

Predicting Poor Outcomes in Heart Failure

David H Smith, RPh, PhD
Eric S Johnson, PhD
Micah L Thorp, DO, MPH
Xiuhai Yang, MS
Amanda Petrik, MS
Robert W Platt, PhD
Kathy Crispell, MD

Abstract

Background: Health plans must prioritize disease management efforts to reduce hospitalization and mortality rates in heart failure patients.

Methods and Results: We developed a risk model to predict the 5-year risk of mortality or hospitalization for heart failure among patients at a large health maintenance organization. We identified 4696 patients who had an echocardiogram and a heart failure diagnosis from 1999 to 2004.

We observed a 56% five-year risk of hospitalization for heart failure or death (95% confidence interval, 54% to 58%). The hazard ratios for echocardiogram data contributed statistically significantly to the model, but echocardiogram findings did not improve our ability to predict risk accurately once we had accounted for demographic characteristics and clinical findings. A more complex model demonstrated a modest capacity to accurately predict risk. Our risk model discriminated the highest- and lowest-risk patients with limited success—the observed risk was 3 times higher in the highest risk quintile, compared with the lowest-risk quintile.

Conclusions: Using data available from electronic health records, we developed a series of risk-prediction models for poor outcomes in patients with heart failure. We found that a relatively simple model is as effective as a more complex model, but that all the models predict with only modest accuracy. Until better prediction variables are available for heart failure patients, our prediction model may be valuable for prioritizing centralized disease management program efforts by stratifying patients according to their absolute risk of poor outcomes.

Introduction

Investigations into risk factors for poor outcomes in patients with heart failure have included patient-reported symptoms, comorbid conditions, and laboratory findings.^{1,2} These studies show that increases in patient-reported symptoms, extreme hemoglobin values, and poor

renal function are important risk factors for poor outcomes. However, findings from echocardiogram data, an important diagnostic tool in heart failure, have been less well studied for their predictive value across the spectrum of heart failure. In those with systolic dysfunction (<50%), lower ejection fraction val-

ues are associated with worse survival.³ Other studies, however, have demonstrated that higher ejection fraction values are associated with worse outcomes than lower values among patients with preserved left ventricular function.⁴ In this study, we examine the predictive value of echocardiogram measurements, including both ejection fraction and left ventricular wall thickness so as to further clarify these risks.

We stress prediction in the analysis and interpretation. Our interest is predicting the outcomes among heart failure patients using information that providers and health plans collect routinely, not hypothesis testing^{5p1-9} (eg, does low ejection fraction cause mortality). This is an important distinction because although every causal factor is a predictor, not every predictor is a cause,⁶ suggesting that some variables included in a prediction model, including the order of their entry into model building, may differ from a causal model. This study sought to predict outcomes among heart failure patients and to develop a prognostic risk model that could separate the higher- and lower-risk patients and determine how effectively ejection fraction and left ventricular wall thickness

David H Smith, RPh, PhD, is a Senior Investigator at the Center for Health Research, Kaiser Permanente Northwest in Portland, OR. E-mail: david.h.smith@kpchr.org.

Eric S Johnson, PhD, is an Investigator at the Center for Health Research, Kaiser Permanente Northwest in Portland, OR. E-mail: eric.s.johnson@kpchr.org.

Micah L Thorp, DO, MPH, is the Chief of Nephrology at Kaiser Permanente Northwest in Portland, OR. E-mail: micah.l.thorp@kp.org.

Xiuhai Yang, MS, is a Senior Research Analyst at the Center for Health Research, Kaiser Permanente Northwest in Portland, OR. E-mail: xiuhai.yang@kpchr.org.

Amanda Petrik, MS, is a Research Associate at the Center for Health Research, Kaiser Permanente Northwest in Portland, OR. E-mail: amanda.f.petrk@kpchr.org.

Robert W Platt, PhD, is an Investigator at McGill University Health Centre Research Institute and Professor at McGill University in Montreal, Quebec, Canada. E-mail: robert.platt@mcgill.ca

Kathy Crispell, MD, is the Chief Medical Officer at the Sunnyside Medical Center in Portland, OR. E-mail: kathy.a.crispell@kp.org.

improved predictions. Whereas other investigators have examined prediction rules in the inpatient setting and for patients with severe heart failure,^{7,8} we included patients more representative of those in a community-based setting.

