Rev Bras Cir Cardiovasc 2012;27(2):187-94

Validation of The 2000 Bernstein-Parsonnet and Euroscore at the Heart Institute - USP

Validação do 2000 Bernstein-Parsonnet e EuroSCORE no Instituto do Coração - USP

Omar Asdrúbal Vilca Mejía[1], Luiz Augusto Ferreira Lisboa[2], Luis Alberto Oliveira Dallan[3], Pablo Maria Alberto Pomerantzeff[4], Luiz Felipe Pinho Moreira[5], Fabio Biscegli Jatene[6], Noedir Antonio Groppo Stolf[7]

DOI: 10.5935/1678-9741.20120033

RBCCV 44205-1370

Abstract

Objective: **To validate the 2000 Bernstein Parsonnet (2000BP) and additive EuroSCORE (ES) to predict mortality in patients who underwent coronary bypass surgery and/or heart valve surgery at the Heart Institute, University of São Paulo (InCor/HC-FMUSP).**

Methods: **A prospective observational design. We analyzed 3000 consecutive patients who underwent coronary bypass surgery and/or heart valve surgery, between May 2007 and July 2009 at the InCor/HC-FMUSP. Mortality was calculated with the 2000BP and ES models. The correlation between estimated mortality and observed mortality was validated by calibration and discrimination tests.**

Results: **There were significant differences in the prevalence of risk factors between the study population, 2000BP and ES. Patients were stratified into five groups for 2000BP and three for the ES. In the validation of models, the ES showed good calibration (*P* = 0.596), however, the 2000BP (*P* = 0.047) proved inadequate. In discrimination, the area under the ROC curve proved to be good for models, ES (0.79) and 2000BP (0.80).**

Conclusion: **In the validation, 2000BP proved questionable and ES appropriate to predict mortality in patients who underwent coronary bypass surgery and/or heart valve surgery at the InCor/HC-FMUSP.**

Descriptors: **Risk factors. Cardiovascular surgical procedures. Risk assessment. Hospital mortality. Validation studies.**

1. PhD in Sciences at Faculty of Faculty of Medicine of the University of São Paulo (FMUSP), São Paulo, SP, Brazil.
2. Full Professor, Physician assistant of the Coronary Heart Diseases Unit at Heart Institute of the Clinics Hospital, Faculty of Medicine, University of São Paulo (InCor/HC-FMUSP), São Paulo, SP, Brazil.
3. Full Professor, Associate Professor at FMUSP, Director of the Coronary Heart Diseases Unit of InCor/HC-FMUSP, São Paulo, SP, Brazil.
4. Full Professor, Associate Professor at FMUSP, Director of the Surgical Unit of Valve Diseases at InCor/HC-FMUSP, São Paulo, SP, Brazil.
5. Full Professor, Associate Professor at FMUSP, Director of the Research Surgical Unit at InCor/HC-FMUSP, São Paulo, SP, Brazil.
6. Titular Professor of the Thoracic Surgery Discipline at FMUSP, São Paulo, SP, Brazil.
7. Titular Professor of the Discipline of Cardiovascular Surgery at FMUSP, São Paulo, SP, Brazil.

This study was carried out at Heart Institute of the Clinics Hospital, Faculty of Medicine, University of São Paulo (InCor/HC-FMUSP), São Paulo, SP, Brazil.

Correspondence address
Omar A. V. Mejía.
Av. Dr. Enéas de Carvalho Aguiar, 44 – Cerqueira César – São Paulo, SP, Brazil – Zip Code 05403-000.
E-mail: omarvilca@incor.usp.br

Support: National Council for Scientific and Technology Development

Article received on March 15th, 2012
Article accepted on May 2nd, 2012

Abbreviations, Acronyms & Symbols	
2000BP	2000 Bernstein Parsonnet
CAPPesq	Ethics Committee for Research Projects Analysis
CRM	CABG
ES	Aditive EuroSCORE
InCor/HC-FMUSP	Heart Institute, University of São Paulo

Resumo

Objetivo: **Validar o 2000 Bernstein Parsonnet (2000BP) e EuroSCORE aditivo (ES) na predição de mortalidade cirúrgica nos pacientes operados de coronária e/ou valva, no Instituto do Coração da Universidade de São Paulo (InCor/HC-FMUSP).**

