Validation of the Meta-Analysis Global Group in Chronic Heart Failure risk score for the prediction of 1-year mortality in a Chinese cohort

Duo Xu[1,2], Yanjing Dong[2], Bo Zhang[2], Li Li[2], Chenyang Jiang[1]

[1]Department of Cardiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 3100016, China;
[2]Department of Cardiology, CHC International Hospital, Cixi, Zhejiang 315310, China.

Abstract

Background: The Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC) risk score was developed in 2013 to predict survival in heart failure (HF) patients. However, it has yet to be validated in a Chinese population. Our study aimed to investigate the ability of the score to predict 1-year mortality in a Chinese population.

Methods: Consecutive patients with HF were retrospectively selected from the inpatient electronic medical records of the cardiology department in a regional hospital in China. A total integer score was calculated for each enrolled patient based on the value of each risk factor in the MAGGIC scoring system. Each enrolled patient was followed for at least 1 year. The observational endpoint of this study was all-cause mortality. The predictive ability of the MAGGIC score was assessed by comparing observed and predicted mortality within 1 year.

Results: Between January 2018 and December 2020, a total of 635 patients were included in the study: 57 (9.0%) of whom died within 1 year after discharge. The average age of all patients was 74.6 ± 11.2 years, 264 of them (41.6%) were male, and the average left ventricular ejection fraction was $50.7\% \pm 13.2\%$. The area under the receiver operating characteristic curve was 0.840 (95% confidence interval: 0.779, 0.901), which indicated a fair discriminatory ability of the score. The Hosmer–Lemeshow test result ($\chi^2 = 12.902$, degree of freedom = 8, $P = 0.115$) indicated that the MAGGIC score had good calibration. The decision curve analysis showed that the MAGGIC score yielded a good clinical net benefit and net reduction in interventions.

Conclusions: This validation of the MAGGIC score showed that it has a good ability to predict 1-year mortality in Chinese patients with HF after discharge. Due to regional and inter-hospital differences, external validation studies need to be further confirmed in other centers.

Keywords: Validation; Risk prediction; MAGGIC; Heart failure; Mortality; Chinese

Introduction

Heart failure (HF) is the severe manifestation or the final stage of various heart diseases. To date, HF remains the leading cause of cardiovascular-related death, despite recent advances in drugs, devices, and care strategies that have led to a decline in mortality.[1] Patients with HF often have a variety of comorbidities and complications. Previous studies have shown that there are many risk factors for short- and long-term mortality in patients with HF.[1-3] Multiple HF risk prediction models have been developed to help clinicians calibrate patient expectations about the future and communicate with patients and to help patients make rational decisions.[4-7] However, patient profiles vary widely across countries and hospitals. Therefore, there is uncertainty about the applicability of risk prediction models in other populations.

The Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC) score was developed in 2013 to predict survival in HF patients. It was based on 39,372 patients with HF, both those with reduced and preserved left ventricular ejection fraction (LVEF), from 30 cohort studies, where six of which were clinical trials.[6,8] The easy-to-use website www.heartfailurerisk.org allows an integer score to be quickly calculated that indicates the risk of the given patient dying within 1 to 3 years. The MAGGIC score was established based on data collected from 1980 to 2006; therefore, it might not be indicative of current or future trends in HF management. The 1- and 3-year risk of mortality could be underestimated or overestimated. In recent studies, the MAGGIC risk score has been validated in specific regional populations and has been shown to be a good tool for predicting all-cause mortality.[9-13] However, it has yet to be validated in a

---

Access this article online

Quick Response Code:

Website:
www.cmj.org

DOI:
10.1097/CM9.0000000000002026

Chinese population. Our study aimed to evaluate its ability to predict post-discharge 1-year mortality in hospitalized HF patients and to provide a basis for its application in a Chinese population.

Methods

Ethical approval

This study was approved by the Ethics Committee of CHC International Hospital (No. 2021-KY-06). Because of the retrospective character of the study and desensitization data, the requirement for obtaining informed consent was waived.

Study design

Between January 2018 and December 2020, 695 consecutive patients were retrospectively selected from the inpatient electronic medical records of the cardiology department of CHC International Hospital, a five hundred-bed general district hospital located in eastern China. The inclusion criteria were primary hospital discharge diagnosis of HF and age >18 years. In addition, the enrolled patients met the 2016 European Society of Cardiology diagnostic criteria for HF, based on a relevant clinical history, physical examination, electrocardiography, natriuretic peptide levels (NPs), and cardiac ultrasonography.[1] Clinical histories included the history of coronary artery disease, arterial hypertension, exposure to cardiotoxic drug/radiation, use of diuretics, or/and orthopnoea/paroxysmal nocturnal dyspnoea. Physical examinations included rales, bilateral ankle edema, heart murmur, jugular venous dilatation, or/and laterally displaced/broadened apical beat. Electrocardiography presented any abnormality. As long as one or more of the above three conditions, HF could be confirmed by combining elevated levels of NPs (brain natriuretic peptide [BNP] >35 pg/mL and/or N-terminal pro brain natriuretic peptide >125 pg/mL) and results of ultrasonography. An HF with preserved ejection fraction was defined as LVEF ≥50%, an HF with mild range ejection fraction as LVEF of 40% to 49%, and an HF with reduced ejection fraction as LVEF <40%.