Methods

Population

We conducted a retrospective study on a prevalence cohort of patients. Patients included in the study were adult (age 18 years and older) members of Kaiser Foundation Health Plan of the Northwest (Health Plan) who had an echocardiogram completed between 1999 and 2004 and a diagnosis of heart failure. Patients were followed for up to 5 years or until April 1 2005, death, or disenrollment from the Health Plan (whichever came first). The patient's first echocardiogram served as the index date. Patients were required to have at least 1 year of Health Plan membership (and prescription benefit coverage) before their index echocardiogram. All patients had 1 to 3 years of baseline data from which baseline covariates were extracted. We included patients with a diagnosis of heart failure (International Classification of Diseases, Ninth Revision [ICD-9] 428) from the inpatient or outpatient setting during the baseline period—or up to 30 days after their index echocardiogram (to account for diagnoses assigned after an echocardiogram, presumably on the basis of the echocardiogram findings). Others have found that ICD-9 428 has a predictive value positive of 82%⁹ for heart failure, but the test performance of that characteristic may not be the same in our setting. However, in order to mimic the data readily available (eg, without chart review) to a centralized population management department we were

Table 1. Baseline characteristics of sample and event rates (heart failure hospitalization or death) by baseline characteristics

Characteristics	Eligible patients N = 4696 (%)	Number of events	Crude event rate (per 100 person-years)
Demographic			
Age			
20 to 59 years	875 (18.6)	177	8.28
60 to 64 years	443 (9.4)	122	11.63
65 to 69 years	551 (11.7)	184	14.08
70 to 74 years	667 (14.2)	252	15.92
75 to 79 years	792 (16.9)	344	19.25
80 to 84 years	689 (14.7)	379	28.04
85 +	679 (14.5)	401	35.29
Sex			
Men	2284 (48.6)	916	18.64
Women	2412 (51.4)	943	17.34
Race			
White	4439 (94.5)	1741	18.78
Nonwhite	257 (5.5)	87	15.14
Body mass index			
<25	1117 (23.8)	580	27.54
25-34	2550 (54.3)	994	17.21
35+	1029 (21.9)	285	11.62
Current smoking			
Yes	456 (9.7)	174	18.28
No	4240 (90.3)	1685	17.92
Clinical			
Ejection fraction percentage			
>65	371 (7.9)	170	19.93
50-65	2500 (53.2)	891	15.84
40-49	593 (12.6)	225	16.31
30-39	550 (11.7)	246	20.93
20-29	472 (10.1)	216	22.93
<20	210 (4.5)	111	29.50
Posterior wall thickness			
≥11	2779 (59.2)	1146	18.93
<11	1917 (40.8)	713	16.59
Coronary artery disease			
Yes	2402 (51.1)	1059	21.07
No	2294 (48.9)	800	15.02
Diabetes			
Yes	1582 (33.7)	721	22.54
No	3114 (66.3)	1138	15.91
Systolic blood pressure			
<120	1393 (29.7)	562	19.46
120-140	1786 (38.0)	681	17.06
140+	1517 (32.3)	616	17.74
Blood pressure treatment			
Yes	4370 (93.1)	1740	18.09
No	326 (6.9)	119	17.46

Table continued on next page.

interested in that pragmatic patient inclusion strategy. The Kaiser Permanente Northwest (KPNW) institutional review board approved this study.

Covariates

The most current covariate baseline value (before the index

date) was extracted from KPNW's electronic medical record (EMR) (including laboratory data from inpatient or outpatient visits, patient registries, and echocardiogram findings; see Table 1 for listing). Covariate laboratory findings were obtained from outpatient laboratory data only because outpatient values

are less likely to be influenced by acute events. Glomerular filtration rate (eGFR) was estimated using the four-variable Modification of Diet in Renal Disease equation.¹⁰

Outcomes

Our outcome of interest was a composite of all-cause mortality or hospitalization (whichever came first) with a primary discharge diagnosis of heart failure.