Métodos: **Desenho prospectivo e observacional. Foram analisados, 3000 pacientes consecutivos operados de coronária e/ou valva, entre maio de 2007 e julho de 2009 no InCor/HC-FMUSP. A mortalidade foi calculada com os escores 2000BP e ES. A correlação entre mortalidade estimada e mortalidade observada foi validada mediante testes de calibração e discriminação.**

Resultados: **Houve diferença significativa na prevalência dos fatores de risco entre as populações do estudo, ES e 2000BP. Os pacientes foram estratificados em cinco grupos para o 2000BP e três para o ES. Na validação dos modelos, o ES apresentou uma boa calibração (*P*=0,596); no entanto, o 2000BP revelou-se inadequado (*P*=0,047). Na discriminação, a área abaixo da curva ROC revelou-se boa para ambos os modelos, ES (0,79) e 2000BP (0,80).**

Conclusão: **Na validação, o 2000BP revelou-se questionável e o ES adequado para predizer mortalidade nos pacientes operados de coronária e/ou valva, no InCor/HC-FMUSP.**

Descritores: **Fatores de risco. Procedimentos cirúrgicos cardiovasculares. Medição de risco. Mortalidade hospitalar. Estudos de validação.**

INTRODUCTION

Risk stratification informs patients and professionals about the likely risk of complications or death for the group of individuals with similar risk profile undergoing the proposed procedure [1]. However, in order to compare results using the same risk score, we would also have similar levels of accuracy and adequacy of the model in the populations studied [2].

Currently, the use of risk scores in decision making in coronary artery bypass surgery is considered IIa recommendation, with level of evidence B [3]. However, to be used, the risk models should be validated. Validating a model means investigating its calibration and discrimination in another population of which was developed [4]. The analysis of calibration requires that the use of the model is strict, without artificially increasing the weight of each variable, and the data is collected from all patients during a given period. Assessment of discrimination power requires no loss of outcome (death) in the calculations. The sample size and number of events are the most important aspects in the validation of a model, where at least 100 deaths should be considered [5]. Unfortunately, several studies performed in order to validate a score involving disabled people, making difficult the applicability of the models and therefore the interpretation of the results.

In Brazil, no score predictor of mortality in cardiac surgery has been adequately validated, although several have already been used. Differences in clinical presentation due to socioeconomic, cultural and geographic reasons, unequal distribution of medical facilities, and high endemicity of subclinical inflammation, infection and rheumatic disease are evident, which could alter the performance of the models. For this, the EuroSCORE [6] and the 2000 Bernstein-Parsonnet [7] in several publications demonstrating its applicability in Brazil [8-11], were finally validated in patients undergoing coronary and/or valve surgery at the Heart Institute of the Clinics Hospital of Faculty of Medicine of the University of São Paulo (InCor/HCFMUSP).

METHODS

Sample

This prospective, observational study was performed at the Division of Cardiovascular Surgery, Department of Cardiology InCor/HCFMUSP.

To calculate the sample size for validation of risk scores

(minimum 100 deaths), the publication of Lisbon et al. [12] on the results of Incor-HCFMUSP in the past 23 years, reports an overall mortality of 6.9% and 4.8% for elective CABG and 8.4% for elective valve surgery. As in our study patients undergoing coronary and/or valve surgery were included, we considered reasonable to use lower mortality as a parameter, in this case the elective coronary surgery, resulting in a minimum sample size of 2084 patients.

Inclusion and exclusion criteria

Inclusion Criteria

We included all consecutive patients who underwent surgery between May 2007 and July 2009, in the modality elective, urgent or emergency:

• Valve surgery (replacement or repair);

• coronary surgery (with or without the use of cardiopulmonary bypass);

• Associated surgery (CABG and valve surgery).

Exclusion Criteria

Other types of associated surgery were excluded.

