

## **MODELOS DE RISCOS COMPETITIVOS PARA ANÁLISE DE SOBREVIVÊNCIA DE PACIENTES COM HIPERTENSÃO ARTERIAL EM TRATAMENTO NO HUCFF/UFRJ**

**Cachimo Combo Assane**

COPPE/UFRJ

Ilha do Fundão, Centro de Tecnologia-Bloco F-Sala 105, Rio de Janeiro

cachimo.assane@gmail.com

**Basílio de Bragança Pereira**

Faculdade de Medicina, HUCFF e COPPE/UFRJ

Ilha do Fundão, Centro de Tecnologia-Bloco F-Sala 105, Rio de Janeiro

basilio@hucff.ufrj.br

**Miranda Albino M. Muualo**

COPPE/UFRJ

Ilha do Fundão, Centro de Tecnologia-Bloco F-Sala 105, Rio de Janeiro

mamuualo@gmail.com

### **RESUMO**

O modelo de riscos competitivos envolve situações nas quais um indivíduo pode sofrer mais do que um tipo de evento de interesse. O uso de métodos padrão de análise de sobrevivência tais como o estimador de Kaplan-Meier e modelos de riscos proporcionais de Cox para descrever os dados de riscos competitivos pode conduzir a estimativas enviesadas.

Neste trabalho utilizamos a função de incidência acumulada para estimar a probabilidade de falha de um paciente por um evento de interesse e o modelo de riscos proporcionais de subdistribuição para investigar os efeitos das covariáveis na ocorrência deste evento, tendo em conta os eventos competitivos. O modelo de risco de subdistribuição permitiu identificar fatores de prognósticos importantes na incidência dos desfechos observados em pacientes com hipertensão arterial.

**PALAVRAS CHAVE.** Análise de sobrevivência, Riscos competitivos, Função de incidência acumulada.

### **ABSTRACT**

The competing risks model involves situations in which an individual can experience more than one type of event of interest. The use of standard survival analysis methods such as the Kaplan-Meier estimator and Cox proportional hazards models to describe the competing risks data can lead to biased estimates.

In this article we use the cumulative incidence function for estimating the probability of failure for an event of interest and the proportional subdistribution hazards model to investigate the effect of covariates on occurrence of this event, taking into account the presence of competing events. The subdistribution hazards model allowed to identify important prognostic factors on incidence of observed outcomes from hypertension patients.

**KEYWORDS.** Survival analysis. Competing risks. Cumulative incidence function.

## 1. Introdução

A análise de sobrevivência clássica refere-se ao estudo de dados relacionados ao tempo até a ocorrência de um determinado evento de interesse, a partir de um tempo inicial até um tempo final, devidamente pré-definidos, conhecido como tempo de estudo ou janela de observação (Collett, 1994). O evento que marca o término de acompanhamento de um indivíduo é também designado por desfecho ou falha.

Uma característica importante inerente aos dados de sobrevivência é a presença de censuras, isto é, informações incompletas sobre o tempo de sobrevivência para um determinado indivíduo. Em estudos médicos e áreas afins, censuras ditas à direita são as mais comuns. Discussão detalhada sobre diferentes tipos de censuras pode ser encontrada em Collett (1994). As censuras à direita podem ocorrer pelas seguintes razões (Putter et al., 2007):

*Término do estudo:* o tempo de sobrevivência de um indivíduo pode ser censurado quando este é acompanhado até o final do tempo de estudo sem ter sofrido ainda o evento de interesse. Este tipo de censura é também conhecido como censura administrativa.

*Perda de seguimento:* durante o acompanhamento, o indivíduo pode abandonar o estudo antes da observação do evento de interesse, devido à, por exemplo, efeitos secundários do tratamento médico ou mudança de localidade.

*Risco competitivo:* o indivíduo pode sofrer um outro tipo de evento que impedirá a observação do evento de interesse em estudo.

A estimativa de probabilidade de um certo desfecho acontecer antes de um determinado tempo é uma medida de extrema importância em análise de sobrevivência aplicada à medicina, e tem sido objeto de muito debate desde a década de 70 (Prentice e Kalbfleisch, 1978). Na análise de sobrevivência padrão, em geral, essa estimativa é definida como complemento da função de sobrevivência que é obtida através do método clássico não paramétrico de Kaplan-Meier, para revisão geral e referências vide Pintilie(2006), Gooley et al.(1999) e Putter et al.(2007).

