ORIGINAL RESEARCH

The prognosis of patients hospitalized with a first episode of heart failure, validation of two scores: PREDICE and AHEAD

Francisco Ruiz-Ruiz[1]
Miguel Menéndez-Orenga[2]
Francisco J Medrano[3–6]
Enrique J Calderón[3–6]
David Lora-Pablos[2]
Maria Asunción Navarro-Puerto[7]
Patricia Rodríguez-Torres[7]
Agustín Gómez de la Cámara[2,4]

[1]Department of Internal Medicine, Quiron Hospital, Seville, Spain; [2]Research Institute, Clinical Research Unit, Hospital Universitario 12 de Octubre, Madrid, Spain; [3]Service of Internal Medicine, Virgen del Rocio University Hospital, Seville, Spain; [4]CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain; [5]Instituto de Biomedicina de Sevilla, Department of Health of the Junta de Andalusia/CSIC/ University of Seville, Seville, Spain; [6]Department of Medicine, University of Seville, Seville, Spain; [7]Department of Internal Medicine, Virgen de Valme University Hospital, Seville, Spain

Purpose: Heart failure (HF) is a chronic, frequent and disabling condition but with a modifiable course and a large potential for improving. The aim of this study was to validate the two available clinical prediction rules for mortality at one year in patients with primo-hospitalization for decompensated HF: PREDICE and AHEAD. The secondary aim was to evaluate in our setting the changes in the clinical pattern of HF in the last decade in patients hospitalized for a first episode of the disease.
Patients and methods: A prospective multicenter cohort study, which included 180 patients hospitalized with "de novo" HF was conducted to validate the PREDICE score. Calibration and discrimination measurements were calculated for the PREDICE model and the PREDICE score (using the validation cohort of the PREDICE) and the AHEAD score (using both the development and the validation cohort of the PREDICE).
Results: For the PREDICE models, the area under the curve (AUC) was 0.68 (95% confidence interval [CI]: 0.57–0.79) and the calibration slope 0.65 (95% CI: 0.21–1.20). For the PREDICE score AUC was 0.59 (95% CI: 0.47–0.71) and slope 0.42 (95% CI: −0.20–1.17). For the AHEAD score the AUC was 0.68 (95% CI: 0.62–0.73) and slope 1.38 (95% CI: 0.62–0.73) when used the development cohort of PREDICE and the AUC was 0.58 (95% CI: 0.49–0.67), and slope 0.68 (95% CI: −0.06 to 1.47) when used its validation cohort.
Conclusion: The present study shows that the two risk scores available for patients with primo-hospitalization for decompensated HF (PREDICE and AHEAD) are not currently valid for predicting mortality at one-year. In our setting the clinical spectrum of hospitalized patients with new-onset HF has been modified over time. The study underscores the need to validate the prognostic models before clinical implementation.
Keywords: heart failure, validation studies as topics, severity of illness index, inpatients

Introduction

Heart failure (HF) is a heterogeneous clinical prevalent syndrome.[1] Its prevalence increases as patients age.[2,3] Furthermore, HF is one of the leading causes of death in developed countries, accounting for 17,000 deaths in Spain in 2014 and it represents a high economic burden in Spain and the rest of the world.[3–5]

Predicting the clinical outcomes of heart failure patients may be useful for decision-making in terms of resources and clinical management of the disease. Prognostic models and clinical prediction rules or scores are tools aimed at helping clinicians assess the risk-benefit balance, so they can make better choices.[6,7] In the care for HF patients, invasive diagnostic tests and potentially severely harmful

Correspondence: Francisco J Medrano
Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocío, Avda. Manuel Siurot s/n, Seville 41013, Spain
Tel +3 465 482 4512
Fax +3 495 501 3292
Email fmedrano@us.es

615

treatments are available, so assessing risk-benefit balance is challenging. At present, more than one hundred prognostic models exist in HF. Their prediction ability is moderate to poor,[1,8–10] yet there are few variables that are consistently kept in some of the models.[9,11]

In 2012, no prognostic model was available to be used in hospitalized patients with newly diagnosed acute HF, so our group developed a one-year prognostic model and clinical prediction score with the aim of being useful at patients' bedsides, this is the PREDICE-score.[12] Since then, only one other risk model (AHEAD-score) has been reported to predict mortality in hospitalized patients with acute HF.[13]

Despite the prediction ability of prognostic models in HF not being perfect, it has a potential role to be helpful for clinicians. However, a lack of external validation compromises its validity.[14] Therefore, developed models should be externally validated or calibrated so clinicians can make proper use of them.[15]

The primary aim of this study was to validate the two available clinical prediction rules for mortality at one year in patients with primo-hospitalization for decompensated HF: PREDICE and AHEAD. The secondary aim was to evaluate, in our setting, the changes in the clinical pattern of HF in the last decade in patients hospitalized after a first episode of the disease.