Collecting, defining and organizing the data

Data were collected preoperatively to clinical assessment and electronic medical records of the institution (SI3) and stored in a single spreadsheet. This spreadsheet has been adapted in order to include all the variables described by the model of the 2000 Bernstein Parsonnet and EuroSCORE. Sixty preoperative variables (demographic, clinical and laboratory) per patient were collected. All definitions assigned to variables for both scores were observed with their respective values, according to their relevance to the death event. Thus, after calculating the value of 2000BP and ES for each patient, they were ordered according to risk groups established by the scores and placed in the database made on Excel for this purpose. All patients were followed until hospital discharge. No patient was excluded from analysis due to missing data. The outcome of interest was in-hospital mortality, defined as death occurring in the time interval between surgery and discharge.

Validation of the 2000 Bernstein Parsonnet and EuroSCORE

To assess the performance of 2000BP and ES in predicting mortality, we performed a validation of predictive models in 3000 patients. The assessment was performed by testing calibration and discrimination.

Calibration

Calibration evaluates the accuracy of the model to predict risk in a group of patients. In other words, the model proposes that mortality in 1000 patients would be 5% and observed mortality is 5% or thereabouts, we say that the model is well calibrated. The force calibration was assessed by testing the goodness of adjustment by the Hosmer-Lemeshow test[13]. The P value > 0.05 indicates that the model fits the data and predicts mortality properly.

Discrimination

Discrimination measures the ability of the model to distinguish between patients at low and high risk. In other words, if the majority of deaths occurring in patients that the model identifies as high risk, we say that the model has good discrimination. Conversely, if the majority of deaths occurring in patients that the model identifies as low risk, we can say that the model has poor discrimination. The discrimination is measured by using the statistical technique called area under the ROC curve (sometimes called c-statistic-index or c). Thus, excellent discrimination refers to values above 0.97, very good discrimination is in the range from 0.93 to 0.96, good discrimination between 0.75 and 0.92; below corresponds to 0.75 models deficient in the ability of discrimination. [14] In practice, the models rarely exceed 0.85.

Statistical analysis

Statistical analysis was performed using SPSS software, version 16.0 for Windows (IBM Corporation Armonk, New York). Continuous variables were expressed as the mean ± standard deviation and categorical variables as percentages. The logistic regression analysis for the outcome of in-hospital mortality was performed by using the value given to each patient by the 2000BP and ES scores. Calibration and discrimination were measured for each value of the score in the patient population. The performance of the models was also measured by comparing the observed mortality and expected mortality in risk groups established by the models. The Fisher exact test was used for contingency tables. The value of P <0.05 was considered significant.

Ethics and Written Informed Consent

This study was approved by the Ethics Committee for Research Projects Analysis (CAPPesq), Clinics Hospital, University of São Paulo, under number 1575.

RESULTS

Casuistry

All patients undergoing coronary and/or valve surgery, between May 2007 and July 2009, at InCor/HCFMUSP, were included in the study. Of the 3000 patients who underwent surgery, 268 (8.9%) died. Of the total procedures, 57.7% (1731) underwent surgery for coronary, 36.8% (1104), valve and 5.5% (165), coronary and valve.

For descriptive purpose, we show in Table 1 the

prevalence of risk factors in the study population and the population of the ES. Similarly, Table 2 shows the prevalence of risk factors in the study population and the population of 2000BP. Because these populations are potentially comparable, we assessed the statistical difference in the prevalence of risk factors in ES and 2000BP, with respect to the study population.

Table 1. Prevalence of risk factors in the study group comparing the risk factors of the EuroSCORE population.

VARIABLES	STUDY	EuroSCORE	*P*
Age	(N=3000)	(N=19030)	
<60 years	44.27%	33.20%	< 0.001
60-64 years	15.80%	17.80%	0.007
65-69 years	13.87%	20.70%	< 0.001
70-74 years	12.20%	17.90%	< 0.001
>75 years	11.50%	9.60%	0.001
Female	35.90%	27.80%	< 0.001
Chronic lung disease	2.60%	3.90%	< 0.001
Extracardiac arteriopathy	4.80%	11.30%	< 0.001
Neurological dysfunction	6.90%	1.40%	< 0.001
Previous cardiac surgery	17.80%	7.30%	< 0.001
Creatinine > 2,3 mg/dl	4.40%	1.80%	< 0.001
Active endocarditis	4.10%	1.00%	< 0.001
Critical preoperative state	10.30%	4.10%	< 0.001
Unstable angina	7.00%	8.00%	0.059
EF 30 – 50	26.10%	25.60%	0.569
EF <30	5.80%	5.80%	0.998
Recent AMI	16.80%	9.70%	< 0.001
Pulmonary hypertension	8.10%	2.00%	< 0.001
Emergency	3.10%	4.90%	< 0.001
Combined surgery	6.90%	36.40%	< 0.001
Thoracic aortic surgery	0.70%	2.40%	< 0.001
Postinfarction VSD	0.50%	0.20%	0.002

