Physicians are becoming enamored of risk prediction models. Cox proportional hazards, logistic regression, and classification and regression tree methods are being widely used to create a wide variety of prediction models. We should be cognizant of the following quote:

“Prediction is very difficult, especially about the future.”

—Niels Bohr (1)

It is easy to develop a risk model to stratify (discriminate) patients who did or did not have an event within a cohort, where the outcomes are already known. It is far more difficult to validate the risk model in another cohort (validation) and, more importantly, to apply the model to assist in making decisions in individual patients. Importantly, validation of a model requires measures of both discrimination and calibration in another cohort.

The perils can be illustrated with coronary artery disease risk models. For example, the Framingham Risk Score when applied to Asian cohorts provided similar discrimination as in the derivation white U.S. cohort (receiver-operating characteristic [ROC] ≥0.75) (2). However, the model overestimated the absolute risk by 276% in men and 102% in women. Thus, without recalibration, the model would have limited utility for treatment decisions in the Asian population. Similar concerns have been raised about the new American College of Cardiology/American Heart Association (ACC/AHA) risk model (3), which has overestimated coronary artery disease risk by 75% to 150% in some cohorts (4).

Do the same concerns apply to heart failure models as well? Natriuretic peptides (NP) are strong predictors of outcomes in patients with heart failure with reduced ejection fraction (HFrEF) and heart failure with preserved ejection fraction (HFpEF) (5,6). Data from patients with HFrEF (Val-HeFT [Valsartan Heart Failure Trial]) and HFpEF (I-PRESERVE [Irbesartan in Heart Failure With Preserved Ejection Fraction Study]) trials show that despite significantly higher baseline levels of NP in HFrEF, the hazard for mortality associated with 1 log unit increase in N-terminal pro-B-type natriuretic peptide (NT-proBNP) is similar in both HFrEF and HFpEF populations (hazard ratio: ~1.70). It would, therefore, appear reasonable to assume that a heart failure risk prediction model NP could be used in a wide variety of heart failure patients to predict outcomes. However, as can be seen in Table 1 (Figure 1), for any given BNP level, the absolute risk of death, adjusted for baseline NT-proBNP, is ~60% higher for HFrEF than for HFpEF patients.

Thus, a model or variable can provide similar discrimination in disparate cohorts, but the calibration of the model may be less than ideal, especially if the cohort is dissimilar from the derivation cohort as has been seen with coronary artery disease risk models and as illustrated in Table 1 with NT-proBNP. Therefore, all heart failure risk prediction models should be validated prior to use as is recommended by the 2013 ACC/AHA heart failure guidelines (7).

In this issue of the JACC: Heart Failure, Ouwerkerk et al. (8) compare the power of various baseline
variables used in models from several studies to predict mortality and/or heart failure hospitalizations in patients with heart failure. This is an enormous undertaking where the investigators analyzed 117 models with 249 different variables (8). Based on the average C-statistic, mortality was easier to predict (0.71) than the combined endpoint of mortality and heart failure hospitalization (0.63). Not surprisingly, the C-statistic was modestly lower in validation cohorts (~0.01) and in randomized clinical trials (~0.03), where the cohorts are much more homogeneous than in observational cohorts. The investigators found that blood urea nitrogen and serum sodium were the strongest individual predictors. For each variable added to a model, the C-statistic increased by 0.0036. Unfortunately, the information necessary for the investigators to perform a true meta-analysis was missing from many publications. For example, with the Seattle Heart Failure Model in ~7,000 patients, the largest area under ROC is daily diuretic dose and New York Heart Association functional class (0.66) compared with creatinine (0.60) and serum sodium (0.56).

Some risk prediction models like the recently published MAGGIC (Meta-analysis Global Group in Chronic Heart Failure) model (9) that was derived in 39,372 patients from 30 heart failure studies do not report a ROC or C-statistic, making those models difficult to compare with other heart failure risk prediction models.

Are there better methods to compare risk models? One solution may be to require an age and sex baseline risk model and compare the incremental value of the final model beyond age and sex. In the Framingham cohort, the C-statistic for all-cause mortality when only age and sex alone were in the model was 0.75 (10). The addition of conventional risk factors increased the C-statistic to 0.80. The incremental value of the conventional risk factors beyond age and sex is +0.05 or an increase in model predictive power of ~20% (0.05/[0.75 – 0.50]=20%). Thus, age and sex alone in Framingham population was able to predict the risk of death (C-statistic: 0.75) as well or better than the average heart failure risk model (C-statistic: 0.71). Why is it so much harder to predict mortality in heart failure patients than in a general population? The major reason is that in heart failure populations a model with only age and sex has a much lower C-statistic. Once a patient has New York Heart Association functional class IV symptoms with EF <25%, the risk of death is much higher irrespective of whether the patient is 40 or 80 years old. In the Seattle Heart Failure Model cohort of ~7,000 patients the model with only age and sex had a ROC of 0.60. The addition of the other Seattle Heart Failure Model variables increased the ROC to ~0.72 or a 120% increase in the model’s predictive power (11).