Statistical Methods

We used Cox regression to predict the combined endpoint of all-cause mortality or first cardiovascular hospitalization. Patients were required to have a measured ejection fraction. As recommended by experts,^{5p41-5} we required at least 20 endpoint events for each degree of freedom reflected in candidate variables. We included candidate variables with <15% missing data in the model and used single imputation on missing data.

We included covariates if they were prevalent in at least 10% of patients. When deciding the variables to include in our model, we tried to balance each variable's ease of electronic extraction and measurement error. To do this we fit a series of models starting with easily obtainable demographic characteristics (ie, Model 1, including age, gender, race, body mass index [BMI]), then added variables readily measured at a primary care office visit (Model 2, adding eGFR, hemoglobin, blood pressure to Model 1 variables). In the next model we added echocardiogram data (Model 3, adding ejection fraction and posterior heart wall thickness, plus an interaction with wall thickness and BMI to standardize for body mass to Model 2 variables). These clinical measures are more expensive to measure, but

Table 1. Continued			
Dyslipidemia			
Yes	2191 (46.7)	819	17.29
No	2505 (53.3)	1040	18.52
Ischemic stroke			
Yes	511 (10.9)	284	31.45
No	4185 (89.1)	1575	16.67
Transient ischemic attack			
Yes	321 (6.8)	152	24.26
No	4375 (93.2)	1707	17.55
Hypothyroidism			
Yes	773 (16.5)	359	22.87
No	3923 (83.5)	1500	17.08
Chronic lung disease			
Yes	2143 (45.6)	917	20.08
No	2553 (54.4)	942	16.28
Aortic/mitral valvular disease			
Yes	1495 (31.8)	673	23.07
No	3201 (68.2)	1183	15.95
Depression			
Yes	1042 (22.2)	435	21.12
No	3654 (77.8)	1424	17.17
Known ventricular tachycardia or fibrillation			
Yes	210 (4.5)	83	19.27
No	4486 (95.5)	1776	17.90
Atrial fibrillation/flutter			
Yes	1726 (36.8)	806	22.20
No	2970 (63.2)	1053	15.67
Peripheral vascular disease			
Yes	752 (16.0)	425	32.21
No	3944 (84.0)	1434	15.88
Laboratory			
Hemoglobin			
<11	508 (10.8)	294	34.54
≥11	4188 (89.2)	1565	16.47
Baseline eGFR			
≥60	2650 (56.4)	821	13.18
<60 to 45	1036 (22.1)	458	19.05
<45 to 30	706 (15.0)	379	29.85
<30	304 (6.5)	201	44.88

eGFR = estimated glomerular filtration rate.

are thought to be closely related to heart failure prognosis. Our next model added a set of comorbid conditions (Model 4, adding smoking, diabetes, dyslipidemia, and hypothyroidism to Model 3 variables) that were readily diagnosed by patient inquiry or laboratory tests (ie, minimal measurement error). Model 5 added comorbid conditions (transient ischemic attack, chronic lung disease, heart valve disease, atrial fibrillation/flutter, depression, coronary artery disease, peripheral vascular disease, and stroke) that required complex clinical interaction to diagnose.

Model fit was assessed using Bayes Information Criterion (BIC). We calculated the concordance statistic^{5p465-507} (C-statistic = 1 is perfect prediction; C-statistic = 0.5 is equal to chance prediction) to evaluate each prediction model's accuracy. Some readers may be more familiar with the Area under the Receiver Operating Curve (AU-ROC) in logistic regression, which

is the analog to Cox regression's C-statistic. To determine the added predictive value of each model's set of additional variables, we evaluated the marginal relative change (%) in the C-statistic in accuracy between models.¹¹ Patients were categorized into quintiles of predicted risk. The mean predicted risk (overall and per quintile) was then compared with the mean observed risk to examine the model calibration (predicted risk/observed risk). A perfectly calibrated model would have the same observed and predicted risk. We plotted the observed risk and the predicted risk (within quintiles of predicted risk) to evaluate the risk model's calibration (ie, the extent to which the predicted risks over- or underestimate the observed risks).¹² We assessed the model's discriminative ability by dividing the predicted risk in the highest quintile by the predicted risk in the lowest quintile.¹²

We used Stata 9.2 (College Station, Texas, USA) and R (version

2.4), an open source software from the R Foundation for Statistical Computing (www.R-project.org).

