Circulation
Heart Failure

American Heart Association®

Risk Prediction Models for Mortality in Ambulatory Patients With Heart Failure: A Systematic Review

Ana C. Alba, Thomas Agoritsas, Milosz Jankowski, Delphine Courvoisier, Stephen D. Walter, Gordon H. Guyatt and Heather J. Ross

Circ Heart Fail. 2013;6:881-889; originally published online July 25, 2013;
doi: 10.1161/CIRCHEARTFAILURE.112.000043
Circulation: Heart Failure is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1941-3289. Online ISSN: 1941-3297

Risk Prediction Models for Mortality in Ambulatory Patients With Heart Failure

A Systematic Review

Ana C. Alba, MD; Thomas Agoritsas, MD; Milosz Jankowski, MD, PhD; Delphine Courvoisier, MSc, PhD; Stephen D. Walter, PhD; Gordon H. Guyatt, MD, MSc; Heather J. Ross, MD, MHSc

Background—Optimal management of heart failure requires accurate assessment of prognosis. Many prognostic models are available. Our objective was to identify studies that evaluate the use of risk prediction models for mortality in ambulatory patients with heart failure and describe their performance and clinical applicability.

Methods and Results—We searched for studies in Medline, Embase, and CINAHL in May 2012. Two reviewers selected citations including patients with heart failure and reporting on model performance in derivation or validation cohorts. We abstracted data related to population, outcomes, study quality, model discrimination, and calibration. Of the 9952 studies reviewed, we included 34 studies testing 20 models. Only 5 models were validated in independent cohorts: the Heart Failure Survival Score, the Seattle Heart Failure Model, the PACE (incorporating peripheral vascular disease, age, creatinine, and ejection fraction) risk score, a model by Frankenstein et al, and the SHOCKED predictors. The Heart Failure Survival Score was validated in 8 cohorts (2240 patients), showing poor-to-modest discrimination (*c*-statistic, 0.56–0.79), being lower in more recent cohorts. The Seattle Heart Failure Model was validated in 14 cohorts (16 057 patients), describing poor-to-acceptable discrimination (0.63–0.81), remaining relatively stable over time. Both models reported adequate calibration, although overestimating survival in specific populations. The other 3 models were validated in a cohort each, reporting poor-to-modest discrimination (0.66–0.74). Among the remaining 15 models, 6 were validated by bootstrapping (*c*-statistic, 0.74–0.85); the rest were not validated.

Conclusions—Externally validated heart failure models showed inconsistent performance. The Heart Failure Survival Score and Seattle Heart Failure Model demonstrated modest discrimination and questionable calibration. A new model derived from contemporary patient cohorts may be required for improved prognostic performance. **(*Circ Heart Fail*. 2013;6:881-889.)**

Key Words: heart failure ■ prediction models ■ prognosis ■ survival

Heart failure (HF) is a frequent health problem with high morbidity and mortality, increasing prevalence and escalating healthcare costs.[1,2] Older patient age, multiple comorbidities, and different patterns of disease progression create important challenges in patient management. Because the impact of these factors and their interactions remain incompletely understood, predicting patients' clinical course is difficult.

Editorial see p 877
Clinical Perspective on p 889

Accurate estimation of prognosis is important for many reasons. Patients are concerned about their probability of future events. Physicians may use prognosis estimates to decide the appropriate type and timing of additional tests or therapies, including heart transplantation and mechanical circulatory support. Accurate prognostic assessment may prevent delays in appropriate treatment of high-risk patients or overtreatment of low-risk patients. Knowledge of prognosis also facilitates research, for instance in the design of randomized trials and the exploration of subgroup effects.

To be usefully applied, prognostic models must be accurate and generalizable. Models may be inaccurate because of omission of important predictors, derivation from unrepresentative cohorts, overfitting or violations of model assumptions.

In the past 3 decades, investigators have developed many models to predict adverse outcomes in patients with HF.[3,4] Clinicians and researchers wishing to use prognostic models would benefit from knowledge of their characteristics and

Received December 5, 2012; accepted July 15, 2013.

From the Heart Failure and Transplantation Program, Toronto General Hospital, University Health Network, Ontario, Canada (A.C.A., H.J.R.); Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada (T.A., S.D.W., G.H.G.); Department of Internal Medicine, Jagiellonian University Medical College, Krakow, Poland (M.J.); and Center for Health Behavior Monitoring and Intervention, University of Rhode Island, Kingston, RI (D.C.).

The online-only Data Supplement is available at http://circheartfailure.ahajournals.org/lookup/suppl/doi:10.1161/CIRCHEARTFAILURE.112.000043/-/DC1.

Correspondence to Ana C. Alba, MD, Toronto General Hospital, University Health Network, 585 University Ave, 11c-1207, Toronto, Ontario, Canada M5G 2N2. E-mail Carolina.alba@uhn.ca

Circ Heart Fail **is available at http://circheartfailure.ahajournals.org** **DOI: 10.1161/CIRCHEARTFAILURE.112.000043**

performance. Therefore, we performed a systematic review to identify studies evaluating the use of risk prediction models for mortality in ambulatory patients with HF and to describe their performance and their clinical applicability.

Methods

Data Sources and Searches

In May 2012, with the assistance of an experienced research librarian, we performed a systematic search of electronic databases, including Medline, Embase, and CINAHL. We used several related terms: (internal cardiac defibrillator [ICD]), (heart or cardiac), (mortality or survival), and (multivariate analysis or regression analysis or risk factor or prediction or prognostic factor). The full search strategy is outlined in Appendix A in the online-only Data Supplement (Methods in the online-only Data Supplement). We identified additional studies by searching bibliographic references of included publications.

Study Selection

Eligible articles enrolled adults (>19 years) who were ambulatory patients with HF; used multivariable analysis (≥2 independent variables) to predict mortality or a composite outcome including mortality; reported >30 deaths; reported results as a score, a prediction rule, or as a set of regression coefficients sufficient to make predictions for individual patients; and reported a measure of discrimination or calibration. We also included studies evaluating the performance of an existing score in a different population to the one from which it was developed, and reported model discrimination and calibration. There were no restrictions on study design, left ventricular ejection fraction (LVEF), language, or date of publication. We excluded studies that enrolled patients during hospital admission or duplicate studies providing no new relevant data.

Two reviewers independently screened titles and abstracts, and then evaluated full-text versions of all articles deemed potentially relevant by either reviewer. During full-text screening, in cases of disagreement, consensus was reached through discussion. If consensus could not be reached, a third reviewer resolved the issue. Agreement between reviewers was assessed using weighted κ (0.92). Appendix B in the online-only Data Supplement (Methods in the online-only Data Supplement) shows the eligibility form.

Data Extraction

From each study, we abstracted data related to eligibility criteria, data source, time frame of recruitment, and characteristics of the population, including age, sex, ischemic cardiomyopathy, LVEF, use of β-blockers and ICD, definition, and number of events. We also identified variables included in the prediction models.

Assessment of Study Quality, Model Adequacy, and Performance

The assessment of study quality and model performance was based on what authors reported in their published articles. The selection of items for the assessment of study quality, model adequacy, and performance was based on the criteria proposed by Concato et al[5] and Moons et al.[6] Items included whether patient selection was consecutive, whether the data were collected prospectively, whether the percentage of missing data were small (<5%) and was correctly managed (ie, using data imputation), whether patients lost to follow-up were infrequent (<1%), and whether predictors were coded clearly.

To assess model adequacy, we abstracted information related to model derivation, including selection of the variables, coding, linearity of the response for continuous variables, overfitting,[7] and model assumptions. To assess model performance, we abstracted data related to discrimination and calibration. Discrimination expresses the extent to which the model is capable of differentiating patients who had events from those who did not. It is commonly assessed using the *c*-statistic, which is equivalent to the area under the receiver-operating characteristic curve.[8] Model discrimination was deemed as poor if the *c*-statistic was between 0.50 and 0.70, modest between 0.70 and 0.80, and acceptable if >0.80.[9] To assess how changes in HF treatment might modify model performance, we evaluated the impact of β-blockers, use of ICD, and study recruitment date on model discrimination graphically including models tested in >1 external cohort.

The calibration and goodness-of-fit of a model involves investigating how close the values predicted by the model are to the observed values. We identified the method used to assess model calibration (ie, Hosmer–Lemeshow test or deviance, Cox–Snell analysis, correlation between observed versus predicted events) and estimate of performance.

Table I in the online-only Data Supplement explains the criteria used to assess model adequacy and performance in more detail. Items that were not relevant (eg, in studies validating a preexisting model) were coded as nonapplicable.

Data Synthesis

We summarized the data, focusing on the characteristics of the population from whence models were derived and validated, and the models' performance. We report findings in 2 sections according to external validation (models that were or were not validated in an independent cohort were summarized separately).

Results

After duplicate citations were removed, we screened 6917 citations and ultimately selected 32 studies evaluating 20 prediction models (Figure 1). Only 5 of these models[10–14] were validated in an independent cohort. Among the remaining 15 models, 6 were internally validated by bootstrap; the remaining models were not validated.

Prediction Models Validated in an Independent Cohort

The Heart Failure Survival Score (HFSS),[10] the Seattle Heart Failure Model (SHFM),[11] the model proposed by Frankenstein et al,[12] the PACE risk score,[13] and the SHOCKED predictors[14] were validated in a different cohort of patients with HF from the model derivation cohort. Tables II and III in the online-only Data Supplement, and the Table summarize the characteristics of studies included, the assessment of study quality and model characteristics, respectively.

Heart Failure Survival Score

The HFSS includes 7 variables to predict a composite outcome of death, urgent (UNOS [United Network for Organ Sharing] status 1) heart transplantation and ventricular assist device implantation. Two predictors are binary: ischemic cardiomyopathy and presence of intraventricular conduction delay (QRS >120 ms); and 5 are continuous: LVEF, resting heart rate, mean blood pressure, peak oxygen consumption, and serum sodium. Scores are then divided into 3 categories: high risk, medium risk, and low risk according to prespecified thresholds.[10] The HFSS was derived from a single center cohort including 268 patients with HF and has been validated in 8 independent single-center cohorts including a total of 2240 HF patients.[10,14–19]

The validation cohorts involve a broad variety of patient populations (Table II in the online-only Data Supplement), with a mean age from 51 to 70 years, mostly males (65%–82%) with a mean LVEF between 20% and 30%. In 3 cohorts, the frequency of use of β-blockers was <30% and in the remaining 4 cohorts was 64% to 80%. In 4 studies reporting ICD status, the frequency of ICD use was 11%, 19%, 49%, and 78%.