If a patient had multiple hospital admissions between 2018 and 2020, the data from the first admission for HF were used to calculate the risk score and the time from discharge to event. Because the MAGGIC score is calculated using 13 variables, patients who had relevant data missing from the electronic medical records were excluded. HF patients presenting with acute coronary syndrome, hemodialysis, or malignant tumors were excluded.

Patient selection and calculation of the MAGGIC score

Patients were selected according to the above inclusion and exclusion criteria, and the MAGGIC score was calculated for each patient. The predicted probability of 1-year mortality according to the MAGGIC score was ascertained based on the following 13 clinical variables: LVEF, age, systolic blood pressure (SBP), body mass index (BMI), creatinine, New York Heart Association (NYHA) functional class, male sex, current smoking status, diabetes, diagnosis of chronic obstructive pulmonary disease (COPD), the first diagnosis of HF in the past 18 months, no use of beta-blockers, and no use of angiotensin-converting enzyme inhibitors/angiotensin receptor blockers (ACEIs/ARBs). In general, the LVEF was calculated by the M-mode ultrasound method. When the left ventricular lumen was too large, too small, deformed, or had obvious segmental movement abnormalities, the modified biplanar Simpson method was used to calculate the LVEF.[14] The SBP values were derived from the patient's vital signs at discharge. The BMI values were taken from the physical examination records at admission. If a patient had smoked within 1 month before admission, the patient was considered to be a current smoker. The time of the first diagnosis of HF and HF duration were not recorded in our case database; therefore, patients with a previous history of hospitalization for HF in the past 18 months were regarded as having HF. The use of beta-blockers and ACEIs/ARBs was determined by checking the discharge medications.

Endpoint of observation

The observational endpoint of this study was 1-year mortality. The estimated risk of mortality was based on the MAGGIC score, and the actual outcome was obtained from the follow-up data available in the medical records. Patients discharged from the hospital were routinely followed up by appointment or telephone call at 1, 3, 6 months, and 1 year after discharge. Death within 1 year was defined as all-cause mortality occurring within 1 year after discharge from the initial hospitalization for HF. Survival status was confirmed in the follow-up records.

Statistical analysis

IBM SPSS Statistics, version 26.0 (International Business Machines Corp., Armonk, NY, USA, RRID: SCR_019096), STATA, version 16.0 (StataCorp LLC, TX, USA, RRID: SCR_012763), and PRISM, version 8.0 (GraphPad Software, San Diego, CA, USA, RRID: SCR_005375) were used to analyze the data and generate the graphics. Continuous variables are expressed as the means ± standard deviations (SDs) or median (interquartile), and classified variables are expressed as percentages. First, differences in baseline clinical variables were compared between surviving and non-surviving patients using the chi-squared test, the continuity correction chi-squared test, or Fisher's exact test for categorical variables and Student's t test, the separate variance estimation t test, or the Mann–Whitney U test for continuous variables, as appropriate. Homogeneity of variance was examined using the F-test. Second, the final score for each patient was calculated according to the MAGGIC model, and then, the histogram showing the distribution of scores and the estimated $vs.$ actual mortality curve was plotted. Third, a receiver operating characteristic (ROC) curve was used to evaluate the discriminatory ability of the model. The larger the area under the curve (AUC), the greater the discriminatory ability of the model. In general, $0.5 < \text{AUC} \leq 0.7$ indicates a low predictive value of the model, $0.7 < \text{AUC} \leq 0.9$ indicates a moderate prediction value, and AUC > 0.9 indicates a high predictive value. The

Hosmer–Lemeshow good-of-fit test was used to evaluate the calibration. The smaller the chi-squared value is, the larger the corresponding *P* value, which indicated better calibration of the prediction model. If the test result showed significance (*P* < 0.05), it indicated that there was a difference between the predicted and observed outcomes, and the model calibration was poor. Decision curve analysis (DCA) was performed to calculate the clinical net benefit and net reduction in interventions. The theoretical correlation between the threshold probability of death and the relative value of false-positive and false-negative results was used to determine the value of the prediction model. All tests were two-sided, and a *P* value < 0.05 indicated statistical significance.

