

## MODELO DE COX PARA EVENTOS CARDIOVASCULARES RECORRENTES EM PACIENTES SOB DIÁLISE COM COVARIÁVEIS MEDIDAS NO TEMPO

Thiago Santos MOTA<sup>1</sup>  
Liciane Vaz de Arruda SILVEIRA<sup>2</sup>  
Aline Araujo ANTUNES<sup>3</sup>

- **RESUMO:** Este estudo foi realizado com o objetivo de avaliar o efeito das covariáveis medidas no tempo até a ocorrência dos eventos cardiovasculares (acidentes vasculares cerebrais) recorrentes em pacientes sob diálise, consistindo de 145 pacientes em tratamento dialítico do hospital das Clínicas da Faculdade de Medicina da UNESP – Campus de Botucatu, SP, com início no ano de 2008. As covariáveis consideradas neste estudo foram variáveis clínicas, nutricionais, laboratoriais e dialíticas. Como os dados tratam de eventos recorrentes do mesmo tipo foi utilizada a modelagem marginal, em especial, os modelos AG (Andersen e Gill, 1982) e PWP (PRENTICE *et al.*, 1981). Foram consideradas nos dois modelos as covariáveis significativas ao nível de 10%, utilizando o critério de seleção de Collett (1994). Foi apresentada também a razão de risco dos pacientes em relação a cada covariável em ambos os modelos e o critério de AKAIKE (AIC). Para o modelo AG o AIC resultou no valor de 499,9 e para o modelo PWP no valor de 425,68, sendo todos os resultados obtidos no software estatístico R. A partir dos resultados optou-se pelo modelo PWP, pois teve menor valor do AIC e pela análise de resíduos mostrou-se ajustar bem aos dados. Este estudo apontou a importância da modelagem marginal como uma forma de modelar eventos recorrentes do mesmo tipo e a utilização dos modelos mencionados em dados que seguem necessariamente uma ordem.
- **PALAVRAS-CHAVE:** Modelagem marginal; AIC; modelo AG e modelo PWP.

<sup>1</sup>Universidade Estadual Paulista – UNESP, Campus de Botucatu, Instituto de Biociências, Programa de Pós-Graduação em Bioestatística, Caixa Postal: 510, CEP: 18618-970, Botucatu, SP, Brasil. E-mail: [thiagosanmota@ibb.unesp.br](mailto:thiagosanmota@ibb.unesp.br)

<sup>2</sup>Universidade Estadual Paulista – UNESP, Campus de Botucatu, Instituto de Biociências, Departamento de Bioestatística, Caixa Postal: 510, CEP: 18618-970, Botucatu, SP, Brasil. E-mail: [liciana@ibb.unesp.br](mailto:liciana@ibb.unesp.br)

<sup>3</sup>Universidade Estadual Paulista – UNESP, Campus de Botucatu, Faculdade de Medicina, Departamento de Fisiopatologia em Clínica Médica, Caixa Postal: 580, CEP:18618-970, Botucatu, SP, Brasil. E-mail: [alineantunes@fmb.unesp.br](mailto:alineantunes@fmb.unesp.br)

## 1 Introdução

A análise de sobrevivência é uma das áreas da estatística que mais tem crescido nos últimos anos segundo Therneau e Grambsch (2000), empregando técnicas e modelos estatísticos com o propósito de analisar dados cuja variável resposta é o tempo até a ocorrência de um evento de interesse (falha), por exemplo, morte de um indivíduo. Estes dados são frequentemente censurados, isto é, apresentam observações incompletas que por algum motivo não foi possível observar a ocorrência do evento de interesse.

