

# Estudo Comparativo de Testes para Avaliação de Novos Marcadores de Risco

Rodrigo Citton Padilha dos Reis - DEs, UFMG<sup>1 2</sup>

Enrico Antonio Colosimo - DEs, UFMG

Maria do Carmo Pereira Nunes - Dep. Clínica Médica, UFMG

**Resumo:** *Recentemente a avaliação de marcadores de predição de risco tem recebido grande atenção. Apresentamos as seguintes técnicas estatísticas para avaliação de marcadores e modelos de predição de risco: a tabela de estratificação de risco, a melhora da reclassificação líquida (NRI) e a melhora da discriminação integrada (IDI). Simulações de Monte Carlo foram utilizadas para avaliar o desempenho dos testes relacionados ao NRI e IDI. Sob a hipótese nula o teste do NRI apresentou taxas de rejeição mais próximas ao nível nominal do que o teste do IDI. Sob a hipótese alternativa, o teste do IDI apresentou mais poder em relação ao teste do NRI e é menos sensível à variação do tamanho de amostra. Avaliamos a inclusão da razão  $E/E'$  no desempenho do modelo de predição de risco de morte precoce em pacientes cardiopatas com doença de Chagas.*

**Palavras-chave:** *desempenho do modelo, doença de Chagas, IDI, NRI, predição de risco, tabela de estratificação de risco.*

## 1 Introdução

Um marcador de predição de risco é qualquer medida obtida de um paciente que é usada para calcular a probabilidade de se ter uma doença ou ocorrer algum evento adverso (1, 2), geralmente num período de tempo estipulado (por exemplo, o aparecimento de um câncer em até cinco anos, morte em dez anos, entre outros). Este marcador pode ser uma medida quantitativa ou qualitativa, tal como: a idade do indivíduo, o sexo, níveis de colesterol, pressão sistólica, entre outros. Modelos estatísticos, como a regressão logística e de Cox, têm sido utilizados para combinar as informações de marcadores e assim prever o risco de um evento ocorrer para cada indivíduo. Neste contexto de prevenção, estes modelos têm o papel de classificar os indivíduos em categorias de risco clinicamente relevantes. Assim, decisões médicas são tomadas em relação a tratamentos ou intervenções preventivas de acordo com a classificação de cada paciente.

---

<sup>1</sup>Agradecimento à FAPEMIG apoio financeiro.

<sup>2</sup>Contato: [rodrigocpdosreis@gmail.com](mailto:rodrigocpdosreis@gmail.com)

A contínua descoberta de novos marcadores pelos cientistas apresenta oportunidades e desafios para estatísticos e clínicos para avaliar esses marcadores e desenvolver novas formulações de risco que os incorporem (3). A questão principal é avaliar a utilidade de um novo marcador para a predição do risco. Necessariamente um marcador deve ser estatisticamente associado a resposta. Porém, significância estatística não implica em significância clínica ou melhora no desempenho do modelo (3).

Neste trabalho apresentamos algumas técnicas desenvolvidas recentemente para avaliação de novos marcadores de risco. A “tabela de estratificação de risco” (1, 4, 5), a “melhora da reclassificação líquida” (NRI) e a “melhora da discriminação integrada” (IDI) (3). Apresentamos resultados de simulações para comparar os testes de hipóteses relacionados ao NRI e IDI. Ilustramos as técnicas com um banco de dados de cardiomiopatia chagásica. Os dados deste estudo se referem a 224 pacientes com cardiomiopatia chagásica, diagnosticados no período de 1999 a 2008 e acompanhados até o mês de outubro de 2009, no Hospital das Clínicas da Universidade Federal de Minas Gerais. Após a exclusão de indivíduos com valores faltantes, 158 pacientes permaneceram na análise. Os pacientes foram acompanhados por pelo menos 13 meses, ou até a ocorrência da morte. O tempo mediano de acompanhamento foi 37,59 meses. Dos 158 pacientes 24 morreram antes de 13 meses. Em particular, avaliamos se a inclusão de um novo marcador, a razão  $E/E'$ , melhora o desempenho do modelo de predição de risco de morte precoce de pacientes com cardiomiopatia chagásica. A razão  $E/E'$  é um marcador acurado das pressões de enchimento (6), com valor prognóstico estabelecido em várias patologias cardíacas, mas não determinado na cardiopatia chagásica.

