JACC: Heart Failure; 2018, 6(6): 452-462

Performance of prognostic risk scores in chronic heart failure patients enrolled in the European Society of Cardiology Heart Failure long-term registry

Marco Canepa, MD, PhD[a], Candida Fonseca, MD, PhD[b], Ovidiu Chioncel, MD[c], Cécile Laroche, MS[d], Maria G. Crespo-Leiro, MD, PhD[e], Andrew J.S. Coats, MD, PhD[f], Alexandre Mebaza[a], MD, PhD[g], Massimo F. Piepoli, MD, PhD[h], Luigi Tavazzi, MD, PhD[i], Aldo P. Maggioni, MD[d, j], on behalf of the ESC HF Long Term Registry Investigators

[a] *Cardiology Unit, Department of Internal Medicine, University of Genova, and Ospedale Policlinico San Martino IRCCS, Genova, Italy*
[b] *Heart Failure Management Programme, S. Francisco Xavier Hospital, CHLO. NOVA Medical School, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisboa, Portugal*
[c] *University of Medicine Carol Davila, Bucuresti, Romania, and the Institutul de Urgente Boli Cardiovasculare C.C. Iliescu, Bucuresti, Romania*
[d] *EURObservational Research Programme (EORP), European Society of Cardiology, Sophia Antipolis, France*
[e] *Heart Failure and Heart Transplant Unit, Complexo Hospitalario Universitario A Coruña (CHUAC), CIBERCV, INIBIC, La Coruña, Spain*
[f] *San Raffaele Pisana Scientific Institute, Rome, Italy*
[g] *University Paris 7; Assistance Publique-Hôpitaux de Paris, U942 Inserm, Paris, France*
[h] *Heart Failure Unit, Cardiology, G da Saliceto Hospital, Piacenza, Italy*
[i] *Maria Cecilia Hospital, GVM Care&Research–ES Health Science Foundation, Cotignola, Italy*
[j] *ANMCO Research Center, Florence, Italy*

Abstract

Objectives. This study compared the performance of major heart failure (HF) risk models in predicting mortality and examined their utilization using data from a contemporary multinational registry.

Background. Several prognostic risk scores have been developed for ambulatory HF patients, but their precision is still inadequate and their use limited.

Methods. This registry enrolled patients with HF seen in participating European centers between May 2011 and April 2013. The following scores designed to estimate 1- to 2-year all-cause mortality were calculated in each participant: CHARM (Candesartan in Heart Failure-Assessment of Reduction in Mortality), GISSI-HF (Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico-Heart Failure), MAGGIC (Meta-analysis Global Group in Chronic Heart Failure), and SHFM (Seattle Heart Failure Model). Patients with hospitalized HF (n = 6,920) and ambulatory HF patients missing any variable needed to estimate each score (n = 3,267) were excluded, leaving a final sample of 6,161 patients.

Results. At 1-year follow-up, 5,653 of 6,161 patients (91.8%) were alive. The observed-to-predicted survival ratios (CHARM: 1.10, GISSI-HF: 1.08, MAGGIC: 1.03, and SHFM: 0.98) suggested some overestimation of mortality by all scores except the SHFM. Overprediction occurred steadily across levels of risk using both the CHARM and the GISSI-HF, whereas the SHFM underpredicted mortality in all risk groups except the highest. The MAGGIC showed the best overall accuracy (area under the curve [AUC] = 0.743), similar to the GISSI-HF (AUC = 0.739; p = 0.419) but better than the CHARM (AUC = 0.729; p = 0.068) and particularly better than the SHFM (AUC = 0.714; p = 0.018). Less than 1% of patients received a prognostic estimate from their enrolling physician.

Conclusions. Performance of prognostic risk scores is still limited and physicians are reluctant to use them in daily practice. The need for contemporary, more precise prognostic tools should be considered.

Key Words
Heart failure; mortality; prognosis; risk score

Abbreviations and Acronyms
AUC, area under the curve; HF, heart failure; ICD, implantable cardioverter-defibrillator; LVEF, left ventricular ejection fraction; SHFM, Seattle Heart Failure Model

Predicting survival in heart failure (HF) has become increasingly important for optimal patient care (1). Accurate assessment of prognosis may allow clinicians to decide whether a patient with chronic HF would most likely benefit from certain diagnostic and therapeutic interventions, especially invasive tests and complex procedures such as ventricular assist device implantation and heart transplantation. With these goals, several prognostic risk models have been developed in the past years, with the Seattle Heart Failure Model (SHFM) being the most popular and most thoroughly validated (2). However, these scores are not routinely calculated in clinical practice (1–4), primarily because of their poor reliability at the individual patient level (5) and also because treatments that specifically fit different levels of risk have not been established. A recent systematic review, which examined the characteristics of 20 prediction models for ambulatory HF patients, highlighted their inconsistent performance and called for new and more contemporary risk models (6). Thereafter, 2 new prognostic scores were proposed, the GISSI-HF (Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico-Heart Failure) (7) and the MAGGIC (Metaanalysis Global Group in Chronic Heart Failure) (8). The latter, in particular, was developed from a dataset of more than 39,000 individual patients' data from 30 studies (8) and recently validated using 2 large administrative datasets, but with a significant number of missing data, which were statistically imputed (5,9).

Between 2011 and 2013, the European Society of Cardiology (ESC) promoted a multinational HF registry, which enrolled more than 9,000 ambulatory HF patients who were followed up for 1 year or more. We tested and compared the performance of main prognostic risk scores in predicting 1- to 2-year survival in this contemporary population, and examined their use in daily clinical practice.

METHODS

STUDY POPULATION.