The addition of exercise variables (exercise duration, peak oxygen consumption, ventilator efficiency) or biomarkers to an existing heart failure model is not simple. It is quite likely that the added variable will alter the hazard ratio (beta coefficients) of other variables in the model. For example, the addition of creatinine clearance is likely to alter the hazard ratio for age and sex as these are used in calculation of creatinine clearance. An example of the appropriate methodology for adding single or multiple variables to a model is shown in the derivation of the Barcelona

See Page 429

---

**TABLE 1** NT-proBNP in HFpEF and HFrEF

<table>
<thead>
<tr>
<th>Trial</th>
<th>n</th>
<th>NT-proBNP Median</th>
<th>Annual Mortality</th>
<th>HR</th>
<th>95% CI</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val-HeFT per log unit</td>
<td>4,066</td>
<td>901</td>
<td>9.0%</td>
<td>1.68</td>
<td>1.58-1.79</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>I-PRESERVE per log unit</td>
<td>3,479</td>
<td>339</td>
<td>4.0%</td>
<td>1.71</td>
<td>1.62-1.81</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

HFrEF = heart failure with reduced ejection fraction; HFpEF = heart failure with preserved ejection fraction; I-PRESERVE = Irbesartan in Heart Failure With Preserved Ejection Fraction Study; NT-proBNP = N-terminal pro-B-type natriuretic peptide; Val-HeFT =Valsartan Heart Failure Trial.

**FIGURE 1** A Cox Model was Derived Within I-PRESERVE (HFpEF) and Val-HeFT (HFrEF) With NT-proBNP

The model was reversed to derive a heart failure risk prediction model for 1-year all-cause mortality. The hazard ratio of the log NT-proBNP was similar in both heart failure with reduced ejection fraction (HFrEF) and heart failure with preserved ejection fraction (HFpEF), but the absolute risk of death for a given NT-proBNP level, after adjustment for the N-terminal pro-B-type natriuretic peptide (NT-proBNP) levels, was 59% higher in HFrEF than HFpEF. Data courtesy of Inder Anand. I-PRESERVE = Irbesartan in Heart Failure With Preserved Ejection Fraction Study; Val-HeFT = Valsartan Heart Failure Trial.
Bio-Heart Failure Risk Model (12), where all variables were derived with the combination of the 3 biomarkers (NT-proBNP, high-sensitivity troponin T, and ST2) alone and in various combinations (12). The model contains many of the Seattle Heart Failure Model variables including heart failure medications, diuretic doses, and so on (11,12) but allows the user to add the biomarkers to alter the estimated mortality. However, none of the biomarkers were individually statistically significant (all p > 0.05) and the incremental value of adding all 3 biomarkers was a change in C-statistic of approximately 0.02 or 0.007/biomarker. These results are similar to the incremental value of adding NT-proBNP and ST2 to the Seattle Heart Failure Model (C-statistic change of 0.02) (13) or the addition of 10 biomarkers to the Framingham cohort (C-statistic change of 0.02) (10).

How can we improve the reporting of heart failure risk models? First, we suggest investigators use an incremental approach starting with the C-statistic for an age and sex model along with addition of incremental approach starting with the C-statistic. Validation requires measures of discrimination (ROC or C-statistic) along with a measure of calibration (e.g., Hosmer-Lemeshow) or mean predicted versus observed mortality/survival. Third, the incremental value of adding variables to a heart failure should report the change in the C-statistic (ROC), which are very difficult changes, and the total chi-square of the model (i.e., a model chi-square increase from 100 to 105 would increase the predictive power of the model by 5%) (14). Newer measures such as integrated discrimination improvement, net reclassification improvement, and so on, are also useful (15). Fourth, online calculators should be made available for multivariable models, wherever possible, to allow for easy provider use.

Validated heart failure risk models are now endorsed by the ACC/AHA guidelines (7). They are likely here to stay and need to be easy to use. Incorporation of the models into electronic health records will enhance their utility for healthcare providers.

REPRINT REQUESTS AND CORRESPONDENCE: Dr. Wayne C. Levy, University of Washington, Division of Cardiology, Box 356422, 1959 NE Pacific Street, Seattle, Washington 98195. E-mail: levywc@uw.edu.

REFERENCES


KEY WORDS heart failure, mortality, prognosis, risk prediction