Despite the remarkable therapeutic advances in the treatment of chronic heart failure, it remains a condition marked by progressive deterioration and premature mortality. Our treatments slow the rate of descent, and some reset the survival curve upward, but decline is inevitable. There is, however, great heterogeneity in the journeys traveled by individual heart failure patients. Heart failure is the final common pathway for a multitude of cardiac insults; individuals first enter the heart failure stream at different severity levels with widely varied clinical characteristics. For some patients, the onset of heart failure is readily defined, whereas for others, the onset is much more insidious and its recognition delayed. The course of illness may be influenced by psychological well-being, environmental factors, and genetic factors with variable expression and penetrance, which may be causative and impact the natural history of illness or the response to pharmacotherapy.1–5 Despite our enhanced understanding of factors that influence plaque rupture, we are far from being able to accurately determine if or when a patient will experience one, whether a myocardial infarction will result, and how large it will be if it does.

Article see p 63

Therefore, prognostication in heart failure is probabilistic, not deterministic, and the upper limit on the accuracy of any model that attempts to predict mortality or morbidity and mortality in all but the most agonal heart failure population will be constrained by the magnitude of this uncertainty. The C-index, a measure of model discrimination that varies from 0.5 (no better than a coin flip) to 1.0 (perfect discrimination), is rarely above 0.8 for published heart failure survival models. It is reasonable to posit that future models are unlikely to perform substantially better. As such, what is the value of these models?

The value of modeling is that of additional knowledge gain about heart failure prognosis and disease trajectory. Such knowledge is useful at both the individual and the group level, especially in settings of medical uncertainty. At the group level, the limitations imposed by the inaccuracy of individual outcome predictions are mitigated, because the goal is usually a measure of central tendency and the confidence intervals around it. Such is the case for a predictive model for 30-day heart failure readmission derived from Medicare claims data6 that will provide risk adjustment for the heart failure component of the Readmission Payment Reduction Program, under which hospitals stand to lose up to 3% of payments if their 30-day readmission rates for heart failure, myocardial infarction, and pneumonia fall above a risk-adjusted threshold. The C-index for the heart failure readmission model is only 0.60. Although this is of modest value for the purpose of predicting individual risk, it may be acceptable for assessing institutional performance, if we assume that misclassification errors are evenly spread across hospitals.

Prognostic information may be very useful in the design of clinical trials. Estimation of event rates by means of prognostic modeling can increase the efficiency of clinical trials by providing more precise sample size estimates. For example, we are using the Seattle Heart Failure Model (SHFM) to identify ambulatory New York Heart Association class III patients at high mortality risk with standard heart failure therapy as candidates for REVIVE-IT (Randomized Evaluation of VAD InterVEntion before Inotropic Therapy), a soon-to-begin randomized clinical trial comparing a strategy of “early” left ventricular assist device therapy to optimal medical management in such patients (http://clinicaltrials.gov/ct2/show/NCT01369407). Prognostic models can also be applied to baseline clinical characteristics of patients in intervention studies to estimate outcomes that would have occurred absent the intervention.7 8

Knowledge of prognostic model information may favorably influence physician prescribing behavior.9 Whether it similarly influences medication adherence among heart failure patients is unknown but merits investigation. Because studies have shown modest but inconsistent effects of telemanagement strategies on reducing heart failure readmissions,10 a study evaluating the effectiveness and cost-effectiveness of selectively applying telemanagement strategies on the basis of the likelihood of death and readmission (versus nontargeted application) should also be pursued.

Most heart failure predictive models were designed for the express purpose of improving outcome prediction in individual patients. Many, but not all, heart failure patients are interested in their prognosis, but they tend to estimate their own prognosis poorly.11 Almost 30 years ago, the President’s Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research proposed that medical decision making for individual patients should be a
process shared between the physician and the patient. To do so ethically, the patient must make an informed decision. Prognostic knowledge, when relevant to the medical decision, is usually considered critical to this process, although individual patient preferences for information and participation, which may reflect different cultural values, should be solicited and respected.

Studies suggest that physicians are most reluctant to share information about disease with patients in conditions of substantial uncertainty, even though it is in these situations that patients most wish to introduce their own values into the decision-making process. Therefore, improving the quality of prognostic information may help physicians provide patients with the information they need to participate most fully in medical decision making.

Thus, while acknowledging their limitations for individual risk assessment, patients need prognostic information, and the information provided by contemporary heart failure prognostic models is the best we can make available to them. These models perform markedly better than do standard clinical assessment tools, such as New York Heart Association class. Risk prediction based on a single variable does not make efficient use of routinely obtained clinical measures of known prognostic significance. Multivariable risk models can incorporate a range of prognostic information, often reflecting different pathophysiological aspects and phenotypic characteristics of the clinical condition, to improve prognostic accuracy.

