



# HHS Public Access

Author manuscript

*Circ Heart Fail.* Author manuscript; available in PMC 2017 September 15.

Published in final edited form as:

*Circ Heart Fail.* 2012 January ; 5(1): 6–9. doi:10.1161/CIRCHEARTFAILURE.111.965848.

## Heart Failure Prognostic Models: Why Bother?

Keith Aaronson, MD, MS and Jennifer Cowger, MD, MS

Division of Cardiovascular Medicine, University of Michigan

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 travelled 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 while 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, impact the natural history of illness or the response to pharmacotherapy.<sup>1–5</sup> 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.

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's 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, since 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 data<sup>6</sup> that will provide the risk adjustment for the heart failure component of Readmission Payment Reduction Program, under which hospitals stand to lose up to 3% of payments if

---

**Correspondence to:** Keith Aaronson, MD, MS, University of Michigan Medical Center, 1500 East Medical Center Drive, Cardiovascular Center, Room 2169, SPC 5853, Ann Arbor, MI 48105-5853, keith@umich.edu, Telephone: 734-764-7440, Fax: 734-615 3326.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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. While this is of modest value for the purpose of predicting *individual* risk, it may be acceptable for assessing institutional performance, assuming 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 clinical trials efficiency 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, a soon to begin randomized clinical trial comparing a strategy of “early” left ventricular assist device therapy to optimal medical management in such patients.<sup>7</sup> 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.<sup>8, 9</sup>

Knowledge of prognostic model information may favorably influence physician prescribing behavior.<sup>10</sup> Whether it similarly influences medication adherence among heart failure patients is unknown but merits investigation. As studies have shown modest but inconsistent effects of telemanagement strategies on reducing heart failure readmissions,<sup>11</sup> a study evaluating the effectiveness and cost-effectiveness of selectively applying telemanagement strategies based on the likelihood of death and readmission (vs. non-targeted 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.<sup>12</sup> 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.<sup>13</sup> 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.<sup>14</sup>

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.<sup>15</sup> 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 NYHA. class.<sup>16, 17</sup> 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 pathophysiologic aspects and phenotypic characteristics of the clinical condition, to improve prognostic accuracy.<sup>18</sup>

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 endpoint) and death alone, in patients with chronic systolic heart failure ( $\approx 2/3$  NYHA class II and  $\approx 1/3$  class III).<sup>19</sup> The models were derived from data collected on 2331 well compensated outpatients enrolled in the HF-ACTION trial, a multicenter randomized controlled trial that evaluated the safety and efficacy of exercise training in this population.<sup>20</sup> 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 build "full" models for the two endpoints. Next, variables were eliminated to generate "simplified" models, containing 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 endpoint (optimism corrected C-index of 0.63 both for the simplified model and the risk score) but moderately good for the death alone endpoint (optimism corrected C-index 0.73 for the simplified model and 0.70 for the risk score). That the primary endpoint model performed less than the mortality model nicely demonstrates the more subjective nature of the decision to hospitalize a patient for heart failure. Other studies show that heart failure hospitalization is a strong predictor of subsequent hospitalization; while 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 employed 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's fair to assume that this would have been spread across a substantially larger number of patients. Missing data poses a substantial challenge in model building as 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 evaluating 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 are not random (i.e., 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 this 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.