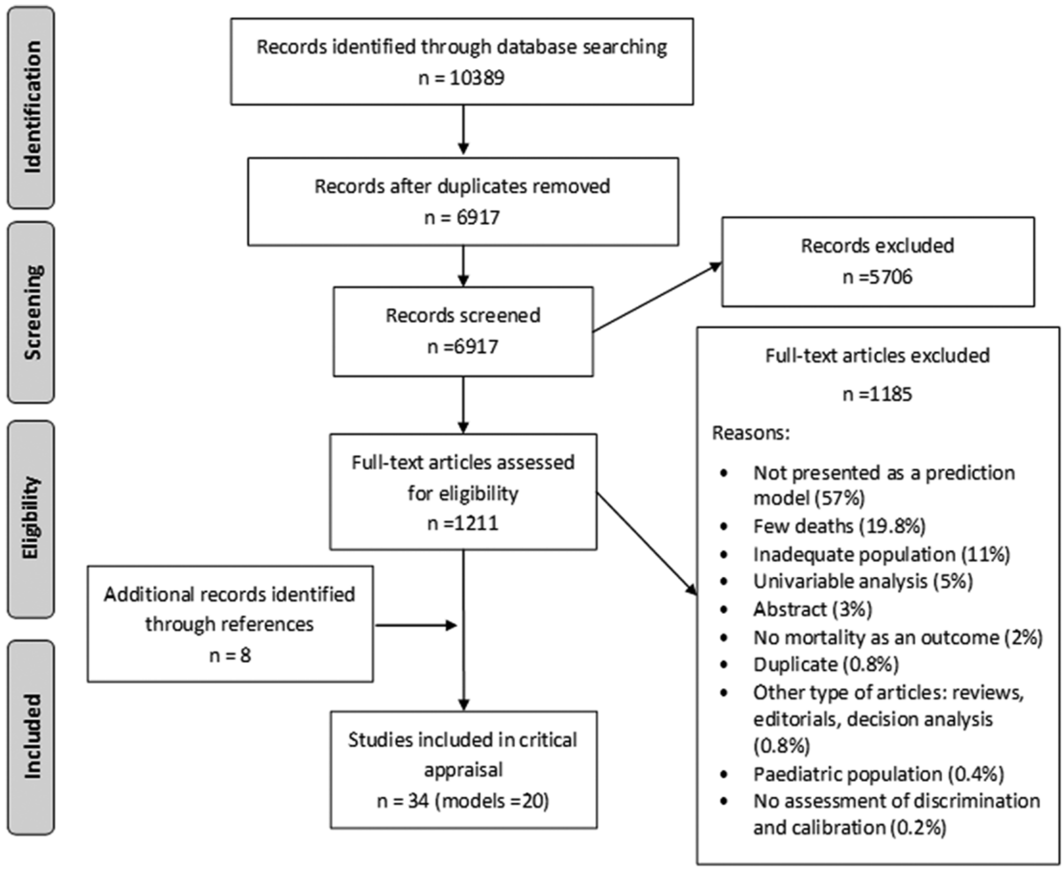

Figure 1. Study selection process. Number of studies during selection.

Model discrimination (assessed by the *c*-statistic at 1 year) in validation cohorts ranged from poor to modest (0.56–0.79), being modest (between 0.70 and 0.79) in 6 (75%) of the 8 validation cohorts. As shown in Figure 2, model discrimination was worse in cohorts with more frequent use of β-blockers or ICDs, and in more recent studies. Discrimination was poor (*c*-statistic, <0.70) in validation cohorts in which the rate of ICD use was >40%, studies with a contemporary recruitment date and in 3 of 4 cohorts in which the use β-blockers was >60%. The study by Zugck et al[15] reported a substantially higher discrimination (*c*-statistic=0.84 at 1 year) when peak oxygen consumption was replaced by the 6-minute walk test.

Table. Model Derivation and Performance

Study	Derivation Validation	Model/Variables	Selection	Linear Gradient	Overfitting	Model Assumptions	Calibration	Discrimination (*c*-Statistic)
Aaronson et al[10]	Derivation	HFSS: • Heart rate • BP • LVEF • Sodium • Ischemic CMP • IVCD • Peak Vo_2	Based on univariable analysis	n.r.	Yes (109 events and 11 variables)	Held	n.r.	At 1 y=0.79 (0.76–0.82)
	Validation in a different cohort	HFSS	n/a	n/a	n/a	n.r.	n.r.	At 1 y=0.76 (0.72–0.80) Overall=0.69 (0.62–0.76)
Zugck et al[15]	Validation	HFSS	n/a	n/a	n/a	n.r.	n.r.	Overall=0.74 (0.70–0.78)
		HFSS replacing peak VO2 by 6′WT	n/a	n/a	No	n.r.	n.r.	Overall=0.83 (0.79–0.87)
Koelling et al[16]	Validation	HFSS	n/a	n/a	n/a	n.r.	n.r.	Not β-blockers: at 1 y=0.76 (0.72–0.80) β-Blockers: at 1 y=0.73 (0.68–0.78)
Parikh et al[17]	Validation	HFSS	n/a	n/a	n/a	n.r.	n.r.	At 1 y=0.76 (0.70–0.83)

(*Continued*)

Table. Continued

Study	Derivation Validation	Model/Variables	Selection	Linear Gradient	Overfitting	Model Assumptions	Calibration	Discrimination (*c*-Statistic)
Gorodeski et al[18]	Validation	HFSS	n/a	n/a	n/a	n/a	Tested graphically: overestimated survival in HT candidates and more pronouncedly in non-HT candidates	At 1 y: In HT candidates=0.53 (0.50–0.63) In non-HT candidates=0.62 (0.55–0.68)
Goda et al[19–21]	Validation	HFSS	n/a	n/a	n/a	n.r.	n.r	*At 1 y: Total cohort=0.72 (0.67–0.76) European American (n=417) =0.69 (0.63–0.75) Black (n=125) =0.73 (0.63–0.84) Hispanic American (n=123) =0.76 (0.66–0.85) ICD/CRT patients (n=382) =0.69 (0.63–0.75)
Levy et al[11]	Derivation	SHFM: • Sex • Age • NYHA • Sodium • Uric acid • Cholesterol • Hemoglobin • Lymphocytes • Systolic BP • LVEF • Ischemic CMP • Statin • Allopurinol • Diuretic dose • β-blockers • ACEI • ARB • K-sparing diuretic • ICD/CRT	Based on univariable analysis, forward elimination effect of some treatments were obtained from previous RCTs or meta-analysis	Checked	No	n.r.	Assessed graphically observed vs predicted survival by deciles and by correlation (*r*=0.97)	At 1 y = 0.73 (0.69–0.76)
	Validation ELITE2	SHFM	n/a	n/a	n/a	n/a	Correlation (*r*=0.97)	At 1 y=0.67 (0.65–0.71)
	Validation RENAISSANCE	SHFM	n/a	n/a	n/a	n/a	Correlation (*r*=0.97)	At 1 y=0.69 (0.68–0.72)
	Validation Val-HeFT	SHFM	n/a	n/a	n/a	n/a	Correlation (*r*=0.98)	At 1 y=0.81 (0.72–0.90)
	Validation IN-CHF	SHFM	n/a	n/a	n/a	n/a	Correlation (*r*=0.99)	At 1 y=0.75 (0.70–0.80)
	Validation UW	SHFM	n/a	n/a	n/a	n/a	Correlation (*r*=0.99)	At 1 y=0.68 (0.63–0.73)
May et al[22]	Validation	SHFM	n/a	n/a	n/a	n/a	Correlation (*r*=0.99)	†At 1 y: Total cohort=0.73 (0.71–0.75) Age >75 y (n=1339) =0.68 (0.65–0.72) LVEF >40% (n=1634)=0.66 (0.62–0.69) ICD patients (n=693)=0.62 (0.56–0.69)
Allen et al[23]	Validation	SHFM	n/a	n/a	n/a	n/a	Assessed graphically. Overestimated survival at 3 y by 8% (72% vs 80%).	At 1 y=0.73
Kalogeropoulos et al[24] and Giamouzis et al[25]	Validation	SHFM	n/a	n/a	n/a	n/a	H-L test, inadequate ($P<0.05$). Graphically, adequate after model recalibration	‡At 1 y: Total cohort (n=445)=0.78 ICD/CRT (n=316)=0.78 No ICD/CRT (n=129)=0.79 White (n=223)=0.78 Black (n=198)=0.79

(*Continued*)

Table. Continued

Study	Derivation Validation	Model/Variables	Selection	Linear Gradient	Overfitting	Model Assumptions	Calibration	Discrimination (*c*-Statistic)
Levy et al[26]	Validation	SHFM and effect of IABP and inotropic support added from effect estimates obtained from previous studies	n/a	n/a	n/a	n/a		At 1 y=0.71
Gorodeski et al[18]	Validation	SHFM	n/a	n/a	n/a	n/a	Tested graphically: overestimated survival in HT candidates and non-HT candidates	§At 1 y: In HT candidates=0.68 (0.63–0.74) In non-HT candidates=0.63 (0.57–0.69)
Goda et al[21]	Validation	SHFM	n/a	n/a	n/a	n/a	n.r.	*At 1 y=0.73
Perrota et al[27]	Validation	SHFM	n/a	n/a	n/a	n/a	H-L test: P>0.2 at 1, 2, and 3 y	At 1 y=0.70 (0.61–0.79)
Haga et al[28]	Validation	SHFM	n/a	n/a	n/a	n/a	n.r.	Overall=0.68 (0.58–0.78)
Frankenstein et al[12]	Derivation	• BNP • 6′WT (different cutoff according to sex and β-blockers)	Based on univariable analysis	n.r.	no	n.r.	n.r.	Overall: Unadjusted=0.76 Sex-adjusted=0.77 β-Blocker-adjusted=0.76 Sex–β-blocker-adjusted=0.77
	Validation	Frankenstein[12]	n/a	n/a	n/a	n/a	n.r.	Unadjusted=0.66 Sex-adjusted=0.66 β-Blockers-adjusted=0.66 Sex–β-blockers-adjusted=0.68
Kramer et al[13]	Derivation	PACE risk score • Age >75 y • LVEF <20% • Creatinine • PVD	Based on univariable analysis	n.r.	no	n.r.	n.r.	At 1 y=0.79
	Validation	PACE risk score	n/a	n/a	n/a	n/a	n.r.	At 1 y=0.69
Bilchick et al[14]	Derivation	SHOCKED predictors • Age • NYHA • LVEF • COPD • Diabetes mellitus • Atrial fibrillation • CKD	Based on clinical importance and statistical analysis	n.r.	no	n.r.	Correlation (r=0.89)	Overall=0.75 (0.75–0.76)
	Validation	SHOCKED predictors	n/a	n/a	n/a	n/a	Correlation (r=0.89) H-L test: P<0.001 at 2 and 3 y	Overall=0.74 (0.74–0.75)

6′WT indicates 6-minute walk test; ACEI, angiotensin-converting enzyme inhibitor; ARB, angiotensin II receptor blocker; BNP, brain natriuretic peptide; BP, blood pressure; CKD, chronic kidney disease; CMP, cardiomyopathy; COPD, chronic obstructive pulmonary disease; CRT, cardiac resynchronization therapy; ELITE2, Losartan Heart Failure Survival Study; HFSS, Heart Failure Survival Score; H-L, Hosmer–Lemeshow; HT, heart transplantation; IABP, intra-aortic balloon pump; ICD, internal cardiac defibrillator; IN-CHF, Italian Congestive Heart Failure Registry; IVCD, intraventricular conduction defect; LVEF, left ventricular ejection fraction; n/a, non applicable; n.r., not reported; NYHA, New York Heart Association; PVD, peripheral vascular disease; RCT, randomized controlled trial; RENAISSANCE, Randomized Etanercept North American Strategy to Study Antagonism of Cytokines; SHFM, Seattle Heart Failure Model; UW, University of Washington HF clinic; Val-HeFT, Valsartan Heart Failure Trial; and Vo_2, oxygen consumption.

*Goda et al[21] reported that *c*-statistic was significantly higher (*c*-statistic=0.77 at 1 y) when HFSS and SHFM were used in a combined manner.

†Authors analyzed the additive discriminative value of creatinine, blood urea nitrogen (BUN), diabetes mellitus, and BNP (*c*-statistic=0.74, 0.74, 0.74, and 0.78, respectively).