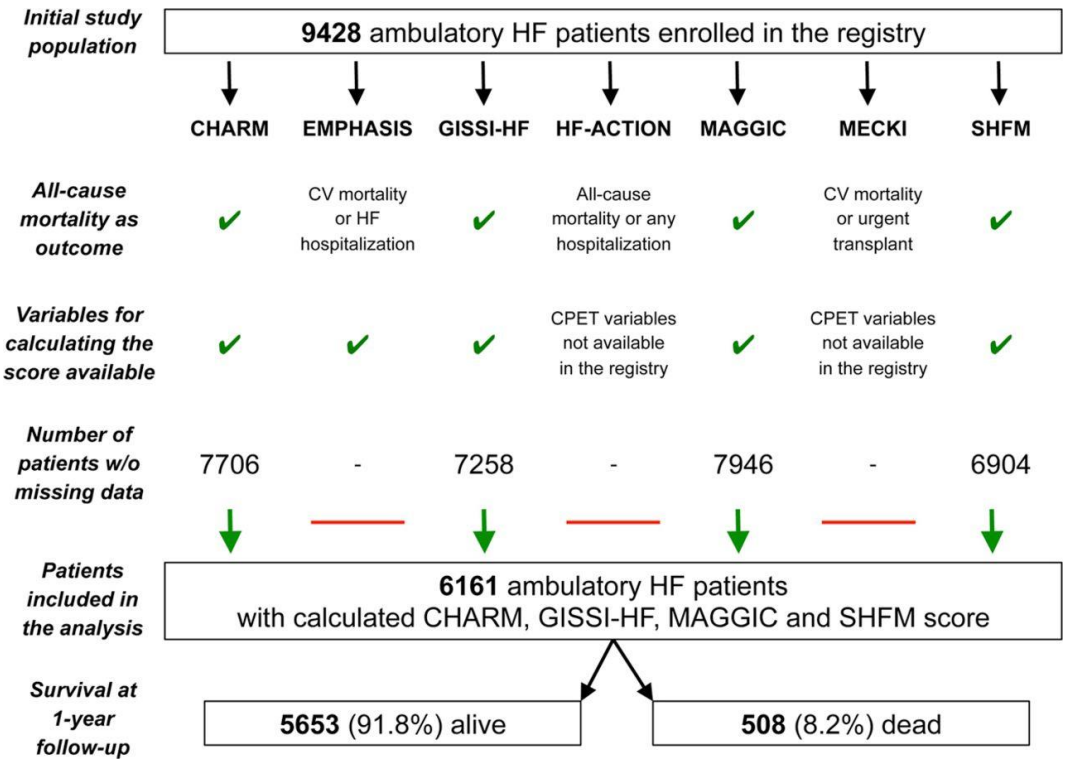
The ESC Heart Failure Long Term Registry has been previously described (10–12). Between May 2011 and April 2013, HF patients presenting to participating European centers with age >18 years were enrolled on a 1-day-per-week basis for 12 consecutive months, regardless of their left ventricular ejection fraction (LVEF). A diagnosis of HF was based on the clinical judgment of study investigators, which were primarily general cardiologists (10). The registry included 6,920 hospitalized HF patients who were excluded from this analysis focused on chronic HF. The majority of ambulatory HF patients were enrolled in Southern Europe (58.4%), and the remaining in Eastern (18.0%), Western (7.5%), and Northern Europe (5.9%), North Africa (6.3%), and the Middle East (3.9%) (10). They were followed up in accordance with the usual practice of each center, but a mandatory follow-up visit at 12 months was requested. At enrollment visit, all study investigators were requested to answer the following question included in the case report form of each patient: "Was prognosis evaluated using a risk score?" Where a positive answer was given, the following risk scores were proposed: SHFM, CHARM (Candesartan in Heart Failure-Assessment of Reduction in Mortality), GISSI-HF, MAGGIC, MECKI (Metabolic Exercise, Cardiac, Kidney Index), HF-ACTION (Heart Failure and A Controlled Trial Investigating Outcomes of Exercise TraiNing), and EMPHASIS (Eplerenone in Mild Patients Hospitalization

and Survival Study in Heart Failure) (see below). Each local Institutional Review Board approved the registry, and all patients enrolled in the study signed an informed consent.

SELECTION OF SCORES.

The score selection process is described in Figure 1. The primary outcome of the ESC HF Long Term Registry was 1-year all-cause mortality. Due to between-country differences in the starting date of enrolment, there were varying follow-up times in the entire study group, with a median follow-up time of 373 days, and 9.7% of patients having more than 2 years of follow-up (10). Thus, we selected those risk scores designed to estimate all-cause mortality at a follow-up of approximately 1 to 2 years in ambulatory HF patients only. These included the GISSI-HF score (7), the MAGGIC score (8), the SHFM score (2) and the CHARM score (13). The latter was primarily designed to predict cardiovascular mortality plus HF hospitalization, but a secondary analysis was performed to estimate predictors of total mortality at approximately 2 years; the prognostic model derived from this secondary analysis was used in our work (13). MECKI (14) and HF-ACTION (15) scores were excluded because they are primarily based on cardiopulmonary exercise test parameters, which were not collected in the present registry. In addition, they were generally designed to predict outcomes different from all-cause mortality, as was the EMPHASIS score (16) (i.e., cardiovascular mortality plus HF hospitalization). Other risk scores obtained from smaller and older cohorts (i.e., <1,000 patients, before the year 2000) (6) or including HF patients discharged from hospital (17) were not considered in this analysis.