De acordo com Colosimo e Giolo (2006) “duas razões justificam considerar a censura na modelagem:

- a) mesmo sendo incompletas, as observações censuradas fornecem informações sobre o tempo de vida de pacientes;
- b) a omissão das censuras no cálculo das estatísticas de interesse pode acarretar conclusões viciadas. ”

Uma das maneiras para modelar dados com essas características na presença de covariáveis é utilizar o modelo de regressão semi-paramétrico proposto por Cox (1972), que também é denominado modelo de riscos proporcionais, pois a razão da taxa de falha de dois indivíduos distintos é constante no tempo. Assumindo que os tempos  $t_i$ , com  $i = 1, \dots, k$ , são independentes o modelo de Cox para o  $i$ -ésimo indivíduo, dado o vetor  $\mathbf{x} = (x_1, \dots, x_p)'$  de covariáveis, é dado por:

$$\lambda(t|\mathbf{x}_i) = \lambda_0(t) \exp \{ \mathbf{x}_i' \beta \},$$

em que  $\lambda(t|\mathbf{x}_i)$  representa a função taxa de risco,  $\lambda_0(t)$  é o componente não paramétrico, conhecido como função de base, pois  $\lambda(t) = \lambda_0(t)$  para  $\mathbf{x} = 0$ ,  $\beta = (\beta_1, \dots, \beta_p)'$  é o vetor de parâmetros associados às covariáveis e  $\mathbf{x}_i$  é o vetor de dimensão  $p$  de covariáveis observadas para o  $i$ -ésimo indivíduo.

Porém, há situações na área médica que o evento de interesse pode ocorrer mais de uma vez para o mesmo indivíduo, como é o caso de eventos cardiovasculares (acidentes vasculares cerebrais) recorrentes em pacientes sob diálise com covariáveis medidas no tempo. Nestas condições não se tem independência entre os tempos de falha o que dificulta a modelagem. De acordo com Carvalho *et al.* (2011) o uso do modelo de Cox na forma usual não é adequado, mesmo se utilizarmos a formulação por processo de contagem, pois os intervalos de tempos para o mesmo indivíduo podem se sobrepor. Além disso, como a mesma pessoa sofre diversos eventos, esses podem ser correlacionados.

Uma das formas de realizar este tipo de modelagem e estimar os parâmetros corretamente é utilizar uma estrutura estocástica, sendo a mais plausível, segundo McLain e Peñay (2008), a proposta de Andersen e Gill (1982), que inseriram o modelo de Cox em um processo de contagem provando a consistência e a normalidade assintótica. Uma das abordagens de modelagem que é inserida neste contexto é a abordagem marginal de eventos recorrentes, que pode ser encontrada

em Therneau e Grambsch (2000), que especifica uma distribuição marginal sobre cada número de eventos, ignorando a dependência entre os eventos e corrige esta dependência na variância. Os modelos mais utilizados nesta abordagem são os modelos AG (Andersen e Gill, 1982), WLW (WEI *et al.*, 1989) e PWP (PRENTICE *et al.*, 1981). Aplicações desses modelos podem ser encontradas, por exemplo, em Therneau e Grambsch (2000), Colosimo e Giolo (2006) e Carvalho *et al.* (2011).

## 2 Material e Métodos

Therneau e Grambsch (2000) fizeram uma explanação dos modelos AG, WLW e PWP de como eles podem ser aplicados. Serão utilizados neste trabalho os modelos AG e PWP, não sendo considerado o modelo WLW pelas seguintes razões:

- (a) Este modelo pode violar a suposição de riscos proporcionais, mesmo quando isso não ocorre para o conjunto de dados em geral, como mostrado em simulações em Therneau e Grambsch (2000);
- (b) Pela forma da estrutura temporal e o tipo de evento estudado neste trabalho considerou-se este modelo inadequado, pois, no modelo WLW o indivíduo ao entrar no estudo está simultaneamente em risco de sofrer o primeiro, o segundo, o terceiro e cessa quando experimenta um número máximo de  $m$  eventos, que podem ocorrer no estudo. Assim, os modelos AG e PWP são mais apropriados para este trabalho, não permitindo que o indivíduo esteja em risco de mais de um evento em um tempo  $t$ .