‡Giamouzis et al[25] analyzed the additive of renal function and reported that renal function (BUN) did not significantly change discriminative capacity.

§Authors analyzed the additive predicted value of BNP, BUN, and peak Vo_2 and reported nonsignificant improvement in *c*-statistic values.

However, this HFSS variant has not been further validated. Only 1 study[18] assessed HFSS model calibration and reported that the model overestimated event-free survival by ≈20% in low-risk patients.

Seattle Heart Failure Model

The SHFM includes 10 continuous variables (age, LVEF, New York Heart Association class, systolic blood pressure, diuretic dose adjusted by weight, lymphocyte count, hemoglobin, serum sodium, total cholesterol, and uric acid) and 10 categorical variables (sex, ischemic cardiomyopathy, QRS>120 ms, use of β-blockers, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, potassium-sparing diuretic, statins and allopurinol, and ICD/cardiac resynchronization therapy [CRT] status) in an equation that provides a continuous risk score for each patient, and which can be expressed as

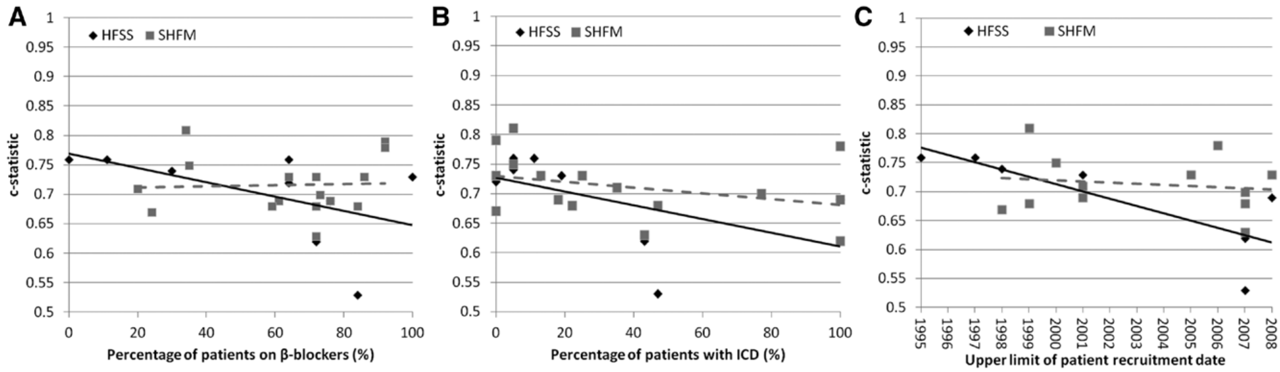

Figure 2. Model discrimination. Model discrimination according to the use of β-blockers (**A**), internal cardiac defibrillator (ICD; **B**), and study patients recruitment date (**C**). HFSS indicates Heart Failure Survival Score; and SHFM, Seattle Heart Failure Model.

predicted mean life expectancy or event-free survival at 1, 2, and 5 years.[11] This model was developed to predict a composite outcome of death, urgent heart transplantation, and ventricular assist device in 1125 patients with HF enrolled in the randomized controlled trial Prospective Randomized Amlodipine Survival Evaluation. The SHFM has been validated in 14 independent cohorts including 16 057 patients with HF (4 cohorts including 8983 patients with HF were selected from randomized controlled trials [Table II in the online-only Data Supplement]).[11,18,22–28] The validation cohorts involve diverse populations with a mean age from 52 to 77 years, a higher proportion of males (61%–82%), and mean LVEF between 17% and 45%. In 4 cohorts, the used of β-blockers was 20% to 35%, and in the remaining cohorts was >60% (maximum of 92%). In 10 studies reporting ICD status, the use of ICD was <25% in 5 cohorts and >65% in 3 cohorts.

Model discrimination varied from poor to acceptable (0.63–0.81), being at least modest (>0.70) in 7 (50%) cohorts of the 14 validation cohorts. There was a slight trend toward poorer discrimination in cohorts with higher use of ICD devices but was only weakly related to β-blocker use and recruitment date (Figure 2). Some studies[18,22,25] have analyzed variations of the SHFM including other predictors, such as renal function, diabetes mellitus, peak oxygen consumption, and brain natriuretic peptide, and reported that discrimination did not improve significantly. However, May et al[22] reported that discrimination was significantly improved from 0.72 to 0.78 when brain natriuretic peptide was added to the model. Model calibration was evaluated in most of the cohorts (Table) and showed a high correlation (r-coefficient >0.97) between observed and predicted survival. In 3 cohorts, calibration was assessed graphically by comparing observed and predicted event-free survival[17,22,24]; the model overestimated event-free survival by ≈2% at 1 year and 10% at 5 years, more significantly in black and patients with ICD/CRT.[22] The study by Kalogeropoulos et al[24] reported inadequate model goodness-of-fit as assessed by the Hosmer–Lemeshow test.

Frankenstein et al's Model

This model includes 2 binary variables: brain natriuretic peptide and 6-minute walk test with different cutoffs depending on sex and use of β-blockers.[12] Patients can then be categorized into 3 groups (scores 0, 1, or 2). This model was derived from 636 patients with HF to predict all-cause mortality and validated in an independent cohort of 676 patients with HF (mean age, 74 years; 76% male; 63% ischemic cardiomyopathy; 54% treated with β-blockers). Model discrimination in the validation cohort was poor, varying from 0.66 to 0.68 (Table). Model calibration was not reported.

PACE Risk Score

This model includes 4 binary variables: the presence of peripheral vascular disease, age >70 years, creatinine >2 mg/dL, and LVEF <20%, and it provides a continuous risk score for an individual patient from 0 to 5.[13] This model was derived from 905 secondary and primary prevention patients with ICD to predict all-cause mortality and validated in an independent cohort of 1812 patients with ICD-HF (mean age, 64 years; 77% male; mean LVEF of 31%; and 58% had ischemic cardiomyopathy [Table II in the online-only Data Supplement]). Model discrimination in the validation cohort was poor with a c-statistic of 0.69 at 1 year (Table). Model calibration was not reported.

SHOCKED Predictors

This model includes 7 binary variables: age >75 years, New York Heart Association class >II, atrial fibrillation, chronic obstructive pulmonary disease, chronic kidney disease, LVEF <20%, and diabetes mellitus.[14] This model provides a continuous risk score from 0 to 400 and estimates 1-, 2-, 3- and 4-year survival using a nomogram. This model was derived and validated from a cohort of Medicare beneficiaries receiving primary prevention ICD. The validation cohort included 27 893 patients (39% of patients were >75 years, 75% male, 31% had LVEF <20%, and 63% had ischemic cardiomyopathy [Table II in the online-only Data Supplement]). Model discrimination in the validation cohort was modest with a c-statistic of 0.74 at 1 year (Table). Overall correlation between observed and predicted survival was high correlation (r-coefficient >0.89). However, model calibration, assessed by Hosmer–Lemeshow test, showed inadequate goodness-of-fit at 2 and 3 years.

Prediction Models Not Validated in an Independent Cohort

We identified 15 prediction models that were not validated in an external cohort. Tables IV, V, and VI in the online-only Data Supplement summarize the characteristics of studies included, the assessment of study quality, and model characteristics, respectively. These models include a wide variety of predictors tested in diverse HF populations. The number of predictors included ranged from 2 to 21. Seven models were derived from patients with reduced LVEF and 1 in patients with preserved LVEF. The remaining studies included patients with clinically diagnosed HF without considering a specific LVEF cutoff as an inclusion criterion. In 6 studies, internally validated by bootstrapping, model discrimination ranged from 0.74 to 0.85. The best discrimination (*c*-statistic, 0.85) was observed in the DSC (Dyssynchrony, posterolateral Scar location and Creatinine) index, a model derived from a selective cohort of patients with HF undergoing CRT implantation, which included some variables that are not routinely available: 1 binary variable, posterolateral scar location evaluated by cardiovascular magnetic resonance; and 2 continuous variables, tissue synchronization index measured by cardiovascular magnetic resonance and serum creatinine. The 5 studies that evaluated model calibration reported adequate performance.

Discussion

In this systematic review, we identified 20 event-free survival prediction models in ambulatory patients with HF. Only 25% (5 of 20 models) have been validated in external cohorts and only 2 models, the HFSS and the SHFM, have been validated in >2 independent cohorts, mostly reporting modest (0.70–0.80)-to-poor (<0.70) discrimination. Studies using the HFSS more frequently reported modest (>0.70) discrimination than cohorts evaluating the SHFM. However, HFSS performance showed a decline over time, whereas the SHFM had a relatively stable performance. Nonetheless, only 2 studies[18,20] have directly compared models within the same population and reported that model discrimination was similar (*c*-statistic of 0.73 and 0.72[20] for the SHFM and 0.68 and 0.63[18] for the HFSS at 1 year).

Model discrimination represents the capacity of the model to differentiate patients who had the event from those who did not. The study by Goda et al[20] reported that discrimination was significantly higher (from 0.72–0.73 to 0.77 at 1 year) when HFSS and SHFM were used in a combined manner within the same model. May et al[22] reported that the discrimination of the SHFM was significantly improved from 0.72 to 0.78 when brain natriuretic peptide was added to the model. As proposed by D'Agostino and Byung-Ho Nam,[9] a model with discriminative capacity >0.70 has acceptable discrimination; a discriminative capacity >0.80 provides strong support to guide medical decision-making. Clearly, HFSS and SHFM have consistently demonstrated that their performance shows only modest discriminative capacity.

One potential reason for suboptimal performance is that the management and treatment of patients with HF has changed substantially in the past 2 decades. These models were derived from cohorts of patients recruited ≈20 years ago (1986–1991 for the HFSS and 1992–1994 for the SHFM).

As proposed by Moons et al,[6] a good model should include variables that are believed to be associated with the outcome of interest. Koelling et al[16] evaluated the association of the 7 predictors included in the HFSS model in patients treated with β-blockers and reported that only peak oxygen consumption and LVEF were factors independently associated with event-free survival. In addition, the directions of association of some predictors are opposite in the validation and derivation cohorts. For instance, the HFSS derivation study reported that the hazard ratio for 1 beat per minute increase in heart rate was 1.02 (95% confidence interval of 1.01–1.04), while in 2 validation cohorts[16,20] including a high proportion of patients treated with β-blockers (>70%), the hazard ratio was 0.98 (95% confidence interval, 0.97–1.01). This may partially explain the decline observed in the HFSS discriminatory capacity in more recent validation cohorts.

A similar situation is found with potassium-sparing diuretic use in the SHFM. Levy et al[11] imputed in the calculus of the score a hazard ratio of 0.74 for patients on potassium-sparing diuretics. Goda et al[20] reported a nonsignificant reverse effect of spironolactone in a contemporary cohort (hazard ratio, 1.20; 95% confidence interval, 0.86–1.48). Importantly, this tells us that predictors that were believed or found to be associated with mortality in patients with HF 20 years ago may not act similarly in contemporary patients with HF. This supports the need to develop and test an up-to-date prediction model.

Discrimination should not be reported in isolation because a poorly calibrated model can have the same discriminative capacity as a perfectly calibrated model.[29] One limitation of calibration is that assessment techniques do not allow for comparisons between models. In the validation cohorts, both the SHFM and the HFSS showed inadequate calibration attributable to the model overestimating survival in some groups of patients, including low-risk patients, blacks, and patients with ICD/CRT therapy.

