

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

*Validation of the 2000 Bernstein-Parsonnet and EuroSCORE at the Heart Institute - USP*

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

DOI: 10.5935/1678-9741.20120033

RBCCV 44205-1370

## 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.

## 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).

Trabalho realizado no Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (InCor/HC-FMUSP), São Paulo, SP, Brasil.

## Endereço para correspondência

Omar A. V. Mejía. Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo. Unidade Cirúrgica de Coronariopatias.  
Av. Dr. Enéas de Carvalho Aguiar, 44 – Cerqueira César – São Paulo, SP, Brasil – CEP 05403-000.  
E-mail: omarvilca@incor.usp.br

Apoio: Conselho Nacional de Desenvolvimento Científico e Tecnológico.

Artigo recebido em 15 de março de 2012  
Artigo aprovado em 2 de maio de 2012

1. Doutor em Ciências pela Faculdade de Medicina da Universidade de São Paulo (FMUSP), São Paulo, SP, Brasil.
2. Livre Docente, médico assistente da Unidade Cirúrgica de Cardiopatias Coronariana do Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (InCor/HC-FMUSP), São Paulo, SP, Brasil.
3. Livre Docente, Professor Associado da FMUSP, Diretor da Unidade Cirúrgica de Cardiopatias Coronariana do InCor/HC-FMUSP, São Paulo, SP, Brasil.
4. Livre Docente, Professor Associado da FMUSP, Diretor da Unidade Cirúrgica de Cardiopatias Valvares do InCor/HC-FMUSP, São Paulo, SP, Brasil.
5. Livre Docente, Professor Associado da FMUSP, Diretor da Unidade Cirúrgica de Pesquisa do InCor/HC-FMUSP, São Paulo, SP, Brasil.
6. Professor Titular da Disciplina de Cirurgia Torácica da FMUSP, São Paulo, SP, Brasil.
7. Professor Titular da Disciplina de Cirurgia Cardiovascular da FMUSP, São Paulo, SP, Brasil.

Abreviaturas, acrônimos & símbolos	
2000BP	2000 Bernstein Parsonnet
CAPPesq	Comissão de Ética para Análise de Projetos de Pesquisa
CRM	Cirurgia de revascularização do miocárdio
ES	EuroSCORE aditivo
InCor/HC-FMUSP	Instituto do Coração da Universidade de São Paulo

**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

## INTRODUÇÃO

A estratificação de risco informa a pacientes e profissionais sobre o provável risco de complicações ou óbito para o grupo de indivíduos com perfil de risco similar, submetidos ao procedimento proposto [1]. No entanto, para poder comparar resultados utilizando o mesmo escore de risco, teríamos que ter também níveis semelhantes de adequabilidade e acurácia do modelo nas populações estudadas [2].

Atualmente, a utilização de escores de risco na tomada de decisões em cirurgia de revascularização miocárdica é considerada recomendação IIa, com grau de evidencia B [3]. Porém, para serem utilizados, os modelos de risco precisam ser validados. Validar um modelo significa investigar sua calibração e discriminação em outra população a da qual foi desenvolvido [4]. A análise da calibração exige que o uso do modelo seja estrito, sem aumento artificial do peso de cada variável, e que os dados sejam coletados de todos os pacientes durante um período determinado. A avaliação do poder de discriminação exige a ausência de perdas do desfecho (óbito) nos cálculos realizados. O tamanho da amostra e o número de eventos são os aspectos mais importante na validação de um modelo, onde no mínimo 100 óbitos devem ser considerados [5]. Infelizmente, muitos trabalhos realizados com o objetivo de validar um escore envolvem populações deficientes, dificultando a aplicabilidade dos modelos e, portanto, a interpretação dos resultados.

No Brasil, nenhum escore preditor de mortalidade em

cirurgia cardíaca foi adequadamente validado, embora vários já sejam utilizados. Diferenças na apresentação clínica devido a razões socioeconômicas, culturais e geográficas; a distribuição desigual de instalações médicas; e alta endemicidade de inflamação subclínica, infecção e doença reumática são evidentes, o que poderia alterar o desempenho dos modelos. Para isso, o EuroSCORE [6] e o 2000 Bernstein-Parsonnet [7], em várias publicações no Brasil demonstrando sua aplicabilidade [8-11], foram finalmente validados nos pacientes operados de coronária e/ou valva no Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (InCor/HC-FMUSP).