FIGURE 1. Flow-Chart for Risk Score Selection



The initial 7 risk scores listed across the top were proposed in the in the case report form of each patient at the enrollment visit. See text for details. CHARM ¼ Candesartan in Heart Failure-Assessment of Reduction in Mortality; CPET = cardiopulmonary exercise test; CV = cardiovascular; EMPHASIS = Eplerenone in Mild Patients Hospitalization and Survival Study in Heart Failure; GISSI-HF = Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico-Heart Failure; HF = heart failure; MAGGIC = Meta-analysis Global Group in Chronic Heart Failure; MECKI = Metabolic Exercise, Cardiac, Kidney Index; SHFM = Seattle Heart Failure Model.

DATA PREPARATION AND MISSING DATA.

Instead of performing multiple imputations of missing data, we favored a more conservative approach and included in the analysis only patients in whom the variables needed to estimate each of the aforementioned scores were available. In accordance with the design of the study, all variables needed to estimate the CHARM, the GISSI-HF, and the MAGGIC scores were collected in the registry. The only uncollected variables for calculating the SHFM score were percentage lymphocytes and ventricular assist device implantation, to which we respectively assigned set values of 20% (normal) and "no," considering the probable low prevalence of ventricular assist devices in this population.

After a first selection of patients, a significant number of data were missing only for the following variables: "uric acid," used to estimate both the SHFM and the GISSI-HF score, and "total cholesterol," used to estimate the SHFM score (Table 1). Instead of imputing these missing data, we created a simplified version of both the SHFM and the GISSI-HF score, which did not comprise these 2 variables, and to be used only for patients in which they were not available. When comparing the complete and simplified SHFM and GISSI-HF scores in patients who had both uric acid and total cholesterol were available, the correlation between the complete and simplified version was very high (SHFM: $r = 0.88$; $p < 0.001$; GISSI-HF: $r = 0.98$; $p < 0.001$).

TABLE 1 Main Characteristics of the Study Population and by Status at 1-Year Follow-Up

	Total (N = 6,161; 100%)	Alive (n = 5,653; 91.8%)	Dead (n = 508; 8.2%)	p Value	C	G	M	S
Characteristics								
Survival probability								
CHARM	83.1 ± 8.4	83.6 ± 8.1	76.5 ± 8.6	<0.001	●			
GISSI-HF	84.9 ± 12.9	86.1 ± 11.6	72.8 ± 18.9	<0.001		●		
MAGGIC	89.4 ± 7.6	90.0 ± 6.9	82.2 ± 10.7	<0.001			●	
SHFM	93.3 ± 9.1	93.8 ± 8.2	87.6 ± 14.1	<0.001				●
Age, yrs	64.9 ± 13.2	64.4 ± 13.1	69.8 ± 13.2	<0.001	✔	✔	✔	✔
Male	71.8	71.3	77.4	0.003	✔	✔	✔	✔
Weight, kg	79.9 ± 16.8	80.4 ± 16.7	74.6 ± 16.4	<0.001				✔
BMI, kg/m^2	28.1 ± 5.1	28.3 ± 5.1	26.5 ± 5.1	<0.001	✔	✔	✔	
HF history								
Overall	91.3	91.4	90.4	0.405				
With previous hospitalization only	43.5	42.8	51.4	<0.001	✔			
>12 months	59.8	60.9	47.3	<0.001	✔		✔	
Heart rate, beats/min	72.8 ± 15.5	72.5 ± 15.4	76.2 ± 17.0	<0.001	✔			
Systolic BP, mm Hg	124.4 ± 20.6	125.0 ± 20.5	118.0 ± 21.0	<0.001		✔	✔	✔
Diastolic BP, mm Hg	73.6 ± 11.9	74.0 ± 11.9	69.1 ± 12.4	<0.001	✔			
Ischemic heart disease primary etiology	42.7	41.8	52.2	<0.001				✔
Smoking status				0.906	✔		✔	
Current	11.4	11.4	10.8					
Former	42.0	41.9	42.5					
Never	46.6	46.6	46.7					
Atrial fibrillation	36.8	35.9	46.3	<0.001	✔			
Diabetes mellitus	32.4	31.9	37.6	0.009	✔	✔	✔	
Diabetes mellitus treated with insulin	11.9	11.5	16.6	<0.001	✔			
Previous MI/angina	40.4	39.5	49.4	<0.001	✔			
COPD	15.1	14.8	18.9	0.014		✔	✔	
Device therapy								✔
CRT-D	10.5	10.1	14.6	0.002				
CRT-P	1.9	1.9	1.4	0.369				
ICD	16.7	17.1	13.2	0.025				
PM	5.6	5.4	7.5	0.047				