Model ability to predict survival has not been compared with intuitive predictions of physicians. A study by Muntwyler et al[30] showed that primary care physicians overestimated mortality risk in patients with HF (1-year observed mortality of 13% versus physician estimate of 26%); this was more pronounced in stable New York Heart Association class II patients (1-year observed mortality of 6% versus physician estimated of 18%).

Whether these models may be used to guide or improve clinical practice remains underexplored. Vickers et al[29] have proposed the use of simple decision analytic techniques to compare prediction models in terms of their consequences. These techniques weight true and false-positive errors differently, to reflect the impact of decision consequences (ie, risks associated with heart transplantation or ventricular assist device versus risks associated with continuing medical therapy). Such decision analytic techniques may assist in determining whether clinical implementation of prediction models would do more good or more harm relative to current practice (physicians' predictions).

Should use and validation of these models continue? Or should we seek better models? There is no consensus on this issue among commentators. Researchers are pursuing both avenues, validating and supporting the use of the SHFM and HFSS as well as developing new models.

The performance of more recent models developed thus far, however, does not provide evidence that they will perform substantially better than older models. The 3 externally validated and recently published models[12–14] have demonstrated poor-to-modest discrimination (between 0.66 and 0.74). Similarly, the 6 models that were validated by bootstrapping showed in general poor-to-modest discrimination. One of these 6 models provided high discriminatory capacity, but it was developed in a selected group of patients with HF undergoing CRT implantation and included 2 variables that are not easily measured (myocardial tissue synchronization index and scar location by cardiovascular magnetic resonance). The lack of external validation makes it difficult to assess how the performance of the model might be generalized to other populations, which clearly limits their clinical use. Discrimination estimated on a first sample is often higher than that on the subsequent samples.[31]

Other reasons potentially explaining the suboptimal performance of existing models may pertain to the presence of missing data and variable selection. For example, in cohorts validating the SHFM, the presence of missing data was as high as 100% for percentage of lymphocytes[26] or 65% for uric acid.[22] Whether frequently missing or not easily available variables should be used to develop a score or should be incorporated to standard clinical practice will depend on the strength of the association between the predictors and outcome, the compromised model performance when the variables are not included in the final score and clinical resources. Nonetheless, adequate methods to deal with missing data, such as multiple imputation techniques, are important when evaluating model performance. The exclusion of cases because of missing information may lead to biased results.[32]

Variable selection based on statistical significance may lead to suboptimal models. Other techniques, such as stability selection and subsampling, have demonstrated to yield more stable models based on a consistent selection of variables decreasing the chances of type I error.[33]

As noticed in this review, the performance of predictive models has been traditionally evaluated by the *c*-statistic, which has been criticized as being insensitive in comparing models and for having limited direct clinical use. Reclassification tables, reclassification calibration statistic, and net reclassification and integrated discrimination improvements are recently developed methods to assess discrimination, calibration, and overall model accuracy. It has been shown that the use of these methods can better guide clinical decision-making by offering prognostic information at different risk strata. The use of these techniques is highly recommended during validation of existing or new models.

Conclusions

Optimal management of patients with HF requires accurate assessment of prognosis; however, making accurate assessment remains challenging. Among 5 externally validated prediction models, the HFSS and SHFM models demonstrated modest discriminative capacity and questionable calibration. The clinical impact of medical decision-making guided by the use of these models has not been explored. Given the limitation of current HF models, the development of a new model derived from contemporary patient cohorts is an appealing option. However, the development and reporting of new models should be optimized by adhering to guidelines to guarantee model adequacy. In addition, new models should seek external validation of their generalizability and performance. Evaluation of the clinical impact of decisions based on models relative to current clinical practice would be enormously informative in determining their use in real-world clinical practice.

Acknowledgements

The authors thank Ani Orchanian-Cheff for her expert assistance in conducting the systematic literature search.

Sources of Funding

Dr Alba was awarded a Vanier Canada Graduate Scholarship, administered by the Canadian Institutes of Health Research, Ottawa, ON, Canada.

Disclosures

None.

References

1. Rosamond W, Flegal K, Furie K, Go A, Greenlund K, Haase N, Hailpern SM, Ho M, Howard V, Kissela B, Kissela B, Kittner S, Lloyd-Jones D, McDermott M, Meigs J, Moy C, Nichol G, O'Donnell C, Roger V, Sorlie P, Steinberger J, Thom T, Wilson M, Hong Y; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics–2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*. 2008;117:e25–e146.
2. Bleumink GS, Knetsch AM, Sturkenboom MC, Straus SM, Hofman A, Deckers JW, Witteman JC, Stricker BH. Quantifying the heart failure epidemic: prevalence, incidence rate, lifetime risk and prognosis of heart failure The Rotterdam Study. *Eur Heart J*. 2004;25:1614–1619.
3. Goldraich L, Beck-da-Silva L, Clausell N. Are scores useful in advanced heart failure? *Expert Rev Cardiovasc Ther*. 2009;7:985–997.
4. Ketchum ES, Levy WC. Multivariate risk scores and patient outcomes in advanced heart failure. *Congest Heart Fail*. 2011;17:205–212.
5. Concato J, Feinstein AR, Holford TR. The risk of determining risk of multivariable models. *Ann Intern Med*. 1993;118:201–210.
6. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, Grobbee DE. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98:683–690.
7. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol*. 1995;48:1503–1510.
8. Cui J. Overview of risk prediction models in cardiovascular disease research. *Ann Epidemiol*. 2009;19:711–717.
9. D'Agostino RB, Byung-Ho N. Evaluation of the performance of survival analysis models: discrimination and calibration measures. In: Balakrishnan N, Rao CR, eds. *Handbook of Statistics v23: Advances in Survival Analysis*. Amsterdam, The Netherlands: Elsevier; 2004:1–26.
10. 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.
11. 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.
12. Frankenstein L, Goode K, Ingle L, Remppis A, Schellberg D, Nelles M, Katus HA, Clark AL, Cleland JG, Zugck C. Derivation and validation of a simple clinical risk-model in heart failure based on 6 minute walk test performance and NT-proBNP status–do we need specificity for sex and beta-blockers? *Int J Cardiol*. 2011;147:74–78.
13. Kramer DB, Friedman PA, Kallinen LM, Morrison TB, Crusan DJ, Hodge DO, Reynolds MR, Hauser RG. Development and validation of a risk score to predict early mortality in recipients of implantable cardioverter-defibrillators. *Heart Rhythm*. 2012;9:42–46.

14. Bilchick KC, Stukenborg GJ, Kamath S, Cheng A. Prediction of mortality in clinical practice for medicare patients undergoing defibrillator implantation for primary prevention of sudden cardiac death. *J Am Coll Cardiol*. 2012;60:1647–1655.
15. Zugck C, Krüger C, Kell R, Körber S, Schellberg D, Kübler W, Haass M. Risk stratification in middle-aged patients with congestive heart failure: prospective comparison of the Heart Failure Survival Score (HFSS) and a simplified two-variable model. *Eur J Heart Fail*. 2001;3:577–585.
16. Koelling TM, Joseph S, Aaronson KD. Heart failure survival score continues to predict clinical outcomes in patients with heart failure receiving beta-blockers. *J Heart Lung Transplant*. 2004;23:1414–1422.
17. Parikh MN, Lund LH, Goda A, Mancini D. Usefulness of peak exercise oxygen consumption and the heart failure survival score to predict survival in patients >65 years of age with heart failure. *Am J Cardiol*. 2009;103:998–1002.
18. 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.
19. Goda A, Lund LH, Mancini DM. Comparison across races of peak oxygen consumption and heart failure survival score for selection for cardiac transplantation. *Am J Cardiol*. 2010;105:1439–1444.
20. Goda A, Lund LH, Mancini D. The Heart Failure Survival Score outperforms the peak oxygen consumption for heart transplantation selection in the era of device therapy. *J Heart Lung Transplant*. 2011;30:315–325.
21. Goda A, Williams P, Mancini D, Lund LH. Selecting patients for heart transplantation: comparison of the Heart Failure Survival Score (HFSS) and the Seattle heart failure model (SHFM). *J Heart Lung Transplant*. 2011;30:1236–1243.
22. May HT, Horne BD, Levy WC, Kfoury AG, Rasmusson KD, Linker DT, Mozaffarian D, Anderson JL, Renlund DG. Validation of the Seattle Heart Failure Model in a community-based heart failure population and enhancement by adding B-type natriuretic peptide. *Am J Cardiol*. 2007;100:697–700.
23. Allen LA, Yager JE, Funk MJ, Levy WC, Tulsky 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.
24. Kalogeropoulos AP, Georgiopoulou VV, Giamouzis G, Smith AL, Agha SA, Waheed S, Laskar S, Puskas J, Dunbar S, Vega D, Levy WC, Butler J. Utility of the Seattle Heart Failure Model in patients with advanced heart failure. *J Am Coll Cardiol*. 2009;53:334–342.
25. Giamouzis G, Kalogeropoulos AP, Georgiopoulou VV, Agha SA, Rashad MA, Laskar SR, Smith AL, Butler J. Incremental value of renal function in risk prediction with the Seattle Heart Failure Model. *Am Heart J*. 2009;157:299–305.
26. Levy WC, Mozaffarian D, Linker DT, Farrar DJ, Miller LW; REMATCH Investigators. 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.
27. Perrotta L, Ricciardi G, Pieragnoli P, Chiostri M, Pontecorboli G, De Santo T, Bellocci F, Vitulano N, Emdin M, Mascioli G, Ricceri I, Porciani MC, Michelucci A, Padeletti L. Application of the Seattle Heart Failure Model in patients on cardiac resynchronization therapy. *Pacing Clin Electrophysiol*. 2012;35:88–94.
28. Haga K, Murray S, Reid J, Ness A, O'Donnell M, Yellowlees D, Denvir MA. Identifying community based chronic heart failure patients in the last year of life: a comparison of the Gold Standards Framework Prognostic Indicator Guide and the Seattle Heart Failure Model. *Heart*. 2012;98:579–583.
29. Vickers AJ, Cronin AM. Traditional statistical methods for evaluating prediction models are uninformative as to clinical value: towards a decision analytic framework. *Semin Oncol*. 2010;37:31–38.
30. Muntwyler J, Abetel G, Gruner C, Follath F. One-year mortality among unselected outpatients with heart failure. *Eur Heart J*. 2002;23:1861–1866.
31. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19:453–473.
32. Janssen KJ, Donders AR, Harrell FE Jr, Vergouwe Y, Chen Q, Grobbee DE, Moons KG. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol*. 2010;63:721–727.
33. Meinshausen N, Buhlmann P. Stability selection. *J R Statist Soc B*. 2010;72:417–473.

CLINICAL PERSPECTIVE

Many models are available to predict adverse outcomes in patients with heart failure. Clinicians and researchers wishing to use prognostic models would benefit from knowledge of their characteristics and performance. Therefore, we performed a systematic review to identify studies evaluating risk prediction models for mortality in ambulatory patients with HF, to describe their performance and clinical applicability. This systematic review included 34 studies testing 20 models. Only 5 models were validated in an independent cohort: the Heart Failure Survival Score, the Seattle Heart Failure Model, the PACE risk score, a model by Frankenstein et al,[12] and the SHOCKED predictors. The Heart Failure Survival Score, validated in 8 cohorts, showed poor-to-modest discrimination (*c*-statistic, 0.56–0.79), being lower in the more recent validation studies possibly because of greater use of β-blockers and implantable cardiac defibrillators. The Seattle Heart Failure Model was validated in 14 cohorts describing poor-to-acceptable discrimination (0.63–0.81), remaining relatively stable over time. Both models reported adequate calibration, although overestimating survival in some specific populations. The other 3 models were validated in a cohort each, with poor-to-modest discrimination (0.66–0.74). There were no studies reporting the clinical impact of medical decision-making guided by the use of these models. In conclusion, externally validated HF models showed inconsistent performance. The Heart Failure Survival Score and Seattle Heart Failure Model demonstrated modest discrimination and questionable calibration. A new model derived from contemporary patient cohorts may be required for improved prognostic performance.