The authors had full access to the data and take responsibility for its integrity. All authors have read and agree to the manuscript as written.

Results

Among 519,383 adults aged 18 years and older, we found 10,265 with a diagnosis of heart failure—8291 of whom had an echocardiogram. Our analysis dataset included 4696 of those patients with at least one year of Health Plan membership and pharmacy coverage before their echocardiogram. Table 1 shows the baseline characteristics and event rates. Older age, low ejection fraction, and cardiovascular disease were associated with higher death rates, while increasing BMI, blood pressure, and dyslipidemia were associated with lower event rates. The overall observed 5-year risk was 56%; 95% Confidence Interval (CI), 54% to 58%.

... easily accessible data from EMRs can be combined to predict patients at risk of poor outcomes from heart failure ...

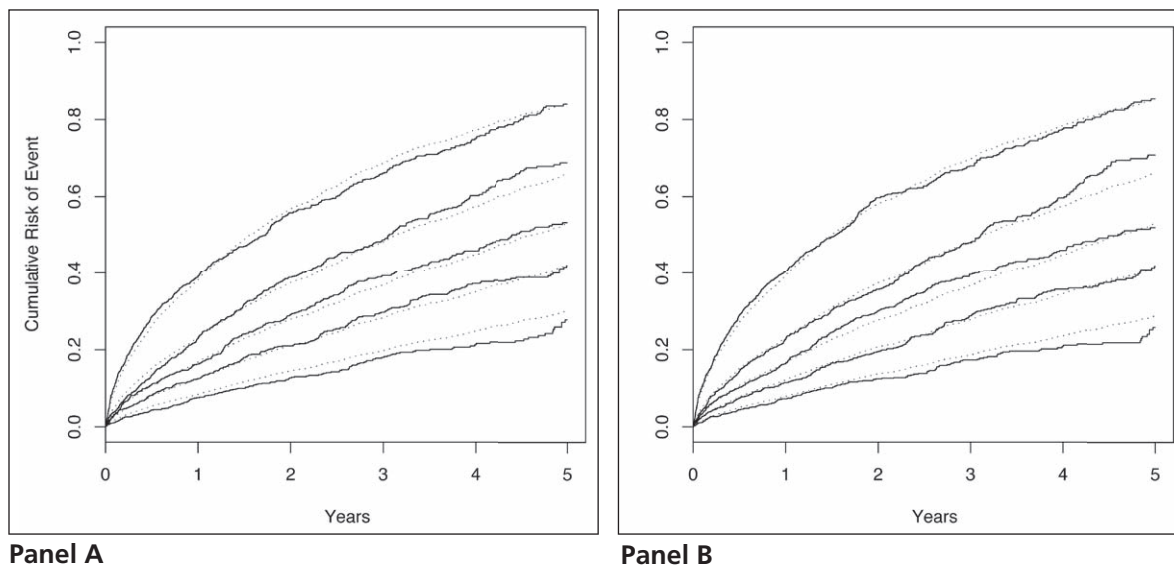


Figure 1. Calibration Plot for Model 2 (Panel A) and Model 3 (Panel B). The curves show the observed risk (solid lines) and predicted risk (dotted lines) of outcome (death or heart failure hospitalization) according to quintiles of predicted risk based on the risk score.

Table 2 shows the main results for the prediction models. The first model, with demographic characteristics only, had a C-statistic of 0.63, suggesting accuracy that is 13% better than chance prediction. In Model 2 (demographics along with hemoglobin, eGFR, and blood pressure), the C-statistic shows a 31% relative improvement: from 0.63 to 0.67. Adding echocardiogram data in Model 3 showed a 6% relative improvement to 0.68. Adding diagnoses that are readily measured in Model 4 improved the accuracy to 0.69 (6% relative improvement),

and adding further comorbidities raised the C-statistic to 0.71 (an 11% relative improvement). The hazard ratios (HRs) on baseline characteristics were consistent across models and model fit is improved (ie, lower BIC), but not appreciably, with added characteristics. Few hazard ratios exceeded 1.5, except older age, anemia, low eGFR, and low ejection fraction. Ejection fraction <20%, however, was a characteristic of <5% of the population. No HRs exceeded 2.5. No interactions were significant at $p < 0.05$.