Clinical presentation								
NYHA functional class				<0.001	✔	✔	✔	✔
I	17.4	18.4	5.7					
II	57.6	58.9	44.1					
III	23.3	21.4	44.5					
IV	1.7	1.4	5.7					
Pulmonary rales	15.0	13.6	30.3	<0.001	✔			
Peripheral edema	20.3	18.9	36.2	<0.001	✔			
Mitral regurgitation	28.1	27.2	38.2	<0.001	✔			
Aortic stenosis	4.3	3.8	9.3	<0.001		✔		
Labs results								
Hemoglobin (g/dl)	13.3 ± 1.9	13.4 ± 1.8	12.3 ± 2.1	<0.001		✔		✔
Serum creatinine (mg/dl)	1.30 ± 2.62	1.28 ± 2.72	1.56 ± 1.04	<0.001			✔	
eGFR (MDRD)	65.2 ± 26.3	66.3 ± 26.2	52.9 ± 23.9	<0.001		✔		
Sodium (mEq/l)	139.4 ± 3.8	139.5 ± 3.6	138.0 ± 4.7	<0.001				✔
Total cholesterol (mg/dl)	167.3 ± 44.8 (n = 4,792)	168.1 ± 44.5 (n = 4,433)	158.3 ± 47.5 (n = 359)	<0.001				✔
Uric acid (mg/dl)	6.9 ± 2.7 (n = 4,175)	6.8 ± 2.5 (n =3,838)	7.6 ± 4.3 (n = 337)	<0.001		✔		✔
Outpatient visit: investigations/procedures								
BBB (LBBB or QRS duration>120)	43.8	42.9	53.5	<0.001	✔			✔
Cardiac enlargement, chest x-ray	29.2	28.7	33.7	0.020	✔			
Pulmonary congestion, chest x-ray	13.4	12.8	19.3	<0.001	✔			
LVEF	37.8 ± 13.7	38.1 ± 13.7	35.3 ± 13.8	<0.001	✔	✔	✔	✔
LVEF <40%	58.3	57.9	63.0	<0.001				
Medications								
ACE inhibitors	66.4	66.8	60.8	0.006			✔	✔
Angiotensin II receptor blockers	24.4	24.9	19.3	0.005			✔	✔
Beta-blockers	89.0	89.7	80.7	<0.001			✔	✔
Mineralocorticoid receptor antagonists	59.6	59.5	61.4	0.396				✔
Statins	61.9	62.3	57.9	0.050				✔
Diuretics oral	83.4	82.7	91.1	<0.001				✔
Furosemide dose equivalence, median (1st, 3rd interquartile)	40 (40,80)	40 (37.5,80)	80 (40,120)	<0.001				✔
Allopurinol	20.8	20.0	29.3	<0.001				✔

Values are mean _ SD or %, unless otherwise noted.
ACE = angiotensin-converting enzyme; BBB = bundle branch block; BMI = body mass index; BP = blood pressure; CHARM = Candesartan in Heart Failure-Assessment of Reduction in Mortality; COPD = chronic obstructive pulmonary disease; CRT-D/P = cardiac resynchronization therapy-defibrillator/pacemaker; eGFR = estimated glomerular filtration rate; GISSI-HF = Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico-Heart Failure; HF = heart failure; ICD = implantable cardioverter-defibrillator; LBBB = left bundle branch block; LVEF = left ventricular ejection fraction; MAGGIC = Meta-analysis Global Group in Chronic Heart Failure; MDRD = modification of diet in renal disease; MI = myocardial infarction; NYHA = New York Heart Association functional class; PM = pacemaker; SHFM = Seattle Heart Failure Model.

STATISTICAL ANALYSIS.

The characteristics of all patients at baseline and the characteristics of patients alive and dead at 1-year follow-up were summarized as means with SDs for continuous variables and as percentages for categorical variables, and compared with the use of Student t tests for continuous variables and chi-square tests for categorical variables. Observed versus model-predicted 1-year all-cause mortality was compared for quintiles of death probability estimated by each score, to evaluate the performance of each score at different levels of risk. Receiver operator characteristic (ROC) curves were generated to estimate the accuracy of each score in predicting 1-year all-cause mortality. The area under each ROC curve (AUC) was calculated and each AUC was compared with the best AUC using the Wald test to highlight significant differences and identify whether a more accurate prognostic score could be identified. A 2-sided p value of <0.05 was considered statistically significant. All analyses were performed using SAS statistical software version 9.3 (SAS Institute, Inc., Cary, North Carolina).

RESULTS

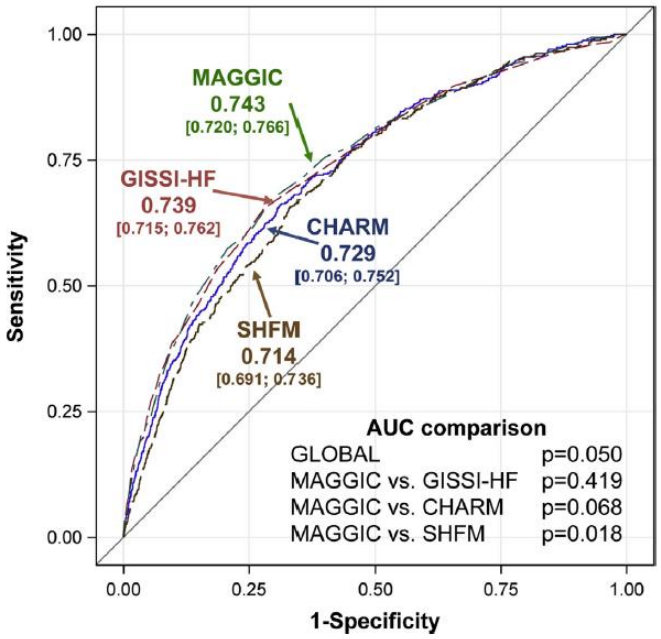
A total of 9,428 chronic HF patients were enrolled in the registry. After applying our inclusion and exclusion criteria (Figure 1), 6,161 patients who completed the 1-year follow-up visit were entered in the final analysis. The clinical characteristics of patients included in the analysis, as compared to those excluded because of missing data, were substantially similar, including age (64.9 ± 13.2 years vs. 65.1 ± 13.7 years; p = 0.244), male sex prevalence (71.8% vs. 71.1%; p = 0.474), systolic blood pressure (124 ± 21 mm Hg vs. 124 ± 22 mm Hg; p = 0.789), serum creatinine (1.30 ± 0.80 mg/dl vs. 1.23 ± 0.68 mg/dl; p = 0. 690), and ischemic heart disease etiology (42.7% vs. 43.4%; p = 0.50).