Materials and methods

A prospective multicenter cohort study was designed for external validation of the PREDICE-score. A total of 180 patients hospitalized with a first episode ("de novo") of HF between 2013 and 2015 in three Spanish hospitals (60 patients for each of the centers) of the National Health Service (12 de Octubre University Hospital, Madrid; Virgen del Rocio University Hospital, Seville; and Valme University Hospital, Seville) were included correlatively to complete the calculated sample size for an estimated mortality at one year of 16% (according to 2013 local data), using 5% precision with a 95% confidence interval (CI). The first two centers provided advanced medical care and the last one (Valme University Hospital), standard medical care. The inclusion criteria were: patient hospitalized at the study center with a primary diagnosis of HF (following Framingham criteria), over 18-year-old, resident in the area of reference of the study center. The exclusion criteria were previous diagnosis of HF, diagnosis of HF not displayed at discharge report and patients not residing in the hospital area. The patients were followed up at one year after discharge. Development[12]

and validation PREDICE datasets were used to externally validate the AHEAD-score.[13]

Outcomes

The outcome of interest was mortality from any cause at one year after discharge. Mortality predictions at 3, 5 and 7.5 years for the AHEAD-score[13] were not validated since this information was not available in datasets.

Validation

To validate the model and score, we use calibration and discrimination. It adds global performance measures. Calibration reflects the extent to which the predicted probabilities and actual probabilities agree. We focus on the calibration-in-the-large (intercept) plus the calibration slope. The concordance statistic (c-statistic) represents the probability that individuals with the outcome will receive a higher predicted probability than those without it. It corresponds to the area under the curve (AUC) and their 95% CI.[16,17] The missing data were excluded from the sample and no method of data input was performed.

The variables included in the PREDICE-score validation are: age, eGFR, serum sodium levels, dependency for daily basic activities and physiopathology diagnosis (systolic function preserved or reduced).[12]

The formula that was constructed in the PREDICE derivation article[12] is called the prognostic index (PI):[7]

The risk of mortality at one year for the PREDICE model[12] were calculated using PI, where the estimated glomerular filtration rate (eGFR) was estimated with the 4-variable version of the Modification of Diet in Renal Disease equation (MDRD4).[18] Age, eGFR, serum sodium levels, dependency for daily basic activities and physiopathology diagnosis (systolic function preserved or reduced) were used to predict the one-year death risk using the clinical prediction rule PREDICE-score.[12]

The probabilities of mortality at one-year for the AHEAD-score were estimated using atrial fibrillation, creatinine levels, hemoglobin levels, age and the existence of diabetes mellitus.[13] The original formula was requested from the correspondence author. Because the exact formula was not provided, the score was firstly obtained and then, probability was assigned using the AHEAD-score table.[13]

The distribution of characteristics of development and validation samples were compared separately. Additionally, relatedness between validation and development PREDICE samples was quantified using the c-statistic.[16] The predictive

ability of the models was evaluated based on calibration and discrimination.

Ethical aspects

The study was conducted in accordance with the Declaration of Helsinki and followed the rules of security and privacy. The Project was approved by the ethical committees of the three participating hospitals (protocol reference: 06/117). According to Spanish law, patient consent is not required for register-based studies. No further ethical permissions were required for the analyses of these anonymized patient-level data.

Results
Sample characteristics

The baseline characteristics and outcome of the different PREDICE and AHEAD populations are presented in Table 1.

First, the PREDICE derivation cohort was compared with the PREDICE validation cohort. Patients in the validation cohort were slightly older (median age, 73.6 vs 75 years), had a slower rate of eGFR (median clearance, 38.6 vs 30.7 ml/min/1.73 m^2), had a higher rate of systolic dysfunction (23.9% vs 37%), had similar serum sodium levels (median, 138.1 vs 138.5 mEq/l) and they had similar levels of independence in basic daily activities (89.9% vs 88.9%). Mortality at one year was 16.3% in the derivation cohort and 22.5% in the validation cohort. The concordance statistic between the development and validation PREDICE dataset (predictor variables and outcome status) was 0.83 (95% CI, 0.79–0.86). The heterogeneity between datasets was due to age and systolic dysfunction (0.77). The characteristics of the PREDICE validation cohort according to the primary outcome (mortality at one year) are detailed in Table 2.

In relation to AHEAD, the derivation cohort of the AHEAD-score was the Acute Heart Failure Database (AHEAD) Network registry that included 6242 primo hospitalizations for acute HF in the Czech Republic between 2012 and 2016. The validation cohort of the AHEAD-score was the GREAT registry that includes 6315 primo hospitalizations for acute HF from nine cohorts of different countries during the same study period.[13] The validation and development PREDICE datasets were similar to the validation and development datasets of the AHEAD study in most of the baseline characteristics and therapeutic management (Table 1). However, PREDICE populations had a lower rate of

ischemic heart disease and the one-year mortality rates of the AHEAD and GREAT cohorts (13.7% vs 14.4%, respectively) were lower than those found in the derivation and validation cohorts of the PREDICE study (16.6 and 22.5%, respectively).

Model evaluation - PREDICE

The discrimination of the PREDICE model in the validation dataset was 0.68 (95% CI 0.57–0.79). The calibration intercept was 0.33 (95% CI −0.77 to 1.45) and the calibration slope was 0.65 (95% CI 0.21–1.20). The calibration plot is shown in Figure 1.

Clinical prediction rule evaluation – PREDICE-score

When we tested the clinical prediction rule, AUC was 0.59 (95% CI 0.47–0.71), the intercept −0.25 (95% CI −1.70 to 0.94) and the slope 0.42 (95% CI −0.20 to 1.17). Figure 2 shows its calibration.

Clinical prediction rule evaluation – AHEAD (using development cohort of PREDICE)

When we used development cohort of PREDICE, the AUC of AHEAD-score was 0.68 (95% CI 0.62–0.73), the intercept −0.60 (95% CI −1.02 to −0.23) and the slope 1.38 (95% CI 0.90–1.90). The calibration plot is shown in Figure 3.