## MÉTODOS

### Amostra

Este estudo prospectivo e observacional foi conduzido na Divisão de Cirurgia Cardiovascular do Departamento de Cardiopneumologia do InCor/HC-FMUSP.

Para o cálculo do tamanho da amostra para validação dos escores de risco (mínimo 100 óbitos), a publicação de Lisboa et al. [12], sobre os resultados do InCor-HCFMUSP, nos últimos 23 anos, relata uma mortalidade global de 6,9%, sendo 4,8% para cirurgia coronária eletiva e 8,4% para cirurgia valvar eletiva. Como no nosso estudo incluímos pacientes operados de coronária e/ou valva, consideramos razoável utilizar a menor mortalidade como parâmetro, nesse caso a cirurgia coronária eletiva, resultando em uma amostra mínima de 2084 pacientes.

### **Critérios de inclusão e exclusão**

#### **Critérios de inclusão**

Foram incluídos todos os pacientes operados consecutivamente, entre maio de 2007 e julho de 2009, na modalidade eletiva, urgência ou emergência de:

- Cirurgia valvar (troca ou plastia);
- Cirurgia de coronárias (com ou sem a utilização de circulação extracorpórea);
- Cirurgia associada (cirurgia de revascularização miocárdica e cirurgia valvar).

#### **Critérios de exclusão**

Outros tipos de cirurgia associada foram excluídos.

### **Coleta, definição e organização dos dados**

Os dados foram coletados pré-operatoriamente à avaliação clínica e do sistema eletrônico de prontuários da instituição (SI3) e armazenados numa planilha única. Essa planilha foi adaptada, de forma a contemplar todas as variáveis descritas pelo modelo do 2000 Bernstein Parsonnet e do EuroSCORE. Sessenta variáveis pré-operatórias (demográficas, clínicas e laboratoriais) por paciente foram coletadas. Todas as definições atribuídas às variáveis por ambos os escores foram respeitadas junto a seus respectivos valores, segundo sua relevância com o evento morte. Assim, após o cálculo do valor do 2000BP e do ES para cada paciente, eles foram ordenados segundo os grupos de risco estabelecidos pelos escores e colocados no banco de dados confeccionado no programa Excel com essa finalidade. Todos os pacientes foram acompanhados até a alta hospitalar. Nenhum paciente foi excluído da análise por falta de dados. O desfecho de interesse foi a mortalidade intra-hospitalar, definida como a morte ocorrida no intervalo de tempo entre a cirurgia e a alta hospitalar.

### **Validação do 2000 Bernstein Parsonnet e do EuroSCORE**

Para avaliar o desempenho do 2000BP e do ES na predição de mortalidade, foi realizada a validação preditiva dos modelos em 3000 pacientes. A análise foi feita mediante teste de calibração e discriminação.

#### **Calibração**

A calibração avalia a acurácia do modelo para prever risco em um grupo de pacientes. Em outras palavras, se o modelo propõe que a mortalidade em 1000 pacientes seria 5% e a mortalidade observada é de 5% ou perto disso, diremos que o modelo está bem calibrado. A força da calibração foi avaliada testando a qualidade do ajuste mediante o teste de Hosmer-Lemeshow [13]. O valor de  $P > 0,05$  indica que o modelo se ajusta aos dados e prediz a mortalidade adequadamente.

#### **Discriminação**

A discriminação mede a habilidade do modelo para distinguir entre pacientes de baixo e alto risco. Em outras palavras, se a maioria dos óbitos ocorre em pacientes que o modelo identifica como de alto risco, diremos que o modelo tem uma boa discriminação. Ao contrário, se a maioria dos óbitos ocorre em pacientes que o modelo identifica como de baixo risco, podemos afirmar que o modelo tem pobre discriminação. A discriminação é medida por meio da utilização da técnica estatística chamada área abaixo da curva ROC (algumas vezes chamada *c-statistic* ou *c-index*). Assim, a excelente discriminação refere-se a valores acima de 0,97; a muito boa discriminação está no intervalo entre 0,93 e 0,96; boa discriminação entre 0,75 e 0,92; abaixo de 0,75 corresponde a modelos deficientes na capacidade de discriminação [14]. Na prática, os modelos raramente excedem 0,85.