O modelo AG considera que o risco basal é igual em todos os intervalos de tempo analisados, sendo que o indivíduo retorna ao grupo de risco após cada evento e assume que os eventos em cada intervalo que não se sobrepõe são independentes. Dependendo da escala de medida do tempo, a primeira observação poderá ou não começar no zero. Se a primeira observação começar no tempo de entrada, o modelo para o  $i$ -ésimo indivíduo fica representado por:

$$\lambda_i(t) = Y_i \lambda_0(t) \exp \{ \mathbf{x}_i'(t) \beta \},$$

em que  $\lambda_i(t)$  é a função de risco para o  $i$ -ésimo indivíduo,  $\lambda_0(t)$  é o componente não paramétrico,  $\beta = (\beta_1, \dots, \beta_p)'$  é o vetor de parâmetros associado às covariáveis,  $\mathbf{x}_i$  é o vetor de dimensão  $p$  de covariáveis observadas para o  $i$ -ésimo indivíduo no tempo  $t$  e a função indicadora de risco:

$$Y_i = \begin{cases} 1, & \text{se o indivíduo } i \text{ estiver sob observação e em risco no tempo } t \\ 0, & \text{c.c.} \end{cases}$$

O uso do modelo AG é apropriado no caso em que se tem independência mútua entre as observações em relação a um mesmo indivíduo.

Por outro lado, no modelo PWP o indivíduo só estará em risco de experimentar o  $m$ -ésimo evento depois que tenha experimentado o evento  $m-1$ . O modelo PWP separa a análise em diferentes estratos, assumindo que existe uma dependência

entre os tempos de falha de um mesmo indivíduo. O uso de estratos dependentes significa que a função de risco pode variar de um evento para outro, ao contrário do que ocorre no modelo AG. Assim, por exemplo, o risco basal para o segundo evento é zero até que o primeiro evento ocorra, enquanto que o risco do terceiro evento é zero até que o segundo evento ocorra, e assim sucessivamente. O modelo pode ser escrito da seguinte forma:

$$\lambda_{i_m}(t) = Y_{i_m} \lambda_{0_m}(t) \exp \{ \mathbf{x}'_i(t) \beta_m \},$$

em que  $\lambda_{i_m}(t)$  é a função de risco do  $m$ -ésimo evento desde que o  $i$ -ésimo indivíduo tenha experimentado os  $m-1$  eventos,  $\lambda_{0_m}(t)$  é a função de base que pode variar de um evento para outro,  $\beta_m = (\beta_1, \dots, \beta_p)'$  é o vetor de parâmetros associados às covariáveis no  $m$ -ésimo evento,  $\mathbf{x}_i$  é o vetor de dimensão  $p$  de covariáveis observadas para o  $i$ -ésimo indivíduo no tempo  $t$  e a função indicadora de risco  $Y_{i_m}$  é zero até ocorrer o evento  $m - 1$ , ocorrendo este evento a função assume o valor 1.

Para ajustar esses dois modelos devem estar claro como é o comportamento dos eventos estudados e a estrutura de entrada dos dados. O conjunto de dados em que serão aplicadas essas técnicas, eventos cardiovasculares (acidentes vasculares cerebrais) recorrentes em pacientes sob diálise com covariáveis medidas no tempo, é um estudo de coorte prospectivo que incluiu pacientes com doença renal crônica em tratamento dialítico há no mínimo três meses, sendo excluídos menores de 18 anos. Os dados fazem parte da tese de Doutorado de Antunes (2012) consistindo de 145 pacientes do Hospital das Clínicas da Faculdade de Medicina da UNESP – Campus Botucatu, SP, que foram coletados a partir do ano de 2008. As covariáveis consideradas neste estudo foram variáveis clínicas, nutricionais, laboratoriais e dialíticas. Nestas situações as covariáveis empregadas foram sexo, método de diálise (diálise peritoneal e hemodiálise), tempo em diálise, diagnóstico de diabetes mellitus, triglicérides, pressão arterial sistólica e diastólica (e a diferença entre essas pressões – DPA), colesterol total e HDL colesterol, risco relativo de ocorrência de doença cardíaca por Framingham, paratormônio, cálcio, fósforo, creatinina, uréia, proteína C-reativa, albumina e índice de massa corpórea. Na análise de bioimpedância (Biodynamics®, modelo 450, 800  $\mu A$ , 50 KHz) obteve-se o valor da relação entre massa extracelular por massa celular corporal (MEC/MCC), que quanto maior pior o estado nutricional e/ou de hidratação, repercutindo em pior prognóstico do paciente, segundo Antunes (2012).