Em algumas situações, o uso do método clássico torna-se inadequado quando os riscos competitivos estão presentes. Por risco competitivo entende-se um evento cuja ocorrência impede a observação ou, fundamentalmente, altera a probabilidade de ocorrência de um outro evento de interesse (Gooley et al., 1999). Nesses casos, o método clássico considera apenas uma causa responsável pela falha do indivíduo, isto é, falha por evento de interesse, e trata como censuras não-informativas todas as outras causas de falhas, incluindo eventos competitivos. Em outras palavras, assume-se que o valor do tempo de sobrevivência para um determinado indivíduo é independente do valor do tempo até a ocorrência de falha por um evento competitivo.

Vários estudos demonstraram que, na presença de riscos competitivos, a suposição da independência entre os eventos provoca um viés do estimador de Kaplan-Meier, vide, por exemplo, Gooley et al.(1999), Kim (2007), Gijbels (2010) e Berry et al. (2010). Na prática, apenas a censura administrativa pode ser razoavelmente considerada independente do evento de interesse, uma vez que o que acontece após esse período não é de conhecimento (e nem de interesse) do pesquisador. No mais, qualquer das outras causas de censura fornece alguma informação relevante sobre evento de interesse em estudo. Para uma discussão detalhada sobre a suposição da independência entre o evento de interesse e censura, vide Putter et al.(2007).

Ao censurar um indivíduo que falha por causa de um evento competitivo, presume-se que este poderá vir a falhar por evento de interesse a qualquer momento no futuro, ou seja, a cada instante, o risco de falha por evento de interesse é o mesmo tanto para um indivíduo que sofreu um certo evento competitivo quanto para aquele que ainda não sofreu qualquer tipo de evento e que está sob acompanhamento neste mesmo instante (Gooley et al., 1999). É claro que, do ponto de vista clínico, entretanto, isto não reflete a realidade. Por exemplo, se um paciente morre por causa do

câncer do pulmão, ele não morreria (novamente) devido à parada cardíaca. Portanto, ao ignorar os riscos competitivos, o método de Kaplan-Meier superestima a probabilidade de falha por um evento de interesse (ou subestima a probabilidade de sobrevivência correspondente) (Berry et al. 2010).

Para minimizar este tipo de viés, Kalbfleisch e Prentice (1980), propuseram uma estimativa alternativa, a *função de incidência acumulada* (FIA), que calcula probabilidade de falha por um evento de interesse tendo em conta a presença de riscos competitivos.

A modelagem dos efeitos das covariáveis sobre a ocorrência de um determinado evento de interesse é também uma questão importante em muitos estudos clínicos. Na análise de sobrevivência padrão, geralmente, este problema é resolvido usando o modelo tradicional de regressão de riscos proporcionais de Cox (Cox, 1972). No entanto, na presença de riscos competitivos este método também pode superestimar o risco de ocorrência do evento de interesse porque trata os demais eventos (incluindo eventos competitivos) como censuras não informativas (Prentice et al., 1978). Como solução para este problema, Fine e Gray (1999) desenvolveram um modelo regressão o qual estima os efeitos das covariáveis diretamente sobre a FIA, tendo em conta a presença de riscos competitivos. Este método é baseado nas suposições do modelo de riscos proporcionais de Cox.

O objetivo principal deste trabalho é aplicar a teoria moderna dos modelos de riscos competitivos para estimar a probabilidade de um certo desfecho acontecer antes de um determinado tempo e identificar fatores de prognóstico importantes para ocorrência desse mesmo desfecho, ou seja, avaliar os efeitos das covariáveis na incidência de cada um dos três principais eventos competitivos relacionados aos desfechos observados em pacientes com hipertensão arterial em tratamento no Hospital Universitário Clementino Fraga Filho da Universidade Federal do Rio de Janeiro (HUCFF/UFRJ).

## 2. Conceitos básicos e notações

Os dados de sobrevivência medem o tempo até a ocorrência de um determinado evento de interesse como, por exemplo, morte, a recidiva ou desenvolvimento de uma doença. A distribuição dos tempos de sobrevivência é, geralmente, caracterizada por três funções: função de sobrevivência, função de densidade de probabilidade e função de risco.