EF = ejection fraction; AMI = acute myocardial infarction; CIV = interventricular communication

Table 2. Prevalence of risk factors in the study group comparing the risk factors of the 2000 Bernstein Parsonnet population.

VARIABLES	STUDY	2000BP	*P*
Age	(N=3000)	(N=10703)	
70-74 years	12.20%	18.50%	< 0.001
>75 years	11.50%	13.70%	0.002
Female	35.90%	31.30%	< 0.001
Chronic lung disease	2.60%	10.80%	< 0.001
Extracardiac arteriopathy	4.80%	9.10%	< 0.001
Neurological dysfunction	6.90%	8.40%	0.008
Previous cardiac surgery	17.80%	7.60%	< 0.001
Creatinine > 2,3 mg/dl	4.40%	4.50%	0.809
EF 30 – 50	26.10%	38.60%	< 0.001
EF <30	5.80%	8.40%	< 0.001
Pulmonary hypertension	8.10%	10.70%	< 0.001

EF = ejection fraction

Outcomes of 2000 Bernstein Parsonnet and EuroSCORE validation.

Calibration Results

2000 Bernstein Parsonnet

Association was found between model 2000BP and death with OR: 1.079 (*P* <0.001). The Hosmer-Lemeshow test showed a goodness-of-fit statistic = 15.678 with 8 degrees of freedom, *P* = 0.0472 (Table 3). For a better suitability analysis, the 2000BP was divided into five categories (Table 4). The 2000BP shows a poor fit in the subgroups established.

EuroSCORE

Association was found between ES model and death with OR: 1.337 (*P* <0.001). The Hosmer-Lemeshow test showed a goodness-of-fit-statistic = 5.5301, with 7 degrees of freedom, *P* = 0.5956 (Table 5). For a better suitability analysis, the ES was divided into three categories (Table 6). The ES presents an appropriate fit in the subgroups established.

Table 3. Observed and expected mortality by use of 2000BP as predictor variable in the groups defined by the Hosmer-Lemeshow test.

		DEATH=1		DEATH=0	
Group	Total	Observed	Expected	Observed	Expected
1	293	6	5.42	287	287.58
2	301	4	7.43	297	293.57
3	300	3	9.17	297	290.83
4	292	12	10.82	280	281.18
5	272	6	12.15	266	259.85
6	286	12	15.52	274	270.48
7	299	25	20.97	274	278.03
8	296	33	27.90	263	268.10
9	299	55	43.60	244	255.40
10	362	113	116.03	249	245.97

** Goodness of fit statistic = 15.678 with 8 DF (P=0.0472)*

Table 4. Percentages of observed and estimated mortality by 2000BP risk group.

Risk	OM/EM	EM	OM	N	%
< 9	0.77	2.19	1.68	594	19.8
9-14	0.75	3.38	2.53	592	19.7
14.1-19.9	0.67	4.84	3.23	558	18.6
20-28.9	1.18	8.24	9.75	595	19.8
≥ 29	1.05	24.21	25.42	661	22

OM / EM = Observed Mortality/Expected Mortality, EM = expected mortality, OM = observed mortality; N = Number of patients; % = Percentage

Table 5. Observed and expected mortality by use of ES as a predictor variable in the groups defined by the Hosmer – Lemeshow test.

		DEATH=1		DEATH=0	
Group	Total	Observed	Expected	Observed	Expected
1	296	5	4.83	291	291.17
2	378	10	8.20	368	369.80
3	287	6	8.27	281	278.73
4	397	12	15.15	385	381.85
5	358	19	18.04	339	339.96
6	269	14	17.83	255	251.17
7	264	19	22.89	245	241.11
8	345	53	43.97	292	301.03
9	406	131	129.83	275	276.17

** Goodness of fit statistic = 5.5301 com 7 DF (P=0.5956)*

Table 6. Percentages of observed and estimated mortality by risk groups of ES.