SUPPLEMENTAL MATERIAL

SUPPLEMENTAL METHODS

Appendix A: Literature Search Results

For: Ana Carolina Alba

Date Completed: 15 May 2012

The databases searched were:

- Ovid MEDLINE
- EMBASE
- CINAHL

RESULTS & STRATEGY USED:

Database: Ovid MEDLINE(R) <1946 to May Week 1 2012>

Search Strategy:

----

1 exp Heart Failure/ (76819)

2 ((heart or cardiac) adj2 failure).mp. (121311)

3 1 or 2 (121859)

4 predict:.mp. (756732)

5 validat:.tw. (180066)

6 scor:.tw. (404761)

7 observ:.mp. (2029286)

8 or/4-7 (3043863)

9 3 and 8 (28134)

10 exp Ambulatory Care/ (42583)

11 Outpatients/ (7351)

12 (ambulatory or stable or chronic or out-patient: or outpatient:).mp. [mp=title, abstract, original title, name of substance word, subject heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier] (1246085)

13 10 or 11 or 12 (1246085)

14 9 and 13 (8814)

15 (mortality or survival or death).mp. [mp=title, abstract, original title, name of substance word, subject heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier] (1266793)

16 14 and 15 (3910)

17 statistics as topic/ or exp regression analysis/ (319979)

18 sn.fs. (425839)

19 statistic:.mp. (727873)

20 (logistic adj2 model:).mp. (85018)

21 (Likelihood adj2 function:).mp. (14814)

22 regression:.mp. (356421)

23 exp mathematical concepts/ (626843)

24 algorithm:.mp. (178754)

25 mathematic:.mp. (122305)

26 multivariate analysis/ (66832)

27 exp models, biological/ or exp models, statistical/ or logistic models/ (743997)

28 area under curve/ (21246)

29 or/17-28 (2456770)

30 "review"/ (1691446)

31 risk assessment/ or risk factors/ (590256)

32 evaluation.mp. (1000618)

33 exp Prognosis/ (930163)

34 prognostic factor:.mp. (47548)

35 8 or 31 or 32 or 33 or 34 (4702602)

36 3 and 13 and 15 and 35 (6181)

37 29 and 36 (2602)

38 30 and 36 (1361)

39 37 or 38 (3762)

Database: Embase <1974 to 2012 May 14>

Search Strategy:

----

1 exp heart failure/ (244924)
2 ((heart or cardiac) adj2 failure).mp. (207214)
3 1 or 2 (278699)
4 predict:.mp. (983853)
5 validat:.tw. (256546)
6 scor:.tw. (563146)
7 observ:.mp. (2609157)
8 risk assessment/ (285564)
9 risk factor/ (519981)
10 evaluation.mp. (1128376)
11 exp prognosis/ (388902)
12 prognostic factor:.mp. (67942)
13 or/4-12 (5511416)
14 3 and 13 (97265)
15 exp ambulatory care/ (35968)
16 outpatient/ (40332)
17 outpatient care/ (18777)
18 (ambulatory or stable or chronic or out-patient: or outpatient:).mp. (1647754)
19 15 or 16 or 17 or 18 (1647754)
20 14 and 19 (24318)
21 (mortality or survival or death).mp. (1806751)
22 20 and 21 (11345)
23 limit 22 to "review" (2010)
24 limit 23 to embase (1656)
25 exp statistics/ (272033)
26 exp regression analysis/ (179182)
27 statistic:.mp. (1196401)
28 (logistic adj2 model:).mp. (31580)
29 (Likelihood adj2 function:).mp. (782)
30 regression:.mp. (461195)
31 exp mathematical phenomena/ (2108262)
32 algorithm:.mp. (176636)
33 mathematic:.mp. (206662)
34 exp multivariate analysis/ (190591)
35 exp biological model/ (805064)

36 statistical model/ (88920)
37 area under the curve/ (55589)
38 or/25-37 (3631278)
39 22 and 38 (5358)
40 limit 39 to embase (4882)
41 24 or 40 (5993)

CINAHL Search Strategy
Tuesday, May 15, 2012 1:44:33 PM

#	Query	Limiters/Expanders	Last Run Via	Results
S29	S18 or S28	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	634
S28	S19 and S27	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	569
S27	S20 or S21 or S22 or S23 or S24 or S25 or S26	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	473798
S26	TX area under curve	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	116
S25	(MH "Models, Theoretical+")	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	52897
S24	(MH "Multivariate Analysis") OR (MH "Multivariate Analysis of Variance") OR (MH "Multivariate Analysis of Covariance")	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	29451
S23	(MH "Mathematics+")	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	291987
S22	TX statistic* or TX logistic N2 model* or TX likelihood N2 function* or TX regression or TX algorithm* or TX mathematic*	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen -	428036

			Advanced Search Database - CINAHL	
S21	(MH "Regression+")	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	109567
S20	(MH "Statistics+")	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	282038
S19	S16 and S17	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	1136
S18	S16 and S17	Limiters - Publication Type: Review Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	73
S17	TX mortality or TX survival or TX death	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	158882
S16	S11 and S15	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	2698
S15	S12 or S13 or S14	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	175366
S14	TX ambulatory or TX stable or TX chronic or TX out-patient* or TX outpatient*	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	171927

S13	(MH "Outpatients") OR (MH "Outpatient Service")	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	29357
S12	(MH "Ambulatory Care") OR (MH "Ambulatory Care Facilities+") OR (MH "Ambulatory Care Nursing")	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	13447
S11	S9 and S10	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	8549
S10	S3 or S4 or S5 or S6 or S7 or S8	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	602415
S9	S1 or S2	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	20275
S8	TX "prognostic factor*"	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	2789
S7	(MH "Prognosis+")	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	119023
S6	TX evaluation	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	263029
S5	(MH "Risk Factors+")	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen -	62487

			Advanced Search Database - CINAHL	
S4	(MH "Risk Assessment")	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	27594
S3	TX predict* or TX validat* or TX scor* or TX observ*	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	276104
S2	TX heart N2 failure or TX cardiac N2 failure	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	20263
S1	(MH "Heart Failure+")	Search modes - Boolean/Phrase	Interface - EBSCOhost Search Screen - Advanced Search Database - CINAHL	

Appendix B. Study eligibility form[1]

Reviewer:	XX	ZZ	NN

Article ID:			
Reference #:	Author:	Journal:	Year:

Population[2]**:**		
• Ambulatory heart failure patients	**YES**	**NO**
• Adults (≥ 19 years old)	**YES**	**NO**

Predictive model[3]**:**		
• ≥ 2 predictors *or* • Validation study of pre-existing score	**YES**	**NO**
• Report of score formula or coefficients and intercept	**YES**	**NO**
• Assessment of discrimination and/or calibration	**YES**	**NO**

Outcomes reported:		
• Mortality or composite outcome including mortality	**YES**	**NO**
• >30 deaths	**YES**	**NO**

Study design:		
• Cohort study (prospective or retrospective) *or* • Randomized control trial *or* • Meta-analysis	**YES**	**NO**

Duplicated population:		
• If duplicated, does this study report new information on model performance?	**YES**	**NO**

Study inclusion:	
• All the answers are YES	**INCLUDE**
• Any answer is NO	**EXCLUDE**

References:

[1] If any response to the above questions is unclear, mark YES.

[2] If a study included hospitalized patients or transplant or VAD patients, consider as NO.

[3] Any type of predictor, including but not limited to clinical characteristics, laboratory values, test results and any other clinical event, such as hospital admissions, ICD shocks, etcetera.

SUPPLEMENTAL TABLES

Supplemental Table 1. Aspects considered in the assessment of model adequacy and performance

Item	Description
Selection of the predictors	A good model should clearly state how predictors were selected. Potential candidate predictors may be chosen according to correlation with the outcome of interest explored in univariable analysis or based on previous knowledge. Whether one approach is better than the other is a matter of unresolved discussion. The former may include predictors that are not necessarily casual while the latter requires robust knowledge on the field of study.
Coding of the predictors	The proper reporting of the coding of variables is important because the effect of an independent variable on the outcome variable depends on the corresponding units of measurement and the manner in which the variable was coded. Articles were considered to properly report the coding of variables if the method of coding for all of the variables that remained in the final statistical model could easily be determined or were referenced anywhere in the article.
Nonconformity to a Linear Gradient	If the manuscript did not report determining the impact of each explanatory variable separately in zones of ranked data or mentioned that conformity to a linear gradient was addressed, this item was coded as not reported.
Over-fitting	Risk estimates may be unreliable if the multivariable model includes too many independent variables and too few outcome events, they may represent spurious associations or the effects may be estimated with low precision. According to Peduzzi et al [1], we categorized the articles with a ratio of < 10:1 (10 outcome events for each single explanatory variable in the final model) as an over-fitted.

Analysis of statistical model assumption	Violation of model assumptions, such as the proportional hazards assumption in the case of Cox method, may lead to unreliable effect estimates. If a manuscript did not state exploring model assumptions and that they were held in the final proposed model, this item was coded as not reporting model assumptions.
Discrimination	Discrimination expresses to what extent the model is capable of differentiating patients who had the event from those who did not. It is commonly assessed using the c-statistic test, which is equivalent to the area under the receiver operating characteristic (ROC) curve [2]. The ROC curve is a plot of sensitivity versus 1-specificity, which are calculated for each value of the predicted risk as a possible cut-off value. A c-statistic of 0.50 indicates that the model performs no better than chance; a c-statistic of 0.50 to 0.70 indicates poor discrimination; a c-statistic of 0.70 to 0.80 indicates modest discriminative ability; and a c-statistic of greater than 0.80 indicates aceptable discriminative ability [2].
Calibration or goodness of fit	The calibration or goodness of fit of a model measures how well the model describes the response variable. Goodness-of-fit involves investigating how close values predicted by the model are to the observed values. It can be assessed using different methods (i.e., Hosmer-Lemeshow test or deviance, Cox-Snell analysis, correlation between observed vs. predicted events).

References of Supplemental Table 1:

1. Peduzzi P, Concato J, Feinsten AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995;48:1503-10.
2. D'Agostino RB, Byung-Ho Nam. Evaluation of the performance of survival analysis models: Discrimination and calibration measures. In: Handbook of Statistics v23: Advances in survival analysis, by Balakrishnan N, Rao CR. 2004.