Utilizamos o modelo de Cox (7) para posteriormente estimarmos os riscos de morte precoce (morte até 13 meses) para cada paciente. Dois modelos foram ajustados, sem e com a covariável  $E/E'$  (“*modelo antigo*” e “*modelo novo*”). As covariáveis comuns aos dois modelos foram o volume do átrio esquerdo (*volae*), a fração de ejeção do ventrículo esquerdo (*fe*), a classe funcional segundo classificação de NYHA (*New York Heart Association*) e o índice de Tei do ventrículo direito (*vdtei*). A classe funcional foi dicotomizada (classe I e II, NYHA = 0; classe III e IV, NYHA = 1).

## 2 Conceitos Básicos e Notação

O objetivo principal das técnicas que abordamos neste trabalho é avaliar se a inclusão de um novo marcador melhora o desempenho do modelo de predição de risco. Dado um vetor  $\mathbf{x}$  de covariáveis observadas para um indivíduo, o risco é definido como a probabilidade condicional a este vetor de covariáveis de ocorrer o evento de interesse no período de tempo especificado:

$$risco(\mathbf{x}) = P(\text{“evento ocorre até o tempo } t\text{”} | \mathbf{x}). \quad (1)$$

Utilizando o modelo de Cox para dados de sobrevivência, 1 é reescrito da seguinte forma:

$$risco(\mathbf{x}_i) = 1 - [S_0(t)]^{\exp\{\mathbf{x}'_i\boldsymbol{\beta}\}}, \quad (2)$$

em que  $\mathbf{x}_i$  é o vetor de covariáveis do  $i$ -ésimo indivíduo. Estimando o vetor  $\beta$  e a função de sobrevivência de base  $S_0(t)$  no tempo  $t$  especificado, temos a estimativa do risco para o  $i$ -ésimo indivíduo.

Quando avaliamos o risco predito em uma amostra, em algumas situações é de interesse estipular limiares que estratificam o risco em categorias. Neste trabalho usamos dois limiares de risco:  $p_L$ , o limiar de baixo risco e  $p_H$ , o limiar de alto risco. As definições de  $p_L$  e  $p_H$  são inteiramente clínicas, ou seja, são obtidas a partir de interesses e experiência da área médica.

### 3 Tabela de Estratificação de Risco, NRI e IDI

Na tabela de estratificação, os riscos são estimados dos modelos sem e com o novo marcador, os indivíduos em estudo são estratificados em categorias de risco clinicamente relevantes, e então estas duas classificações são cruzadas. Janes, Pepe e Gu (1) concluíram que a informação que pode ser extraída da tabela de estratificação de risco é um enorme avanço sobre medidas geralmente utilizadas para avaliar o desempenho do modelo de predição de risco. Apresentamos a tabela de estratificação de risco utilizando os dados de cardiomiopatia chagásica (Tabela 1).

Tabela 1: Tabela de estratificação: Risco de morte em um até um ano pelos modelos sem e com E/E'.

		<b>Risco 1 ano modelo sem E/E'</b>			<b>Risco 1 ano modelo com E/E'</b>		
		0-2%	2-10%	$\geq 10\%$	Total		
0-2%	N	57	5	0	62		
	Eventos	0	1	0	1		
	Não-eventos	57	4	0	61		
	Prop. Eventos	0,00	20,00	-	1,61		
2-10%	N	4	46	4	54		
	Eventos	0	1	0	1		
	Não-eventos	4	45	4	53		
	Prop. Eventos	0,00	2,17	0,00	1,85		
$\geq 10\%$	N	0	5	37	42		
	Eventos	0	0	22	22		
	Não-eventos	0	5	15	20		
	Prop. Eventos	-	0,00	59,46	52,38		
Total	N	61	56	41	158		
	Eventos	0	2	22	24		
	Não-eventos	61	54	19	134		
	Prop. Eventos	0,00	3,57	53,66	15,19		

Na Tabela 1, para cada classe de risco, temos o número de indivíduos alocados na classe (N), o número de eventos na classe, o número de não-eventos na classe, e a proporção de eventos na classe (Prop. Eventos). Janes, Pepe e Gu (1) descrevem como é possível acessar informações a respeito da calibração do modelo, capacidade de estratificação e acurácia da classificação através da tabela de estratificação. Observando as proporções de eventos nas marginais da Tabela 1, percebemos que o modelo sem a razão E/E' não está bem calibrado, pois 1,85% dos indivíduos da

classe “2 a 10%” são eventos. Neste aspecto o modelo com a razão E/E’ leva pequena vantagem em relação ao modelo que não a inclui. Quanto a capacidade de estratificação, o desempenho dos dois modelos foi muito semelhante. O modelo sem E/E’ designa aproximadamente 34% dos indivíduos à classe intermediária (“2 a 10%”), enquanto o modelo com E/E’ designa 35% dos indivíduos à classe intermediária. A acurácia da classificação também foi praticamente igual para os dois modelos. Da Tabela 1 estimamos  $TPF(p_H) = 92\%$  para os dois modelos e  $FPF(p_H) = 15\%$  para o modelo sem E/E’ e  $FPF(p_H) = 14\%$  para o modelo com E/E’.