Clinical prediction rule evaluation - AHEAD (using validation cohort of PREDICE)

We evaluated the model of the AHEAD-score also using the validation cohort of PREDICE. It has an AUC of 0.58 (95% CI 0.49–0.67), an intercept of −0.78 (95% CI −1.39 to −0.24) and a slope of 0.68 (95% CI −0.06 to 1.47). The calibration plot it is shown in Figure 4.

Discussion
Main findings

Our main results are that the performance of the model and the clinical prediction rule are worse in the validation cohort than in the development cohort. We have not been able to validate the PREDICE-score or AHEAD-score. In addition, high mortality and comorbidity of patients was found in the validation PREDICE-cohort. The PREDICE-model showed an acceptable discrimination with low calibration. The AHEAD-SCORE exhibited similar discrimination to development of the AHEAD study, but poor calibration.

Table 1 Baseline characteristics and one-year mortality in the PREDICE and AHEAD cohorts

	PREDICE, derivation cohort (N=600)	PREDICE, validation cohort (N=180)	p	AHEAD, derivation cohort (N=5846)	AHEAD, validation cohort* (N=6315)	p
Age [years], mean (± SD)	73.6 (±12.3)	75 (±9.7)	0.16	74	77	
Gender [female], n (%)	305 (50.8)	88 (48.9)	0.64	2536 (43.4)	2813 (44.5)	0,20
Ischemic cardiopathy, n (%)	98 (16.3)	18 (10.1)	0.04	1422 (24.3)	2255 (35.7)	<0.001
Atrial fibrillation, n (%)	177 (29.5)	56 (31.1)	0.67	1792 (30.7)	2462 (39)	<0.001
Valvulopathy, n (%)	160 (26.7)	23 (15)	<0.001			
HFrEF, n (%)	106 (23.9)	54 (37)	<0.001			
Hemoglobin [g/l], mean (± SD)	12.8 (±2.3)	12.3 (±2.2)	0.01	132	124	
Sodium [mEq/l], mean (± SD)	138.1 (±4.6)	138.5 (±5)	0.32			
Potassium [mEq/l], mean (± SD)	4.3 (±0.7)	4.4 (±0.7)	0.09			
eGFR [ml/min], mean (± SD)	73.5 (±38.6)	72.0 (±30.7)	0.63			
Beta-blockers, n (%)	188 (31.3)	90 (50.6)	<0.001	2931 (52)	2129 (36)	<0.001
ACE inhibitors, n (%)	328 (54.7)	96 (53.9)	0.75	3327 (59.3)	3357 (56.3)	<0.001
ARBs, n (%)	104 (17.3)	27 (15.3)	0.46	1276 (23)	635 (18)	<0.001
Diuretics, n (%)	439 (73.2)	137 (77)	0.43			
Nitrates, n (%)	70 (11.7)	8 (4.9)	0.005			
Digoxin, n (%)	152 (25.3)	26 (14.6)	0.002			
Statins, n (%)	150 (25)	86 (48.3)	<0.001			
Antiaggregants, n (%)	263 (43.8)	83 (46.6)	0.59			
Oral anticoagulants, n (%)	203 (33.8)	75 (42.1)	0.05			
IBDA, n (%)	537 (89.5)	160 (88.9)	0.82			
Charlson Index >2 points, n (%)	77 (12.8)	162 (90)	<0.001			
One-year mortality, n (%)	98 (16.3)	40 (22.5)	0.07	800 (13.7)	909 (14.4)	

Notes: *GREAT registry. eGFR was estimated with the 4-variable version of the Modification of Diet in Renal Disease equation.
Abbreviations: SD, standard deviation; HFrEF, heart failure with reduced ejection fraction; eGFR, estimated glomerular filtration rate; ACE, angiotensin-converting-enzyme; ARBs, angiotensin-II receptor blockers; IBDA, independent basic daily activities.

Mathematical model - PREDICE

In the development cohort, the AUC of the mathematical model was 0.76 while when we used validation cohort for validation the AUC is 0.68. The values of the AUC operate between 0.5 (no discrimination; it is considered to predict just as chance) and 1.0 (perfect discrimination).[16,19,20] As expected, the performance of the model is worse in the validation cohort than in the development cohort. However, in both populations the lower CI does not contain 0.5; suggesting that discrimination is acceptable in both populations. By definition, the intercept is always optimal (0) in the development sample. Consequently, it is a useful statistic for identifying whether unexplained differences exist in the outcome frequency of the validation sample. Values below (or above) indicate that the model overestimates (or respectively underestimates) the outcome.[16,21,22] The value of the intercept in the mathematical model is 0.33, which overestimates the risk, but it works thanks to its confidence interval (results above). The calibration slope can be used as a statistic for evaluating to what

extent the model's predictive mechanisms remain valid in the validation sample. Slope>1 occurs when predicted probabilities do not vary enough (eg, predicted risks are systematically too low) and 0<slope<1 occurs when they vary too much (predicted risks are too low for low outcome risks and too high for high outcome risks).[19,23,24] The slope in the mathematical model is 0.65, which is poor calibration.

Clinical prediction rule - PREDICE

The AUC of the clinical prediction rule is 0.59. Discrimination is worse in the validation cohort (see above). In the calibration, the intercept is −0.25 and its CI does not contain 1. In addition, the slope is lower than 1. This means poor calibration.