Supplemental Table 2. Characteristics of the population of studies included

Study	Model's name	Derivation/ Validation study	Population										Events	
			Source	Inclusion criteria	Time frame	N	Mean Age	% male	Mean LVEF	% ischemic	% β-blocker	% ICD	Definition	n
Aaronson [1] 1997 USA	HFSS	Derivation	Single center	LVEF <40% Age <70 years	1986-1991	268	50	80	20	45	10	n.r.	Death and urgent HTx	109
		Validation	Single centre		1993-1995	199	52	81	22	47	11	n.r.		~60
Zugck [2] 2001 Germany	HFSS	Validation	Single center	NYHA I-III LVEF <40% Age <70 years	1995-1998	208	54	82	22	29	30	n.r.	Death	52
Koelling [3] 2004 USA	HFSS	Validation	Single center	LVEF <40% CP study	1994-1997	320	52	74	23	52	10	11	Death, urgent HTx and VAD	64
					1999-2001	187	54	76	21	56	72	19		30
Parikh [4] 2009 USA	HFSS	Validation	Single center	HF Age >65 years CP study	n.r.	396	70	75	30	50	64	n.r.	Death, urgent HTx and VAD	111
Gorodeski [5] 2010 USA	SHFM HFSS	Validation	Single centre	Referred for HTx assessment	2004-2007	215	55	77	20	55	80	78	Death, urgent HTx and VAD	157

Supplemental Table 2. Continued

Study	Model's name	Derivation/ Validation study	Population										Events	
			Source	Inclusion criteria	Time frame	N	Mean Age	% male	Mean LVEF	% ischemic	% β-blocker	% ICD	Definition	n
Goda [6-8] 2010 USA	HFSS SHFM 3 papers	Validation	Single center	Referred for HTx assessment	1993-2008	715	54	65	22	40	71	49	Death, urgent HTx and VAD	354
Levy [9] 2006 USA	SHFM	Derivation	PRAISE-1 Trial	LVEF <30%	1992-1994	1125	65	76	21	64	0	0	Death, urgent HTx and VAD	403
		Validation[#]	ELITE2 Trial	LVEF <40% Age >60 years	1997-1998	2987	71	69	31	74	24	0		505
			RENAISSA NCE trial	LVEF <30% NYHA II-IV	1999-2001	925	62	78	22	61	61	18		179
			Val-HeFT Trial	LVEF <40% NYHA II-IV	1997-1999	5010	63	80	27	58	34	n.r.		979
			IN-CHF Registry	HF patients	1995-n.r.	872	64	76	35	47	35	n.r.		115
			UW Cohort	HF patients	n.r.	148	53	78	27	34	72	22		48

Supplemental Table 2. Continued

Study	Model's name	Derivation/ Validation study	Population										Events	
			Source	Inclusion criteria	Time frame	N	Mean Age	% male	Mean LVEF	% ischemic	% β-blocker	% ICD	Definition	n
May [10] 2007 USA	SHFM	Validation	Single centre	Hospitalized HF patients	1993-2005	4077	67	61	45	60	77	13	Death, urgent HTx and VAD	2142
Allen [11] 2008 USA	SHFM	Validation	Single centre	HF patients	2004-2008	122	61	62	26	38	86	25	Death	35
Kalogeropoulos [12] Giamouzis [13] 2009 USA	SHFM	Validation	Single centre	LVEF <30% NYHA II-IV	2000-2006	445	52	69	18	38	92	68	Death, urgent HTx and VAD	109
Levy [14] 2009 Atlanta, USA	SHFM	Validation	REMATCH trial	HF non-HTx candidates (medical treatment arm)	1998-2001	61	68	82	17	69	20	35	Death	56
Perrota [15] 2012 Italy	SHFM	Validation	Single centre	NYHA I-III LVEF <35% CRT implant	2000-2007	342	71	79	26	52	73	77	Death and urgent HTx	86

Supplemental Table 2. Continued

Study	Model's name	Derivation/ Validation study	Population										Events	
			Source	Inclusion criteria	Time frame	N	Mean Age	% male	Mean LVEF	% ischemic	% β-blocker	% ICD	Definition	n
Haga [16] 2012 UK	SHFM	Validation	Single centre	NYHA III-IV No HF admissions for 6 weeks	n.r.	138	77	66	n.r.	68	59	n.r	Death	43
Frankenstein [17] 2011 Germany	-	Derivation	Single center	LVEF <40%	1995-2005	636	56	81	28	32	78	n.r	Death	151
		Validation			2001-2005	676	74	76	34	63	54	n.r.		160
Kramer [18] 2012 USA	PACE risk score	Derivation	Multi-center	Primary and secondary prevention ICD patients	2001-2008	905	65	78	31	59	n.r.	100	Death	125
		Validation			2001-2008	1812	64	77	31	58	n.r.	100		296
Bilchick [19] 2012 USA	SHOCKED predictors	Derivation	Multi-center (Medicare database)	Primary prevention ICD patients	2005-2006	17991	n.r.	77	n.r.	59	79	100	Death	6741
		Validation			2005-2007	27893	n.r.	75	n.r.	63	n.r.	100		8595

HFSS, Heart Failure Survival Score; LVEF, left ventricular ejection fraction; HTx, heart transplantation; NYHA, New York Heart Association; CP, cardio-pulmonary; VAD, ventricular assist device; SHFM, Seattle Heart Failure Model; MI; myocardial infarction; PRAISE, Prospective Randomized Amlodipine Survival Evaluation; ELITE2, Losartan Heart Failure Survival Study; RENAISSANCE, Randomized Etanercept North American Strategy to Study Antagonism of Cytokines; IN-CHF, Italian Congestive Heart Failure Registry; UW, University of Washington HF clinic; CRT, cardiac resynchronization therapy; HF, heart failure; ICD, internal cardiac defibrillator; n.r., not reported.

Supplemental Table 3. Assessment of study quality

Study	Derivation Validation	Model	Patient selection	Data collection	Missing data	Loss of follow up
Aaronson 1997 [1]	Derivation	**HFSS**	n.r.	Retrospective	n.r.	1-3%
	Validation	**HFSS**	n.r.	Retrospective	n.r.	1-3%
Zugck 2001 [2]	Validation	**HFSS**	n.r.	Retrospective	n.r.	0%
Koelling 2004 [3]	Validation	**HFSS**	n.r.	Retrospective	0%	0%
Parikh 2009 [4]	Validation	**HFSS**	n.r.	Retrospective	36% of patients excluded	0%
Gorodeski 2010 [5]	Validation	**HFSS**	Consecutive	Retrospective	Peak VO_2 = 36%. Imputed by multiple imputation	n.r.
Goda 2010 [6] and 2011 [7,8]	Validation	**HFSS**	Consecutive	Retrospective	18 patients excluded	0%
Levy 2006 [9]	Derivation PRAISE-1	**SHFM**	RCT	Prospective	n.r.	n.r.
	Validation ELITE2	**SHFM**	RCT	Prospective	n.r.	n.r.
	Validation Val-HeFT	**SHFM**	RCT	Prospective	n.r.	n.r.

Supplemental Table 3. Continued.

Study	Derivation Validation	Model	Patient selection	Data collection	Missing data	Loss of follow up
Levy 2006 [9]	Validation UW	**SHFM**	n.r.	Prospective	n.r.	n.r.
	Validation RENAISSANCE	**SHFM**	RCT	Prospective	n.r.	n.r.
	Validation IN-CHF	**SHFM**	Registry	Prospective	n.r.	n.r.
May 2007 [10]	Validation	**SHFM**	Consecutive	Prospective	NYHA=72% Lymphocytes=35% Uric acid=66% LVEF=25% Cholesterol=20% Imputed using multiple regression	0%
Allen 2008 [11]	Validation	**SHFM**	Consecutive	Prospective	Imputed with the mean	0%

Supplemental Table 3. Continued

Study	Derivation Validation	Model	Patient selection	Data collection	Missing data	Loss of follow up
Kalogeoropoulos [12] and Giamouzis [13] 2009	Validation	**SHFM**	Consecutive	Retrospective	Exclusion of patients with >2 missing variables. The rest were imputed with the mean (lymphocytes=71%).	0%
Levy 2009 [14]	Validation	**SHFM**	RCT	Prospective	Lymphocytes imputed by multiple regression. Uric acid, cholesterol and diuretic dose were imputed from a comparable group of patients from SHFM cohort.	0%
Gorodeski 2010 [5]	Validation	**SHFM**	Consecutive	Retrospective	Uric acid = 64% Cholesterol = 11% Lymphocytes = 10% Imputed by multiple imputation	n.r.
Goda 2011 [8]	Validation	**SHFM**	Consecutive	Retrospective	In 38% patients, imputed with the mean	0%
Perrota 2012 [15]	Validation	**SHFM**	n.r.	Retrospective	Imputed with the mean	n.r.

Supplemental Table 3. Continued.

Study	**Derivation Validation**	**Model**	**Patient selection**	**Data collection**	**Missing data**	**Loss of follow up**
Haga 2012 [16]	Validation	**SHFM**	n.r.	Retrospective	n.r.	n.r.
Frankenstein 2011[17]	Derivation	-	Consecutive	Retrospective	n.r.	n.r.
	Validation		Consecutive	Retrospective	n.r.	n.r.
Kramer 2012 [18]	Derivation	**PACE risk score**	Consecutive	Retrospective	n.r.	n.r.
	Validation		Consecutive	Retrospective	n.r.	n.r.
Bilchick 2012 [19]	Derivation	**SHOCKED predictors**	Consecutive	Prospective	n.r.	n.r.
	Validation		Consecutive	Prospective	n.r.	n.r.

HFSS, Heart Failure Survival Score; peak VO_2, peak oxygen consumption; RCT, randomized controlled trial; SHFM, Seattle Heart Failure Model; PRAISE, Prospective Randomized Amlodipine Survival Evaluation; ELITE2, Losartan Heart Failure Survival Study; RENAISSANCE, Randomized Etanercept North American Strategy to Study Antagonism of Cytokines; IN-CHF, Italian Congestive Heart Failure Registry; UW, University of Washington HF clinic; LVEF, left ventricular ejection fraction; n.r., not reported.

References of Supplemental Tables 2 and 3:

1. Aaronson KD, Schwartz JS, Chen T-M, Wong K-L, 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;95:2660–7.
2. Zugck C, Kruger C, Kell R, Korber S, Schellberg D, Kubler W, Haass M.. Risk stratification in middle-aged patients with congestive heart failure: prospective comparison of the Heart Failure Survival Score (HFSS) and a simplified two-variable model. Eur J Heart Fail 2001;3:577-85.
3. Koelling TM, Joseph S, Aaronson KD. Heart Failure Survival Score continues to predict clinical outcomes in patients with heart failure receiving β-blockers. J Heart Lung Transplant 2004;23:1414-22.
4. Parikh MN, Lund LH, Goda A, Mancini D. Usefulness of peak exercise oxygen consumption and the heart failure survival score to predict survival in patients >65 years of age with heart failure. Am J Cardiol 2009;103:998-1002.
5. 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;3:706-14.
6. Goda A, Lund LH, Mancini D. Comparison across races of peak oxygen consumption and heart failure survival score for selection for cardiac transplantation. Am J Cardiol 2010; 105:1439-44.
7. Goda A, Lund LH, Mancini D. The Heart Failure Survival Score outperforms the peak oxygen consumption for heart transplantation selection in the era of device therapy. J Heart Lung Transplant 2011;30:315-25.
8. Goda A, Williams P, Mancini D, Lund LH. Selecting patients for heart transplantation: comparison of the Heart Failure Survival Score (HFSS) and the Seattle heart failure model (SHFM). J Heart Lung Transplant 2011;30:1236-43.
9. 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;113:1424–33.
10. May HT, Horne BD, Levy WC, Kfoury AG, Rasmusson KD, Linker DT, Mozaffarian D, Anderson JL, Renlund DG. Validation of the Seattle Heart Failure Model in a community-based heart failure population and enhancement by adding B-type natriuretic peptide. Am J Cardiol 2007;100:697-700.
11. Allen LA, Yager JE, Funk MJ, Levy WC, Tulsky 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;299:2533-42.
12. Kalogeropoulos AP, Georgiopoulou VV, Giamouzis G, Smith AL, Agha SA, Waheed S, Laskar S, Puskas J, Dunbar SA, Vega, D, Levy WC, Butler J. Utility of the Seattle Heart Failure Model in patients with advanced heart failure. J Am Coll Cardiol 2009;53:334-42.
13. Giamouzis G, Kalogeropoulos AP, Georgiopoulou VV, Agha SA, Rashad MA, Laskar SR, Smith AL, Butler J. Incremental value of renal function in risk prediction with the Seattle Heart Failure Model. Am Heart J 2009;157:299-305.