Para quantificar a melhora na reclassificação devido ao acréscimo de uma nova variável ao modelo, Pencina *et al.* (3) propõem o NRI, que é um índice que pode ser estimado diretamente da tabela de estratificação. Desta maneira, a melhora na reclassificação pode ser quantificada como a soma da “proporção de indivíduos movendo para cima” menos a “proporção de indivíduos movendo para baixo” na população de eventos, e a “proporção de indivíduos movendo para baixo” menos a “proporção de indivíduos movendo para cima” na população de não-eventos. O NRI é estimado por

$$\hat{NRI} = (\hat{p}_{cima,E} - \hat{p}_{baixo,E}) - (\hat{p}_{cima,NE} - \hat{p}_{baixo,NE}), \quad (3)$$

em que  $n_E$  e  $n_{NE}$  são o número de indivíduos eventos e número de indivíduos não-eventos, respectivamente.

Assumindo independência entre os indivíduos das populações de eventos e não-eventos, Pencina *et al.* (3) obtiveram a seguinte estatística de teste para a hipótese nula  $NRI = 0$  ( $H_0 : NRI = 0$ ):

$$z_{NRI} = \frac{\hat{NRI}}{\sqrt{\frac{\hat{p}_{cima,E} + \hat{p}_{baixo,E}}{n_E} + \frac{\hat{p}_{cima,NE} + \hat{p}_{baixo,NE}}{n_{NE}}}}. \quad (4)$$

A estatística  $z_{NRI}$  tem distribuição assintótica normal padrão, sob  $H_0$ . Para o estudo de cardiomiopatia chagásica o NRI foi estimado em 0,0491 e não foi significativo (valor  $p = 0,3429$ ).

A alta dependência na escolha das classes de risco foi reconhecida e apontada como uma possível desvantagem do NRI (3). Para superar esta limitação, Pencina *et al.* (3) propuseram o IDI (*integrated discrimination improvement*). Imagine que o risco estimado para cada indivíduo é a sua própria categoria de risco, então qualquer mudança no risco estimado proporciona um movimento para baixo ou para cima. Pencina *et al.* (3) mostraram que isto equivale a quantificar melhoras conjuntas nas curvas de sensibilidade (fração de verdadeiros positivos) e especificidade (um menos a fração de falsos positivos). Assim o IDI é estimado por

$$\hat{IDI} = (\bar{r}_{novo,E} - \bar{r}_{antigo,E}) - (\bar{r}_{novo,NE} - \bar{r}_{antigo,NE}). \quad (5)$$

Como os eventos atuais não dependem do modelo (“antigo” ou “novo”), o desvio padrão de  $(\bar{r}_{novo,E} - \bar{r}_{antigo,E})$  pode ser calculado como o erro padrão das diferenças pareadas dos riscos dos dois modelos para indivíduos da população de eventos,  $\hat{e}p_E$  (3). Da mesma forma o correspondente estimador do desvio padrão de  $(\bar{r}_{novo,NE} - \bar{r}_{antigo,NE})$  é  $\hat{e}p_{NE}$ . Assumindo independência entre eventos e não-eventos, Pencina *et al.* (3) propuseram a seguinte estatística

de teste para hipótese nula  $IDI = 0$  ( $H_0 : IDI = 0$ ):

$$z_{IDI} = \frac{\hat{IDI}}{\sqrt{(\hat{e}p_E)^2 + (\hat{e}p_{NE})^2}}. \quad (6)$$

A estatística  $z_{IDI}$  tem distribuição assintótica normal padrão, sob  $H_0$ . Para os dados de cardiomiopatia chagásica o IDI foi estimado em 0,0873 e foi significativo (valor  $p = 0,0145$ ).