#### **Análise estatística**

A análise estatística foi realizada com auxílio do *software* SPSS versão 16.0 para Windows (IBM Corporation Armonk, New York). As variáveis contínuas foram expressas como a média  $\pm$  desvio-padrão e as variáveis categóricas, como porcentagens. A análise de regressão logística para o desfecho mortalidade intra-hospitalar foi realizada por meio da utilização do valor dado para cada paciente pelos escores 2000BP e ES. A calibração e a discriminação foram medidas para cada valor do escore na população de pacientes. O desempenho dos modelos foi também medido por meio da comparação entre a mortalidade observada e a mortalidade esperada nos grupos de risco estabelecidos pelos modelos. O teste exato de Fisher foi utilizado para as tabelas de contingência. O valor  $P < 0,05$  foi considerado significativo.

#### **Ética e Termo de Consentimento**

Este trabalho foi aprovado pela Comissão de Ética para Análise de Projetos de Pesquisa (CAPPesq) do Hospital das Clínicas da Universidade de São Paulo, com o número 1575.

### **RESULTADOS**

#### **Casuística**

Todos os pacientes operados de coronária e/ou valva, entre maio de 2007 e julho de 2009, no InCor/HC-FMUSP, foram incluídos no estudo. Dos 3000 pacientes operados, 268 (8,9%) morreram. Do total de procedimentos, 57,7% (1731) foram operados de coronária; 36,8% (1104), de valva e 5,5% (165), de coronária e valva.

Para efeito descritivo, demonstramos na Tabela 1 a prevalência dos fatores de risco na população de estudo e na população do ES. Do mesmo modo, na Tabela 2 é apresentada a prevalência dos fatores de risco na população

de estudo e na população do 2000BP. Pelo fato dessas populações serem potencialmente comparáveis, foi analisada a diferença estatística das prevalências dos fatores de risco do ES e do 2000BP, em relação à população de estudo.

Tabela 1. Prevalência dos fatores de risco no grupo estudo, comparado aos fatores de risco da população do EuroSCORE.

VARIÁVEIS	ESTUDO (N=3000)	EuroSCORE (N=19030)	P
Idade			
<60 anos	44,27%	33,20%	< 0,001
60-64 anos	15,80%	17,80%	0,007
65-69 anos	13,87%	20,70%	< 0,001
70-74 anos	12,20%	17,90%	< 0,001
>75 anos	11,50%	9,60%	0,001
Feminino	35,90%	27,80%	< 0,001
Doença pulmonar crônica	2,60%	3,90%	< 0,001
Arteriopatia extracardíaca	4,80%	11,30%	< 0,001
Disfunção neurológica	6,90%	1,40%	< 0,001
Cirurgia cardíaca prévia	17,80%	7,30%	< 0,001
Creatinina > 2,3 mg/dl	4,40%	1,80%	< 0,001
Endocardite ativa	4,10%	1,00%	< 0,001
Estado pré-operatório crítico	10,30%	4,10%	< 0,001
Angina instável	7,00%	8,00%	0,059
FE 30 – 50	26,10%	25,60%	0,569
FE <30	5,80%	5,80%	0,998
IAM recente	16,80%	9,70%	< 0,001
Hipertensão pulmonar	8,10%	2,00%	< 0,001
Emergência	3,10%	4,90%	< 0,001
Cirurgia associada	6,90%	36,40%	< 0,001
Cirurgia na aorta torácica	0,70%	2,40%	< 0,001
CIV pós-infarto	0,50%	0,20%	0,002

FE = Fração de ejeção; IAM = Infarto agudo do miocárdio; CIV = Comunicação interventricular

Tabela 2. Prevalência dos fatores de risco no Grupo Estudo, comparado aos fatores de risco da população do 2000 Bernstein Parsonnet.