Risc	OM/EM	EM	OM	N	%
0 – 2	1	2.19	2.19	961	32.03
3 – 5	0.88	4.98	4.39	1024	34.13
≥ 6	1.03	19.41	20.00	1015	33.83

OM / EM = Observed Mortality/Expected Mortality, EM = expected mortality, OM = observed mortality; N = Number of patients; % = Percentage

Outcomes of Discrimination (ROC curves, Figure 1).

2000 Bernstein Parsonnet

Assessing the discriminative power of 2000BP, we observe that the area under the ROC curve was 0.800 (95% CI, 0.772 to 0.827, *P*=0.014) (Table 7).

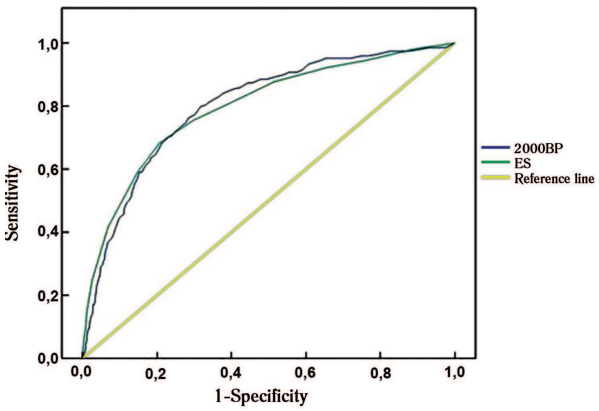

Fig 1 - ROC curve for 2000BP and ES in assessing the power of discrimination performed in 3000 patients. 2000BP ROC ROC curve = 2000 presented by Bernstein Parsonnet score, ES ROC = ROC curve resulting from the EuroSCORE

Table 7. Area under the ROC curve for the 2000BP and ES from the analysis performed in 3000 patients.

	Area	CI 95%	ep	*P*
2000BP	0.800	0.772 – 0.827	0.014	< 0.001
ES	0.796	0.766 – 0.826	0.015	< 0.001

EuroSCORE

Assessing the discriminative power of ES, we observe that the area under the ROC curve was 0.796 (95% CI, 0.766 to 0.826, *P*=0.015) (Table 7).

DISCUSSION

Only predictive models consisting of preoperative variables can be used in making decisions, by not including variables per and/or postoperative. Thus, several publications for the use of 2000BP and ES models in predicting mortality in cardiac surgery, even in our country [8], consolidated the importance of these models.

One of the first analysis was published in Saudi Arabia in 2004 by Syed et al. [15] comparing the ES model with the initial Parsonnet, in 194 patients. The areas under the ROC curve were 0.77 for the ES model and 0.69 for the initial Parsonnet. However, the sample size, with only 13 deaths loses in credibility and statistical power. In the same year, an assessment made in Taiwan by Chen et al. [16] used the ES in 801 consecutive patients with coronary artery bypass graft (CABG). With just over 80 deaths, the area under the ROC curve reached 0.75.

One of the best study was performed by Berman et al. [17] in Israel in 2006. The 2000BP was compared to ES. They assessed 1639 consecutive patients with coronary and/or valve disease. The area under the ROC curve was 0.83 for 2000BP and 0.73 for the ES. This result was similar to our initial analysis performed in InCor/HCFMUSP [8] in 744 patients using the same risk scores. In this analysis, the Hosmer-Lemeshow test for 2000BP (*P* = 0.70) and the ES (*P* = 0.39) indicated good calibration. Also, the ROC curve for the 2000BP = 0.84 and = 0.81 for the ES was suitable for predicting mortality. However, the interpretation of model validation was limited by the number of deaths. In the final analysis, with 3000 patients, we can see that for this sample size the calibration is reversed, the 2000BP no longer calibrates (*P* = 0.047) and ES increases its calibration (*P* = 0.597). However, good discrimination persists with an area under the ROC curve of 0.80 for ES and 0.81 for 2000BP.

