European Journal of Internal Medicine 66 (2019) 35–40

Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim

Original Article

External validation of the CACE-HF risk score for mortality in patients with heart failure

Quirós-López R.[a,b,c,*], Manzano-Espinosa L.[c,d], Bilbao A.[b,g], Montero Pérez-Barquero M.[c,e], Cepeda-Rodrigo J.M.[c,f], Escobar-Martínez A.[b,g]

[a] Internal Medicine Unit, Costa del Sol Hospital, Marbella, Spain
[b] Health Services Research on Chronic Patients Network (REDISSEC), Spain
[c] RICA Registry, Spain
[d] Heart Failure and Vascular Risk Unit, Department of Internal Medicine, Ramón y Cajal University Hospital, Universidad of Alcalá, Madrid, Spain
[e] Department of Internal Medicine, IMIBIC/Hospital Reina Sofía, University of Cordoba, Spain
[f] Internal Medicine Department, Vega Baja Hospital, Orihuela, Valencia, Spain
[g] Research Unit, Basurto University Hospital, Bilbao, Spain

ARTICLE INFO

Keywords:
Heart failure
Clinical prediction rule

ABSTRACT

Aims: To validate externally the CACE-HF clinical prediction rule, which predicts 1-year mortality in patients with heart failure (HF).

Methods: We performed an external validation of the CACE-HF risk score in patients included in the RICA heart failure registry who had completed 1 year of follow-up, comparing the characteristics of the derivation and validation cohorts. The performance of the risk score was evaluated in terms of calibration, using calibration-in-the-large (a), calibration slope (b), and the Hosmer-Lemeshow test, and in terms of discrimination, using the area under the ROC curve.

Results: In total, 3337 patients were included in the validation cohort. There were no significant differences between the derivation and validation cohorts in 1-year mortality (24.63% vs. 22.98%) or in the risk score and risk classes. The discrimination capacity in the validation cohort was slightly lower, 0.67 (95% CI: 0.65, 0.69), compared to that of the derivation cohort. Calibration results were a −0.05 (95% CI: −0.14, 0.03), indicating that the average predictions did not differ from the average outcome frequency, and b = 0.75 (95% CI: 0.64, 0.86), indicating a modest inconsistency in predictor effects. Observed mortality versus predicted mortality according to the deciles and risk classes were very similar in both cases, indicating good calibration.

Conclusions: The results of the external validation of the CACE-HF risk score show that although the capacity for discrimination was slightly lower than in the derivation cohort, the calibration was excellent. This tool, therefore, can assist in decision-making in the management of these patients.

1. Introduction

Heart failure (HF) is the paradigm of chronic diseases, characterised by patients evolving steadily through different stages of progression until death [1]. Its high prevalence, estimated at 16.1% of the population over 75 years of age, is a real challenge for health providers and even for society as a whole: if the prevalence of HF is combined with the expected population figures, in Spain alone the number of individuals with this disease will increase by 21% between 2020 and 2030. By 2050, the prevalence of HF will have increased by 98.5%.

If we are to respond to this threat, we need to modify our management of HF, with the implementation of HF programs and tools that help us optimize the available resources, so that patients can be offered solutions conferring potential benefits in quality of life and prognostic expectations. Tools that help us differentiate the life expectancy of patients include clinical prediction rules, also known as prognostic models. Although a great variety of models have been described, these have not been widely used for different reasons: few have been validated externally in populations other than those in which they were developed, and to date we have not identified a highly accurate model with a small number of variables that is easy to apply in routine practice [2]. External validation should be a preliminary step in the implementation of newly created models, and the procedure should be carried out in a patient population other than the one used for internal

* Corresponding author at: Internal Medicine Unit, Costa del Sol Hospital, Marbella, Spain
E-mail address: quiroslopez77@gmail.com (R. Quirós-López).

https://doi.org/10.1016/j.ejim.2019.05.010
Received 27 December 2018; Received in revised form 6 May 2019; Accepted 13 May 2019
Available online 11 June 2019

development and validation [3].

In these cases, the discrimination and calibration of the model are studied. Discrimination is the ability of the model to distinguish patients who present the result of interest, mortality in our case, from those who do not. Calibration refers to the agreement between the probabilities observed and predicted by the model.