As discussed above, the results

from the C-statistic show that Models 2 and 3 have similar discriminatory power. Another dimension of discrimination is the ratio of observed risks across the highest and lowest quintiles predicted risk. The risk model from Model 2 showed that patients in the highest quintile were about 3 times more likely to have the outcome as patients in the lowest quintile of risk: 84% (highest quintile); 66% (60th to 79th percentiles); 53% (middle quintile); 42% (20th to 39th percentiles); 30% (lowest quintile). Similar results were found for Model 3 (see Figure 1).

Table 2. Cox model results of heart failure hospitalization or death adjusting using baseline characteristics^a

Characteristic	Model 1 HR; 95% CI	Model 2 HR; 95% CI	Model 3 HR; 95% CI	Model 4 HR; 95% CI	Model 5 HR; 95% CI
Age					
20 to 59 years	0.54 (0.44 to 0.65)	0.60 (0.49 to 0.73)	0.57 (0.47 to 0.69)	0.57 (0.46 to 0.69)	0.58 (0.49 to 0.72)
60 to 64 years	0.75 (0.60 to 0.93)	0.79 (0.64 to 0.97)	0.75 (0.61 to 0.94)	0.76 (0.61 to 0.95)	0.75 (0.60 to 0.94)
65 to 69 years	0.89 (0.74 to 1.08)	0.93 (0.77 to 1.13)	0.93 (0.77 to 1.13)	0.92 (0.76 to 1.11)	0.93 (0.77 to 1.13)
70 to 74 years (reference)					
75 to 79 years	1.18 (1.00 to 1.39)	1.21 (1.02 to 1.42)	1.21 (1.02 to 1.42)	1.25 (1.06 to 1.47)	1.19 (1.01 to 1.40)
80 to 84 years	1.64 (1.40 to 1.93)	1.64 (1.39 to 1.93)	1.67 (1.42 to 1.97)	1.85 (1.57 to 2.18)	1.82 (1.54 to 2.14)
85 plus years	1.96 (1.66 to 2.30)	1.87 (1.59 to 2.20)	1.92 (1.63 to 2.27)	2.16 (1.82 to 2.55)	2.16 (1.82 to 2.56)
Sex					
Men	1.16 (1.05 to 1.27)	1.24 (1.13 to 1.36)	1.15 (1.04 to 1.26)	1.14 (1.04 to 1.26)	1.14 (1.03 to 1.26)
Race					
Non-white (white reference)	0.99 (0.80 to 1.23)	0.91 (0.73 to 1.13)	0.92 (0.74 to 1.14)	0.90 (0.72 to 1.12)	0.93 (0.75 to 1.16)
BMI					
<25 (reference)					
25-34	0.73 (0.68 to 0.84)	0.76 (0.69 to 0.85)	0.81 (0.69 to 0.95)	0.80 (0.68 to 0.94)	0.81 (0.69 to 0.95)
>34	0.70 (0.60 to 0.81)	0.72 (0.61 to 0.84)	0.82 (0.63 to 1.08)	0.76 (0.58 to 1.00)	0.80 (0.60 to 1.05)
C=0.63 BIC=29066					
Baseline eGFR					
≥60 (reference)					
<60 to 45		1.25 (1.11 to 1.40)	1.22 (1.09 to 1.37)	1.18 (1.05 to 1.33)	1.14 (1.01 to 1.28)
<45 to 30		1.66 (1.46 to 1.89)	1.59 (1.40 to 1.81)	1.51 (1.33 to 1.71)	1.43 (1.26 to 1.63)
<30		2.60 (2.22 to 3.06)	2.50 (2.13 to 2.93)	2.34 (1.99 to 2.75)	2.22 (1.88 to 2.61)
Baseline anemia					
Hemoglobin <11		1.70 (1.49 to 1.93)	1.75 (1.54 to 1.99)	1.71 (1.50 to 1.94)	1.59 (1.40 to 1.82)
Systolic blood pressure					
<120 (reference)					
120-140		0.82 (0.52 to 1.29)	0.81 (0.52 to 1.28)	0.81 (0.51 to 1.28)	0.84 (0.53 to 1.33)
>140		0.67 (0.41 to 1.08)	0.68 (0.42 to 1.10)	0.65 (0.40 to 1.05)	0.67 (0.41 to 1.08)
BP treatment (90 days prior to BP)		0.84 (0.58 to 1.22)	0.80 (0.55 to 1.15)	0.78 (0.54 to 1.13)	0.73 (0.50 to 1.06)
BP treatment x systolic blood pressure (p value)		0.44	0.40	0.34	0.29
C=0.67 [relative improvement = 31%] BIC=28893					