Seja  $T$  uma variável aleatória que representa o tempo de sobrevivência de um indivíduo por um determinado evento de interesse. A função de sobrevivência no instante  $t$  é definida como a probabilidade de um indivíduo sobreviver além do tempo  $t$ , ou por no mínimo um tempo igual a  $t$ :

$$S(t) = P(T > t).$$

Se  $F(t)$  representa a função de distribuição da variável  $T$ , então a função de sobrevivência pode ser expressa como

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t).$$

Observe que  $S(t)$  é uma função contínua em  $t$  e monótona decrescente, tal que  $S(0) = 1$  e  $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$ .

A função de densidade de probabilidade (fdp),  $f(t)$ , é definida como probabilidade de um indivíduo sofrer um evento de interesse em um intervalo instantâneo de tempo. Ela pode ser expressa em termos da função de sobrevivência ou da função de distribuição através da diferenciação:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}.$$

A função de risco representa a taxa instantânea de ocorrência do evento de interesse no tempo  $T = t$  condicionada à sobrevivência do indivíduo até o tempo  $t$ . Matematicamente, a função de risco é a probabilidade condicional de um indivíduo sofrer o evento de interesse dentro de um intervalo de tempo pequeno  $(t, t + \Delta t)$ , sabendo que o mesmo sobreviveu até o início deste intervalo:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = -\frac{d \log S(t)}{dt}. \quad (1)$$

A função de risco acumulado do tempo de sobrevivência  $T$ , denotada por  $\Lambda(t)$ , é definida por

$$\Lambda(t) = \int_0^t \lambda(u) d(u) \quad (2)$$

e, como consequência,

$$S(t) = \exp\{-\Lambda(t)\}. \quad (3)$$

Assim, quando  $t = 0$ ,  $S(t) = 1$  e  $\Lambda(t) = 0$ , e quando  $t \rightarrow \infty$ ,  $S(t) = 0$  e  $\Lambda(t) = \infty$ , isto é, a função de risco acumulado pode assumir qualquer valor de zero à infinito.

Um método bastante utilizado em estudos clínicos para estimar a função de sobrevivência é o estimador não-paramétrico de Kaplan-Meier, por vezes chamado de estimador limite-produto, proposto por Kaplan e Meier (1958). Na ausência de censuras, ou seja, se todos os indivíduos são acompanhados até a observação do evento de interesse, o estimador limite-produto é idêntico à estimativa da função de sobrevivência empírica, definida como proporção de indivíduos que sobreviveram além do tempo  $t$ .

Sejam  $t_1 < t_2 < \dots < t_m$  os  $m$  diferentes tempos em que ocorreram os eventos (ou censuras) em uma amostra com  $n$  indivíduos. Caso ocorra empate entre os tempos da censura e do evento de interesse, assume-se, por convenção, que o evento precede a censura (Gijbels, 2010). Seja  $R(t_j)$  um conjunto de indivíduos em risco (indivíduos em acompanhamento e que ainda não sofreram qualquer evento) no tempo  $t_j$  e  $n_j$  o número de indivíduos no grupo de risco no tempo  $t_j$ . Para cada  $t_j$ , denota-se  $d_j$  o número total de eventos ocorridos precisamente em  $t_j$ . O estimador de Kaplan-Meier assume que  $S(t)$  é uma função discreta com probabilidade maior que zero somente nos tempos em que ocorrem os eventos  $t_j$ . Assim, para os  $m$  tempos em que ocorrem os eventos, a probabilidade de sobrevivência de um indivíduo em  $t_j$  será estimada pelo número de sobreviventes no tempo  $t_j$  ( $n_j - d_j$ ) sobre os que estavam em risco naquele momento ( $n_j$ ). Assumindo que os eventos são independentes, o estimador de Kaplan-Meier da função de sobrevivência é obtido empiricamente pelo produto das probabilidades de sobrevivência a cada tempo  $t_j \leq t$ :

$$\hat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j}\right).$$

O efeito das covariáveis para ocorrência de um desfecho é frequentemente estimado utilizando o modelo de riscos proporcionais de Cox. Na sua forma mais simples, o risco para um indivíduo com valores de covariáveis  $\mathbf{x} = (x_1, x_2, \dots, x_p)$  é assumido como sendo