Our group developed and internally validated a clinical prediction rule, the CACE-HF risk score for 1-year mortality in patients admitted for HF [4]. Two models were developed and validated in the study. The variables selected for the first included data on health-related quality of life (HRQOL), using the specific Minnesota Living with Heart Failure (MLHFQ) questionnaire, and the second included the same variables as the first, with the exception of the HRQOL questionnaire.

The objective of this study was to conduct an external validation of the CACE-HF risk score model in a sample from the RICA registry (National Registry of Cardiac Insufficiency) [5].

2. Methods

2.1. CACE-HF risk score derivation cohort

This cohort has been described elsewhere [4]. It is, in brief, a prospective, multicentre cohort study, conducted in 13 hospitals in Spain: 3 in the Canary Islands, 4 in Andalusia, 1 in Catalonia, and 5 in the Basque Country. The cohort comprised patients admitted consecutively between January 2009 and May 2013 to the cardiology or internal medicine departments. The first admission within the study period was the index admission, and subsequent hospitalizations were considered re-admissions. Inclusion criteria were: primary diagnosis of HF [International Classification of Diseases (ICD), 9th Revision, Clinical Modification code 428], age over 18 years, and agreement to participate in the study. We excluded patients who developed the HF episode during their hospital stay, were transferred from other healthcare centres, or who died during the stay. All patients were followed up for 1 year. The study complies with the ethical principles of the Declaration of Helsinki and all patients provided signed informed consent. In that study, we developed (derivation cohort) and internally validated (validation cohort) a clinical prediction rule by which patients admitted for HF can be classified according to their 1-year mortality risk. A total of 1283 patients were included in the final model.

2.2. External validation cohort

The RICA Registry was created by the Heart Failure Working Group of the Spanish Society of Internal Medicine [6]. The study protocol was approved by the Ethics Committee of the Reina Sofía University Hospital of Cordoba. All patients give their informed consent to participate in the study.

This is a multicentre, prospective study that began in March 2008, designed to study the characteristics and progress of patients hospitalized with heart failure in the Internal Medicine departments of participating hospitals with a 1-year follow-up.

The objective of this registry is to analyse the characteristics of patients admitted for heart failure in Spanish hospitals and to assess their clinical progress after hospital discharge. Patients are included if they are older than 50 years and have a diagnosis of HF according to the criteria of the European Society of Cardiology [1]. The criteria for including patients in the RICA registry were similar to those of the CACE study. As of March 2018, the RICA registry had included 4188 patients with HF, of which 3725 had already completed the full 1-year follow-up. 30 patients were excluded due to being in-hospital deaths. Of the remaining 3695 patients, we lost 358 (9.69%) because they had some variables necessary to define the risk score missing. Specifically, the missing percentage was as follows: 0.60% (n = 22) for NYHA, 0.73% (n = 27) for cause of valvular HF, and 8.44% (n = 312) for urea. In the rest of the variables we did not have any missing data. Finally, 3337 patients were included in our external validation study.

2.3. Variables

Two models were used in the creation and validation of the CACE-HF score [4]. Model I contained the following variables: age, systolic blood pressure, NYHA classification at admission, valvular cause of HF, dementia, previous hospitalizations, haemoglobin, sodium, urea, average stay and physical dimension of the specific MLHFQ HRQOL questionnaire. Model II, also developed and validated internally in the previous study, contains the same variables, with the exception of the MLHFQ questionnaire. The outcome variable was mortality 1 year after discharge.

Since the RICA registry does not include HRQOL data, we chose Model II for external validation. For this study, we used the original derivation cohort composed of 1283 patients.

2.4. Statistical analysis

Data are displayed as means with standard deviation for continuous variables, and frequencies and proportions for categorical variables.