The calibration of Models 2 and 3 were also very similar, as shown in Figure 1. Specifically, the calibration was excellent at the highest level of predicted risk (ie, 84% observed vs 84% predicted for Model 2 and 85% observed vs 85% predicted for Model 3), and was within 5% for all quintiles in both Model 2 and Model 3.

Discussion

We found that easily accessible data from EMRs can be combined to predict patients at risk of poor outcomes from heart failure, and

that they predict as well as models using less easily accessible clinical data. From the perspective of the Health Plan, our prediction model may be most valuable for prioritizing centralized disease management program efforts by stratifying patients according to their absolute risk of poor outcomes. Care Managers might then use that risk data to focus coordination of care efforts on those at highest risk, or perhaps to deliver specific health prevention information to patients not yet at the highest risk. Information on individual patient risk level could

also be provided to physicians as part of their decision making as they identify patients most likely to benefit from case management, a complex patient medical home or referral for other heart failure specific services. Unlike many previous efforts at risk prediction for patients with heart failure, our analysis was not restricted to any particular subgroup of patients; instead, the population we used is representative of the community setting. This is an important point for disease management efforts because those responsible for population man-

Table 2. Continued

Characteristic	Model 1 HR; 95% CI	Model 2 HR; 95% CI	Model 3 HR; 95% CI	Model 4 HR; 95% CI	Model 5 HR; 95% CI
Ejection fraction					
>65%			1.17 (0.99 to 1.38)	1.14 (0.96 to 1.34)	1.10 (0.93 to 1.30)
50-65 (reference)					
40-49			1.00 (0.86 to 1.16)	0.99 (0.85 to 1.14)	0.98 (0.84 to 1.14)
30-39			1.30 (1.13 to 1.51)	1.26 (1.09 to 1.46)	1.29 (1.12 to 1.50)
20-29			1.54 (1.32 to 1.81)	1.50 (1.28 to 1.76)	1.56 (1.32 to 1.83)
<20			2.18 (1.77 to 2.68)	2.09 (1.70 to 2.58)	2.31 (1.87 to 2.84)
Posterior wall thickness			1.29 (1.10 to 1.53)	1.27 (1.07 to 1.50)	1.23 (1.04 to 1.45)
Posterior wall thickness x BMI (p value)			0.74	0.63	0.75
C=0.68 [relative improvement = 6%] BIC=28879					
Current Smoking				1.46 (1.24 to 1.72)	1.34 (1.14 to 1.58)
Diabetes				1.55 (1.40 to 1.71)	1.46 (1.32 to 1.61)
Dyslipidemia				0.97 (0.88 to 1.07)	0.90 (0.81 to 1.00)
Hypothyroidism				1.19 (1.06 to 1.34)	1.13 (1.00 to 1.27)
C=0.69 [relative improvement = 6%] BIC=28816					
Transient ischemic attack					0.97 (0.81 to 1.15)
Chronic lung disease					1.24 (1.13 to 1.36)
Aortic/mitral valvular disease					1.14 (1.04 to 1.26)
Atrial fibrillation/flutter					1.18 (1.06 to 1.30)
Depression					1.32 (1.18 to 1.48)
Coronary artery disease					1.07 (0.97 to 1.19)
Peripheral vascular disease					1.41 (1.26 to 1.59)
Ischemic stroke					1.40 (1.23 to 1.61)
C=0.71 [relative improvement = 11%] BIC=28736					

^a The concordance statistic (C=) shows the overall accuracy of the prediction as sets of characteristics were evaluated cumulatively. Bayes Information Criterion (BIC) was used to evaluate model fit.