Para selecionar as covariáveis que serão inseridas no modelo final utilizou-se a estratégia de seleção de modelos proveniente da proposta de Collett (1994) e que também pode ser vista em Colosimo e Giolo (2006). A análise estatística foi feita pelo programa estatístico R (R DEVELOPMENT CORE TEAM, 2012), cujas rotinas encontram-se no Apêndice.

### 3 Resultados e discussões

Inicialmente foi feita a seleção das covariáveis dos modelos AG e PWP, utilizando a proposta de Collett (1994), considerando um nível de significância 10%

e analisando a significância das covariáveis pelo teste de Wald robusto, que utiliza a variância robusta em sua estatística para lidar com a dependência dos tempos de falha observados para o mesmo indivíduo.

Os resultados obtidos para o modelo AG são apresentados na Tabela 1, pode-se observar que todas as covariáveis incluídas no modelo foram significativas ao nível de 5%. Em seguida, obteve-se o valor de 499,9 do critério de AKAIKE (AIC) na presença das covariáveis no modelo.

Tabela 1 - Estimativa dos parâmetros do modelo AG e razão de risco, ajustados aos dados de eventos cardiovasculares recorrentes em pacientes sob diálise

Covariável	Estimativas	Erro Padrão	Wald	Valor p	Razão de risco
Cálcio	0,4466	0,1920	2,326	0,020010	1,563
Fósforo	0,3696	0,1556	2,375	0,017550	1,447
MEC/MCC	2,9777	0,7732	3,851	0,000118	19,643

Os resultados obtidos para o modelo PWP são apresentados na Tabela 2, observa-se também que todas as covariáveis incluídas no modelo foram significativas ao nível de 5%. Para este modelo, obteve-se o valor de 425,68 do AIC na presença das covariáveis no modelo. Pelos valores do AIC optou-se pelo modelo PWP, pois apresentou menor valor deste critério.

Tabela 2 - Estimativa dos parâmetros do modelo PWP e razão de risco, ajustados aos dados de eventos cardiovasculares recorrentes em pacientes sob diálise

Covariável	Estimativas	Erro Padrão	Wald	Valor p	Razão de risco
Cálcio	0,3780	0,1843	2,050	0,04033	1,459
Fósforo	0,3396	0,1584	2,144	0,03206	1,404
MEC/MCC	2,1552	0,7557	2,852	0,00434	8,630

A partir destes resultados, o próximo passo foi verificar a qualidade do ajuste do modelo PWP e decidiu-se em utilizar os resíduos de Schoenfeld e os resíduos de Martingale.

Na Tabela 3, pode-se observar que os valores do coeficiente de correlação de Pearson ( $\rho$ ) são todos próximos de zero. Além disso, observa-se que tanto o teste global quanto os testes para cada covariável não apresentam evidências para rejeitar a hipótese nula de riscos proporcionais.

Observando os gráficos dos resíduos padronizados de Schoenfeld versus os tempos para as covariáveis consideradas no modelo PWP (Figura 1), confirma-se o que foi exposto no teste de proporcionalidade, onde não se consegue visualizar tendências evidentes ao longo do tempo. Pela análise do gráfico de resíduos de Martingale para o modelo PWP versus índice (Figura 2), nota-se que os resíduos

Tabela 3 - Teste de proporcionalidade dos riscos no modelo PWP ajustado

Covariável	$\hat{\rho}^a$	$T^b$	valor p
Cálcio	0,0702	0,219	0,639
Fósforo	-0,0708	0,558	0,455
MEC/MCC	-0,0975	0,531	0,466
GLOBAL	NA	0,953	0,813

<sup>a</sup> Coeficiente de correlação de Pearson estimado entre os resíduos padronizados de Schoenfeld e a variável resposta tempo até a ocorrência do evento .