Results

Characteristics of the enrolled patients

A total of 695 patients were originally identified from the electronic medical records; 60 of whom were excluded due to a variety of reasons. Of those 60 patients, 25 patients were excluded because of the missing data, which was necessary for calculating the total scores [Supplementary Figure 1, http://links.lww.com/CM9/A961]. Ultimately, 635 patients were included. Of the included patients, 57 (9.0%) died within 1 year after discharge. The baseline clinical characteristics of the patients are shown in Table 1. In brief, the average age of all patients was 74.6 ± 11.2 years, 264 (41.6%) were male, the average LVEF was 50.7% ± 13.2%, the median (interquartile) serum creatinine level was 80.1 (62.7−108.9) μmol/L, 258 (40.6%) were classified as NYHA class IV, and 154 (24.3%) had diabetes mellitus. A comparison between the two groups revealed that the non-surviving group had a higher proportion of men (66.7% *vs.* 39.1%, $\chi^2 = 16.231$, *P* < 0.001), the elderly (78.9 ± 8.7 *vs.* 74.2 ± 11.4 years, $t = 3.016$, *P* = 0.003), a higher proportion of patients with COPD (22.8% *vs.* 8.7%, $\chi^2 = 11.635$, *P* = 0.001), higher creatinine levels (137.6 [86.9−220.8] *vs.* 77.5 [62.3−104.0] μmol/L, $Z = −5.934$, *P* < 0.001), and lower LVEF (46.0% ± 12.9% *vs.* 51.1% ± 13.2%, $\chi^2 = 2.818$,

Table 1: Baseline clinical characteristics of the study cohort.

Variables	All (*n* = 635)	Survival (*n* = 578)	Death (*n* = 57)	Statistics value	*P* value
Male	264 (41.6)	226 (39.1)	38 (66.7)	16.231[†]	<0.001
Age (years)	74.6 ± 11.2	74.2 ± 11.4	78.9 ± 8.7	3.016[*]	0.003
BMI (kg/m^2)	21.6 ± 4.0	21.7 ± 4.0	20.4 ± 3.4	0.016[*]	0.191
HF duration >18 months	180 (28.3)	170 (29.4)	10 (17.5)	3.598[†]	0.058
Current smoking	12 (1.9)	11 (1.9)	1 (1.8)	0.000[†]	1.000
Current drinking	31 (4.9)	28 (4.8)	3 (5.3)	0.000[†]	1.000
Hypertension	407 (64.1)	376 (65.1)	31 (54.4)	2.565[†]	0.109
Diabetes mellitus	154 (24.3)	138 (23.9)	16 (28.1)	0.497[†]	0.481
Cerebrovascular disease	17 (2.7)	15 (2.6)	2 (3.5)	0.000[†]	1.000
Liver disease	19 (3.0)	14 (2.4)	5 (8.8)	5.185[†]	0.023
COPD	63 (9.9)	50 (8.7)	13 (22.8)	11.635[†]	0.001
NYHA class				17.848[†]	<0.001
II	36 (5.7)	35 (6.1)	1 (1.8)		
III	341 (53.7)	323 (55.9)	18 (31.6)		
IV	258 (40.6)	220 (38.1)	38 (66.7)		
Drugs					
β-blocker	532 (83.8)	484 (83.7)	48 (84.2)	0.009[†]	0.926
ACEI/ARB	544 (85.7)	498 (86.2)	46 (80.7)	1.259[†]	0.262
Diuretic	556 (87.6)	504 (87.2)	52 (91.2)	0.774[†]	0.379
SBP (mmHg)	133.7 ± 25.0	134.4 ± 24.8	126.8 ± 26.8	2.182[*]	0.029
DBP (mmHg)	78.3 ± 14.1	78.6 ± 14.1	75.4 ± 13.8	1.608[*]	0.108
NT-proBNP	710.6 (268.5−1722.3)	679.3 (262.8−1623.9)	970.4 (410.9−4319.9)	−2.832[‡]	0.005
Creatinine (μmol/L)	80.1 (62.7−108.9)	77.5 (62.3−104.0)	137.6 (86.9−220.8)	−5.934[‡]	<0.001
Hemoglobin (g/L)	118.0 (102.0−131.0)	119.0 (103.0−132.0)	108 (84.0−123.5)	−3.068[‡]	0.002
Potassium (mmol/L)	3.91 ± 0.65	3.89 ± 0.65	4.13 ± 0.62	−2.757[*]	0.006
Sodium (mmol/L)	138.8 ± 4.35	139.0 ± 4.30	137.2 ± 4.60	2.999[*]	0.003
LVEF (%)	50.7 ± 13.2	51.1 ± 13.2	46.0 ± 12.9	2.818[*]	0.005
HFpEF	351 (55.3)	328 (56.7)	23 (40.4)	5.754[†]	0.056
HFmrEF	152 (23.9)	133 (23.0)	19 (33.3)		
HFrEF	132 (20.8)	117 (20.2)	15 (26.3)	9.630[*]	
MAGGIC score	25.6 ± 5.8	25.0 ± 5.4	32.1 ± 4.9		<0.001

Data are presented as *n* (%), mean ± standard deviation or median (interquartile). [*] The statistics were calculated using Student's t test or separate variance estimation t test, as appropriate. [†] The statistics were calculated using chi-squared test, continuity correction chi-squared test, or Fisher's exact test, as appropriate. [‡] The statistics were calculated using Mann–Whitney *U* test. All the data were from the results of the first tests after admission. ACEI: Angiotensin-converting enzyme inhibitor; ARB: Angiotensin receptor blocker; BMI: Body mass index; COPD: Chronic obstructive pulmonary disease; DBP: Diastolic blood pressure; HF: Heart failure; HFmrEF: HF with mild range ejection fraction; HFpEF: HF with preserved ejection fraction; HFrEF: HF with reduced ejection fraction; LVEF: Left ventricular ejection fraction; MAGGIC: Meta-Analysis Global Group in Chronic Heart Failure; NYHA: New York Heart Association; NT-proBNP: N-terminal pro brain natriuretic peptide; SBP: Systolic blood pressure.