We quantified the extent to which the development and validation samples are related. Firstly, patient characteristics, including predictor variables and the 1-year mortality (outcome), were compared between the derivation and validation cohorts. Chi-square or Fisher's exact test were performed for the comparison of categorical variables, and the Student's t-test or nonparametric Wilcoxon tests were performed for continuous variables. Secondly, we used the two statistical approaches proposed by Debray et al. [8], to calculate an overall measure of their (dis)similarity. The first one calculates a summary measure of relatedness based on how well the study individuals from both samples can be distinguished. We considered a logistic regression model, henceforth referred to as the membership model, to predict the probability that an individual belongs to the original derivation sample as compared with the RICA validation sample. Therefore, the dependent variable is belonging or not belonging to the derivation sample, and the independent variables are both the predictors and the outcome of the original prediction model. If the membership model discriminates poorly, both samples are strongly related, and vice versa. The discriminative ability is measured by concordance (c) statistics. The second approach assesses the extent to which the predicted risk distribution of the derivation and validation samples diverge. For this purpose, the mean and standard deviation (SD) of the linear predictor (LP) of the model based on the predictor variables and the model based on the risk score was calculated. The LP is the logit transformation of the predicted risk in the logistic regression. An increased SD of the LP reflects more heterogeneity, that is, a larger variety of patient characteristics. On the other hand, the difference in the mean of the LP reflects the differences in the outcome frequency.

Moreover, both the original model based on the predictors and the model based on the risk score were refitted in the validation sample by means of logistic regression models. In both cases, the dependent variable was mortality at 1 year and the independent variables were the predictive variables in the first case, and the developed risk score in the second case. We analysed the significance of each variable, and the predictive accuracy of the model by calculating the area under the ROC curve (AUC) for discrimination [9], and by comparing predicted and observed mortality at 1 year using the Hosmer and Lemeshow test for calibration [10].

We then evaluated the performance of the developed risk score in the validation sample in terms of calibration and discrimination [8,11,12]. Calibration refers to the agreement between observed outcome and predictions. For this purpose, calibration-in-the-large, also called calibration intercept, and calibration slope were calculated. These values are ideally 0 and 1, respectively (perfect calibration). They were calculated as follows: 1) calibration-in-the-large is given as the

intercept term a from the recalibration model, based on the risk score, $logit(y) = a + logit(\hat{y})$ [13] [1]. Because the calibration intercept is difficult to interpret when the calibration slope is not equal to 1, the calibration intercept is estimated with the calibration slope fixed at 1. This is done by fitting the logistic regression model with the regression coefficient of the linear predictor fixed at 1 [12]. The calibration-in-the-large quantifies whether the average of predictions corresponds with the average outcome frequency and ideally equals 0. Values below 0 indicate that the model overestimates the outcome, and values above 0 indicate that the model underestimates the outcome. 2) Calibration slope, denoted as b, can be estimated from the recalibration model $logit(y) = a + b \cdot logit(\hat{y})$. Values of $b = 1$ indicate that predicted risks are appropriately scaled with respect to each other over the entire range of predicted probabilities. $b > 1$ occurs where predicted probabilities do not vary enough, and values $0 < b < 1$ occurs when they vary too much (that is, predicted risks are too low for low outcome risks and too high for high outcome risks). A poor calibration slope ($0 < b < 1$) usually reflects overfitting of the model in the development sample but may also indicate inconsistency of predictor effects between the development and validation samples. Moreover, the Hosmer and Lemeshow goodness-of-fit test was used to study calibration [10], comparing the observed and predicted outcome according to the deciles and according to the 4 risk classes [4].

Regarding discrimination, the concordance statistic was estimated. The concordance statistic, c, corresponds to the area under the ROC curve, and can range from 0.5 (no discrimination) to 1 (perfect discrimination) [9].

All effects were considered significant at $p < .05$, unless otherwise stated. All statistical analyses were performed using SAS for Windows statistical software, version 9.2 (SAS Institute, Inc., Carey, NC).

3. Results

The comparison of predictor variables and the outcome between the derivation and validation cohort are shown in Table 1. We found statistical differences in age, NYHA, valvular cause of HF and prior HF hospitalization. The percentage of patients with age ≥ 75 was higher in the validation sample ($p < .0001$), while the percentage of patients with NYHA III-IV, cause of valvular HF, or prior HF hospitalization were higher in the derivation cohort ($p < .01$). There was no significant difference between the groups for 1-year mortality. The overall rates of mortality were 24.63% in the derivation cohort and 22.98% in the validation cohort.

Table 2 shows the comparison of the distribution of the developed risk score and the risk classes between the derivation and validation cohorts. There were no significant differences in either of the two.