In this issue of Circulation: Heart Failure, O’Connor and colleagues present a family of multivariable risk models for the prediction of death and hospitalization (the primary end point) and death alone in patients with chronic systolic heart failure (approximately two thirds in New York Heart Association class II and approximately one third in class III). The models were derived from data collected on 2331 well-compensated outpatients enrolled in HF-ACTION (Heart Failure: A Controlled Trial Investigating Outcomes of Exercise Training), a multicenter randomized controlled trial that evaluated the safety and efficacy of exercise training in this population. The 48 candidate variables represented a broad range of baseline characteristics, including demographics, medical history, laboratory values, exercise parameters from maximal treadmill cardiopulmonary exercise testing, and measures of quality of life and depression. Using a backward-selection method, they first built “full” models for the 2 end points. Next, variables were eliminated to generate “simplified” models, which contained a parsimonious set of variables that continued to provide good discrimination. Risk scores were then derived from the simplified model coefficients.

Discrimination was only modest for the primary end point (optimism-corrected C-index of 0.63 both for the simplified model and the risk score) but moderately good for the death-alone end point (optimism-corrected C-index 0.73 for the simplified model and 0.70 for the risk score). That the primary end-point model performed less well than the mortality model nicely demonstrates the more subjective nature of the decision to hospitalize a patient for heart failure. Other studies have shown that heart failure hospitalization is a strong predictor of subsequent hospitalization; although the threshold to hospitalize differs among physicians, individual physicians are likely fairly consistent in their own thresholds. Calibration of the model was not formally assessed but appears to be reasonably good, at least for the higher-risk deciles of risk.

There are a number of unique strengths to these models. As we have come to expect from this analysis team, there was a high level of statistical rigor used in model development. There was a relatively small amount of missing data on the candidate variables for a trial as large as HF-ACTION (<5%), but it is fair to assume that this would have been spread across a substantially larger number of patients. Missing data pose a substantial challenge in model building, because the statistical routines for regression analyses can only analyze complete data sets. If, for example, the 5% missing data were spread across 25% of the patients, a model that evaluated all 48 variables in a backward-selection algorithm would first eliminate all of the data from 25% of the patients before beginning the analysis. Along with the reduction in statistical power, restricting the analysis to 75% of the sample could introduce substantial bias if the missing data were not randomly missing (ie, if patients with missing data differed in important ways from those with complete data). Single imputation of mean values is the more common approach to the problem but does not reflect the uncertainty about the prediction of the missing value, nor does it utilize unique information about that patient from nonmissing values. The multiple-imputation method used here took the nonmissing values from all other patients (for all 48 clinical characteristics) and nonmissing values for the remaining 47 variables for the “index” patient to replace the missing value with a set of plausible values. This process was iteratively applied to all missing values to create multiple complete data sets for further analysis.

Because the investigators did not have an independent data set to externally validate their models, they used a bootstrapping method for internal validation. The relatively small differences between the C-indexes and the optimism-corrected C-indexes are good (although the authors caution that the optimism corrections do not incorporate the variability caused by the variable selection process used in developing the models), but internal validation is a relatively weak test of model performance. A model will generally perform best in the data set from which it was derived. The statistical programs used to develop the models try to make sense of data sets that contain both useful information and “noise” (measurement error, nonrandom variation, etc). Random sampling of the population of interest is assumed, but in fact, the samples used to develop these models are generally anything but random (clinical trial participants, patients presenting to an advanced heart failure group of a particular health system, etc). Measurement techniques, variable definitions, and patterns of care may differ. Therefore, enthusiasm for any new model must be tempered until it has stood multiple tests of external validation.

At this point, the SHFM remains the most thoroughly validated model for heart failure mortality prognostication, having now been validated externally in more than 25 000 patients in samples derived from 8 large clinical trials, 4
multicenter registries, and numerous single-center observational studies. Whether the ACTION-HF survival and survival/hospitalization models or risk scores will perform as well awaits similar investigations. The HF-ACTION investigators point out that the SHFM performed poorly in 1 small, single-center sample of patients with advanced heart failure undergoing evaluation for left ventricular assist device or heart transplantation. In a small sample from a single center, referral patterns, physician behaviors, and chance could all result in anomalous model performance.