14. Levy WC, Mozaffarian D, Linker DT, Farrar DJ, Miller LW, REMATCH Investigators. 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;28:231-6.
15. Perrotta L, Ricciardi G, Pieragnoli P, Chiostri M, Pontecorboli G, De Santo T, Bellocci F, Vitulano N, Endin M, Mascioli G, Ricceri I, Porciani MC, Michelucci A, Padeletti L. Application of the Seattle Heart Failure Model in patients on cardiac resynchronization therapy. PACE 2012;35:88-94.
16. Haga K, Murray S, Reid J, Ness A, O'Donnell M, Yellowlees D, Denvir MA. Identifying community based chronic heart failure patients in the last year of life: a comparison of the Gold Standards Framework Prognostic Indicator Guide and the Seattle Heart Failure Model. Heart 2012;98:579-83.
17. Frankenstein L, Goode K, Ingle L, Remppis A, Schellberg D, Nelles M, Katus HA, Clark AL, Cleland JG, Zugck C. Derivation and validation of a simple clinical risk-model in heart failure based on 6 minute walk test performance and NT-proBNP status--do we need specificity for sex and beta-blockers? Int J Cardiol 2011;147:74-8.
18. Kramer DB, Friedman PA, Kallinen LM, Morrison TB, Crusan DJ, Hodge DO, Reynolds MR, Hauser RG.. Development and validation of a risk score to predict early mortality in recipients of an implantable Cardioverter-defibrillator. Heart Rhythm 2012;9:42-6.
19. Bilchick KC, Stukenborg GJ, Kamath S, Cheng A. Prediction of mortality in clinical practice for Medicare patients undergoing defibrillator implantation for primary prevention of sudden cardiac death. J Am Coll Cardiol 2012;60:1647-55.

Supplemental Table 4. Characteristics of the population of studies included

Study	Model name	Derivation/ Validation study	Population										Events	
			Source	Inclusion criteria	Time frame	N	Mean Age	% male	Mean LVEF	% ischemic	% β-blocker	% ICD	Definition	n
Kearney 2003 [1] UK	-	Derivation	Heart study	Clinically diagnosed HF NYHA I-III	1993-1995	553	63	76	42	79	8	n.r.	Death	201
Rickli 2003 [2] Switzerland	-	Derivation	Single center	LVEF<40% CP study	n.r.	202	52	86	28	53	45	n.r	Death and urgent HTx	59
Adlam 2005 [3] UK	-	Derivation	Single centre	Clinically diagnosed HF	1995-1998	532	75	41	45	41	14	n.r.	Death	190
Pocock 2006 [4] UK	CHARM	Derivation	CHARM trial	Clinically diagnosed HF	1999-2003	7599	65	68	39	57	n.r.	n.r.	Death	1831
Myers 2008 [5] Italy	CPX score	Derivation	Multi-center	Clinically diagnosed HF	1993-2007	710	56	80	34	39	63	n.r.	Death, urgent HTx and VAD *	110

Supplemental Table 4. Continued.

Study	Model name	Derivation/ Validation study	Population										Events	
			Source	Inclusion criteria	Time frame	N	Mean Age	% male	Mean LVEF	% ischemic	% β-blocker	% ICD	Definition	n
Huynh 2008 [6] USA	-	Derivation	Single center	HF patients Age >70 years	1990-1994	282	80	34	42	54	n.r.	n.r.	Death	43
Wedel 2009 [7] Europe	CORONA score	Derivation	CORONA trial	LVEF <40% NYHA II-IV	2003-2005	3342	72	73	32	100	78	2.3	Death *	934
Leyva 2009 [8] UK	DSC index	Derivation	Single center	LVEF<35% NYHA III-IV CRT implant	2001-2008	148	68	77	23	62	55	0	CV Death	37
Vazquez 2009 [9] Spain	MUSIC score	Derivation	Multi-centre	Clinically diagnosed HF NYHA II-IV	2003-2004	992	65	72	37	46	68	n.r.	Death *	267
Komajda 2011 [10] France	-	Derivation	I-PRESERVE trail	LVEF >45% NYHA II-IV Age >50 years	2003-2007	4128	72	40	59	25	n.r.	n.r.	Death *	881

Supplemental Table 4. Continued

Study	Model's name	Derivation/ Validation study	Population										Events	
			Source	Inclusion criteria	Time frame	N	Mean Age	% male	Mean LVEF	% ischemic	% β-blocker	% ICD	Definition	N
Subramanian 2011 [11] USA	VEST score	Derivation	VEST trail	LVEF <30% NYHA III-IV	1995-1996	963	62	78	21	57	n.r.	n.r.	Death *	172
O'Connor 2012 [12] USA	HF-ACTION score	Derivation	HF-ACTION trail	LVEF <35% NYHA II-IV	2003-2007	2331	59	72	25	54	95	40	Death *	387
Herrmann 2012 [13] UK		Derivation	Single centre	LVEF <40% HF symptoms	n.r.	114	63	n.r.	29	n.r.	4	n.r.	Death	31
Scrutinio 2012 [14] Italy		Derivation	Single centre	LVEF <40% HF symptoms	2001-2007	802	64	79	28	50	73	n.r.	Death	301
Pocock 2012 [15] Europe		Derivation	Multi-centre	Clinically diagnosed HF	n.r.	39372	67	67	35	53	34	n.r.	Death	15851

HF, heart failure; NYHA, New York Heart Association; CP, cardio-pulmonary; LVEF, left ventricular ejection fraction; HTx, heart transplantation; VAD, ventricular assist device; CV, cardiovascular; n.r., not reported.

Supplemental Table 5. Assessment of study quality

Study	**Derivation Validation**	**Model**	**Patient selection**	**Data collection**	**Missing data**	**Loss of follow up**
Kearney 2003 [1]	Derivation		n.r.	Prospective	Multiple regression	n.r.
Rickli 2003 [2]	Derivation		Consecutive		n.r.	n.r.
Adlam 2005 [3]	Derivation		Consecutive	Prospective	Excluded	0%
Pocock 2006 [4]	Derivation	**CHARM**	RCT cohort	Prospective	n.r.	n.r.
Myers 2008 [5]	Derivation	**CPX score**	n.r.	Prospective	n.r.	n.r.
Huynh 2008 [6]	Derivation		RCT cohort	Prospective	n.r.	n.r.
Wedel 2009 [7]	Derivation	**CORONA**	RCT cohort	Prospective	Excluded	n.r.
Leyva 2009 [8]	Derivation	**DSC index**	Consecutive	Prospective	0%	0%
Vazquez 2009 [9]	Derivation	**MUSIC score**	Consecutive	Prospective	Imputed with the mean	1.1%
Komajda 2011 [10]	Derivation		RCT cohort	Prospective	Excluded	n.r.
Subramanian 2011 [11]	Derivation	**VEST**	RCT cohort	Prospective	19% of patients excluded	n.r.

Supplemental Table 5. Continued

Study	**Derivation Validation**	**Model**	**Patient selection**	**Data collection**	**Missing data**	**Loss of follow up**
O'Connor 2012 [12]	Derivation	**HF-ACTION**	RCT cohort	Prospective	Hemoglobin= 24% Urea= 13% Sodium= 11% Creatinine= 10% MR= 8% Multiple imputation	n.r.
Herrmann 2012 [13]	Derivation		n.r.	Prospective	n.r.	n.r.
Scrutinio 2012 [14]	Derivation		Consecutive	Prospective	0%	0%
Pocock 2012 [15]	Derivation		Meta-analysis on RCT and observational studies	Prospective and retrospective	Multiple imputation	0%

LVEF, left ventricular ejection fraction; ICD, internal cardiac defibrillator; HFSS, Heart Failure Survival Score; HTx, heart transplantation; VAD, ventricular assist device; NYHA, New York Heart Association; MFH; metabolic, functional, hemodynamic; CPX, cardiopulmonary exercise test; MRT, mean response time; SHFM, Seattle Heart Failure Model; MI; myocardial infarction; DSC, Dyssynchrony, posterolateral Scar location and Creatinine; CRT, cardiac resynchronization therapy; CV, cardiovascular; n.r., not reported.

Supplemental Table 6. Model derivation and performance

Study	Derivation Validation	Model/ Variables	Selection	Linear Gradient	Over-fitting	Model assumptions	Calibration	Discrimination (c-statistic)
Zugck 2001 [15]	Derivation	• LVEF • Peak VO_2 or 6'WT	n.r.	n.r.	No	n.r.	n.r.	Overall = 0.84 (0.80-0.88) or 0.83 (0.79-0.87)
Kearney 2003 [1]	Derivation	• Sodium • Creatinine • CT ratio • QRS dispersion • QT • Non-sustained VT • LVH by ECG • SDNN	Based on univariable analysis	n.r.	Yes (201 events and 30 variables tested)	Held	n.r.	* Binary predictors= 0.74 (0.70-0.78) Continuous predictors= 0.78 (0.74-0.82)
	Validation by bootstrap	**Kearney 2003**	n/a	n/a	n/a	n/a	n.r	n.r.

Supplemental Table 6. Continued.

Study	Derivation Validation	Model/ Variables	Selection	Linear Gradient	Over- fitting	Model assumptions	Calibration	Discrimination (c-statistic)
Rickli 2003 [2]	Derivation	• Predicted peak VO_2 • MRT >50 seconds • Systolic BP	Based on univariable analysis	n.r.	No	n.r.	n.r.	At 1 year=0.86 (0.82-0.90)
Adlam 2005 [3]	Derivation	• BNP • Age • Sex • Diabetes • CVA • Abnormal ECG	Based on univariable analysis using bootstrap estimated	n.r.	No	Held	n.r.	Overall = 0.76
	Validation by bootstrap	**Adlam 2005**	n/a	n/a	n/a	n/a	n.r.	Overall = 0.75

Supplemental Table 6. Continued.

Study	Derivation Validation	Model/ Variables	Selection	Linear Gradient	Over- fitting	Model assumptions	Calibration	Discrimination (c-statistic)
Pocock 2006 [4]	Derivation	**CHARM:** • Age • Sex • Diabetes • LVEF • NYHA • Cardiomegalia • Time HF diagnose • Prior HF admission • BMI • Diastolic BP • Smoking • BBB • Previous MI	Probably on clinical importanc e. Forward selection	n.r.	No	n.r.	Graphically observed vs. predicted survival by deciles. Under-estimated survival at 3 years	At 2 years = 0.75 In preserved EF = 0.74 In low-EF=0.76

		• Pulmonary crackles • Edema • Pulmonary edema • Heart Rate • Mitral regurgitation • Atrial fibrillation • Rest dyspnea • Candesartan						
	Validation by bootstrap	**CHARM**	n/a	n/a	n/a	n/a	n.r.	At 2 years = 0.75

Supplemental Table 6. Continued.

