>
</tr>
<tr>
<td>Neoplasms of unspecified nature (29)</td>
<td>10,062 (1.36%)</td>
<td>2.30 (2.23–2.43)</td>
<td>0.53</td>
</tr>
<tr>
<td>Cancer of colon (10)</td>
<td>9,131 (1.23%)</td>
<td>1.34 (1.28–1.42)</td>
<td>0.83</td>
</tr>
<tr>
<td>Leukemia (15)</td>
<td>6,184 (0.83%)</td>
<td>2.58 (2.47–2.75)</td>
<td>0.57</td>
</tr>
<tr>
<td>Non-Hodgkin’s lymphomas (26)</td>
<td>5,909 (0.8%)</td>
<td>2.38 (2.18–2.43)</td>
<td>0.31</td>
</tr>
<tr>
<td>Head &amp; neck cancer (22)</td>
<td>5,339 (0.72%)</td>
<td>1.96 (1.86–2.1)</td>
<td>0.28</td>
</tr>
<tr>
<td>Cancers of kidney &amp; renal pelvis (4)</td>
<td>4,385 (0.59%)</td>
<td>1.51 (1.42–1.63)</td>
<td>0.73</td>
</tr>
</tbody>
</table>
observed for diagnoses within the categories. In many instances, the OR of the diagnoses ranged from less than one to more than one, indicating that these categories contain diagnoses that both reduce and increases odds of mortality. For example, the anemia category had an OR of 1.62 (95% CI: 1.59–1.64). This category included 25 diagnoses, some of which reduced the odds of mortality by 1/0.58 = 1.72-fold and others which increased the odds of mortality by 4.84-fold. Clearly, including these diagnoses into the same category distorts the impact of the category on mortality.

It is illustrative to think through the OR associated with the “secondary malignancies.” The category itself had an OR of 9.85, with 95% CI: 9.56–10.15. This category included 28 diagnoses, including secondary malignant neoplasm of brain and spinal cord with OR of 17.28 and secondary neuroendocrine tumor of distant lymph nodes with OR of 2.43. There was a nearly nine fold difference in the OR associated with diagnoses in the category of secondary malignancies. Some diagnoses within this category were more fatal than other diagnoses. These data show that grouping many different diagnoses into a common category may reduce accuracy. A strategy of scoring each diagnosis separately, if possible, may be more accurate.

The MM index was developed to score each diagnosis separately. Figure 1 shows the distribution of the MM scores (converted from odds to probability) and its relationship to observed rates of mortality. These data show that the MM scores were distributed evenly across most of the range, except at start and at end of the scale. In addition, the figure shows that the MM score had a monotone relationship with the observed rate of mortality; increases in MM probability scores were associated with increases in observed rate of mortality.

Figure 2 shows overall performance of the MM index using AROC. The AROC for the MM index was 0.784 (95% CI: 0.781–0.786). At the cutoff for doubling of the odds of
Discussion

Our findings support that including a comprehensive list of comorbidities and avoidance of broad disease categories improves the performance and accuracy of prediction. Broad disease categories, of the type used in Charlson, Elixhauser, or CAN, were not homogeneous. These broad categories included diagnoses that differed significantly in their association with mortality. These data suggest that relying on these broad categories could reduce the overall accuracy of predictions. In order to test this hypothesis, we constructed the MM index. The MM index separately scored each diagnosis and therefore did not rely on broad disease categories. The MM index expanded the number of diagnoses scored from 17 broad categories in the Charlson Index to 8704 separate diagnoses. The MM index outperformed Charlson, the Quan variant of the Charlson Index, the von Walraven variant of the Elixhauser Index, and the comorbidity component of the CAN Index. To the best of our knowledge, the current study is the first to highlight the impact of including additional comorbidities on improving the performance and accuracy of HF prognostic indices.

In most patients, and particularly in elderly patients, HF is associated with a spectrum of comorbidities that play an important role in trajectory of illness and response to treatment. Given the tremendous heterogeneity of that age group, developing an index that incorporates comorbid conditions is more accurate than age alone. Most of the previous indices examined only a limited number of comorbidities in each study, and the specific comorbidities evaluated varied across studies. Understanding the role of comorbid conditions may result in improved shared decision making as well as improved health outcomes.

A comprehensive approach to comorbidity that scores diagnoses separately may improve statistical predictions but has limited clinical use unless it is automated as a clinical decision support tool that is embedded in the electronic health record. It can function in the background to regularly generate predictions using existing patients’ records. Clinicians and patients can then use the automated scores available through the electronic health record to guide disease modifying treatment and shared decision making for patients with serious and advanced chronic illness and multiple comorbid conditions.

This study has several limitations inherent to retrospective observational studies. VA population is predominantly men with many comorbidities. Study findings may not be readily generalizable either to women or to other nonveteran populations.

This study has not demonstrated how the MM index and heart-specific variables could be combined. It is possible that when the MM index and HF-specific variables are used together, then the accuracy level improves further. However, our findings support that clinicians should pay careful attention to comorbid conditions affecting heart failure patients in order to improve outcomes of care and quality of life.

Author Disclosure Statement

Financial support for this study was provided by the Washington DC VA Medical Center. The sponsor had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; or the decision to submit the manuscript for publication.

All authors are employed by the sponsor. This paper reflects the views of the authors and does not represent the views of the Department of Veterans Affairs or the U.S. government. None of the authors have a conflict of interest with publication of this manuscript. All authors meet the criteria for authorship stated in the Uniform Requirements for Manuscripts Submitted to Biomedical Journals.

References


Address correspondence to:
Raya E. Kheirbek, MD
Washington DC VA Medical Center
50 Irving Street, NW
Washington, DC 20422
E-mail: Raya.elfadelkheirbek@va.gov