BIC = Bayes Information Criterion, BMI = body mass index, BP = blood pressure, C = concordance statistic, CI = Confidence Interval, eGFR = estimated glomerular filtration rate, HR = hazard ratio

agement are concerned with the totality of the Health Plan members with heart failure. We also purposefully focused on demographic and clinical findings that are available from routinely collected EMR data, making our results more immediately applicable to centralized care management programs.

Some data like ejection fraction measurements are inconvenient to obtain from automated data sources because they don't exist in easily extractable data fields. However, our analysis showed that there was little increase in predictive ability

... predicting absolute risk requires a clinician to balance competing risk factors simultaneously ... an extremely difficult cognitive challenge.

when we added increasingly complex clinical measurements. Our findings illustrate that the added predictive ability of knowing ejection fraction is small when compared with a model that includes demographics, blood pressure, renal function and anemia status. The small increase in accuracy apparent when measurements from echocardiogram were added is particularly interesting and has potentially significant implications for disease management prioritization efforts. Specifically, whereas ejection fraction was available to us electronically at KPNW, not all managed care environments enjoy such access.

Some may find it surprising that ejection fraction did not add more predictive ability in our study. To put this into context, previous investigations have shown that each 10 point decrease in ejection fraction below 45% is related to a 30% increase in death rate (HR 1.31, 95% CI, 1.24 to 1.38), after adjusting for other potent risk factors (including, for example, the New York Heart Association heart failure classifica-

tion) over a median of 38 months of follow-up.³ However, it is important to note that our findings do confirm this strong relation between ejection fraction and increasing poor outcomes. Additionally, our findings are dependent on the specific time frame of prediction (ie, 5 years) that we used; a shorter time frame may have suggested stronger predictive value for echocardiogram findings. The HR for individuals with poor ejection fraction was greater than 2.0 (eg, Model 3, HR = 2.18, 95% CI, 1.77 to 2.68)—meaning that they were twice as likely to suffer the outcome as patients with better ejection fraction—so we found that ejection fraction does have *independent* predictive ability. Thus, taken on its own, ejection fraction is an important risk factor, but as pointed out by Guyatt,¹³ predicting absolute risk requires a clinician to balance competing risk factors simultaneously. Doing so is an extremely difficult cognitive challenge. A regression-based approach like the one taken here can help providers and health plans avoid problems with double-counting the contribution of correlated risk factors. It should be noted that we used ejection fraction findings from a specific point in time, and that additional prognostic value may be available for example, from serial ejection fraction findings. We purposefully included the entire spectrum of heart failure patients (ie, both preserved and reduced systolic function) in our model. Thus, our findings do not speak to the issue of whether ejection fraction measurements would add to predictive ability if stratified models were built for patients with preserved systolic function, and, separately, patients with systolic dysfunction.

A study⁸ of four clinical prediction rules in hospitalized heart

failure patients examined the predictive ability related to outcomes of inpatient death, complications, and 30-day mortality. The measure of accuracy used in that study (area under the receiver operating characteristic (ROC) curve, analogous to the C-statistic used in our study) was below 0.62 for inpatient death or complications, and went as high as 0.74 for inpatient death. Our study is different in important ways: 1) we used outpatient instead of inpatient characteristics; 2) we were able to assess the usefulness of echocardiogram data instead of physical examination findings such as pulse and respiratory rate; and 3) we had a 5-year follow-up period, not a 30-day period. In spite of these differences, however, 2 studies have remarkably similar findings and illustrate the difficulty in predicting outcomes in patients with heart failure. They further suggest investigators are missing important characteristics in existing models of heart failure prognosis. For example, prognosis may depend on a patient's willingness to adhere with medications and other daily disease management efforts that are difficult to capture at baseline.

Our study's main limitation is the lack of a protocol for measuring the characteristics completely and reliably. We had to impute 11% of BMI values, for example, because BMI was not collected during the baseline period. It is possible that BMI would be a stronger predictor if we had measured it more completely, instead of assigning patients' values according to their other (known) characteristics. The other characteristics that contributed to the risk prediction model had far less missing data (for example, 99.8% of patients had a recorded systolic blood pressure value), but may have been measured unreliably

on the basis of a single value. Other investigators have shown that predicting cardiovascular events on the basis of a single baseline value for blood pressure underestimates the strength of the relation by as much as 60%, a statistical problem known as regression dilution bias.¹⁴ The characteristics that we evaluated might have discriminated patients' risk more effectively if we were able to reduce regression dilution bias through repeated baseline measurements. Although a prognostic risk model based on repeated baseline measurements would be better in theory, it would be impractical for most health plans, as they lack repeated baseline measures collected according to a protocol. Our findings should be subject to validation efforts in other cohorts.