Table 1
Predictive variables and outcome in the derivation and validation cohorts.

Characteristics	Derivation Cohort	RICA Validation Cohort	*p*-Value
	n = 1283	n = 3337	
	n (%)	n (%)	
Predictors			
Age ≥ 75	887 (69.13)	2509 (75.19)	< 0.0001
SBP < 135	604 (47.94)	1594 (47.77)	0.9185
NYHA III-IV	557 (43.41)	1245 (37.31)	0.0001
Cause of valvular HF	275 (21.43)	598 (17.92)	0.0063
Dementia	60 (4.68)	179 (5.36)	0.3447
Prior HF hospitalization	428 (33.36)	852 (25.53)	< 0.0001
Haemoglobin < 13	856 (67.61)	2238 (67.07)	0.7234
Sodium < 136	245 (19.46)	646 (19.36)	0.9383
Urea ≥ 86	329 (26.03)	783 (23.46)	0.0697
Length of stay ≥ 14	255 (19.88)	654 (19.60)	0.8321
Outcome			
Mortality at 1 year	316 (24.63)	767 (22.98)	0.2372

HF: Heart failure; SBP: Systolic blood pressure.

Table 2
Mortality risk score and risk classes in the derivation and validation cohorts.

Characteristics	Derivation Cohort	RICA Validation Cohort	*p*-value
	n = 1283	n = 3337	
Risk score, mean (SD)	11.21 (5.14)	10.92 (4.88)	0.0826
Risk classes, n (%)			0.0967
Score ≤ 7	293 (23.76)	825 (24.72)	
7 < Score ≤ 13	535 (43.99)	1518 (45.49)	
13 < Score ≤ 16	205 (16.63)	547 (16.39)	
Score > 16	200 (16.22)	447 (13.40)	

SD: Standard deviation. Score range of values: 0–32.

Regarding (dis)similarities and differences between both models, the concordance statistic of the membership model based on predictors and the outcome was $c_m = 0.58$ with a 95% confidence interval (CI) of (0.56, 0.59), and the c_m based on the risk score and the outcome was 0.52 (95% CI: 0.50, 0.53), indicating in both cases that both samples are largely similar. Regarding the distribution of the LP of the model based on predictor variables, SD was similar in both the derivation and validation samples, 0.86 vs. 0.82, respectively, and the mean of the LP was slightly lower in the validation cohort (−1.27 vs. -1.34). Similar results are obtained in the distribution of the LP of the model based on the risk score. SD was similar in both cohorts (0.87 vs. 0.82) and the mean was also slightly lower in the validation cohort (−1.28 vs. -1.32). Both approaches therefore indicate that the derivation and validation samples had a similar distribution.

Table 3 shows the results of the logistic regression model based on the predictors and based on the developed risk score in the derivation cohort and the refitted models in the validation cohort. In the refitted model based on the predictors, all variables remained statistically significant. The discrimination ability was somewhat lower, $c = 0.68$ (95% CI: 0.65, 0.70), and the model was well calibrated (Hosmer and Lemeshow, $p = .6020$). In the refitted model based on the risk score the results were similar. The model was also well calibrated (Hosmer and Lemeshow, $p = .8954$) and the discrimination ability was 0.67 (95% CI: 0.65, 0.69).

With respect to the performance of the developed risk score in the validation sample, the calibration and discrimination results are shown in Table 4 and Fig. 1. As stated previously, the discrimination ability was slightly lower, 0.67 (95% CI: 0.65, 0.69), compared with that of the derivation cohort. Calibration-in-the-large was $a = -0.05$ (95% CI: −0.14, 0.03, indicating that the average of the predictions did not differ from the average outcome frequency. The calibration slope was $b = 0.75$ (95% CI: 0.64, 0.86), indicating overfitting of the model or inconsistency of predictor effects between the development and validation samples (Table 4). The observed vs. predicted 1-year mortality according to the deciles and according to the risk classes are shown in Fig. 1. Very similar results are observed in both cases, indicating good calibration.