In summary, we have been able to validate the model PREDICE-score. Conversely we have not been able to validate the clinical prediction rule PREDICE. We cannot reject the null hypothesis (model predicts better than chance).

Table 2 Baseline characteristics of the PREDICE validation cohort according one-year mortality

	Overall [N=180]	Death [N=40]	Non death [N=140]
Age [years], mean (± SD)	75 (±9.7)	79.9 (±7.8)	73.7 (±9.9)
Gender [female], n (%)	88 (48.9)	17 (42.5)	71 (51.4)
Ischemic cardiopathy, n (%)	18 (10.1)	3 (7.5)	14 (10.2)
Atrial fibrillation, n (%)	56 (31.1)	18 (45)	38 (27.5)
Valvulopathy, n (%)	23 (15)	6 (18.2)	17 (14.4)
HFrEF, n (%)	54 (37)	14 (38.9)	40 (37)
Hemoglobin [g/l], mean (± SD)	12.3 (±2.2)	12.4 (±2.1)	12.3 (±2.2)
Sodium [mEq/l], mean (± SD)	138.5 (±5.0)	137.5 (±6.6)	138.9 (±4.4)
Potassium [mEq/l], mean (± SD)	4.4 (±0.7)	4.6 (±0.8)	4.4 (±0.6)
eGFR, [ml/min], mean (± SD)	72.0 (±30.7)	72.3 (±35.7)	70.8 (±27.9)
Beta-blockers, n (%)	90 (50.6)	14 (36.8)	75 (54.3)
ACE inhibitors, n (%)	96 (53.9)	19 (50)	75 (54.3)
ARBs, n (%)	27 (15.3)	6 (15.8)	20 (14.6)
Diuretics, n (%)	137 (77)	32 (84.2)	103 (74.6)
Nitrates, n (%)	8 (4.9)	5 (13.2)	2 (1.4)
Digoxin, n (%)	26 (14.6)	7 (18.4)	19 (13.8)
Statins, n (%)	86 (48.3)	15 (39.5)	69 (50)
Antiagregants, n (%)	83 (46.6)	17 (44.7)	64 (46.4)
Oral anticoagulants, n (%)	75 (42.1)	20 (52.6)	55 (39.9)
IBDA, n (%)	160 (88.9)	28 (70)	130 (94.2)
Charlson Index >2 points, n (%)	162 (90)	40 (100)	122 (87.1)

Note: eGFR was estimated with the 4-variable version of the Modification of Diet in Renal Disease equation.
Abbreviations: SD, standard deviation; HFrEF, heart failure with reduced ejection fraction; eGFR, estimated glomerular filtration rate; ACE, angiotensin-converting-enzyme; ARBs, angiotensin-II receptor blockers; IBDA, independent basic daily activities.
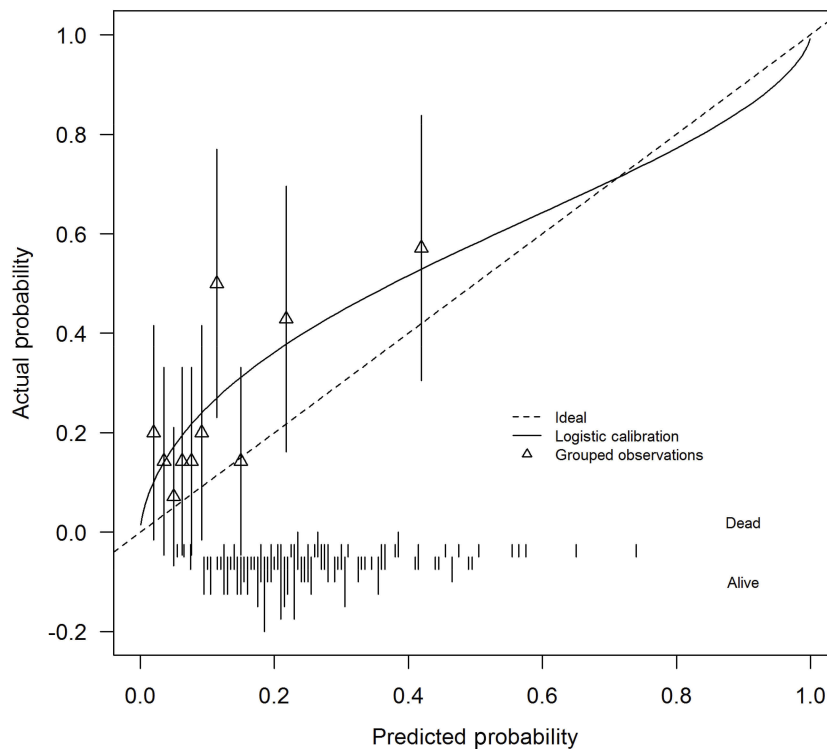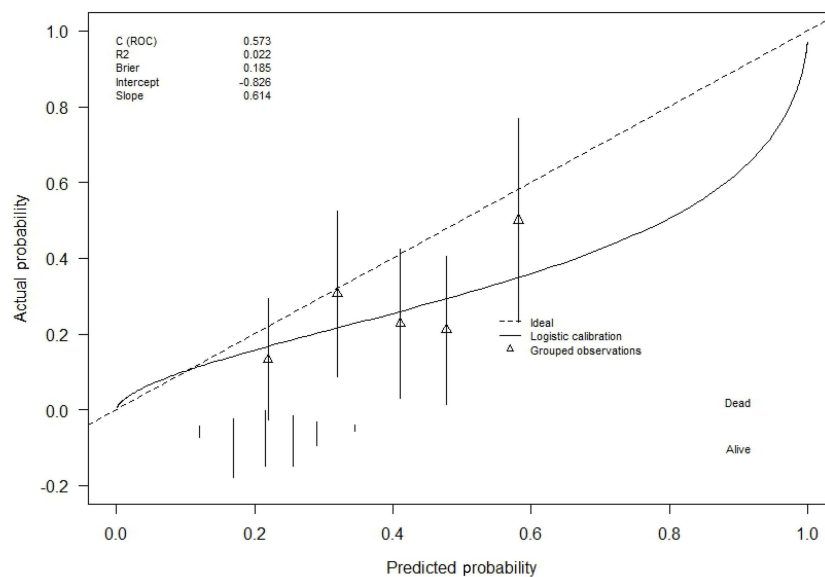