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}^\top \mathbf{x}),$$

onde  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$  é um vetor dos coeficientes de regressão e  $\lambda_0(t)$  é uma função de risco básica (ou basal) não especificada quando o vetor das covariáveis  $\mathbf{x} = 0$ . A bem da simplicidade, de agora em diante usaremos  $\boldsymbol{\beta}^\top \mathbf{x}$  como uma notação para  $\sum_{k=1}^p \beta_k \times x_k$ . Supondo que em uma amostra de  $n$  indivíduos, existam  $m < n$  falhas distintas nos tempos  $t_1 < t_2 < \dots < t_m$ , o vetor do parâmetro  $\boldsymbol{\beta}$  é encontrado através da maximização da função de verossimilhança parcial. Esta função é o produto, em todos os tempos de ocorrência dos eventos, da razão entre o risco de o indivíduo  $j$  apresentar o evento no tempo  $t_j$  e a soma dos riscos de ocorrência do evento para todos os indivíduos em risco no mesmo período:

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{j=1}^m \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x}_j)}{\sum_{l \in R(t_j)} \exp(\boldsymbol{\beta}^\top \mathbf{x}_l)}. \quad (4)$$

### 3. Modelos de riscos competitivos

O modelo de riscos competitivos envolve situações nas quais um indivíduo pode sofrer mais do que um tipo de evento, ou seja, existem  $K$  diferentes causas responsáveis pela falha do indivíduo, porém somente a primeira delas pode ser observada.

Os dados observados para cada indivíduo num modelo de riscos competitivos são representados como um par de variáveis aleatórias,  $(T, C)$ . A variável  $C$  assume valor zero se a observação do indivíduo é censurada, caso contrário,  $C = k$ , onde  $k$  é o tipo da primeira causa de falha observada ( $k = 1, 2, \dots, K$ ). Se  $C = k$  então  $T$  corresponde ao tempo até a ocorrência da falha por causa  $k$ ; caso contrário,  $T$  é o tempo até a censura.

O conceito fundamental em modelos de riscos competitivos é a função de *risco da causa específica*, dada por

$$\lambda_k(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, C = k | T \geq t)}{\Delta t}, k = 1, 2, \dots, K. \quad (5)$$

Ela representa a taxa instantânea da ocorrência da  $k$ -ésima falha no tempo  $t$ , condicionada a sobrevivência até o tempo  $t$ , na presença de todas as outras causas de falha possíveis.

Qualquer quantidade que pode ser determinada exclusivamente pela função de risco da causa específica (Por exemplo, equações 2 e 3) é estimável (Putter et al., 2007). Portanto, o risco acumulado da causa específica é definido por

$$\Lambda_k(t) = \int_0^t \lambda_k(u) du,$$

e

$$S_k(t) = \exp\{-\Lambda_k(t)\}. \quad (6)$$

Note-se que, embora  $S_k(t)$  possa ser estimada, ela não deve ser interpretada como uma função marginal sobrevivência para causa  $k$  (Pintilie, 2006, Putter et al., 2007, Lim et al., 2010). Só terá esta interpretação numa “situação ideal” em que as outras causas de falhas não estão presentes, ou seja, se as distribuições dos tempos para os eventos competitivos forem independentes.

A função de sobrevivência (3) pode ser definida como

$$S(t) = \exp\left\{-\sum_{k=1}^K \Lambda_k(t)\right\}.$$

Esta sim, tem uma interpretação prática; representa a probabilidade de um indivíduo não sofrer falha por qualquer causa até o tempo  $t$ .

A *função de incidência acumulada* (FIA) para causa  $k$  (Kalbfleisch e Prentice, 1980) é definida como probabilidade de ocorrência de falha por causa  $k$  num tempo menor ou igual a  $t$ ,

$$F_k(t) = P(T \leq t, C = k), k = 1, \dots, K. \quad (7)$$

Ela pode ser expressa em termos dos riscos para todas as causas de falha,

$$F_k(t) = \int_0^t \lambda_k(u) S(u) du = \int_0^t \lambda_k(u) \exp\left\{-\sum_{k=1}^K \Lambda_k(u)\right\} du. \quad (8)$$

A função de distribuição global é igual a soma das FIA's, para todos as causas de falha:

$$F(t) = P(T \leq t) = \sum_{k=1}^K P(T \leq t, C = k) = \sum_{k=1}^K F_k(t).$$

Observe que a FIA para causa  $k$  não é exatamente uma função de distribuição, uma vez que  $\lim_{t \rightarrow \infty} F_k(t) = P(C = k) \neq 1$  (se houver pelo menos duas causas de falha). Por essa razão ela também é conhecida como *função de subdistribuição*.