$P = 0.005$). The non-surviving group had a higher average MAGGIC score than the surviving group (32.1 ± 4.9 *vs.* 32.1 ± 4.9, $\chi^2 = 9.630$, $P < 0.001$).

Distribution of the MAGGIC scores and subgroup analysis

We calculated a total integer MAGGIC score for each patient based on the risk factors. The bell-shaped distribution of the risk scores for all 635 patients is shown in Figure 1. The median was 25 points, and the range was 5 to 42 points, with 90% of patients in the range from 15 to 36 points. The trend line with the 95% confidence interval (CI) in Figure 1 shows the patients' scores, corresponding to their probability of dying within 1 year. Figure 2 shows the trend of 1-year mortality increasing as the MAGGIC integer score increases in different types of HF. When the MAGGIC score exceeds

29 points, 1-year mortality in all three groups increases significantly.

Discriminatory ability analysis

Subsequently, we used SPSS software to generate ROC curves according to the MAGGIC scores and 1-year survival outcomes and calculate the AUC. The ROC curve analysis is shown in Figure 3. The AUC was 0.840 (95% CI: 0.779, 0.901), indicating that the model had fair discriminatory ability. The Youden index (sensitivity + specificity − 1), is commonly used to determine the optimal cut-off value.[15,16] The maximum value of this index corresponds to the optimal cut-off value. According to this formula, a MAGGIC score of 32 points, with a sensitivity of 66.7% and a specificity of 91.0%, was confirmed as the optimal cut-off value for the prediction of mortality within 1 year.

Calibration analysis

We also assessed the calibration of the MAGGIC score. The Hosmer–Lemeshow goodness-of-fit test was used to evaluate the calibration. The Hosmer–Lemeshow test result ($\chi^2 = 12.902$, degree of freedom = 8, $P = 0.115$) indicated that the difference between the predicted value and the observed value was not statistically significant, and the MAGGIC score had good calibration. In addition, the Hosmer-Lemeshow test contingency table listed ten decile sets of the observed values and predicted values, which could be intuitively compared, to help determine the calibration of the model. A calibration curve [Figure 4] was obtained by drawing a scatter plot of the observed mortality and mortality predicted by the model and fitting a linear
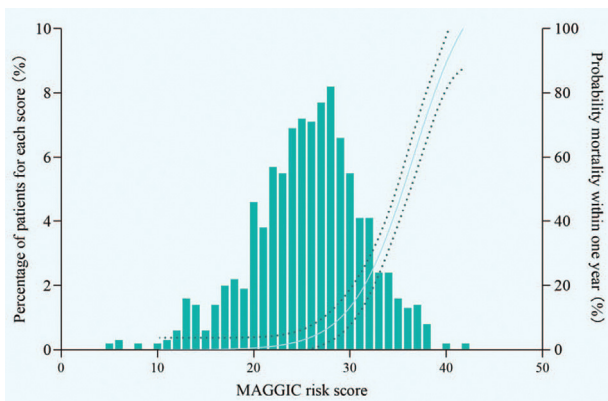
Figure 1: Distribution of the integer risk score for all 635 patients and Its association with the probability mortality (the dotted line represents a 95% confidence interval) within 1 year. MAGGIC: Meta-Analysis Global Group in Chronic Heart Failure.
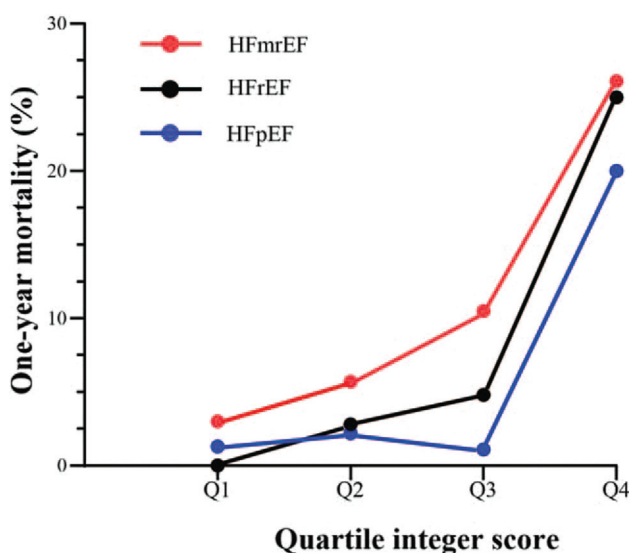