At 1-year follow-up, 508 patients (8.2%) had died. Table 1 shows the baseline characteristics of the overall population, and the comparison between those who were alive and those dead at 1-year follow up. Most covariates used in the risk score models were significantly different between the 2 groups (Table 1). In particular, at baseline, patients who had died at the time of scheduled follow-up, as compared to those who were alive, were significantly older, and had a more recent onset of HF with previous hospitalizations, a worse clinical presentation, and a lower LVEF. In addition, the use of HF-recommended medications was significantly lower in those who had died at follow-up, with a significantly lower prevalence of an implanted implantable cardioverterdefibrillator (ICD) (Table 1).

The CHARM, GISSI-HF, MAGGIC, and SHFM risk scores were retrospectively calculated for each patient and the distribution of each risk score in the study population is shown in Online Figure 1. Regardless of the risk score used, the average probability of survival was significantly higher in those who were alive at follow-up (Table 1). The observed- to- predicted survival ratios were the following: CHARM: 1.10, GISSI-HF: 1.08, MAGGIC: 1.03, and SHFM: 0.98, suggesting some over prediction of mortality by all scores except the SHFM.

The performance of each score is shown in Figure 2. Discrimination at the individual patient level was good overall (AUC >0.700), with the MAGGIC (AUC = 0.743) and the SHFM (AUC = 0.714), respectively, appearing to have the best and the worst accuracy, with a significant statistical difference (Figure 2).

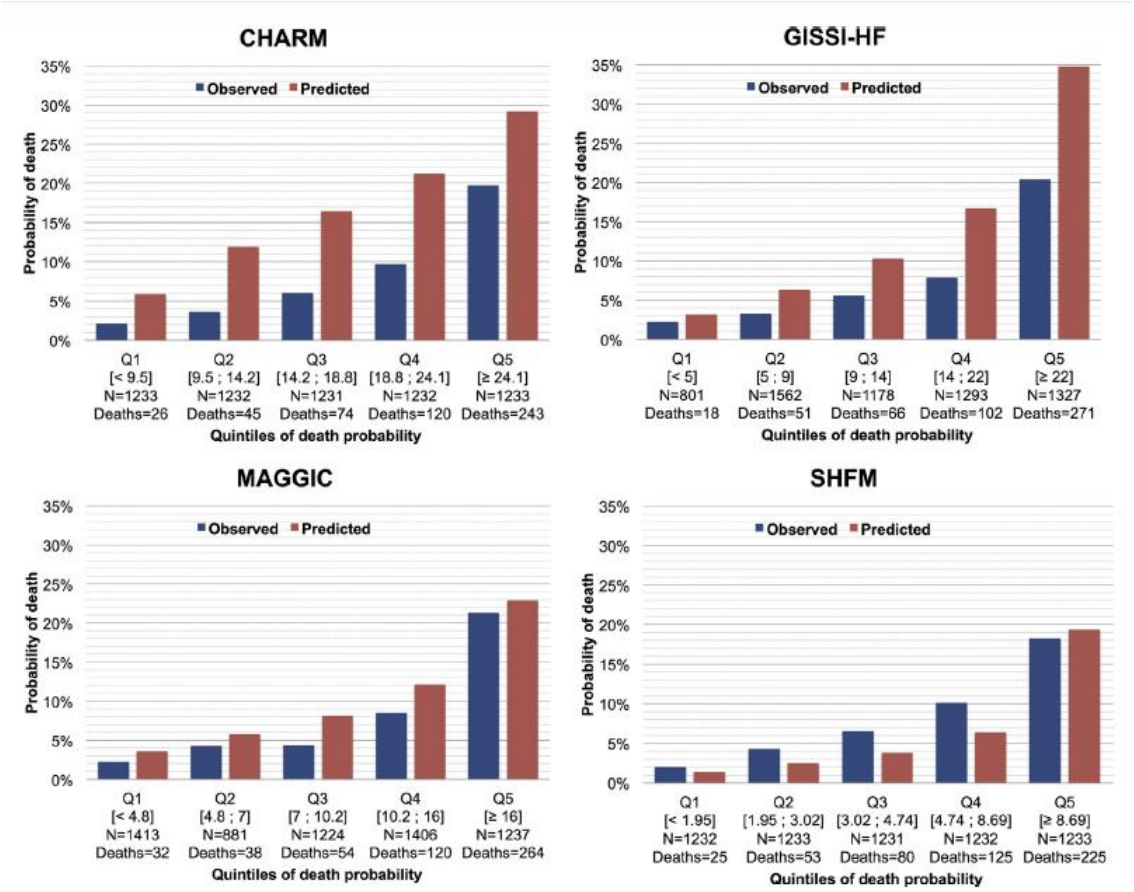
FIGURE 2. Receiver Operating Characteristic Curves Comparing the Areas Under the Curve of Risk Scores



AUC = area under the curve; other abbreviations as in Figure 1.

Histograms in Figure 3 show the observed versus model-predicted 1-year mortality for patients classified into quintiles of death probability estimated by each score. There was an over prediction of mortality for most risk groups using either the CHARM or the GISSI-HF score, whereas the SHFM score under predicted mortality in all risk groups except the highest. Despite some modest overestimation in the mid-probability range, the MAGGIC score showed the best overall accuracy in predicting 1-year mortality. Nonetheless, the Hosmer-Lemeshow goodness- of-fit test was <0.001 for all scores, which confirms the lack of fits of the models.

FIGURE 3. Observed Versus Predicted 1-Year Mortality by Risk Score in 5 Risk Categories



Abbreviations as in Figure 1.

Less than 4% of the case report forms (n = 241 of 6,161; 3.91%) had a response to the question "Was prognosis evaluated using a risk score?", and in front of a list of 7 prognostic risk scores, the majority had a negative answer. Indeed, probability of risk was estimated in only 16 patients overall using the SHFM.