As 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 due to 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 containing 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 Seattle Heart Failure Model (SHFM)<sup>17</sup> remains the most thoroughly validated model for heart failure mortality prognostication, having now been externally validated in over 25,000 patients in samples derived from eight large clinical trial samples, four multicenter registries and numerous single center observational studies. Whether the ACTION-HF survival and survival and hospitalization models or risk scores will perform as well awaits similar investigations. The HF-ACTION investigators point out that the SHFM performed poorly in one small, single center sample of patients with advanced heart failure undergoing evaluation for LVAD or heart transplantation.<sup>21</sup> 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 CPX testing (using a modified Naughton protocol) is known to be strongly associated with peak  $V_{O_2}$ , and, in one study from the Cleveland Clinic (but not another<sup>22</sup>), predicted mortality or urgent transplantation very nearly as well as peak  $V_{O_2}$ .<sup>23</sup> It is surprising though that  $V_E/V_{CO_2}$  slope, which has repeatedly been found to be a stronger mortality predictor than peak  $V_{O_2}$ , was not so in this fairly large sample. While it is simpler, cheaper and easier to perform a treadmill test without respiratory gas measurement, that's not what was done in either HF-ACTION or in the Cleveland Clinic study cited, and it's 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. While 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 (KCCQ) 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 endpoint but not for the mortality alone endpoint. While 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’s a bit surprising that HF-ACTION investigators choose to include this in their models and risk scores. Creating a successful predictive model requires decisions about which candidate variables to include and how they contribute to the model.<sup>24</sup> While 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 BMI at 25 kg/m<sup>2</sup>, suggesting that in the HF-ACTION data set mortality decreased as BMI rose to 25 but then was constant. An obesity paradox, in which obese subjects (BMI ≥ 30 kg/m<sup>2</sup>) have similar or lower mortality than normal weight individuals (BMI 18.5 to < 25 kg/m<sup>2</sup>), 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. Since physician 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 partly 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 they were derived and validated, the variables used for risk stratification, their utility in predicting mortality at varying time points, and in 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 that of the sample from which the tool was derived. Validated risk prediction tools should be utilized. 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 frequently reassessed, particularly in patients for whom critical treatment decisions may hinge on the results.

## Acknowledgments

### 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 activities as a site principal investigator for REVIVE-IT are supported by funds contracted to the University of Michigan by HeartWare, Inc.

## References

- Andersson B, Sylven C. The DD genotype of the angiotensin-converting enzyme gene is associated with increased mortality in idiopathic heart failure. *J. Am. Coll. Cardiol.* 1996; 28:162–167. [PubMed: 8752809]
- Brook RD, Rajagopalan S, Pope CA 3rd, Brook JR, Bhatnagar A, Diez-Roux AV, Holguin F, Hong Y, Luepker RV, Mittleman MA, Peters A, Siscovick D, Smith SC Jr, Whitsel L, Kaufman JD. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation.* 2010; 121:2331–2378. [PubMed: 20458016]
- Burkett EL, Hershberger RE. Clinical and genetic issues in familial dilated cardiomyopathy. *J. Am. Coll. Cardiol.* 2005; 45:969–981. [PubMed: 15808750]
- McNamara DM, Holubkov R, Janosko K, Palmer A, Wang JJ, MacGowan GA, Murali S, Rosenblum WD, London B, Feldman AM. Pharmacogenetic interactions between beta-blocker therapy and the angiotensin-converting enzyme deletion polymorphism in patients with congestive heart failure. *Circulation.* 2001; 103:1644–1648. [PubMed: 11273991]
- Nair N, Farmer C, Gongora E, Dehmer GJ. Commonality Between Depression and Heart Failure. *Am. J. Cardiol.* 2011 Dec 5. [Epub ahead of print].
- Keenan PS, Normand SL, Lin Z, Drye EE, Bhat KR, Ross JS, Schuur JD, Stauffer BD, Bernheim SM, Epstein AJ, Wang Y, Herrin J, Chen J, Federer JJ, Mattera JA, Krumholz HM. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. *Circ Cardiovasc Qual Outcomes.* 2008; 1:29–37. [PubMed: 20031785]
- [ClinicalTrial.gov](http://clinicaltrials.gov) [Internet]. Bethesda (MD): National Library of Medicine (US); 2000. University of Michigan. Randomized Evaluation of VAD InterVEention before Inotropic Therapy (REVIVE-IT). Available: from <http://clinicaltrials.gov/ct2/show/NCT01369407> NLM Identifier: NCT01369407 [cited 2011 Dec 13]
- Levy WC, Lee KL, Hellkamp AS, Poole JE, Mozaffarian D, Linker DT, Maggioni AP, Anand I, Poole-Wilson PA, Fishbein DP, Johnson G, Anderson J, Mark DB, Bardy GH. Maximizing survival benefit with primary prevention implantable cardioverter-defibrillator therapy in a heart failure population. *Circulation.* 2009; 120:835–842. [PubMed: 19704100]
- Levy WC, Mozaffarian D, Linker DT, Farrar DJ, Miller LW, Investigators R. Can the Seattle heart failure model be used to risk-stratify heart failure patients for potential left ventricular assist device therapy? *J. Heart Lung Transplant.* 2009; 28:231–236. [PubMed: 19285613]
- Prasad H, Sra J, Levy WC, Stapleton DD. Influence of predictive modeling in implementing optimal heart failure therapy. *Am. J. Med. Sci.* 2011; 341:185–190. [PubMed: 21233695]
- Clarke M, Shah A, Sharma U. Systematic review of studies on telemonitoring of patients with congestive heart failure: a meta-analysis. *J. Telemed. Telecare.* 2011; 17:7–14. [PubMed: 21097564]