4. Discussion

We have externally validated a risk model to predict 1-year mortality in patients admitted for HF in a cohort of patients similar to that of the developed model. The RICA cohort is a real-world observational registry of HF admissions without restrictions with respect to age, comorbidity and value of left ventricular ejection fraction, similar to the population in which the risk score was developed. The model has an acceptable number of variables and is comprised of information routinely collected at discharge.

Before they are used in clinical practice, risk prediction models should be seen to perform well in a similar but separate population

Table 3
Estimated regression coefficients and corresponding 95% confidence interval and standards errors (SE) for the derivation and validation samples, considering the prediction model based on the predictors and based on the risk score.

	Derivation Cohort			RICA Validation Cohort		
	n = 1283			n = 3337		
	β (95% CI)	SE	p-value	β (95% CI)	SE	p-value
Predictive model						
Intercept	−3.14 (−3.59, −2.68)	0.23	< 0.0001	−2.47 (−2.73, −2.21)	0.13	< 0.0001
Age ≥ 75	0.79 (0.46, 1.13)	0.17	< 0.0001	0.37 (0.16, 0.58)	0.11	0.0006
SBP < 135	0.34 (0.06, 0.62)	0.14	0.0169	0.21 (0.04, 0.38)	0.09	0.0160
NYHA III-IV	0.40 (0.12, 0.68)	0.14	0.0049	0.40 (0.23, 0.57)	0.09	< 0.0001
Cause of valvular HF	0.39 (0.07, 0.71)	0.16	0.0171	0.38 (0.17, 0.58)	0.10	0.0003
Dementia	0.71 (0.13, 1.29)	0.30	0.0160	0.85 (0.53, 1.18)	0.17	< 0.0001
Prior HF hospitalization	0.56 (0.28, 0.85)	0.15	0.0001	0.34 (0.15, 0.52)	0.09	0.0003
Haemoglobin < 13	0.45 (0.13, 0.78)	0.17	0.0062	0.21 (0.02, 0.40)	0.10	0.0314
Sodium < 136	0.45 (0.13, 0.78)	0.17	0.0064	0.30 (0.10, 0.50)	0.10	0.0037
Urea ≥ 86	0.67 (0.38, 0.97)	0.15	< 0.0001	0.69 (0.51, 0.88)	0.09	< 0.0001
Length of stay ≥ 14	0.48 (0.15, 0.81)	0.17	0.0040	0.40 (0.20, 0.60)	0.10	0.0001
Hosmer and Lemeshow, p-value	0.6907			0.6020		
AUC (95% CI) model	0.72 (0.69, 0.76)			0.68 (0.65, 0.70)		
Risk score						
Intercept	−3.17 (−3.58, −2.75)	0.21	< 0.0001	−2.69 (−2.92, −2.45)	0.12	< 0.0001
Score	0.17 (0.14, 0.20)	0.02	< 0.0001	0.13 (0.11, 0.14)	0.01	< 0.0001
Hosmer and Lemeshow, p-value	0.7226			0.8954		
AUC (95% CI) model	0.72 (0.69, 0.76)			0.67 (0.65, 0.69)		

AUC: Area under the ROC curve; CI: Confidence interval; SE: standard error.
Risk score range of values: 0–32.

Table 4
Assessment of the performance of the model in the validation cohort, considering the prediction model based on the risk score: calibration and discrimination.

	Derivation Cohort	RICA Validation Cohort
	n = 1283	n = 3337
Predictive risk score		
Calibration-in-the-large, *a* (95% CI)	0 (−0.22, 0.22)	−0.05 (−0.14, 0.03)
Calibration slope, *b* (95% CI)	1 (0.82, 1.18)	0.75 (0.64, 0.86)
Concordance statistic, *c* (95% CI)	0.72 (0.69, 0.76)	0.67 (0.65, 0.69)

CI: confidence interval.

from the one used for derivation and internal validation [14].