Figure 1 Model calibration – PREDICE.
Notes: Calibration plot: predicted outcome probabilities (on the x-axis) plotted against observed outcome frequencies (on the y-axis). Perfect predictions should be on the 45° line.
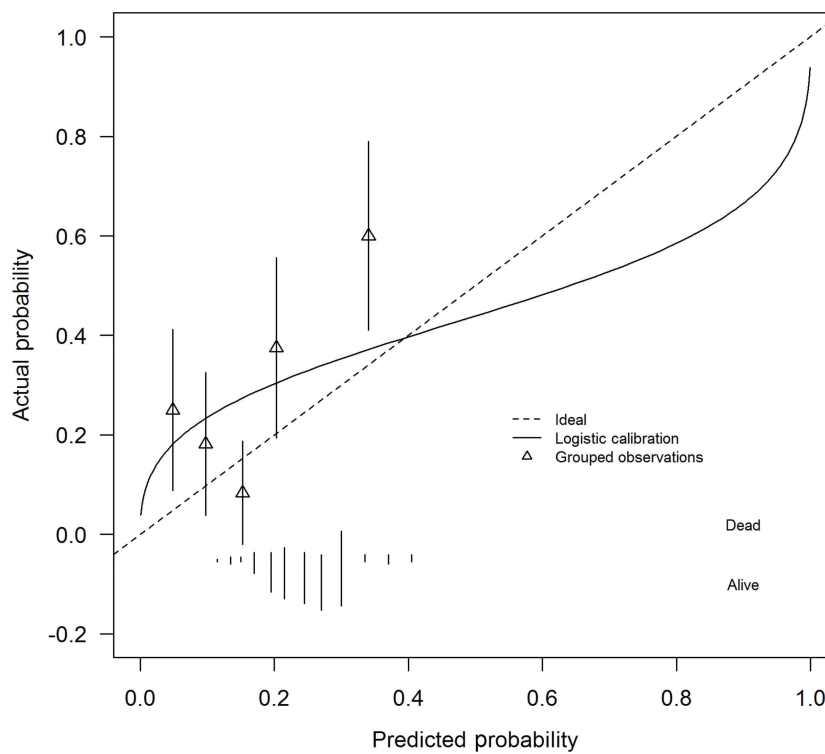
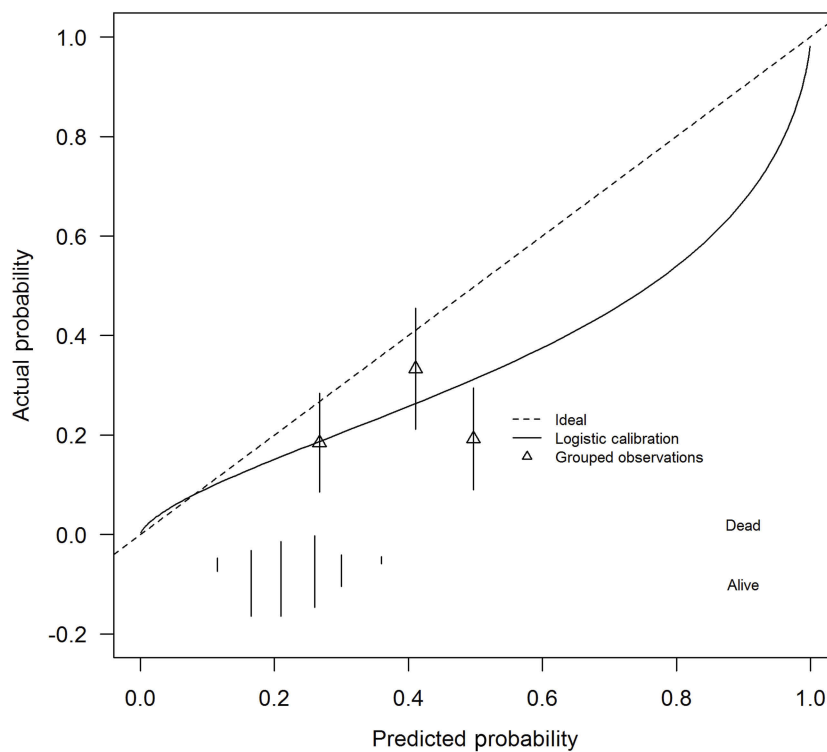**Figure 2** SCORE calibration – PREDICE.
Notes: Calibration plot: predicted outcome probabilities (on the x-axis) plotted against observed outcome frequencies (on the y-axis). Perfect predictions should be on the 45° line.