Como os eventos por outras causas de falha (diferentes de  $k$ ) são tratados como censuras não informativas, a estimativa de Kaplan-Meier de probabilidade de falha por causa  $k$  num tempo menor ou igual a  $t$  é definida como

$$1 - S_k(t) = \int_0^t \lambda_k(u) S_k(u) du.$$

A diferença entre esta equação e a função de incidência acumulada  $F_k(t)$  dada pela equação 8 é que  $S(t)$  é substituída por  $S_k(t)$ . Observe que  $S_k(t)$  representa a probabilidade de um indivíduo sobreviver livre somente por causa  $k$  no tempo  $t$ , ignorando as demais causas, enquanto que  $S(t)$  indica a sobrevivência livre de todas as causas de falha em  $t$ . Portanto,  $S_k(t) \geq S(t)$ , para qualquer

$t$ . A igualdade acontece em  $t$  quando não há eventos competitivos, isto é,  $\sum_{i=1, i \neq k}^K \Lambda_i(t) = 0$ . Dai

resulta que  $F_k(t) \leq 1 - S_k(t)$ . Este resultado evidencia o viés no estimador de Kaplan-Meier quando é usado em situações de riscos competitivos. Ele superestima a probabilidade de falha por evento de interesse.

Agora vejamos mais formalmente como a função de incidência acumulada FIA (7) é definida na presença de censuras. sejam  $0 < t_1 < t_2 < \dots < t_m$  os  $m$  distintos tempos ordenados da ocorrência de qualquer um dos eventos e  $d_{kj}$  o número de indivíduos que sofreram o evento  $k$

no tempo  $t_j$ . O número total de eventos (por qualquer causa) ocorridos no tempo  $t_j$  é  $d_j = \sum_{k=1}^K d_{kj}$ .

Tal como na seção 2, seja  $n_j$  o número de indivíduos no grupo de risco no tempo  $t_j$ .

A FIA pode então ser definida como soma em todos os tempos  $t_j \leq t$  das probabilidades de se observar o evento  $k$  no tempo  $t_j$  entre indivíduos que não sofreram qualquer dos possíveis eventos imediatamente antes deste instante. A probabilidade de estar livre de qualquer evento antes do tempo  $t_j$  é dada por  $\hat{S}(t_{j-1})$ . Logo, a probabilidade conjunta de um indivíduo estar livre de qualquer evento imediatamente antes de  $t_j$  e sofrer o evento  $k$  no tempo  $t_j$  é

$$\hat{F}_k(t) = \sum_{\forall j, t_j \leq t} \hat{\lambda}_k(t_j) \hat{S}(t_{j-1}),$$

onde  $\lambda_k(t_j)$  é o risco da causa específica para o evento  $k$  no tempo  $t_j$ . Como os indivíduos no conjunto de risco ainda não sofreram qualquer tipo de evento (e nem foram censurados) até o tempo  $t_{j-1}$ , o risco pode ser, intuitivamente, estimado como  $\frac{d_{kj}}{n_j}$ . Segue que

$$\hat{F}_k(t) = \sum_{\forall j, t_j \leq t} \frac{d_{kj}}{n_j} \hat{S}(t_{j-1}).$$

Assim, o estimador da FIA para um evento do tipo  $k$  depende não só do número de indivíduos que experimentaram este tipo de evento, mas também do número de indivíduos que não tenham sofrido qualquer outro tipo de evento até aquele momento.

Na análise de riscos competitivos, muitas vezes os pesquisadores têm interesse em estimar o efeito das covariáveis para uma causa específica de falha. Na abordagem clássica, isto é feito modelando a função do risco da causa específica para o evento  $k$  (5), considerando os outros eventos como censuras não-informativa. Sob o modelo clássico de riscos proporcionais de Cox, o risco da causa específica do evento  $k$  para um indivíduo com vetor de covariáveis  $\mathbf{x} = (x_1, x_2, \dots, x_p)$  é definido como

$$\lambda_k(t|\mathbf{x}) = \lambda_{k,0}(t) \exp(\beta_k^\top \mathbf{x}), \quad (9)$$

onde  $\lambda_{k,0}(t)$  é o risco basal da causa específica do evento  $k$ , e o vetor  $\beta_k$  representa os efeitos das covariáveis na ocorrência do evento  $k$ . Pelas mesmas razões discutidas anteriormente, esta abordagem não é adequada para estimar o efeito de covariáveis quando os eventos competitivos estão presentes.