12. Allen LA, Yager JE, Funk MJ, Levy WC, Tulskey JA, Bowers MT, Dodson GC, O'Connor CM, Felker GM. Discordance between patient-predicted and model-predicted life expectancy among ambulatory patients with heart failure. *JAMA*. 2008; 299:2533–2542. [PubMed: 18523222]
13. Washington, DC: US Government Printing Office; 1982. President's Commission for the Study of Ethical Problems in Medicine and Biomedical Research. Making Health Care Decisions.
14. Fetters MD, Masuda Y. Japanese patients' preferences for receiving cancer tests results while in the United States: Introducing an advanced care directive for cancer disclosure. *J Palliat Med*. 2000; 3:361–374. [PubMed: 15859685]
15. Parascandola M, Hawkins J, Danis M. Patient autonomy and the challenge of clinical uncertainty. *Kennedy Inst. Ethics J*. 2002; 12:245–264. [PubMed: 12472078]
16. Aaronson KD, Schwartz JS, Chen TM, Wong KL, Goin JE, Mancini DM. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. *Circulation*. 1997; 95:2660–2667. [PubMed: 9193435]
17. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, Anand I, Maggioni A, Burton P, Sullivan MD, Pitt B, Poole-Wilson PA, Mann DL, Packer M. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation*. 2006; 113:1424–1433. [PubMed: 16534009]
18. Aaronson, KD., Mancini, DM. Risk factors for mortality with heart failure. In: Kormos, RL., Miller, LW., editors. *Mechanical Circulatory Support: A Companion to Braunwald's Heart Disease*. Philadelphia: Elsevier Saunders; 2012. p. 56-71.
19. O'Connor CM, Whellan DJ, Wojdyla D, Leifer E, Clare RM, Ellis SJ, Fine LJ, Fleg JL, Zannad F, Keteyian SJ, Kitzman DW, Kraus WE, Rendall D, Pina IL, Cooper LS, Fiuzat M, Lee KL. Factors Related to Morbidity and Mortality in Patients with Chronic Heart Failure with Systolic Dysfunction: The HF-ACTION Predictive Risk Score Model. *Circ Heart Fail*. 2012; 5 XX-XX.
20. O'Connor CM, Whellan DJ, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, Leifer ES, Kraus WE, Kitzman DW, Blumenthal JA, Rendall DS, Miller NH, Fleg JL, Schulman KA, McKelvie RS, Zannad F, Piña IL, Investigators HA. Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. *JAMA*. 2009; 301:1439–1450. [PubMed: 19351941]
21. Gorodeski EZ, Chu EC, Chow CH, Levy WC, Hsich E, Starling RC. Application of the Seattle Heart Failure Model in ambulatory patients presented to an advanced heart failure therapeutics committee. *Circ Heart Fail*. 2010; 3:706–714. [PubMed: 20798278]
22. Myers J, Gullestad L, Vagelos R, Do D, Bellin D, Ross H, Fowler MB. Clinical, hemodynamic, and cardiopulmonary exercise test determinants of survival in patients referred for evaluation of heart failure. *Ann. Intern. Med*. 1998; 129:286–293. [PubMed: 9729181]
23. Hsich E, Gorodeski EZ, Starling RC, Blackstone EH, Ishwaran H, Lauer MS. Importance of treadmill exercise time as an initial prognostic screening tool in patients with systolic left ventricular dysfunction. *Circulation*. 2009; 119:3189–3197. [PubMed: 19528334]
24. Harrell FE Jr, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat. Med*. 1984; 3:143–152. [PubMed: 6463451]