VARIÁVEIS	ESTUDO (N=3000)	2000BP (N=10703)	P
Idade			
70-74 anos	12,20%	18,50%	< 0,001
>75 anos	11,50%	13,70%	0,002
Feminino	35,90%	31,30%	< 0,001
Doença pulmonar crônica	2,60%	10,80%	< 0,001
Arteriopatia extracardíaca	4,80%	9,10%	< 0,001
Disfunção neurológica	6,90%	8,40%	0,008
Cirurgia cardíaca prévia	17,80%	7,60%	< 0,001
Creatinina > 2,3 mg/dl	4,40%	4,50%	0,809
FE 30 – 50	26,10%	38,60%	< 0,001
FE <30	5,80%	8,40%	< 0,001
Hipertensão pulmonar	8,10%	10,70%	< 0,001

FE = Fração de ejeção

## Resultados da Validação do 2000 Bernstein Parsonnet e EuroSCORE.

### Resultados da Calibração

#### 2000 Bernstein Parsonnet

Foi encontrada associação entre o modelo 2000BP e óbito com OR: 1,079 ( $P < 0,001$ ). O teste de Hosmer-Lemeshow demonstrou uma *goodness-of-fit statistic* = 15,678, com 8 graus de liberdade,  $P = 0,0472$  (Tabela 3). Para uma melhor adequação da análise, o 2000BP foi dividido em cinco categorias (Tabela 4). O 2000BP apresenta um ajuste inadequado nos subgrupos estabelecidos.

#### EuroSCORE

Foi encontrada associação entre o modelo ES e óbito com OR: 1,337 ( $P < 0,001$ ). O teste de Hosmer-Lemeshow demonstrou uma *goodness of fit statistic* = 5,5301, com 7 graus de liberdade,  $P = 0,5956$  (Tabela 5). Para uma melhor adequação da análise, o ES foi dividido em três categorias (Tabela 6). O ES apresenta um ajuste adequado nos subgrupos estabelecidos.

Tabela 3. Mortalidade observada e esperada mediante uso do 2000BP como variável preditora nos grupos definidos pelo teste Hosmer-Lemeshow.

Grupo	Total	ÓBITO=1		ÓBITO=0	
		Observada	Esperada	Observada	Esperada
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 com 8 DF ( $P = 0,0472$ )

Tabela 4. Porcentagens da mortalidade observada e estimada por grupos de risco do 2000BP.

Risco	MO/ME	ME	MO	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

MO/ME = Mortalidade observada/Mortalidade esperada; ME = Mortalidade esperada; MO = Mortalidade observada; N = Número de pacientes; % = Porcentagem

Tabela 5. Mortalidade observada e esperada mediante uso do ES como variável preditora nos grupos definidos pelo teste Hosmer – Lemeshow.

Grupo	Total	ÓBITO=1		ÓBITO=0	
		Observada	Esperada	Observada	Esperada
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)

Tabela 6. Porcentagens da mortalidade observada e estimada por grupos de risco do ES.

Risco	MO/ME	ME	MO	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

MO/ME = Mortalidade observada/Mortalidade esperada; ME = Mortalidade esperada; MO = Mortalidade observada; N = Número de pacientes; % = Porcentagem

### Resultados da Discriminação (Curvas ROC, Figura 1). 2000 Bernstein Parsonnet

Avaliando o poder de discriminação do 2000BP, observamos que a área abaixo da curva ROC foi de 0,800 (IC95%, 0,772 – 0,827, P=0,014) (Tabela 7).