Study	Derivation Validation	Model/ Variables	Selection	Linear Gradient	Over-fitting	Model assumptions	Calibration	Discrimination (c-statistic)
Myers 2008 [5]	Derivation	**CPX score:** • OUES>1.4 • VE/VCO2 >34 • peak VO2<14 • HR recovery <6 beats at 1minute • PetCO2 <33mmHg	Not clearly stated	n.r.	No	Held	n.r.	n.r.
	Validation by bootstrap	**CPX score**	n/a	n/a	n/a	n/a	n.r.	‡ Overall = 0.77
Huynh 2008 [6]	Derivation	• Urea • Systolic BP • PVD • Sodium	Based on univariable analysis.	n.r.	Yes (43 events and 15 variables)	n.r.	n.r.	At 6 months=0.80

Supplemental Table 6. Continued.

Study	Derivation Validation	Model/ Variables	Selection	Linear Gradient	Over-fitting	Model assumptions	Calibration	Discrimination (c-statistic)
	Validation by bootstrap	**Huynh 2008**	n/a	n/a	n/a	n/a	n.r.	n.r.
Wedel 2009 [7]	Derivation	**CORONA:** • BNP • Age • Diabetes • LVEF • BMI • Sex • CABG • Atrial fibrillation • NHYA • Apo-A1 • Creatinine • PVD	Not clearly stated	n.r.	No	n.r.	n.r.	Overall mortality=0.72 HF mortality=0.80

		• Heart rate • MI						
Leyva 2009 [8]	Derivation	**DSC index:** • Dyssynchrony • Scar location • Creatinine	Based on previous reports	Checked by martingale residuals	No	Held	Correlation (r=0.93)	At 1 year = 0.88 At 1 year = 0.87
	Validation by bootstrap	**DSC index**	n/a	n/a	n/a	n/a	****	Overall=0.85
Vazquez 2009 [9]	Derivation	**MUSIC score:** • Prior MI, stroke or limb ischemia • Left atrium size>26mm/m2 • LVEF<35% • LBBB or IVCD (QRS>110) • non-sustained VT or frequent	Based on previous knowledge and <5% missing data	n.r.	No	n.r	Correlation (r=0.99)	Overall mortality=0.76 Cardiac mortality=0.78 HF mortality=0.80 Sudden death=0.77

		extra-beats • GFR <60ml/min • BNP>1000pg/dl • Troponin posit • Sodium <138meq/L						
	Validation by bootstrap	**MUSIC score**	n/a	n/a	n/a	n/a	n.r.	Overall mortality=0.77 Cardiac mortality=0.78 HF mortality=0.80 Sudden death=0.78

Supplemental Table 6. Continued.

Study	Derivation Validation	Model/ Variables	Selection	Linear Gradient	Over-fitting	Model assumptions	Calibration	Discrimination (c-statistic)
Komajda 2011 [10]	Derivation	• BNP • Age • Diabetes • LVEF • Heart rate • Previous hospital admission • Quality of life • COPD or asthma • Ischemic CMP • MI	Based on univariable analysis	n.r.	No	n.r.	Graphically observed vs. predicted = Adequate	Overall=0.74
	Validation by bootstrap	**Kornajda 2011**	n/a	n/a	n/a	n/a	n.r.	Overall=0.74

Supplemental Table 6. Continued.

Study	**Derivation Validation**	**Model/ Variables**	**Selection**	**Linear Gradient**	**Over-fitting**	**Model assumptions**	**Calibration**	**Discrimination (c-statistic)**
Subramanian 2011 [11]	Derivation	**VEST:** Model:1 • BUN • LVEF • Lymphocytes • CT radio Model 2: 1+ • TNFR • Interleukin 6 Model 3: 2+ • Serial measurement of cytokines	Based on univariable analysis	n.r.	Yes (172 events and 19 variables tested)	n.r.	n.r.	Overall= Model 1: 0.73 Model 2: 0.74 Model 3: 0.81

Supplemental Table 6. Continued.

Study	**Derivation Validation**	**Model/ Variables**	**Selection**	**Linear Gradient**	**Over-fitting**	**Model assumptions**	**Calibration**	**Discrimination (c-statistic)**
O'Connor 2012 [12]	Derivation	**HF-ACTION:** • Exercise duration • Urea • Sex • BMI	Based on univariable analysis	Checked by restrictive cubic spline	No	n.r.	Correlation (r=0.99 at 1,2 and 3 years and 0.98 at 5 years)	Overall=0.73
Herrmann 2012 [13]	Derivation	• Peak VO_2 <14ml/kg/min • Uric acid >565µmol/L • LVEF<22% • Cholesterol <5.27mmol/L • sTNF-R1 >1016pg/L	Based on previous knowledge	n.r.	Yes (31 deaths and 5 variables tested)	n.r.	n.r.	† Overall=0.91

Supplemental Table 6. Continued.

Study	**Derivation Validation**	**Model/ Variables**	**Selection**	**Linear Gradient**	**Over-fitting**	**Model assumptions**	**Calibration**	**Discrimination (c-statistic)**
Scrutinio 2012 [14]	Derivation	• Age • Ischemic CMP • Anemia • LVEF • Renal function	Based on univariable analysis	n.r.	No	n.r.	H-L test (p>0.45)	Overall=0.74
Pocock 2012 [15]	Derivation	• Age • Gender • BMI • Current smoker • Systolic BP • Diabetes • NYHA class • LVEF • COPD • HF duration	Based on statistical significance	n.r.	No	n.r.	Graphically observed vs. predicted = Adequate	n.r.

		• Creatinine • β-blockers • ACE-I/ARB						

* This model was validated by bootstrapping but discrimination capacity on bootstrapping is not reported.

‡ Authors conducted a subgroup analysis based on underlying etiology and LVEF and reported that c-index was equal in ischemic, non-ischemic CMP and patients with LVEF <30%, but lower (c-statistic = 0.73) in patients with LVEF ≥30%.

† Authors reported that a model excluding cholesterol has similar c-statistic and that a model including uric acid, sTNF-R1, LVEF and NYHA class (<3) instead of peak VO_2 had an overall c-statistic of 0.84.

LVEF, left ventricular ejection fraction; VO2, oxygen consumption; CT, cardio-thoracic; VT, ventricular taqui-arrhythmia; LVH, left ventricular hypertrophy; ECG, electro-cardiogram; SDNN, standard deviation of all R-to-R intervals on 24-h; MRT, mean response time; BP, blood pressure; CVA, cerebro-vascular accident; NYHA, New York Heart Association; BMI, body mass index; BBB, bundle branch block; MI, myocardial infarction; PVD, peripheral vascular disease; ICD, internal cardiac defibrillator; MFH; metabolic, functional, hemodynamic; CPX, cardiopulmonary exercise test; MRT, mean response time; MI; myocardial infarction; DSC, Dyssynchrony, posterolateral Scar location and Creatinine; CRT, cardiac resynchronization therapy; CV, cardiovascular; BNP, brain natriuretic peptide; COPD, chronic obstructive pulmonary disease; CMP, cardiomyopathy; sTNF-R1, soluble tumor necrosis factor alpha receptor 1; H-L, Hosmer and Lemeshow; ACE-I, angiotensin converting enzyme inhibitor; ARB, angiotensin receptor blocker; n.r., not reported; n/a, not applicable.

References of Supplemental tables 4, 5 and 6

1. Kearney MT, Nolan J, Lee AJ, Brooksby PW, Prescott R, Shah AM, Zaman AG, Eckberg DL, Lindsay HS, Batin PD, Andrews R, Fox KA. A prognostic index to predict long-term mortality in patients with mild to moderate chronic heart failure stabilised on angiotensin converting enzyme inhibitors. Eur J Heart Fail 2003;5:489-97.
2. Rickli H, Kiowski W, Brehm M, Weilenmann D, Schalcher C, Bernheim A, Oechslin E, Brunner-La Rocca HP. Combining low-intensity and maximal exercise test results improves prognostic prediction in chronic heart failure. J Am Coll Cardiol 2003; 42:116-22.
3. Adlam D, Silcocks P, Sparrow N. Using BNP to develop a risk score for heart failure in primary care. Eur Heart J 2005;26:1086-93.
4. Pocock SJ, Wang D, Pfeffer MA, Yusuf S, McMurray JJ, Swedberg KB, Ostergren J, Michelson EL, Pieper KS, Granger CB. Predictors of mortality and morbidity in patients with chronic heart failure. Eur Heart J 2006;27:65-75.
5. Myers J, Arena R, Dewey F, Bensimhon D, Abella J, Hsu L, Chase P, Guazzi M, Peberdy MA. A cardiopulmonary exercise testing score for predicting outcomes in patients with heart failure. Am Heart J 2008;156:1177-83.
6. Huynh BC, Rovner A, Rich MW. Identification of older patients with heart failure who may be candidates for hospice care: development of a simple four-item risk score. J Am Geriatr Soc 2008;56:1111-5.
7. Wedel H, McMurray JJ, Lindberg M, Wikstrand J, Cleland JG, Cornel JH, Dunselman P, Hjalmarson A, Kjekshus J, Komajda M, Kuusi T, Vanhaecke J, Waagstein F. CORONA Study Group. Predictors of fatal and non-fatal outcomes in the Controlled Rosuvastatin Multinational Trial in Heart Failure (CORONA): incremental value of apolipoprotein A-1, high-sensitivity C-reactive peptide and N-terminal pro B-type natriuretic peptide. Eur J Heart Fail 2009;11:281-91.
8. Leyva F, Foley PW, Stegemann B, Ward JA, Ng LL, Frenneaux MP, Regoli F, Smith RE, Auricchio A. Development and validation of a clinical index to predict survival after cardiac resynchronisation therapy. Heart 2009;95:1619-25.

9. Vazquez R, Bayes-Genis A, Cygankiewicz I, Pascual-Figal D, Grigorian-Shamagian L, Pavon R, Gonzalez-Juanatey JR, Cubero JM, Pastor L, Ordonez-Llanos J, Cinca J, de Luna AB. MUSIC Investigators. The MUSIC Risk score: a simple method for predicting mortality in ambulatory patients with chronic heart failure. Eur J Heart Fail 2009;30:1088-96.
10. Komajda M, Carson PE, Hetzel S, McKelvie R, McMurray J, Ptaszynska A, Zile MR, Demets D, Massie BM. Factors associated with outcome in heart failure with preserved ejection fraction: findings from the Irbesartan in Heart Failure with Preserved Ejection Fraction Study (I-PRESERVE). Circ Heart Fail 2011;4:27-35.
11. Subramanian D, Subramanian V, Deswal A, Mann DL. New predictive models of heart failure mortality using time-series measurements and ensemble models. Circ Heart Fail 2011;4:456-62.
12. 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;5:63-71.
13. Herrmann R, Sandek A, von Haehling S, Doehner W, Schmidt HB, Anker SD, Rauchhaus M. Risk stratification in patients with chronic heart failure based on metabolic-immunological, functional and haemodynamic parameters. International J Cardiol 2012;156:62-8.
14. Scrutinio D, Passantino A, Catanzaro R, Guida P. Clinical utility of different estimates of renal function for predicting mortality in chronic heart failure. International J Cardiol 2012;157:24-30.
15. Pocock SJ, Ariti CA, McMurray JJV, Maggioni A, Kober L, Squire IB, Swedberg K, Dobson J, Poppe KK, Whaley GA, Doughty RN. Predicting survival in heart failure: a risk score based on 39372 patients from 30 studies. Eur Heart J 2012; doi: 10.1093/eurheartj/ehs337.