## 4 Estudo de Simulação

Avaliamos através de simulações Monte Carlo a qualidade das inferências para o NRI e o IDI introduzidos por Pencina *et al.* (3). Fizemos comparações em relação as taxas de rejeição dos testes de hipóteses do NRI e IDI sob a hipótese nula ( $H_0 : NRI = 0$  e  $H_0 : IDI = 0$ ) e alguns cenários sob a hipótese alternativa, considerando o nível nominal de 5%. As comparações foram feitas utilizando dois modelos para a predição do risco de um determinado evento. O primeiro modelo com uma única covariável ( $X_1$ ), denominado de *modelo antigo*, e o segundo modelo com as covariáveis  $X_1$  e  $X_2$ , o *modelo novo*. Desta forma o NRI e o IDI avaliam se o novo marcador (a covariável  $X_2$ ) melhora o desempenho do modelo. Utilizamos o modelo exponencial para dados de sobrevivência para o ajuste dos modelos, pois este é um caso particular do modelo de Cox. As simulações foram realizadas utilizando a linguagem de programação R 2.9.0 (8). Em todos os casos simulados, utilizamos 3.000 réplicas de Monte Carlo. Apresentamos os resultados para o cenário em que simulamos sob a hipótese nula na Tabela 2.

Tabela 2: Taxas de rejeição (%) das hipóteses nula ( $H_0 : NRI = 0$  e  $H_0 : IDI = 0$ ) com base em 3.000 réplicas de Monte Carlo sob  $H_0$ .

n	Amostras		
	Mal Sucedidas	NRI	IDI
5.000	38	4,63	5,64
2.000	57	4,04	11,45
1.000	85	4,08	7,72
500	164	4,37	6,73
300	209	3,33	5,05
100	257	3,86	6,05

## 5 Conclusões

Os resultados que encontramos nos mostram que o teste relacionado ao NRI é mais conservador. Quando simulamos sob  $H_0$ , o teste do NRI mostrou taxas de rejeição mais próximas ao nível nominal de 5%. Quando não temos condições de afirmar que estamos sob  $H_0$ , mas  $X_2$  claramente não acrescenta ao modelo de predição de risco o teste relacionado ao NRI também teve taxas de rejeição mais próximas do nível nominal. Porém, quando simulamos sob a hipótese

alternativa, o teste do NRI perdeu poder conforme variamos os cenários. Nos casos que  $X_2$  claramente adiciona valor preditivo ao modelo, o teste relacionado ao NRI mostrou alto poder, mas este decresce rapidamente conforme o tamanho da amostra diminui. Para os cenários que  $X_2$  acrescenta no desempenho do modelo de forma moderada para fraca, o teste do NRI apresentou baixo poder para praticamente todos os tamanhos de amostra.

O teste relacionado ao IDI mostrou leve desvantagem em relação ao NRI, quando simulamos sob  $H_0$ . Neste caso, para todos os tamanhos de amostra o teste do IDI apresentou taxas de rejeição acima do nível nominal, demonstrando ser um teste menos conservador, no sentido que rejeita mais do que o valor nominal. Para os cenários que consideramos estar simulando muito próximos da situação sob  $H_0$ , o teste do IDI rejeitou mais do que o esperado. Assim, qualquer mudança no comportamento do risco estimado, quando incorporado um novo marcador ao modelo, parece não passar despercebida pelo teste do IDI. Quando simulamos sob a hipótese alternativa, o teste do IDI apresentou os melhores resultados do que o teste do NRI, mostrando ser muito poderoso.

## Referências

- 1 JANES, H.; PEPE, M. S.; GU, W. Accessing the value of risk predictions by using risk stratification tables. *Annals of Internal Medicine*, v. 149, p. 751–760, 2008.
- 2 GU, W.; PEPE, M. S. Estimating the capacity for improvement in risk prediction with a marker. *Biostatistics*, v. 10, p. 172–186, 2008.
- 3 PENCINA, M. J.; D’AGOSTINO Sr., R. B.; D’AGOSTINO Jr., R. B.; VASAN, R. S. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, v. 27, p. 157–172, 2008.
- 4 COOK, N. R.; BURING, J. E.; RIDKER, P. M. The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Annals of Internal Medicine*, v. 145, p. 21–29, 2006.
- 5 COOK, N. R. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, v. 115, p. 928–935, 2007.
- 6 OMMEN, S. R.; NISHIMURA, R. A.; APPLETON, C. P.; MILLER, F. A.; OH, J. K.; REDFIELD, M. M.; TAJIK, A. J. Clinical utility of doppler echocardiography and tissue doppler imaging in the estimation of left ventricular filling pressures: a comparative simultaneous doppler catheterization study. *Circulation*, v. 102, p. 1788–1794, 2000.
- 7 COX, D. R. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, B*, v. 34, p. 187–220, 1972.
- 8 R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2009. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>.