Figure 2: One-year mortality and quartile integer score in HFpEF, HFmrEF, and HFrEF patients. Q1: <22 points; Q2: ≥22 and <26 points; Q3: ≥26 and <29 points; Q4: ≥29 points; HFpEF: HF with preserved ejection fraction; HFmrEF: HF with mild range ejection fraction; HFrEF: HF with reduced ejection fraction.
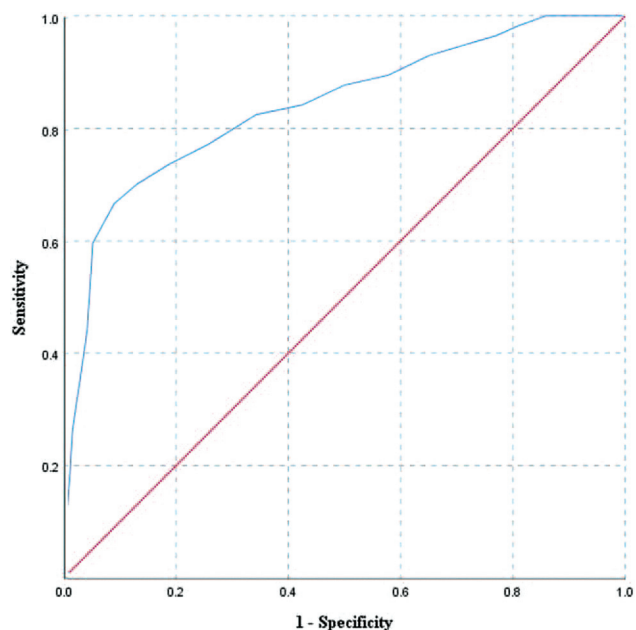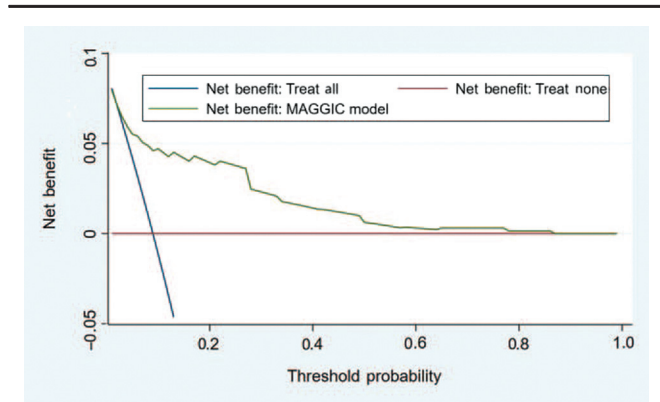

Figure 3: ROC curve for the MAGGIC score to predict one-year mortality. The red line represents a reference curve that connects the extremes at each end. The ROC area (0.840, 95% CI: 0.779−0.901) indicated a fair discrimination. MAGGIC: Meta-Analysis Global Group in Chronic Heart Failure; ROC: Receiver operating characteristic.
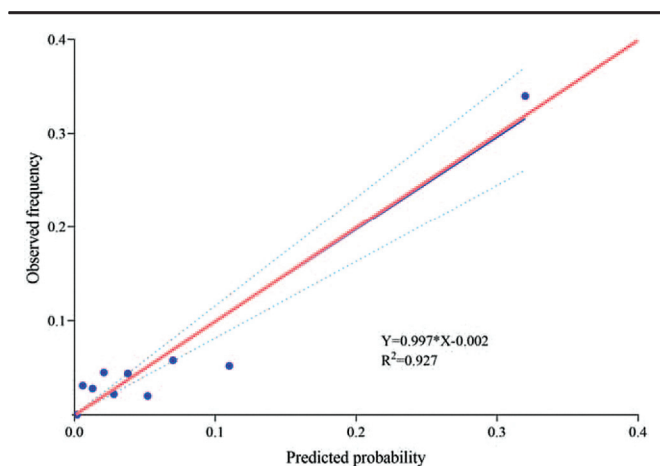
Figure 4: Calibration curve for observed vs. model-predicted 1-year mortality by risk decile. The blue scatters were based on the actual observed values and the predicted values of the MAGGIC score. The blue line represented the linear trend line based on these scatters, and the dotted line represented its 95% confidence interval. The blue calibration curve was very close to the red standard curve, which indicated a good calibration ability of the MAGGIC score. MAGGIC: Meta-Analysis Global Group in Chronic Heart Failure.

Figure 6: Net reduction of the decision curve in interventions for the MAGGIC score to predict one-year mortality. MAGGIC: Meta-Analysis Global Group in Chronic Heart Failure.

score could help reduce unnecessary treatment. The curve of net reduction in interventions provides the relevant information as an extension of DCA, as shown in Figure 6. For example, at a probability threshold of 20%, the net reduction in interventions was approximately 70 per 100 patients. In other words, at this probability threshold, treating patients based on the MAGGIC score was the equivalent of a strategy that reduced unnecessary treatment by 70%, without missing the HF patients predicted to die within 1 year.