In this study we have followed the methodology recently proposed by Debray et al. [8], which comprises several steps. In the first step to examine the relatedness of the development and validation samples, our validation cohort showed differences in some predictors, but there were no differences in the distribution of 1-year mortality. The 2 approaches to investigate the extent of relatedness of both cohorts provided satisfactory results: 1) the membership models showed concordance statistics close to 0.5, indicating that both samples were largely similar; and 2) the distribution of the LP of the models in both cohorts was also similar. Therefore, both approaches indicated that the derivation and validation samples had a similar distribution. Results from step 2, performed to assess the model's performance in the validation cohort, were also satisfactory. All predictor variables remained statistically significant in the validation sample, although the discrimination ability was slightly lower. The calibration slope was < 1, indicating a modest inconsistency of the predictor effects between the development and validation samples. However, our other calibration methods, such as calibration-in-the-large or the Hosmer and Lemeshow goodness-of-fit test, demonstrated excellent calibration, strengthening the validation of our model.

Why is it important to know the life expectancy of patients with HF? In the first place, we can try to modify the natural course of the disease, optimizing treatment and adapting it to patient needs, and secondly, the early identification of patients with a worse life expectancy can assist in decision-making, and help us to avoid the use of drugs with no benefit and focus on the control of symptoms.

As mentioned above, there are several reasons why HF forecast models have not been sufficiently implemented in routine practice, and it is clear that we need precise, simple and reproducible models, based on variables that can be obtained at the bedside.

A recent meta-analysis [15] identified 117 models studying outcomes of mortality or hospitalizations due to heart failure, in which 249 variables were used. Our model consists of 10 variables, several of which are among the most commonly used, including age, systolic blood pressure, sodium, NYHA functional class, urea and aetiology. We have also included other variables that are easy to obtain, such as haemoglobin, length of stay and prior hospitalizations.

Another review [16] found 43 main models for the prediction of death and 11 for the prediction of death or hospitalization, developed in a range of settings: 25 included hospitalized patients only, including those presenting to emergency departments. The timing of data collection also varied from admission to the pre-discharge visit, as in our case. In this review, 40% (8/20) of the models that were developed in hospitalized patients with reduced and preserved systolic function were validated externally, but half of these did not describe the calibration of the model. We believe that this is a strength of our study.

Our study also has limitations [17]. Firstly, the same authors developed the model and carried out the external validation. An independent evaluation conducted by authors not involved in its development would be desirable. We included 90% of the total RICA cohort, and some data corresponding to different variables were missing from a few subjects, so we did not perform multiple imputation. The external validation of our prediction model only in Spanish population may be a limitation, as well as the fact that patients were always attended by internal medicine doctors. Similarly, the differences in annual mortality due to heart failure between different European countries could condition the reliability of the results of our score in them.

In conclusion, this risk score is quite a simple tool for stratifying patients according to the risk of death at 1 year. Knowledge of this risk would allow the different treatments to be adjusted to the specific needs