DISCUSSION

This is one of the few analyses in the literature that has compared the performance of multiple existing prognostic risk scores in a real-world population of ambulatory HF patients enrolled in a multinational and contemporary registry. We showed a significant difference in the accuracy of 4 different risk scores predicting all-cause mortality, with the MAGGIC risk score outperforming others, particularly the more popular SHFM. Nonetheless, calibration at different levels of risk was still imperfect for most scores, with a general trend toward overestimation of risk, which could in part justify why <1% of patients received a prognostic estimate from their enrolling physician.

Table 2 summarizes the main characteristics of prognostic risk scores tested in this analysis. In particular, despite its subsequent validation and wide use, the SHFM score was originally derived from a sample of only approximately 1,000 chronic HF patients with reduced LVEF enrolled more than 20 years ago (2). Since then, the management and treatment of HF has changed substantially. Thus, the SHFM discrimination capacity in external validation cohorts has been recently estimated between 0.63 and 0.81 (6), with an AUC as low as 0.66 in a more contemporary sample of more than 10,000 chronic HF patients enrolled between 2005 and 2008 (5). It is no surprise, therefore, to find that it performed more poorly than other scores in our analysis (AUC = 0.714). However, we unexpectedly found that it generally determined a modest underestimation and not overestimation of the risk of death, as did the other scores. The explanation for this finding is unclear, but it most likely relates to the excessive protective weight assigned to ICDs in the SHFM. Despite being the oldest risk score among those used in this analysis (Table 2), the SHFM is the only 1 that accounts for the presence of cardiac implantable electronic devices in risk estimation (Table 1), attributing a substantial reduction in the risk of death when the implantation of an ICD (with or without biventricular pacemaker capability) has already been performed or has been planned (2). At a closer look, device information (as well as beta-blocker medications) were not available in the original SHFM derivation database, and hazard ratios for these variables in the SHFM score were estimated from published literature "with the use of effects seen in large published trials" (2). These effects were obtained before the widespread implementation of triple neuro humoral blockade in HF, and have been nowadays largely downsized in patients receiving optimal HF medical treatment (18). Thus, when applied to a well-treated population such as ours, in which more than one-fourth of the patients had an ICD in place at the time of enrollment, and approximately one-half of the remaining had an ICD implantation planned in the near term (data not shown), they may have determined a reduction in the estimation of risk in a considerable proportion of patients.

TABLE 2 Main Characteristics of Prognostic Risk Scores for Chronic HF Patients*

	CHARM	GISSI-HF	MAGGIC	SHFM
Characteristics of score development population				
Sample size	7,599 from 3 RCT	6,975 from 1 RCT	39,372 from 30 studies (6 RCT)	1,125 from 1 RCT
Mean age, yrs	66	68	68	65
Women	31	~22	35	34
Ischemic etiology	~70	40	54	63
Preserved/reduced LVEF	Both, ~60% reduced	Both, ~90% reduced	Both, ~75% reduced	Reduced only
Acute/chronic	HF Chronic	Chronic	Both	Chronic
HF treatments				
ACEi or ARB	41	93	67	99
Beta-blockers	55	62	34	0
MRA	17	40	21	3
Diuretics	83	90	82	100
ICD	0	7	NA	0
Follow-up information				
Median time	38 months	48 months	30 months	14 months
Total mortality	24.1	28.2	40.2	35.8
Main risk score outcome	All-cause mortality at 2 yrs	All-cause mortality at 2 and 4 yrs	All-cause mortality at 1 and 3 yrs	All-cause mortality at 1, 2, and 5 yrs
Time period				
Years of data collection	1999 to 2001	2002 to 2005	1992 to 2009	1992 to 1994
Year of publication	2006	2013	2012	2006
Methodology				
Initial validation in independent cohort	No, only internal	No, only internal	No, only internal	Yes, 9,942 patients from 5 cohorts (1997 to 2001)
Subsequent validation in independent cohort	No	No	Yes, 2 cohorts (5,9)	Yes, widely (6)
Number of variables included	NA	25	31	31
Number of variables in final model	24	12	13	24
Imputation of missing variables	No	Yes	Yes	Only in the validation cohorts

Values are %, unless otherwise noted. *The characteristics of the whole population were derived from the original publication of the CHARM (27), the MAGGIC (28), the GISSI-HF (7) and the SHFM (29) studies. ACEi = angiotensin-converting enzyme inhibitor; ARB = angiotensin II receptor blocker; MRA = mineralocorticoid receptor antagonists; NA = not available; RCT = randomized controlled trial; other abbreviations as in Table 1.

On the other hand, we found that the MAGGIC score had the best overall accuracy in predicting 1-year survival (AUC = 0.743), with some modest overestimation of risk only in the mid-probability range (Figures 2 and 3). This finding is in line with what has been observed in a similar analysis testing the performance of the MAGGIC score in a large administrative Swedish registry of more than 51,000 HF patients, which reported an AUC as high as 0.781 and a slight overestimation and underestimation of risk of 1-year mortality in the lowest and highest risk groups, respectively (9). Recent clinical trials of device (18) and drug therapies (19) have shown a progressive improvement with time in the prognosis of chronic HF with reduced LVEF treated with optimal medical therapy, with an overall mortality as low as 5% per year. In this contemporary multinational European registry, we observed a relevant improvement in the use of HF treatments (11), with percentages of patients treated with renin-angiotensin system blockers and beta-blockers as high as 90%, and a similar 1-year mortality of 8.2%. This mortality rate remains slightly higher than the 5% mortality observed in contemporary clinical trials, but it is significantly lower than that recorded in older derivation cohorts of prognostic risk score, particularly the SHFM (Table 2). Thus, the general over prediction of risk shown by the CHARM, the MAGGIC, and the GISSI-HF scores is most likely due to the significant improvement in the treatment of HF patients since the conception of these risk scores (Table 2).