Figure 3 SCORE calibration – AHEAD (using the development cohort of PREDICE).
Notes: Calibration plot: predicted outcome probabilities (on the x-axis) plotted against observed outcome frequencies (on the y-axis). Perfect predictions should be on the 45° line.

Clinical prediction rule - AHEAD

Additionally, we tested the AHEAD-score, although target populations are different. In the AHEAD-score we included patients with primo-hospitalizations for HF,[13] but not only "de novo" HF. Figures 1 and 3 show that

the PREDICE model underestimates the risk of mortality while the PREDICE-score underestimates the real risk of mortality. We see that the AHEAD-score overestimates the risk too. We have verified that the calibration in the AHEAD-score (using development and validation cohorts

Figure 4 SCORE calibration – AHEAD (using the validation cohort of PREDICE).
Notes: Calibration plot: predicted outcome probabilities (on the x-axis) plotted against observed outcome frequencies (on the y-axis). Perfect predictions should be on the 45° line.

of PREDICE) is good but the discrimination in the AHEAD-score is poor.

Differences between populations of the PREDICE study

There are important differences between samples. On the one hand, this penalizes the model because it is examined on a sample with difference characteristics. But on the other hand, if one model wants to be useful in clinical practice it has to behave appropriately in any situation. This increases the pragmatic value of the study.[25]

If we focus on variables include in the PREDICE-score, we can find some differences.

Age has an influence on the prognosis. Furthermore, old patients have an increased prevalence of comorbidities and higher dependency. Advanced age is included in several prognostic models for mortality after hospitalization as in the PREDICE-score.[13,26] However, both PREDICE cohorts have similar mean age.

In several previous studies, social factors (living alone, family support, follow-up medical visits, etc.) had been proposed as predictors of mortality in chronic diseases, but dependence had never been identified as an important predictor.[8,27,28] In the original paper, we underlined that

being dependent for daily basic activities was found to be a relevant non-biological prognostic factor.[12]

The ejection fraction divides patients with HF into two large groups: HF preserved and HF reduced. Both diseases are very different in terms of etiology, prognosis and treatment.[1,29] There may also be differences in the pathophysiological diagnosis, despite the similarity between both populations. In the development cohort we did not obtain a pathophysiological diagnosis in an important proportion of cases,[12] which limits the comparison with the validation cohort.

Additionally if the Charlson-Index (CIN) is used to measure comorbidities,[30] we see that there is an important difference between both populations ($p<0.001$, Table 1). A Charlson-score above 2 indicates the presence of multiple comorbidities and/or advanced age, which generates worse life expectancy in the following decade.[30] In the PREDICE validation cohort, we identify that 90% of patients have a Charlson-Index higher than 2 points. However, in PREDICE derivation cohort only 12.8% of patients have a CIN higher than 2 points. These differences between both populations explain that patients who are currently hospitalized are more complex and that it is expected that they will have higher mortality. Lastly, the use of beta-blockers was higher in the validation cohort.

In summary, we found that patients belonging to the validation cohort have high comorbidity and more frequent systolic disfunction. These findings reveal that clinical spectrum of hospitalized patients with HF has been modified in the last decade. Additionally, there is a tendency to hospitalize more severe patients which, despite the better adherence to pharmacological treatment, involves an increase of overall mortality in the hospital setting.

Differences between the PREDICE and AHEAD populations

Age, gender, ischemic cardiopathy, atrial fibrillation, hemoglobin, beta-blockers, angiotensin converting enzyme (ACE) inhibitors and angiotensin receptors blockers (ARB) were the only data available for the four populations (Table 1). The most important differences between the four populations are found in the use of beta-blockers and in ischemic cardiopathy. Its prevalence is 35% in GREAT, 24% in AHEAD, 16% in the PREDICE development cohort and 10% in the PREDICE validation cohort.

Strengths and limitations of PREDICE

A strength of PREDICE is that the model includes a limited number of non-invasive and widely available variables. Generally, in predictive models there is a conflict between the complexity of the models and their clinical applicability. Models should be easy and quick to use, which should promote their use in normal clinical practice.[31,32] An example is Seattle Heart Failure, which is a predictive model of survival in chronic HF.[33] This model is widely used but has more than twenty variables so it is impossible to calculate it manually. For this reason, Levy et al provide a calculator to facilitate the use of the model.[7,13] Our model does not include a calculator, which is a relative limitation, since the score is calculated only with five variables.

The most important strength of the PREDICE study is the homogeneity of the cohorts used to develop and validate the score in which only hospitalized patients with new-onset HF are included. We excluded patients with a previous diagnosis of HF to delimit the incident cases so our model is very exclusive. It is known that "de novo" HF and acute decompensating of chronic HF are different entities so the models are not interchangeable.[34] In fact, multiple studies have found that acute decompensations of chronic HF are more frequent while mortality in "de novo"

HF is higher.[5,25] All this implies that it is not possible to validate others' models within our population because another prognostic models refers to patients with chronic HF. After initiating this study, another clinical prediction rule in HF was published called AHEAD but it refers to patients with chronic HF too, so we cannot use it to validate our model. This study includes patients from two European registries with first hospitalization for HF where we also found patients with "de novo" HF and patients with chronic HF decompensation.[13]

The study's main limitation could be its sample size. Only 180 were included, less than recommended in validation studies,[16,35] although, prior to the study, a sample size calculation was performed. Differentiating HF in systolic or diastolic is another limitation of our study, since the model could have been used as a numerical variable.