In this context, the literature shows that the origin, nature and evolution of these models are in favor of better performance of ES in larger populations, while the 2000BP makes smaller groups. One explanation for this is known in

statistics as overfitting of the models when presenting many variables [5.18]. This was confirmed by an analysis of 1000 patients in the same sample, where 2000BP presented a calibration with $P = 0.157$ and $P = 0.593$ with ES (unpublished data). Therefore, since the sample size increases, the calibration of ES improves, and the 2000BP, worsens. Thus, we recommend that in the calibration, the 2000BP is chosen to populations up to 744 patients, and from that, the ES is preferred. Even then, the persistence of discrimination adequate for both models indicates that the variables contained in the models are true predictors of mortality [19].

One of the first studies performed in Brazil, Pernambuco, is authored by Moraes et al. [9] who in 2006, retrospectively assessed the applicability of the ES in 752 patients undergoing CABG. With only 13 deaths, the sample had an area under the ROC curve of 0.70, lower than that found in our study, which also assessed valve and associated surgery. In 2008, Campagnucci et al. [10] published in Brazil, a retrospective analysis with ES in 100 consecutive patients undergoing CABG. The sample size limited the statistical analysis, hampering proper conclusion of the study. In 2009, Ranucci et al. [18] published in Italy, an analysis of 11,150 patients undergoing cardiac surgery, demonstrating that limiting the number of variables used by EuroSCORE would decrease the risk of overfitting, multicollinearity and human error. The best accuracy was obtained with five variables (age, ejection fraction, creatinine, emergency surgery and CABG combined), with an area under the ROC curve of 0.76 compared to 0.75 of the logistic EuroSCORE. In this study it was shown that models of few variables, but with strong association with mortality, could provide good calibration, obviously the expense of proper discrimination.

In 2010, Malik et al. [20] published in India, the validation of the ES in 1000 consecutive patients after cardiac surgery. The area under the ROC curve was 0.827. In calibration, the Hosmer-Lemeshow test showed $P = 0.73$. The difference in the clinical profile of patients between both populations was marked by a high prevalence of variables associated with late presentation of the disease. The data from this analysis are very similar to those of our study, both in methodology, results and prevalence of risk factors. In March 2011, Shih et al. [21] published in Taiwan, the performance of the ES in 1240 patients undergoing cardiac surgery. The area under the ROC curve was 0.839. In calibration, all subgroups except for CABG, demonstrated good application of the model. A study published in Pakistan in April 2011, by Qadir et al. [22] retrospectively assessed the ES in 2004 patients undergoing CABG. The area under the ROC curve was 0.866. In calibration, it was yielded a P value = 0.424. The model overestimated mortality in the group of low and medium risk.

Currently, the use of risk scores is made preoperatively, to aid in making decisions (questionable in indicating new technologies) and postoperatively, for the prevention of adverse effects and cost control, mainly in intensive care unit. It is logical to think that, in time and space, variations in the systems of prevention, diagnosis and treatment of risk factors can alter the accuracy of the models. Thus, in order to use these mathematical models, we must first validate them with the principles of proper statistical analysis.

In our reality, the lack of proper validation of external models, required in developing a population with high prevalence of rheumatic disease and Chagas, was impairing the knowlegement about the risk assessment of patients undergoing cardiac surgery in Brazil. As can be evidenced in Tables 2 and 3, in most validation studies, there is a significant difference in the prevalence of risk factors between the study population and the populations of the analyzed models. Even so, the appropriate application (respecting the statistical principles) for accurate risk models consisting of variables strongly predictive of mortality can succeed. Certainly, the fact of recalibrating the model (adapting the weights of the variables according to their importance in the study population), or better yet, reshaping the model (adding new variables related to mortality or removing variables that may hinder the stability of the model), would lead to a more sophisticated and accurate for the population under study, with larger areas under the ROC curve [23].

A Brazilian model (even without external validation), published in 2010 by Cadore et al. [24], in Rio Grande do Sul, brings a proposal for a local model to predict outcomes in coronary artery bypass surgery. This model demonstrates a practical and simple good area under the ROC curve of 0.86. However, it is derived from a retrospective database of patients who have undergone surgery in 1996-2007 (> 5 years) and that with a mortality of 10% could overestimate results.