It has been argued that prognostic risk scores, by incorporating multiple clinical variables, may be of greater help in estimating the risk of death from competing causes rather than sudden cardiac death (20). For example, it has been shown that with the increase in SHFM score, the rise in the risk of HF death is greater than that in the risk of sudden cardiac death (21). Unfortunately, a centralized validation of causes of death was not performed, but local investigators reported that approximately one-half of deaths in our registry were due to cardiovascular reasons (10). Thus, we foresee that the role of prognostic risk scores will become increasingly relevant in the near future, considering that the annual rate of sudden death has been significantly falling over the past 20 years (22), whereas non cardiac mortality has become the most relevant issue, particularly in patients with HF with preserved LVEF (23). Future prognostic risk models should be designed to also include this specific group of HF patients, which were a minority in previous derivation cohorts (Table 2) and accounted for approximately 40% of our study population (Table 1).

A final finding, although apparently minor, is the confirmation that such risk scores are not routinely used in daily clinical practice (1). Web-based applications have been developed for the SHFM and MAGGIC risk score to facilitate their use, and the MAGGIC software also allows for 2 of the 13 variables to be unknown. The GISSI-HF investigators proposed a practical nomogram usable at the desk to estimate the risk of death in individual patients (7). Despite these efforts, <1% of patients enrolled in this contemporary European registry received a prognostic estimate. Unfortunately, risk scores are known to perform very poorly for short-term medical decisions about individual patients (1), and this is probably discouraging their current application. A recent analysis of 10,930 ambulatory HF patients confirmed this open issue, showing that only 8 of 1,661 patients who died in the year after study enrollment had a >50% mortality rate predicted using the SHFM (sensitivity, 0.5%), and only 52 using the MAGGIC score (sensitivity, 3.1%), with the majority of deaths occurring in those with an estimated probability of survival >80% (5). Critical medical decisions (in cardiology as in other disciplines of medicines) are based on life expectancy, and prognostication remains essential to developing appropriate treatment plans and to relaying truthful information to the patient and his/her family members. Thus, further research is needed in search of increased accuracy and precision at the individual, more than at the population level.

One major limitation of risk score models in HF is the absence of impact analyses (24). Studies to evaluate the effect of using a prognostic model on current medical practice and on patient outcome would be informative and could lead to clinical implementation of such a model. An impact analysis could determine whether the use of the model is better than the usual care, and this remains an unmet need. In addition, the present prognostic scores, mostly generated by selected populations included in randomized controlled trials, might be used to assess the severity of HF in trial populations, either to characterize and compare previous trials targets or to assess the HF severity in populations enrolled in current prospective trials. However, as far as we know, this potentially interesting application of available scores has never been considered.

STUDY LIMITATIONS.

This work has several strengths and limitations to be acknowledged. Our analysis included ambulatory HF patients only, and tested prognostic scores which were specifically designed for this population. On the contrary, a previous analysis similar to ours also included HF patients discharged from hospital (9), for whom different and specific prognostic models should be applied (25). In addition, there was a relatively small amount of missing data on the candidate variables considering the multinational and voluntary nature of the registry. Thus, we refrained from using multiple imputation of missing data, and at the same time we made sure that the characteristics of included versus excluded patients were similar. Finally, we used real-world data collected in recent years (2011–2013), providing the most recent validation of the 4 risk scores used to date. This approach appears more consistent than the extrapolation of data from administrative electronic health records, which frequently do not record clinical variables that are very relevant in the prediction of outcomes, such as New York Heart Association (NYHA) functional class (5).

Among the limitations, we unfortunately could not test additional prognostic scores due to the lack of candidate variables in our registry, particularly those obtained from cardiopulmonary exercise tests. We created a simplified version of the SHFM and GISSI-HF for patients who were missing uric acid and total cholesterol values, but reported an excellent correlation between the complete and simplified versions in those who had both these variables available (see Methods section).

Our registry namely recorded a 1-year follow-up time point only. This might have determined an underestimation of the number of events against the GISSI-HF and CHARM scores, which were originally designed to predict 2-year all-cause mortality (Table 2). Nonetheless, as explained in the Methods section, approximately 10% of our patients had a follow-up of 2-years or longer, which might have partially mitigated this limitation. Our present analysis accounted neither for hospitalizations nor for specific causes of death, which were missing in about one-third of our ambulatory HF population (10). Thus, we elected to examine only prognostic risk scores specifically designed for total mortality.

The average age of the studied population was relatively young (<70 years of age): separate analysis in an older population would be clinically relevant because this subgroup of patients presents more challenging issues in prognostication and in the choice of the most appropriate therapy. In addition, only approximately one-third of patients in this registry were female, as was the case in the datasets from which these scores were developed (Table 2). This seems to be a general limitation inherent to the cardiology setting that does not include the universe of HF patients (26). The application and reliability of these score in female HF patients remains an open issue that may warrant further research. Finally, we can only speculate that the wider racial/ethnic composition of the MAGGIC derivation sample may have contributed to its greater accuracy in our multinational European cohort.

CONCLUSIONS

Prognostication in chronic HF patients has become increasingly important. We used a large contemporary and multinational ambulatory HF population to confirm that the performance of available prognostic risk scores is still limited. In our analysis, the most recent MAGGIC risk model appeared to be more accurate than the older CHARM and SHFM models in predicting 1-year mortality at the population level. However, investigators in this European registry were reluctant to use these scores, likely because their reliability at the individual patient level is known to be very poor. With the progressive improvement in HF therapeutic management and resulting decrease in mortality new, more precise, and thereby more clinically useful, prognostic tools are needed.