Additionally, the last guidelines of ESC show HF with a moderately reduced ejection fraction, although this is not well established.[1] In our study, we do not contemplate that entity but it should be taken into account for future studies.

The population of validation cohort is recruited in the same geographical and care areas as the development cohort, but ten years later, which increases the transportability. In addition, transportability increases if it is another investigator group that carries out the validation.[16]

It is important to note the difference between HF with a reduced ejection fraction (HFrEF) and HF with preserved ejection fraction (HFpEF). The prevalence of HFpEF has increased in western societies over recent years. The medical knowledge is poor and no specific treatments have been shown to reduce mortality. There are only expert opinions for the management of this pathology.[1,36] Both in the development population and in the validation population, the incidence of HFpEF is more frequent than HFrEF,[12] which is similar to findings in the literature. This is a strength versus other predictive models for HF, which included populations with similar incidences between HFrEF and HFpEF.[37]

Some recent prognostic models include atrial fibrillation (AF) as a variable with predictive power,[12] but other earlier models do not include this variable.[37] The prevalence of AF collected during the development and validation phases is around 30% similar to that found in other studies. However, in the original article AF was not included in the PREDICE prognostic model because it didn't reach sufficient predictive power.[12] Given the high prevalence of AF, it could be a limitation compared to other studies.

In our study, the results are reported after a one-year follow-up. This can be considered a limitation because in one year new clinical situations may appear and these may modify the course of the disease. On the other hand, this can also be considered as strength, because it has a long-term forecast although there are some articles that carry out a longer follow-up of patients, up to five years.[13] A longer follow-up allows us to know more outcomes of the same pathology, which can be valuable information.

Models that predict mortality in HF underestimate risk in elderly patients.[9,37] We have verified that our model underestimates risk in elderly patients too because both populations were elderly.[12]

Another choice when studying prognostic models lies in the moment where the prognosis is to be estimated. If an estimate is produced with data from the first evaluation, information about early evolution is neglected, probably losing accuracy but gaining interest (clinicians usually want to estimate risk as early as possible). We produced a prediction score which needs a physiopathology diagnosis by echocardiography and which does therefore not allow for the earliest risk assessment, but does not need to wait for a therapy response.

Conclusion

The present study shows that the two risk scores available for patients with primo-hospitalization for decompensated HF (PREDICE and AHEAD) are currently not valid for predicting mortality at one year. Both models performed poorly in the validation cohort compared to the development cohort.

In our setting the clinical spectrum of hospitalized patients with new-onset HF has been modified over time. In the last decade there has been an increase in the average age, comorbidities and dependence of patients, partly because hospital admission criteria have become more restrictive, which determines a worsening of their cardiac and general functional situation and, consequently, of the prognosis of the disease after hospitalization. The prognosis of "de novo" HF remains uncertain. For this reason, more research is needed to develop more accurate prognosis rules to be able to identify high-risk patients who require closer control and follow up.

The study underscores the need to validate the prognostic models before clinical implementation. Prognostic models developed should not be used if they are not validated, especially if the target population differs from the development one.

Disclosure
The authors developed a PREDICE-score in 2012. The authors report no other conflicts of interest in this work.

References

1. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*. 2016;37:2129–2200. doi:10.1093/eurheartj/ehw128

2. Ceia F, Fonseca C, Mota T, et al. Prevalence of chronic heart failure in Southwestern Europe: the EPICA study. *Eur J Heart Fail*. 2002;4:531–539. doi:10.1016/s1388-9842(02)00034-x

3. Van Riet EES, Hoes AW, Wagenaar KP, Limburg A, Landman MAJ, Rutten FH. Epidemiology of heart failure: the prevalence of heart failure and ventricular dysfunction in older adults over time. A systematic review. *Eur J Heart Fail*. 2016;18:242–252. doi:10.1002/ejhf.483

4. Defunciones según la Causa de Muerte 2014. Instituto Nacional de Estadística. [Internet, Spanish]. Available from: http://www.ine.es/jaxi/Datos.htm?path=/t15/p417/a2014/l0/&file=01001.px. Accessed 15 January, 2019

5. Harjola VP, Follath F, Nieminen MS, et al. Characteristics, outcomes, and predictors of mortality at 3 months and 1 year in patients hospitalized for acute heart failure. *Eur J Heart Fail*. 2010;12:239–248. doi:10.1093/eurjhf/hfq002

6. Gómez de la Cámara A, Tapia PM, Rivas FJP, Rodríguez-Moñino AP. ¿Qué pretenden estimar los estudios de pronóstico? Revisión de la metodología en la investigación pronóstica. *Med Clin (Barc)*. 2010;135:456–461. doi:10.1016/j.medcli.2009.03.020

7. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162:W1–W73. doi:10.7326/M14-0698

8. Alba AC, Agoritsas T, Jankowski M, et al. Risk prediction models for mortality in ambulatory patients with heart failure: a systematic review. *Circ Heart Fail*. 2013;6:881–889. doi:10.1161/CIRCHEARTFAILURE.112.000043

9. Ouwerkerk W, Voors AA, Zwinderman AH. Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure. *JACC Heart Fail*. 2014;2:429–436. doi:10.1016/j.jchf.2014.04.006

10. Stevenson LW, Davis RB. Model building as an educational hobby. *Circ Heart Fail*. 2016;9:pii: e003457. doi:10.1161/CIRCHEARTFAILURE.116.003457

11. Rahimi K, Bennett D, Conrad N, et al. Risk prediction in patients with heart failure: a systematic review and analysis. *JACC Heart Fail*. 2014;2:440–446. doi:10.1016/j.jchf.2014.04.008

12. Gómez de la Cámara A, Guerra Vales JM, Tapia PM, et al. Role of biological and non biological factors in congestive heart failure mortality: PREDICE-SCORE: a clinical prediction rule. *Cardiol J*. 2012;19:578–585.