We suggest that Model 2 could be used for disease management prioritization efforts, provided that patients who are included in the population to be stratified all have a recent echocardiogram and a diagnosis of heart failure. We feel this is important because all patients who contributed to our risk model had an echocardiogram, which may have influenced the spectrum of heart failure patients. Patients without an echocardiogram, for example, may have been less severe or at least less symptomatic compared with patients who had an echocardiogram. Because we excluded patients without an echocardiogram, we cannot evaluate how the effectiveness of our risk model predictions varies across the entire spectrum of patients with a diagnosis of heart failure.¹⁵ So decision makers who use our risk model to prioritize patients for disease management in their populations may elect to only calculate predictions for patients who have had an echocardiogram. Including lower-risk heart failure patients

would probably compromise the risk model's accuracy and reduce its transportability, as successful transportability of a risk model to other clinical populations depends on the comparability of their disease spectrum.¹⁶ ♦

Acknowledgments

This study was funded by Amgen Pharmaceuticals. Amgen sponsored the study through a contract. Although Amgen had an opportunity to comment on the manuscript, they had no role in the study design, analysis, or writing of the manuscript; the authors bear full responsibility for this final version.

The author(s) have no other conflicts of interest to disclose.

References

- Go AS, Yang J, Ackerson LM, et al. Hemoglobin level, chronic kidney disease, and the risks of death and hospitalization in adults with chronic heart failure: the Anemia in Chronic Heart Failure: Outcomes and Resource Utilization (ANCHOR) Study. *Circulation* 2006 Jun 13;113(23):2713-23.
- Felker GM, Shaw LK, Stough WG, O'Connor CM. Anemia in patients with heart failure and preserved systolic function. *Am Heart J* 2006 Feb;151(2):457-62.
- Solomon SD, Anavekar N, Skali H, et al; Candesartan in Heart Failure Reduction in Mortality (CHARM) Investigators. Influence of ejection fraction on cardiovascular outcomes in a broad spectrum of heart failure patients. *Circulation* 2005 Dec 13;112(24):3738-44.
- Ahmed A, Aronow WS, Fleg JL. Higher New York Heart Association classes and increased mortality and hospitalization in patients with heart failure and preserved left ventricular function. *Am Heart J* 2006 Feb;151(2):444-50.
- Harrell FE. *Regression Modeling Strategies: Introduction*. New York: Springer; 2001.
- Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009 Feb 23;338:b375.
- Frankel DS, Piette JD, Jessup M, Craig K, Pickering F, Goldberg LR. Validation of prognostic models among patients with advanced heart failure. *J Card Fail* 2006 Aug;12(6):430-8.
- Auble TE, Hsieh M, McCausland JB, Yealy DM. Comparison of four clinical prediction rules for estimating risk in heart failure. *Ann Emerg Med* 2007 Aug;50(2):127-35.
- Roger VL, Weston SA, Redfield MM, et al. Trends in heart failure incidence and survival in a community-based population. *JAMA* 2004 Jul 21;292(3):344-50.
- Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med* 1999 Mar 16;130(6):461-70.
- Royston P, Altman DG. Risk stratification for in-hospital mortality in acutely decompensated heart failure. *JAMA* 2005 May 25;293(20):2467-8.
- Lumley T, Kronmal RA, Cushman M, Manolio TA, Goldstein S. A stroke prediction score in the elderly: validation and Web-based application. *J Clin Epidemiol* 2002 Feb;55(2):129-36.
- Guyatt GH. Determining prognosis and creating clinical decision rules. In: Haynes RB, Sackett DL, Guyatt GH, Tugwell P, eds. *Clinical Epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2006. p 323-55.
- MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990 Mar 31;335(8692):765-74.
- Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 2002 Oct 1;137(7):598-602.
- Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999 Mar 16;130(6):515-24.