Discussion

We evaluated the applicability of the MAGGIC score in HF patients by externally validating it in a Chinese cohort from a regional general hospital. After a series of statistical analyses, the MAGGIC score was found to have good discriminatory ability for the risk of mortality within 1 year. In addition, the risk of mortality predicted by the MAGGIC score was close to the observed outcomes. The calibration was excellent. Further, based on DCA, the MAGGIC score could help physicians make appropriate decisions, increasing the net benefit of treatment while theoretically reducing unnecessary treatment.

The prediction of prognosis is particularly challenging for HF patients. The variable and complex clinical characteristics of patients with HF mean that risk stratification for these patients is very important for helping physicians make optimal decisions and assist patients and their relatives with regard to maintaining reasonable expectations and selecting treatment targets. In the past decades, a variety of predictive models have been developed for HF patients, but few are externally validated.[20] The use of the predictive models that have not been widely validated can lead to make the inappropriate medical decisions that deviate from reality in clinical practice. One well-validated risk score, the Seattle Heart Failure Model (SHFM), was developed in a cohort of 1125 HF patients using a multivariate Cox model to predict 1-, 2-, and 3-year survival in 2006. In the following decade years, >15 studies have verified the good predictive ability to predict
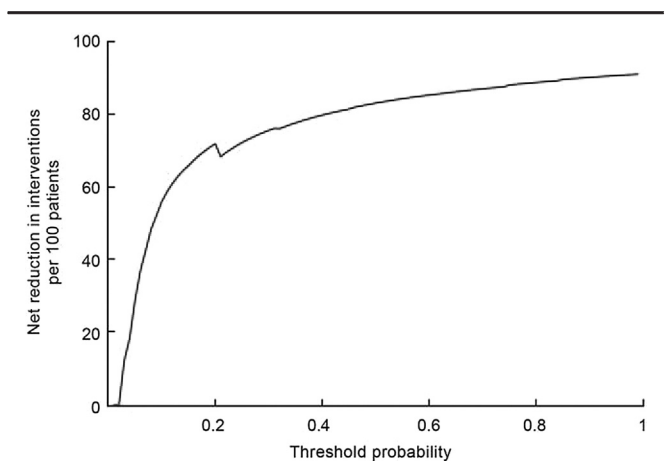

Figure 5: Net benefit of the decision curve for the MAGGIC score to predict one-year mortality. MAGGIC: Meta-Analysis Global Group in Chronic Heart Failure.

trend line with a 95% CI. The linear regression formula for the calibration plot was as follows: observed mortality $= 0.997 \times$ (predicted mortality) $- 0.002$, $R^2 = 0.927$. In Figure 4, the red curve indicates the standard curve, and the blue curve indicates the calibration curve. The closer the blue calibration curve is to the red standard curve, the better the calibration of the model. As shown in the figure, the blue curve was nearly identified to the red curve, indicating that the MAGGIC score model has good calibration.

DCA

Finally, we performed DCA to evaluate the net benefit and net reduction in the number of false positives.[17-19] As shown in Figure 5, the blue line indicates the net benefit for treating all patients, and the red line indicates the net benefit for treating none of the patients. When the MAGGIC score predicted HF patients with a threshold probability between 0.05 and 0.6, positive treatment strategies led to the largest clinical benefit. At the same time, we were interested in whether using the MAGGIC

long-term mortality in HF patients. However, the complex calculation of estimated survival containing 14 continuous variables and 10 categorical values limits its clinical use. In 2013, a predictive model, the Metabolic Exercise test data combined with Cardiac and Kidney Indexes score, based on cardiopulmonary exercise test (CPET), was developed in Italy to predict cardiovascular death and cardiac depression. Subsequently, the score has been validated in several studies. However, it is not easy to obtain CPET parameters, which also limits its clinical application. The MAGGIC score, which was established in 2012, was based on 39,372 patients from 30 cohort studies (six randomized clinical trials and 24 observational registry studies) in 13 countries. The MAGGIC score with the appropriate easy-to-get variables was built based on the largest multinational population in currently available HF scores and had been validated by several studies in recent decades. However, these data did not include HF patients in China, which limited the application of the MAGGIC score in the Chinese population. Differences among ethnicities and regions could lead to bias in the predictive power of the model. Therefore, external validation in a contemporary HF population in China was needed. To summarize, these characteristics led our team to select the MAGGIC score for verification in the first place. In addition, the data used in this study were obtained from the electronic medical records of a regional general hospital. The results are based on real-world data, making them more relevant to clinical practice. The MAGGIC score was based on six randomized clinical trials and 24 observational registries. Initially, the studies were designed for different purposes; therefore, there was a wide variation in the included population. Our study did not set up strict inclusion criteria, except for the obvious causes of death, such as acute coronary syndrome, hematodialysis, and malignant tumor. It is highly likely that the good predictive ability in the validation study results from the general population on whom the MAGGIC score is based.