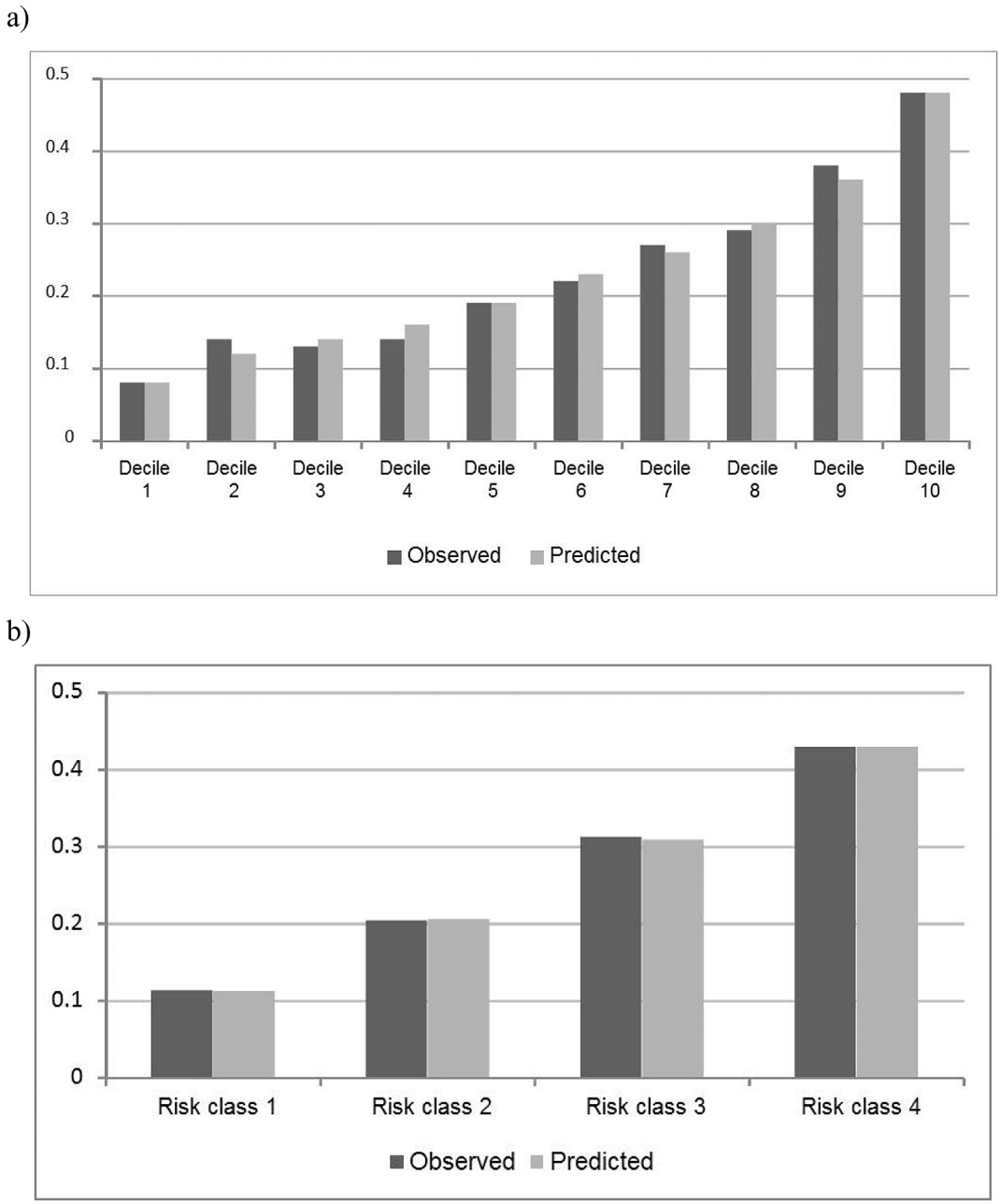

Fig. 1. Observed vs. model-predicted 1-year mortality based on the risk score (Fig. 1a) and according to the risk classes (Fig. 1b). Risk classes are defined as follows: score ≤ 7 (risk class 1), 7 < score ≤ 13 (risk class 2), 13 < score ≤ 16 (risk class 3), and score > 16 (risk class 4).
Decil: Decile.

of each patient.

References

[1] Ponikowski P, Voors AA, Anker SD, et al. 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 J Heart Fail 2016;2016(18):891–975. https://doi.org/10.1002/ejhf.592.

[2] 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;2:429–36.

[3] Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;338:b605.

[4] Escobar A, Garcia-Perez L, Navarro G, et al. A one-year mortality clinical prediction rule for patients with heart failure. Eur J Intern Med 2017;44:49–54.

[5] Trullas JC, Miro O, Formiga F, et al. The utility of heart failure registries: a descriptive and comparative study of two heart failure registries. Postgrad Med J 2016;92:260–6.

[6] Trullas JC, Formiga F, Montero M, et al. Paradox of obesity in heart failure: results from the Spanish RICA Registry. Med Clin (Barc) 2011;137:671–7.

[8] Debray TP, Vergouwe Y, Koffijberg H, et al. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 2015;68:279–89.

[9] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36.

[10] Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley; 1989.

[11] Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128–38.

[12] Janssen KJ, Moons KG, Kalkman CJ, et al. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol 2008;61:76–86.

[13] Steyerberg EW. Clinical prediction models. A practical approach to development, validation, and updating. Springer; 2009.

[14] Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External

validation, model updating, and impact assessment. Heart 2012;98:691–8.
[15] Fonarow GC. Clinical risk prediction tools in patients hospitalized with heart failure. Rev Cardiovasc Med 2012;13:14–23.
[16] Rahimi K, Bennett D, Conrad N, et al. Risk prediction in patients with heart failure: a systematic review and analysis. JACC Heart Fail 2014;2:440–6.
[17] 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.