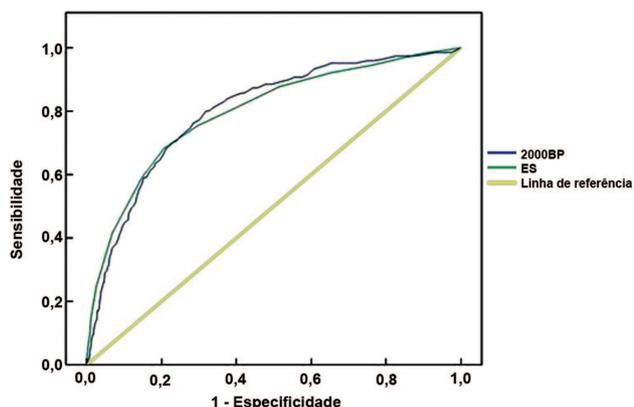


Fig. 1 - Curva ROC para o 2000BP e o ES na avaliação do poder de discriminação realizada em 3000 pacientes. 2000BP ROC = Curva ROC apresentada pelo escore 2000 Bernstein Parsonnet; ES ROC = Curva ROC resultante do EuroSCORE

Tabela 7. Área abaixo da curva ROC para o 2000BP e o ES da análise realizada em 3000 pacientes.

	Área	IC 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

Avaliando o poder de discriminação do ES, observamos que a área abaixo da curva ROC foi de 0,796 (IC95%, 0,766 – 0,826, P=0,015) (Tabela 7).

### DISCUSSÃO

Unicamente modelos preditivos constituídos de variáveis pré-operatórias podem ser utilizados na tomada de decisões, por não incluir variáveis per e/ou pós-operatórias. Dessa forma, várias publicações a favor da utilização dos modelos 2000BP e ES na predição de mortalidade em cirurgia cardíaca, inclusive em nosso meio [8], consolidaram a importância desses modelos.

Uma das primeiras análises foi a publicada na Arábia Saudita, em 2004, por Syed et al. [15], comparando o ES com o modelo inicial do Parsonnet, em 194 pacientes. As áreas abaixo da curva ROC foram de 0,77 para o ES e 0,69 para o modelo inicial do Parsonnet. Contudo, o tamanho da amostra, com apenas 13 óbitos, perde em credibilidade e em força estatística. No mesmo ano, uma avaliação feita em Taiwan, por Chen et al. [16], utilizaram o ES em 801 pacientes consecutivos operados de cirurgia de revascularização do miocárdio (CRM). Com pouco mais de 80 óbitos, a área abaixo da curva ROC alcançou 0,75.

Um dos melhores trabalhos foi realizado por Berman et al. [17], em Israel, em 2006. O 2000BP foi comparado ao ES. Foram analisados 1639 pacientes operados consecutivamente de coronária e/ou valva. A área abaixo da curva ROC foi 0,83 para o 2000BP e 0,73 para o ES. Esse resultado foi similar à nossa análise inicial realizada no InCor/HC-FMUSP [8], em 744 pacientes utilizando os mesmos escores de risco. Nessa análise, o teste de Hosmer-Lemeshow para o 2000BP (P = 0,70) e para o ES (P = 0,39) indicou boa calibração. Assim mesmo, a curva ROC para o 2000BP = 0,84 e para o ES = 0,81 foi adequada para prever mortalidade. No entanto, a interpretação da validação dos modelos ficou limitada pelo número de óbitos. Na análise final, com 3000 pacientes, podemos observar que para esse tamanho de amostra a calibração se inverte; o 2000BP já não calibra (P=0,047) e o ES aumenta sua calibração (P=0,597). Contudo, a boa discriminação persiste com uma área abaixo da curva ROC de 0,80 para o ES e de 0,81 para o 2000BP.

Nesse contexto, a literatura atesta que a origem, natureza e evolução desses modelos são favoráveis a melhor desempenho do ES em populações maiores, enquanto o 2000BP o faz com grupos menores. Uma explicação para isso é conhecida na estatística como sobreajuste ou “*overfitting*” dos modelos quando apresentam muitas variáveis [5,18]. Isso ficou confirmado por uma análise feita em 1000 pacientes da mesma amostra, onde o 2000BP apresentou uma calibração com  $P = 0,157$  e o ES com  $P = 0,593$  (dados não publicados). Portanto, na medida em que a amostra aumenta de tamanho, a calibração do ES melhora, e a do 2000BP, piora. Assim, podemos recomendar que, na calibração, o 2000BP é o escolhido para populações de até 744 pacientes e, a partir disso, o ES é preferido. Mesmo assim, a persistência de adequada discriminação para ambos os modelos indica que as variáveis contidas nos modelos são verdadeiras preditoras de mortalidade [19].