<sup>b</sup> Estatística do teste com aproximadamente distribuição qui-quadrado com 3 graus de liberdade.

estão relativamente bem distribuídos acima e abaixo da linha que passa pelo zero e não tem nenhum padrão detectável. Desse modo, o modelo PWP é uma opção satisfatória para análise desses dados.

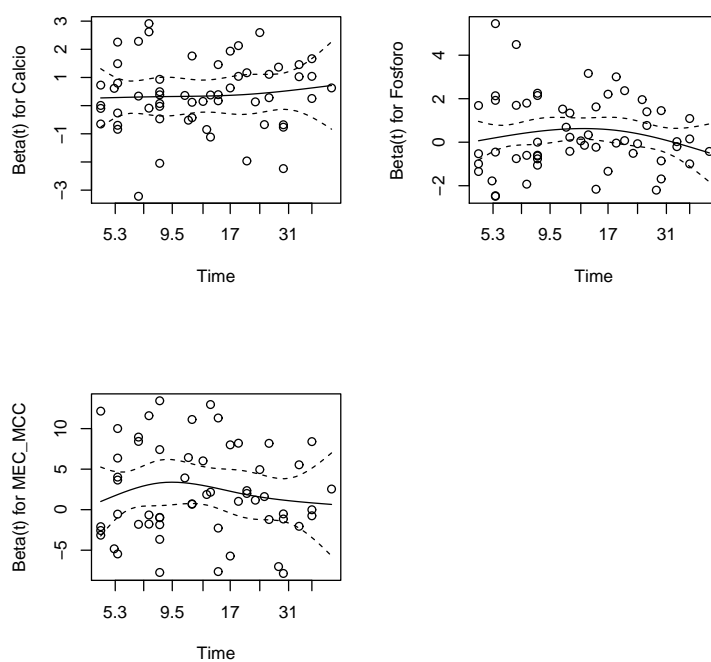


Figura 1 - Resíduos padronizados de Schoenfeld versus os tempos para as covariáveis consideradas no modelo PWP.

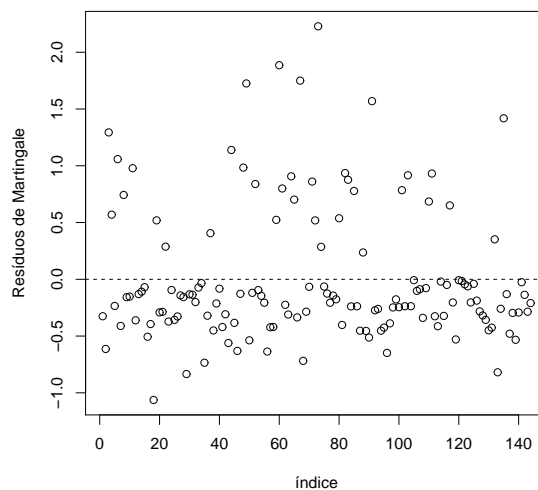


Figura 2 - Gráfico dos resíduos de Martingale para o modelo PWP versus índice.

## Conclusões

A partir dos resultados apresentados nas tabelas, figuras e os valores do AIC, neste estudo considerou-se que o modelo PWP teve melhor ajuste do que o modelo AG. Este estudo aponta a importância da modelagem marginal como uma forma de modelar dados de eventos recorrentes e a utilização dos modelos AG e PWP em dados que seguem necessariamente uma ordem.

## Agradecimentos

Agradeço a Capes e a FAPESP (Processo: 2012/03599-8) pelo apoio financeiro, e aos revisores pelas correções e sugestões.

MOTA, T. S; SILVEIRA, L. V. A; ANTUNES, A. A. Cox model for recurring cardiovascular events in patients under dialysis with covariates measured over time. *Rev. Bras. Biom.*, São Paulo, v.30, n.1, p.150-159, 2012.