13. Spinar J, Jarkovsky J, Spinarova L, et al. AHEAD score–long-term risk classification in acute heart failure. *Int J Cardiol*. 2016;202:21–26. doi:10.1016/j.ijcard.2015.08.187

14. Levy WC, Anand IS. Heart failure risk prediction models: what have we learned? *JACC Heart Fail*. 2014;2:437–439. doi:10.1016/j.jchf.2014.05.006

15. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19:453–473.

16. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68:279–289. doi:10.1016/j.jclinepi.2014.06.018

17. Steyerberg E Clinical prediction models: a practical approach to development, validation, and updating [Internet]. Springer Science & Business Media; 2008 [Internet]. Available from: https://books.google.es/books?hl=es&id=kHGK58cLsMIC&oi=fnd&pg=PR2&dq=Clinical+Prediction+Models:+A+Practical+Approach+to+Development,+Validation,+and+Updating&ots=TMXdI0eLhi&sig=G1pavyIK0wWc141D3kZlQEERlNw. Accessed 15 January, 2019.

18. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of diet in renal disease study group. *Ann Intern Med*. 1999;130:461–470. doi:10.7326/0003-4819-130-6-199903160-00002

19. Harrell F Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer; 2015 [Internet, Spanish]. Available from: https://books.google.es/books?hl=es&id=94RgCgAAQBAJ&oi=fnd&pg=PR7&dq=Regression+Modeling+Strategies:+With+Applications+to+Linear+Models,+Logistic+Regression,+and+Survival+Analysis&ots=Zzq7Uo7S2m&sig=DuEFLtS8jUTOcoCwaI-oizMZZeI. Accessed 15 January, 2019.

20. Núñez E, Steyerberg EW, Núñez J. Regression modeling strategies. *Rev Esp Cardiol*. 2011;64:501–507. doi:10.1016/j.recesp.2011.01.019

21. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98:691–698. doi:10.1136/heartjnl-2011-301247

22. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605. doi:10.1136/bmj.b902

23. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10:e1001381. doi:10.1371/journal.pmed.1001381

24. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol*. 2005;58:475–583. doi:10.1016/j.jclinepi.2004.06.017

25. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14:40. doi:10.1186/1471-2288-14-40

26. Lazzarini V, Mentz RJ, Fiuzat M, Metra M, O'Connor CM. Heart failure in elderly patients: distinctive features and unresolved issues. *Eur J Heart Fail*. 2013;15:717–723. doi:10.1093/eurjhf/hft028

27. Nielsen FE, Mard S. Single-living is associated with increased risk of long-term mortality among employed patients with acute myocardial infarction. *Clin Epidemiol*. 2010;2:91–98.

28. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review Brayne C, editor. *PLoS Med*. 2010;7:e1000316. doi:10.1371/journal.pmed.1000316

29. Rudski LG, Lai WW, Afilalo J, et al. Guidelines for the echocardiographic assessment of the right heart in adults: a report from the American society of echocardiography. *J Am Soc Echocardiogr*. 2010;23:685–713. doi:10.1016/j.echo.2010.05.010

30. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40:373–383.

31. Ely JW. Answering physicians' clinical questions: obstacles and potential solutions. *J Am Med Inform Assoc*. 2004;12:217–224. doi:10.1197/jamia.M1608

32. Kawamoto K. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005;330:765–770. doi:10.1136/bmj.38398.500764.8F

33. Levy WC, Mozaffarian D, Linker DT, et al. The Seattle heart failure model: prediction of survival in heart failure. *Circulation*. 2006;113:1424–1433. doi:10.1161/CIRCULATIONAHA.105.584102

34. Cohen-Solal A, Laribi S, Ishihara S, et al. Prognostic markers of acute decompensated heart failure: the emerging roles of cardiac biomarkers and prognostic scores. *Arch Cardiovasc Dis*. 2015;108:64–74. doi:10.1016/j.acvd.2014.10.002

35. Miralda GP, Soriano N, Brotons C, et al. Baseline characteristics and determinants of outcome in a patient population admitted for heart failure to a general hospital. *Rev Esp Cardiol*. 2002;55:571–578.

36. Jondeau G, Arnoult F, Caligiuri G, et al. Practical management of heart failure with preserved ejection fraction. A modest proposal. *Arch Cardiovasc Dis*. 2013;106:345–348. doi:10.1016/j.acvd.2013.04.005

37. Nutter AL, Tanawuttiwat T, Silver MA. Evaluation of 6 prognostic models used to calculate mortality rates in elderly heart failure patients with a fatal heart failure admission. *Congest Heart Fail*. 2010;16:196–201. doi:10.1111/j.1751-7133.2010.00180.x