Um dos primeiros trabalhos realizados no Brasil, em Pernambuco, é de autoria de Moraes et al. [9] que, em 2006, avaliaram retrospectivamente a aplicabilidade do ES em 752 pacientes operados de CRM. Com apenas 13 óbitos, a amostra teve uma área abaixo da curva ROC de 0,70, inferior à encontrada no nosso estudo, que também avaliou cirurgia valvar e associada. Em 2008, Campagnucci et al. [10] publicaram, no Brasil, uma análise retrospectiva com o ES, em 100 pacientes consecutivos operados de CRM. O tamanho da amostra limitou a análise estatística, prejudicando a conclusão adequada do estudo. Em 2009, Ranucci et al. [18] publicaram, na Itália, uma análise em 11150 pacientes submetidos à cirurgia cardíaca, demonstrando que a limitação do número de variáveis usadas pelo EuroSCORE diminuiria o risco de sobreajuste, multicolinearidade e erro humano. A melhor acurácia foi obtida com cinco variáveis (idade, fração de ejeção, creatinina, cirurgia de emergência e cirurgia de revascularização miocárdica associada), com uma área abaixo da curva ROC de 0,76 em comparação a 0,75 do EuroSCORE logístico. Aqui ficou demonstrado que, modelos de poucas variáveis, mas de forte associação com mortalidade, poderiam apresentar boa calibração, logicamente às custas de uma adequada discriminação.

Em 2010, Malik et al. [20] publicaram, na Índia, a validação do ES em 1000 pacientes consecutivos operados de cirurgia cardíaca. A área abaixo da curva ROC foi de 0,827. Na calibração, o teste de Hosmer-Lemeshow apresentou  $P=0,73$ . A diferença do perfil clínico dos pacientes entre ambas as populações foi marcada por alta prevalência de variáveis relacionadas com a apresentação tardia da doença. Os dados dessa análise são muito parecidos aos de nosso estudo, tanto na metodologia, resultados e prevalência dos fatores de risco. Em março de 2011, Shih et al. [21] publicaram, em Taiwan, o desempenho do ES em 1240 pacientes submetidos a cirurgia cardíaca. A área abaixo da

curva ROC foi de 0,839. Na calibração, todos os subgrupos, exceto para CRM, demonstraram boa aplicação do modelo. Um trabalho publicado no Paquistão, em abril de 2011, por Qadir et al. [22], analisou retrospectivamente o ES em 2004 pacientes operados de CRM isolada. A área abaixo da curva ROC foi de 0,866. Na calibração, obteve-se um valor de  $P=0,424$ . O modelo superestimou a mortalidade no grupo de baixo e médio risco.

Atualmente, a utilização dos escores de risco é feita, no pré-operatório, para auxílio na tomada de decisões (questionável na indicação de novas tecnologias) e no pós-operatório, na prevenção dos efeitos adversos e controle dos custos, principalmente, na unidade de cuidados intensivos. É lógico pensar que, no tempo e espaço, variações nos sistemas de prevenção, diagnóstico e tratamento dos fatores de risco podem alterar a acurácia dos modelos. Assim, para poder utilizar esses modelos matemáticos, devemos primeiro validá-los respeitando os princípios de uma adequada análise estatística.

Em nossa realidade, a falta de correta validação de modelos externos, obrigatória em uma população em desenvolvimento e com alta prevalência de doença reumática e Chagas, estava dificultando o conhecimento sobre a avaliação de risco dos pacientes operados de cirurgia cardíaca no Brasil. Como pode ser evidenciado nas Tabelas 2 e 3, na maioria dos estudos de validação, existe uma diferença significativa na prevalência dos fatores de risco entre a população do estudo e as populações dos modelos analisados. Mesmo assim, a adequada aplicação (respeitando os princípios estatísticos) de modelos de risco acurados e constituídos por variáveis fortemente preditoras de mortalidade consegue ter sucesso. Certamente, o fato de recalibrar o modelo (adaptar os pesos das variáveis segundo a importância delas na população de estudo), ou melhor ainda, de remodelar o modelo (adicionar variáveis novas relacionadas à mortalidade ou retirando variáveis que possam dificultar a estabilidade do modelo), levariam a um modelo mais sofisticado e acurado para a população em estudo, com áreas maiores abaixo da curva ROC [23].