As noted by O’Connor and colleagues, exercise duration during baseline cardiopulmonary exercise testing (with a modified Naughton protocol) is known to be strongly associated with peak VO2, and, in 1 study from the Cleveland Clinic but not another, it predicted mortality or urgent transplantation very nearly as well as peak VO2. It is surprising, however, that the VE/VCO2 slope, which repeatedly has been found to be a stronger mortality predictor than peak VO2, was not so in this fairly large sample. Although it is simpler, cheaper, and easier to perform a treadmill test without respiratory gas measurement, that is not what was done in either HF-ACTION or in the Cleveland Clinic study cited, and it is not clear that the results would be the same. Unencumbered by the breathing apparatus, some patients may exercise longer on the treadmill; as the authors alluded, without seeing respiratory gas information, exercise physiologists may not encourage as much time on the treadmill. Although there is no question that exercise testing provides a wealth of important prognostic information, the requirement for exercise data makes the HF-ACTION model less accessible than the SHFM, which does not.

The inclusion of the Kansas City Cardiomyopathy Questionnaire Symptom Stability Score is the most unique aspect of this risk model and, to us, the most intriguing. Responses to the simple question, “Compared with 2 weeks ago, have your symptoms of heart failure changed?” had substantial prognostic value for the primary end point but not for the mortality-alone end point. Although observational studies can only identify associations and not causation, the notion that a patient’s perception of worsening symptoms is a major driver of heart failure hospitalization is intuitively attractive. Just as pain thresholds differ among patients, so may heart failure symptom tolerance thresholds.

The impact of female sex on outcomes has not been consistent across studies, so it is a bit surprising that the HF-ACTION investigators chose to include this in their models and risk scores. The creation of a successful predictive model requires decisions about which candidate variables to include and how they contribute to the model. Although in this sample, sex substantially improved the predictive ability of the models, the inconsistent impact it has had in other analyses would argue for accepting a bit poorer discrimination in the internal validation for the likelihood of greater preservation of both models’ discrimination when reexamined in other cohorts.

The simplified mortality prediction model truncates body mass index at 25 kg/m2, which suggests that in the HF-ACTION data set, mortality decreased as body mass index rose to 25 kg/m2 but then was constant. An obesity paradox, in which obese subjects (body mass index ≥30 kg/m2) have similar or lower mortality than normal-weight individuals (body mass index 18.5 to <25 kg/m2), has been observed in heart failure patients. Was the obesity paradox absent in this data set?

The authors chose to include only patient-level information in their models. Because physicians’ actions are not random but rather are guided by their assessment of patient needs, physicians’ decisions, such as medications prescribed, can be powerful predictors of outcome. Loop diuretic dosing is a component of the SHFM and has a major impact on its mortality predictions; however, physicians differ in their recognition of excess volume in heart failure patients and in the aggressiveness with which they treat it, so identical patients might have different SHFM risk assessments based on decisions that are in part independent of their severity of illness. In choosing to limit their candidate variables to patient-level characteristics, the authors potentially enhance their models’ generalizability across a wider range of physicians whose practice styles differ. However, generalizability may have been reduced by studying clinical trial patients with a high prevalence of evidence-based therapy use, a group likely to have better outcomes than unselected heart failure outpatients.

There have been many heart failure risk stratification tools developed, each differing in the type of sample from which it was derived and validated, the variables used for risk stratification, their utility in predicting mortality at varying time points, and their ease of use. Risk prediction tools are invaluable for determining heart failure prognosis. They can be useful in helping clinicians, patients, and families make informed decisions in the setting of end-of-life discussions and to help guide the implementation of further medical or surgical interventions. However, risk prediction tools must be selected carefully, matching the clinical characteristics of the patient of interest with those of the sample from which the tool was derived. Validated risk prediction tools should be used. No tool can encompass all of the relevant information crucial for informed decision making. Therefore, these tools should not be used in isolation but rather should be used to enhance clinical decision making. Because heart failure is a dynamic condition with high morbidity and mortality, HF prognosis should be reassessed frequently, particularly in patients for whom critical treatment decisions may hinge on the results.

Disclosures
Dr Aaronson’s activities as national principal investigator for REVIVE-IT are supported by funds contracted to the University of Michigan by the National Heart, Lung, and Blood Institute (contract number HHSN268201100026C) and HeartWare, Inc. The University of Michigan Medical School Conflict of Interest Board monitors Dr Aaronson’s relationship with HeartWare. Dr Cowger’s activities as a site principal investigator for REVIVE-IT are supported by funds contracted to the University of Michigan by HeartWare, Inc.

References


*Key Words:* Editorials ■ heart failure ■ systolic dysfunction ■ exercise capacity ■ risk model ■ risk score
Heart Failure Prognostic Models: Why Bother?
Keith D. Aaronson and Jennifer Cowger

Circ Heart Fail. 2012;5:6-9
doi: 10.1161/CIRCHEARTFAILURE.111.965848

Circulation: Heart Failure is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-3289. Online ISSN: 1941-3297

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circheartfailure.ahajournals.org/content/5/1/6

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Heart Failure can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Heart Failure is online at:
http://circheartfailure.ahajournals.org//subscriptions/