ACKNOWLEDGMENTS The authors thank the Registry Executive Committee and Steering Committee of the EURObservational Research Programme (EORP). Data collection was conducted by the EORP department from the ESC by Emanuela Fiorucci as Project Officer, Gerard Gracia and Maryna Andarala as Data Manager. Statistical analyses were performed by Cecile Laroche. Overall activities were coordinated and supervised by Doctor Aldo P. Maggioni (EORP Scientific Coordinator). All investigators are listed in the Online Appendix.

REFERENCES

1. Howlett JG. Should we perform a heart failure risk score? Circulation Heart failure 2013;6:4–5.
2. Levy WC, Mozaffarian D, Linker DT, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. Circulation 2006;113:1424–33.
3. Aaronson KD, Cowger J. Heart failure prognostic models: why bother? Circulation Heart failure 2012;5:6–9.
4. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013;10:e1001381.
5. Allen LA, Matlock DD, Shetterly SM, et al. Use of risk models to predict death in the next year among individual ambulatory patients with heart failure. JAMA Cardiol 2017;2:435–41.
6. Alba AC, Agoritsas T, Jankowski M, et al. Risk prediction models for mortality in ambulatory patients with heart failure: a systematic review. Circ Heart Fail 2013;6:881–9.
7. Barlera S, Tavazzi L, Franzosi MG, et al. Predictors of mortality in 6975 patients with chronic heart failure in the Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico-Heart Failure trial: proposal for a nomogram. Circ Heart Fail 2013;6:31–9.
8. Pocock SJ, Ariti CA, McMurray JJ, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. Eur Heart J 2013;34:1404–13.
9. Sartipy U, Dahlstrom U, Edner M, Lund LH. Predicting survival in heart failure: validation of the MAGGIC heart failure risk score in 51,043 patients from the Swedish heart failure registry. Eur J Heart Fail 2014;16:173–9.
10. Crespo-Leiro MG, Anker SD, Maggioni AP, et al. European Society of Cardiology Heart Failure Long-Term Registry (ESC-HF-LT): 1-year follow-up outcomes and differences across regions. Eur J Heart Fail 2016;18:613–25.
11. Maggioni AP, Anker SD, Dahlstrom U, et al. Are hospitalized or ambulatory patients with heart failure treated in accordance with European Society of Cardiology guidelines? Evidence from 12,440 patients of the ESC Heart Failure Long-Term Registry. Eur J Heart Fail 2013;15:1173–84.
12. Canepa M, Straburzynska-Migaj E, Drozdz J, et al. Characteristics, treatments and 1-year prognosis of hospitalized and ambulatory heart failure patients with chronic obstructive pulmonary disease in the European Society of Cardiology Heart Failure Long-Term Registry. Eur J Heart Fail 2018;20:100–10.
13. Pocock SJ, Wang D, Pfeffer MA, et al. Predictors of mortality and morbidity in patients with chronic heart failure. Eur Heart J 2006;27:65–75.
14. Agostoni P, Corra U, Cattadori G, et al. Metabolic exercise test data combined with cardiac and kidney indexes, the MECKI score: a multiparametric approach to heart failure prognosis. Int J Cardiol 2013;167:2710–8.
15. O'Connor CM, Whellan DJ, Wojdyla D, et al. 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;5:63–71.
16. Collier TJ, Pocock SJ, McMurray JJ, et al. The impact of eplerenone at different levels of risk in patients with systolic heart failure and mild symptoms: insight from a novel risk score for prognosis derived from the EMPHASIS-HF trial. Eur Heart J 2013;34:2823–9.
17. Senni M, Parrella P, De Maria R, et al. Predicting heart failure outcome from cardiac and comorbid conditions: the 3C-HF score. Int J Cardiol 2013;163:206–11.
18. Kober L, Thune JJ, Nielsen JC, et al. Defibrillator implantation in patients with nonischemic systolic heart failure. N Engl J Med 2016;375:1221–30.
19. McMurray JJ, Packer M, Desai AS, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med 2014;371:993–1004.
20. Halliday BP, Cleland JGF, Goldberger JJ, Prasad SK. Personalizing risk stratification for sudden death in dilated cardiomyopathy: the past, present, and future. Circulation 2017;136:215–31.
21. Mozaffarian D, Anker SD, Anand I, et al. Prediction of mode of death in heart failure: the Seattle Heart Failure Model. Circulation 2007;116:392–8.
22. Shen L, Jhund PS, Petrie MC, et al. Declining risk of sudden death in heart failure. N Engl J Med 2017;377:41–51.
23. Vaduganathan M, Patel RB, Michel A, et al. Mode of death in heart failure with preserved ejection fraction. J Am Coll Cardiol 2017;69:556–69.
24. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 2009;338:b375.

25. Lagu T, Pekow PS, Shieh MS, et al. Validation and comparison of seven mortality prediction models for hospitalized patients with acute decompensated heart failure. Circ Heart Fail 2016;9:e002912.
26. Maggioni AP, Orso F, Calabria S, et al. The real-world evidence of heart failure: findings from 41 413 patients of the ARNO database. Eur J Heart Fail 2016;18:402–10.
27. Pfeffer MA, Swedberg K, Granger CB, et al. Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the CHARM-Overall programme. Lancet 2003;362:759–66.
28. Meta-analysis Global Group in Chronic Heart F. The survival of patients with heart failure with preserved or reduced left ventricular ejection fraction: an individual patient data meta-analysis. Eur Heart J 2012;33:1750–7.
29. Packer M, O'Connor CM, Ghali JK, et al. Effect of amlodipine on morbidity and mortality in severe chronic heart failure. Prospective Randomized Amlodipine Survival Evaluation Study Group. N Engl J Med 1996;335:1107–14.