The MAGGIC score, which has been validated in several studies, is a robust model that can be used to predict 1-year mortality in a high-risk population. In Chicago, Rich et al[21] validated the ability of the MAGGIC score and the SHFM score to predict the mortality, the HF hospitalizations, and the combined endpoint of cardiovascular-related hospitalizations or death in patients with preserved EF. After a mean follow-up time of 3.6 ± 1.8 years, the analysis showed that the MAGGIC score and the SHFM score had a similar predictive ability in predicting HF with preserved EF outcomes, but the MAGGIC score demonstrated better calibration for hospitalization outcomes. Actually, the MAGGIC score was originally performed to predict the 1- and 3-year mortality in ambulatory HF patients. The study by Rich et al[21] additionally validated the predictive ability of hospitalization outcomes in preserved EF patients. The difference with the study of Rich et al[21] was that our study just validated the predictive ability of predicting 1-year mortality for the MAGGIC score. In Japan, using the MAGGIC score, Sawano et al[9] evaluated 2215 consecutive acute HF patients from a prospective multi-centre registry. They verified that the model showed

modest discriminatory ability (C-index = 0.71, 95% CI: 0.67–0.74) and good calibration ($R^2 = 0.97$) for the prediction of 1-year mortality. In addition, when the BNP level was integrated into the model and the discriminatory ability improved, while the calibration remained adequate. In Korea, Khanam et al[10] externally validated the use of the MAGGIC score to predict 1-year mortality after discharge in 5625 hospitalized patients from the Korean acute HF registry. The MAGGIC risk score had a C-index of 0.734, indicating moderate discriminatory ability, and the calibration was found to be good. Therefore, they concluded that the MAGGIC score performed well in a large nationwide contemporary external validation cohort. In contrast, Michaels et al[11] validated the MAGGIC score for the prediction of 1-year mortality using 4138 HF patients from two groups, including discharged patients from an administrative database ($n = 2503$, 60.5%) and a prospective registry of ambulatory HF patients ($n = 1635$, 39.5%) in Detroit. The C-statistic was 0.668 in the discharge population and 0.784 in the ambulatory cohort. They concluded that the MAGGIC score had poor discriminatory ability when used in HF patients at hospital discharge; it had better discriminatory ability among ambulatory HF patients. There were discrepancies in baseline age, sex, LVEF, and 1-year mortality between our study and the three aforementioned studies, as shown in Supplementary Figure 2, http://links.lww.com/CM9/B152. We selected the hospitalized cohort for comparison with the cohort in the study by Michaels et al[11]. The population in the study by Khanam et al[10] was younger than the populations in other studies and had a lower average LVEF. The proportion of males was the highest in the study by Sawano et al.[9] Moreover, the 1-year mortality in the study by Michaels et al[11] from the USA was significantly higher than that in the three studies from Asia. We speculated that different admission criteria might be responsible for this result because the MAGGIC score had better discriminatory ability in ambulatory HF patients. Furthermore, the baseline MAGGIC scores were also compared among the four studies. The average score in our study was 25.6 ± 5.75, the median score in the study by Sawano et al[9] was 25 (21–29), and the average score in the study by Michaels et al[11] was 29.5 ± 6.52. The study by Khanam et al[10] did not report the overall average MAGGIC score. However, the average score in the non-surviving group was 30.61 ± 6.32, and the average score in the surviving group was 24.80 ± 6.81. The average score in the study by Michaels et al[11] was close to 30 points, which indicated a more severe condition in those patients than in the patients in the other studies. This could be one of the reasons for the poor discriminatory ability of the MAGGIC score in the study by Michaels et al.[11]

There were some limitations of our study. First, compared with other validation studies, the number of participants in our study was relatively small, which could have led to selection bias. Second, there is marked regional variability among hospitals in China, which suggests that our conclusions cannot be generalized to other parts of China. The data used in our study came from a relatively small regional hospital in eastern China. The admission criteria

and treatment conditions among patients with HF are different from those in other regions, which could have affected the final results of the validation analyses. Third, according to the traditional local culture, patients are not willing to die in the hospital; therefore, some deaths that would have occurred in the hospital were counted in the 1-year mortality rate. Fourth, most of the LVEF values were calculated by the M-mode ultrasound method. However, due to its limitations, the value could have been overestimated, which would have resulted in an underestimation of the MAGGIC score. Fifth, the MAGGIC score was used to predict 1- and 3-year all-cause mortality among people with HF. However, because the follow-up data did not reach 3 years, the ability of the MAGGIC score to predict 3-year mortality was not validated. Sixth, after we used "HF" as the main diagnosis to retrieve the patients from the in-hospital electronic medical record system, some patients were excluded because some of the data needed to calculate the MAGGIC score were missing, which could have biased the results.