In 2010, Sá et al. [11] published in Pernambuco, a retrospective analysis of 500 patients on the ES, including 65 deaths. In addition to the limited number of events in the study, difficulty in defining and collecting data for the retrospective nature of the analysis, may have hindered the allocation of patients to the risk group established by ES. In our study, the mortality of 20% (including coronary and/or valve surgery) in high-risk patients was similar to that expected of 19.41%, even though between 75 and 80% of patients who have undergone surgery at InCor/HCFMUSP are served by the Unified Health System (SUS). This high mortality was confirmed by ES, considering the high prevalence of risk factors in this group of patients.

Therefore, controlling and decreasing the prevalence of risk factors result in lower values of observed mortality. Moreover, for the calibration of the models, it is

recommended to use the Hosmer-Lemeshow test. The Kappa index depends on the prevalence of the disease under study. A high prevalence results in high level of agreement expected by chance, resulting in lower k value. Therefore, we may make the mistake of basing this index on a comparison of two studies with different prevalences.

Limitations of this study were: first, although as unicentric, the most important limitation is the generalizability of the results, about 50% of patients attending the hospital are from different states of Brazil. Secondly, because of its nature, the additive EuroSCORE tends to underestimate risk in high-risk patients, although this has not been shown in multicenter studies [25]. Finally, although hospital mortality (up to 30 days after surgery) appears to be more complete than the in-hospital mortality, the current definitions suggest that both have equivalent accuracy, and in-hospital mortality was more practical and easy to use [26].

Thus, also the advances in perioperative care in cardiac surgery could be better assessed with the remodeled EuroSCORE (EuroSCORE II) [27], especially in places where the EuroSCORE lost calibration. But even so, we should be careful about the limitations of the new model, because it was the inappropriate use of the first version which led to a dramatic expansion of the market for transcatheter aortic valve implantation.

Finally, it is important to clarify that the scores assess only a tiny part of the multiple variables known and unknown to the patient and the health care structure, which directly influence the outcome of the process. Therefore, the conclusions derived from its application must be carefully assessed.

CONCLUSIONS

The 2000 Bernstein Parsonnet proved to be poor in calibration and good discrimination, being questionable in validating to predict mortality in patients undergoing coronary and/or valve surgery at InCor/HCFMUSP.

The EuroSCORE proved to good both in calibration and discrimination, with appropriate validation to predict mortality in patients undergoing coronary and or valve surgery in InCor/HCFMUSP.

REFERENCES

1. Daley J. Criteria by which to evaluate risk-adjusted outcomes programs in cardiac surgery. Ann Thorac Surg. 1994;58(6):1827-35.

2. Shahian DM, Normand SL. Comparison of "risk-adjusted" hospital outcomes. Circulation. 2008;117(15):1955-63.

3. Hillis LD, Smith PK, Anderson JL, Bittl JA, Bridges CR, Byrne JG, et al. 2011 ACCF/AHA guideline for coronary artery bypass graft surgery: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Thorac Cardiovasc Surg. 2012;143(1):4-34.

4. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med. 2000;19(4):453-73.

5. Harrell FE. Regression modeling strategies. Berlin: Springer-Verlag; 2001.

6. Nashef SAM, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R, the EuroSCORE study group. European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardio-thorac Surg. 1999;16(1):9-13.

7. Bernstein AD, Parsonnet V. Bedside estimation of risk as an aid for decision-making in cardiac surgery. Ann Thorac Surg. 2000;69(3):823-8.

8. Mejía OA, Lisboa LA, Puig LB, Dias RR, Dallan LA, Pomerantzeff PM, et al. The 2000 Bernstein-Parsonnet score and EuroSCORE are similar in predicting mortality at the Heart Institute, USP. Rev Bras Cir Cardiovasc. 2011;26(1):1-6.

9. Moraes F, Duarte C, Cardoso E, Tenório E, Pereira V, Lampreia D, et al. Assessment of the EuroSCORE as a predictor for mortality in myocardial revascularization surgery at the Heart Institute of Pernambuco. Rev Bras Cir Cardiovasc. 2006;21(1):29-34.