■ **ABSTRACT:** The aim is to evaluate the effect of the covariates measured over time for the occurrence of the (stroke) cardiovascular events recurring in patients under dialysis, consisting of 145 hemodialysis patients from the clinics hospital of the Medicine College, UNESP – Botucatu, SP, starting in 2008. The covariates considered in this study were clinical, nutritional, laboratory and dialysis. As the data treat the recurring events of the same type, the marginal modeling employing, in particular, the models AG (Andersen and Gill, 1982) and PWP (PRENTICE et al., 1981). The covariates were considered significant at 10% in the two models, using the selection criteria Collett (1994). It also presented the risk ratio of patients in relation to each covariate in both models and the AKAIKE information criteria (AIC). For the AG model the result of AIC was 499.9 and in the PWP model it was 425.68. All the results obtained in the R software. From the results, the PWP model was chosen because it had value lowest of the AIC and by the residuals analysis it showed good fit to the data. This study showed the importance of the marginal modeling as a way to model recurring events of the same type, and the utilization of the models mentioned in data which necessarily follow an order.

■ **KEYWORDS:** Marginal modeling; AIC; AG model and PWP model.

## Referências

ANDERSEN, P. K.; GILL, R. D. Cox's regression model for counting processes: a large sample study. *Annals of Statistical*, v.10, p.1100-1120, 1982.

ANTUNES, A. A. *Impacto do acompanhamento com bioimpedância na predição de eventos cardiovasculares em diálise*. 2012. Tese (Doutorado em Fisiopatologia em Clínica Médica) – Faculdade de Medicina, Campus de Botucatu, Universidade Estadual Paulista, Botucatu, 2012.

CARVALHO, M. S.; ANDREOZZI, V. L.; CODEÇO, C. T.; CAMPOS, D. P.; BARBOSA, M. T. S.; SHIMAKURA, S. E. *Análise de sobrevivência: teoria e aplicações em saúde*. Rio de Janeiro: Editora Fiocruz, 2011. 432p.

COLLETT, D. *Modelling survival data in medical research*. New York: Chapman and Hall, 1994. 347p.

COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. São Paulo: Edgard Blücher, 2006. 392p.

COX, D. R. Regression Models and Life-Tables. *Journal of the Royal statistical Society-B*, v.34, p.187-220, 1972.

McLAIN, A.; PEÑAY, E. Some Issues in Marginal Recurrent Event Cox type Models. *Institute of Statistical Mimeo Series*, série.2618, 2008.

PRENTICE, R. L.; WILLIAMS, B. J.; PETERSON, A. V. On the regression analysis of multivariate failure time data. *Biometrika*, v.68, p.373-379, 1981.



R DEVELOPMENT CORE TEAM. *R: A language and environment for statistical computing*. R: Foudation for Statistical Computing. Viena, Austria. Disponível em:<<http://www.R-project.org>>. Acesso em 30 mar. 2012.

THERNEAU, T. M.; GRAMBSCH, P. M. *Modeling survival data: extending the Cox Model*. New York: Springer, 2000. 350p.

WEI, L. J.; LIN, D. Y.; WEISSFELD, L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, v.84, p.1065-1073, 1989.

Recebido em 27.02.2012.

Aprovado após revisão em 22.06.2012.

## APÊNDICE

### Rotinas do programa estatístico R utilizadas

#### Entrada dos dados:

```
dados <- read.csv2("dados1.csv")
require(survival)
require(MASS)
```

#### Ajuste dos modelos:

```
agfit <- coxph(Surv(start, stop, status)~
Fosforo+Calcio+MEC_MCC+cluster(id), data=dados, method="breslow")
summary(agfit)
stepAIC(agfit)
pwpfit <- coxph(Surv(start, stop, status)~
Calcio+Fosforo+MEC_MCC+cluster(id)+strata(stratum), data=dados, method="breslow")
summary(pwpfit)
stepAIC(pwpfit)
```

#### Avaliação da suposição de riscos proporcionais no modelo PWP - Resíduos padronizados de Schoenfeld:

```
cox.zph(pwpfit, transform="identity")
par(mfrow=c(2,2))
plot(cox.zph(pwpfit))
```

#### Identificando pontos aberrantes - Resíduos de Martingale:

```
res.mart <- residuals(pwpfit, type="martingale", collapse=dados$id)
plot(res.mart, ylim=c(-3,3), xlab="índice", ylab="Resíduos de Martingale" )
abline(0,0, lty= 2)
```