Conclusions

The MAGGIC score was found to have a good ability to predict 1-year mortality in patients with HF after hospital discharge in a Chinese cohort. Due to regional and interhospital differences, external validation studies need to be performed in other centers.

Conflicts of interest

None.

References

1. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2015;37:2129–2200. doi: 10.1093/eurheartj/ehw128.
2. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Drazner MH, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;62:e147–e239. doi: 10.1016/j.jacc.2013.05.019.
3. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Colvin MM, et al. 2017 ACC/AHAHFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Circulation 2017;136:e137–e161. doi: 10.1161/CIR.0000000000000509.
4. Mahajan SM, Heidenreich P, Abbott B, Newton A, Ward D. Predictive models for identifying risk of readmission after index hospitalization for heart failure: a systematic review. Eur J Cardiovasc Nurs 2018;17:675–689. doi: 10.1177/1474515118799059.
5. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. Circulation 2006;113:1424–1433. doi: 10.1161/CIRCULATIONAHA.105.584102.
6. Pocock SJ, Ariti CA, McMurray JJ, Maggioni A, Køber L, Squire IB, et al. Predicting survival in heart failure: a risk score based on

7. 39,372 patients from 30 studies. Eur Heart J 2013;34:1404–1413. doi: 10.1093/eurheartj/ehs337.
7. Lee DS, Austin PC, Rouleau JL, Liu PP, Naimark D, Tu JV. Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical model. JAMA 2003;290:2581–2587. doi: 10.1001/jama.290.19.2581.
8. Meta-analysis Global Group in Chronic Heart Failure (MAGGIC). 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;33:1750–1757. doi: 10.1093/eurheartj/ehr254.
9. Sawano M, Shiraishi Y, Kohsaka S, Nagai T, Goda A, Mizuno A, et al. Performance of the MAGGIC heart failure risk score and its modification with the addition of discharge natriuretic peptides. ESC Heart Fail 2018;5:610–619. doi: 10.1002/ehf2.12278.
10. Khanam SS, Choi E, Son JW, Lee JW, Youn YJ, Yoon J, et al. Validation of the MAGGIC (meta-analysis global group in chronic heart failure) heart failure risk score and the effect of adding natriuretic peptide for predicting mortality after discharge in hospitalized patients with heart failure. PLoS One 2018;13: e0206380. doi: 10.1371/journal.pone.0206380.
11. Michaels A, Aurora L, Peterson E, Liu B, Pinto YM, Sabbah HN, et al. Risk prediction in transition: MAGGIC score performance at discharge and incremental utility of natriuretic peptides. J Card Fail 2020;26:52–60. doi: 10.1016/j.cardfail.2019.11.016.
12. Szczurek W, Gasior M, Skrzypek M, Szyguła-Jurkiewicz B. Apelin improves prognostic value of HFSS (heart failure survival score) and MAGGIC (meta-analysis global group in chronic heart failure) scales in ambulatory patients with end-stage heart failure. J Clin Med 2020;9:2300. doi: 10.3390/jcm9072300.
13. Kouwert IJ, Bakker EA, Cramer MJ, Snoek JA, Eijsvogels TM. Comparison of MAGGIC and MECKI risk scores to predict mortality after cardiac rehabilitation among Dutch heart failure patients. Eur J Prev Cardiol 2020;27:2126–2130. doi: 10.1177/2047487319865730.
14. Potter E, Marwick TH. Assessment of left ventricular function by echocardiography: the case for routinely adding global longitudinal strain to ejection fraction. JACC Cardiovasc Imaging 2018;11:260–274. doi: 10.1016/j.jcmg.2017.11.017.
15. Carter JV, Pan J, Rai SN, Galandiuk S. ROC-ing along: evaluation and interpretation of receiver operating characteristic curves. Surgery 2016;159:1638–1645. doi: 10.1016/j.surg.2015.12.029.
16. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. Biochem Med (Zagreb) 2016;26:297–307. doi: 10.11613/BM.2016.034.
17. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–574. doi: 10.1177/0272989X06295361.
18. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res 2019;3:18. doi: 10.1186/s41512-019-0064-7.
19. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and interpreting decision curve analysis: a guide for investigators. Eur Urol 2018;74:796–804. doi: 10.1016/j.eururo.2018.08.038.
20. Sahle BW, Owen AJ, Chin KL, Reid CM. Risk prediction models for incident heart failure: a systematic review of methodology and model performance. J Card Fail 2017;23:680–687. doi: 10.1016/j.card-fail.2017.03.005.
21. Rich JD, Burns J, Freed BH, Maurer MS, Burkhoff D, Shah SJ. Meta-analysis global group in chronic (MAGGIC) heart failure risk score: validation of a simple tool for the prediction of morbidity and mortality in heart failure with preserved ejection fraction. J Am Heart Assoc 2018;7:e009594. doi: 10.1161/JAHA.118.009594.