10. Campagnucci VP, Pinto e Silva AM, Pereira WL, Chamlian EG, Gandra SM, Rivetti LA. EuroSCORE and the patients undergoing coronary bypass surgery at Santa Casa de São Paulo. Rev Bras Cir Cardiovasc. 2008;23(2):262-7.

11. Sá MP, Soares EF, Santos CA, Figueredo OJ, Lima RO, Escobar RR, et al. EuroSCORE and mortality in coronary artery bypass graft surgery at Pernambuco Cardiologic Emergency Medical Services. Rev Bras Cir Cardiovasc. 2010;25(4):474-82.

12. Lisboa LAF, Moreira LFP, Dallan LAO, Pomerantzeff PMA, Costa R, Puig LB, et al. Vinte e três anos de evolução da cirurgia cardiovascular em serviço de atendimento terciário no Brasil: análise de 67000 procedimentos registrados no banco de dados. Arq Bras Cardiol. 2007;89(1):205.

13. Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. New York: John Wiley and Sons; 2000.

14. Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg. 2005;79(1):16-20.

15. Syed AU, Fawzy H, Farag A, Nemlander A. Predictive value of EuroSCORE and Parsonnet scoring in Saudi Population. Heart Lung Circ. 2004;13(4):384-8.

16. Chen CC, Wang CC, Hsieh SR, Tsai HW, Wei HJ, Chang Y. Application of European system for cardiac operative risk evaluation (EuroSCORE) in coronary bypass surgery for Taiwanese. Interact Cardiovasc Thorac Surg. 2004;3(4):562-5.

17. Berman M, Stamler A, Sahar G, Georghiou GP, Sharoni E, Brauner R, et al. Validation of the 2000 Bernstein-Parsonnet score versus the EuroSCORE as a prognostic tool in cardiac surgery. Ann Thorac Surg. 2006;81(2):537-41.

18. Ranucci M, Castelvecchio S, Menicanti L, Frigiola A, Pelissero G. Accuracy, calibration and clinical performance of the EuroSCORE: can we reduce the number of variables? Eur J Cardiothorac Surg. 2010;37(3):724-9.

19. Nashef SA. Applying and evaluating risk models. Eur J Cardiothorac Surg. 2012;41(2):314-5.

20. Malik M, Chauhan S, Malik V, Gharde P, Kiran U, Pandey RM. Is EuroSCORE applicable to Indian patients undergoing cardiac surgery? Ann Card Anaesth. 2010;13(3):241-5.

21. Shih HH, Kang PL, Pan JY, Wu TH, Wu CT, Lin CY, et al. Performance of European system for cardiac operative risk evaluation in Veterans General Hospital Kaohsiung cardiac surgery. J Chin Med Assoc. 2011;4(3):115-20.

22. Qadir I, Perveen S, Furnaz S, Shahabuddin S, Sharif H. Risk stratification analysis of operative mortality in isolated coronary artery bypass graft patients in Pakistan: comparison between additive and logistic EuroSCORE models. Interact Cardiovasc Thorac Surg. 2011;13(2):137-41.

23. Ivanov J, Tu JV, Naylor CD. Ready-made, recalibrated, or remodeled? Issues in the use of risk indexes for assessing mortality after coronary artery bypass graft surgery. Circulation. 1999;99(16):2098-104.

24. Cadore MP, Guaragna JC, Anacker JF, Albuquerque LC, Bodanese LC, Piccoli JC, et al. A score proposal to evaluate surgical risk in patients submitted to myocardial revascularization surgery. Rev Bras Cir Cardiovasc. 2010;25(4):447-56.

25. Jin R, Grunkemeier GL. Does the logistic EuroSCORE offer an advantage over the additive model? Interact Cardiovasc Thorac Surg. 2006;5(1):15-7.

26. Likosky DS, Nugent WC, Clough RA, Weldner PW, Quinton HB, Ross CS, et al. Comparison of three measurements of cardiac surgery mortality for the Northern New England Cardiovascular Disease Study Group. Ann Thorac Surg. 2006;81(4):1393-5.

27. Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. EUROSCORE II. Eur J Cardiothorac Surg. 2012;41(4):734-45.