Um modelo brasileiro (ainda sem validação externa), publicado em 2010 por Cadore et al. [24], no Rio Grande do Sul, traz a proposta de um modelo local para predição de resultados na cirurgia de revascularização miocárdica. Esse modelo prático e simples demonstra uma boa área abaixo da curva ROC, de 0,86. No entanto, ele é derivado de um banco de dados retrospectivo com pacientes operados no período 1996-2007 (>5 anos) e que junto a uma mortalidade geral de 10% poderia superestimar resultados.

Em 2010, Sá et al. [11] publicaram, em Pernambuco, uma análise retrospectiva com o ES em 500 pacientes, incluindo 65 óbitos. Além do número limitado de eventos no estudo, dificuldade na definição e coleta dos dados pela natureza retrospectiva da análise, podem ter dificultado a alocação

dos pacientes nos grupo de risco estabelecidos pelo ES. No nosso estudo, a mortalidade de 20% (incluindo coronária e/ou valva) nos pacientes de alto risco foi similar à esperada de 19,41%, mesmo sabendo que entre 75 e 80% dos pacientes operados no InCor/HC-FMUSP são atendidos pelo Sistema Único de Saúde (SUS). Essa mortalidade alta foi corroborada pelo ES, considerando-se a alta prevalência de fatores de risco nesse grupo de pacientes.

Portanto, o controle e a diminuição da prevalência dos fatores de risco resultariam em valores mais baixos da mortalidade observada. Por outro lado, para a calibração dos modelos, o recomendável é utilizar o teste de Hosmer-Lemeshow. O índice de Kappa depende da prevalência da doença em estudo. Uma grande prevalência resulta em alto nível de concordância esperada pelo acaso, o que resultará em valor de k mais baixo. Portanto, poderemos cometer o erro de basear esse índice na comparação de dois estudos com prevalências distintas.

As limitações do estudo foram: primeiro, embora como estudo unicêntrico, a limitação mais importante seja a generalização dos resultados, cerca de 50% dos pacientes atendidos no hospital são oriundos de diversos estados do Brasil. Em segundo lugar, em virtude de sua natureza, o EuroSCORE aditivo tende a subestimar o risco nos pacientes de alto risco; embora isso não tenha sido demonstrado em trabalhos multicêntricos [25]. Por último, embora a mortalidade hospitalar (até 30 dias após cirurgia) pareça ser mais completa que a mortalidade intra-hospitalar, as definições atuais sugerem que ambas têm acurácia equivalente, sendo a mortalidade intra-hospitalar mais prática e fácil de usar [26].

Assim, também os avanços nos cuidados perioperatórios em cirurgia cardíaca poderiam ser mais bem avaliados com o EuroSCORE remodelado (EuroSCORE II) [27], principalmente, nos locais onde o EuroSCORE perdeu calibração. Mas, mesmo assim, devemos ter cuidado com as limitações do novo modelo, porque foi o uso inadequado da primeira versão o que levou a uma expansão dramática do mercado do implante transcater de valva aórtica.

Finalmente, é importante esclarecer que os escores avaliam somente uma parte ínfima das múltiplas variáveis conhecidas e desconhecidas do paciente e da estrutura assistencial, que influenciam diretamente no resultado final do processo. Por essa razão, as conclusões derivadas da sua aplicação devem ser cuidadosamente analisadas.

## CONCLUSÕES

O 2000 Bernstein Parsonnet revelou-se ruim na calibração e bom na discriminação, sendo questionável na validação para prever a mortalidade nos pacientes operados de coronária e/ou valva no InCor/HC-FMUSP.

O EuroSCORE revelou-se bom, tanto na calibração como na discriminação, sendo adequado na validação para prever a mortalidade nos pacientes operados de coronária e/ou valva no InCor/HC-FMUSP.

## REFERÊNCIAS

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, Castelvechio 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.