Como alternativa, Fine e Gray (1999) introduziram uma forma de estimar o efeito de covariáveis modelando diretamente as FIA's, sem supor independência entre os tempos dos eventos. Em analogia com a relação 1 entre o risco e a função de sobrevivência, eles definiram uma função de *risco de subdistribuição* como

$$\begin{aligned}\bar{\lambda}_k(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{P}\{t \leq T < t + \Delta t, C = k | T \geq t \cup (T \leq t \cap C \neq k)\} \\ &= -\frac{d}{dt} \log\{1 - F_k(t)\}.\end{aligned}\quad (10)$$

Fine e Gray impuseram a suposição de riscos proporcionais sobre os riscos de subdistribuição:

$$\bar{\lambda}_k(t|\mathbf{x}) = \bar{\lambda}_{k,0}(t) \exp(\beta_k^\top \mathbf{x}). \quad (11)$$

A forma da função de verossimilhança parcial para os coeficientes da (11) é similar àquela usada no modelo de riscos proporcionais de Cox (4), porém diferem no conceito do conjunto de risco para cada evento. Para o risco da causa específica,  $\lambda_k(t|\mathbf{x})$ , o conjunto de risco diminui a cada instante em que falha por outra causa ( $\neq k$ ) ocorre. Para  $\bar{\lambda}_k(t|\mathbf{x})$ ,

$$\mathcal{L}(\beta_k) = \prod_{k=1}^K \prod_{i=1}^{d_j} \frac{\exp(\beta_k^\top \mathbf{x}_{(k)i})}{\sum_{l \in R(t_{j(i)})} w_{il} \exp(\beta_j^\top \mathbf{x}_l)}. \quad (12)$$

o conjunto de risco,  $R(t_{j(i)})$  inclui, além de todos os indivíduos que não sofreram qualquer evento em  $t$ , todos os que falharam por outros eventos competitivos em  $t$ , e que até a censura administrativa permanecerão nesse conjunto. Entretanto, não faz sentido que esses indivíduos contribuam para o conjunto de risco da mesma forma que os demais. Por isso se atribui a cada indivíduo  $l$  que sofreu qualquer evento competitivo um peso,  $w_{jl}$ , que muda a cada  $t_j$  no qual ocorre o evento de interesse ( $k$ ).

Assim, indivíduos que sofreram os evento competitivos antes da ocorrência de qualquer evento de interesse permanecem no grupo de risco, mas não contribuem integralmente na função de verossimilhança parcial. O peso é definido como

$$w_{jl} = \begin{cases} 1, & \text{se } l \text{ não sofreu evento ou censura;} \\ \frac{\hat{G}(t_j)}{\hat{G}(\min(t_j, t_l))}, & \text{se } l \text{ sofreu evento competitivo;} \\ 0, & \text{quando o indivíduo é retirado do estudo.} \end{cases} \quad (13)$$

$\hat{G}$  é a estimativa de Kaplan-Meier da função de sobrevivência da distribuição de censura, o que significa que, no cálculo de  $G(\cdot)$ , censura vira evento e o evento (qualquer que seja) vira censura (Pintilie, 2006).

#### 4. Aplicação numérica

Nesta seção apresentamos os resultados numéricos baseados na metodologia de análise de sobrevivência com riscos competitivos discutida na seção anterior. Para isso, utilizamos a *função de incidência acumulada* para calcular a probabilidade de um paciente sofrer um certo desfecho antes de um determinado tempo, e o modelo de regressão sobre *risco de subdistribuição* para estimar os efeitos das covariáveis na ocorrência desse desfecho, tendo em conta a presença de eventos competitivos. Os dados usados para aplicação desses métodos são descritos a seguir.

A análise estatística dos dados, incluindo a construção dos gráficos referentes às estimativas de probabilidade de falha por um evento, foi feita utilizando exclusivamente *software* estatístico R-versão 3.1.3, disponível grátis na página da internet.

#### 4.1. Descrição dos dados

O banco de dados utilizado para mostrar a aplicação de modelos para eventos competitivos contém variáveis referentes a uma coorte de 572 pacientes com hipertensão arterial, acompanhados pelo Programa de Hipertensão Arterial do Hospital Universitário Clementino Fraga Filho da UFRJ, no período entre Janeiro de 1999 e Dezembro de 2007. O tempo de observação para cada paciente foi o número de meses a partir da data do primeiro exame clínico até a data da última visita clínica ou data do primeiro evento. O acompanhamento dessa coorte permitiu identificar os fatores de prognóstico importantes para a ocorrência de três eventos (desfechos) distintos para os quais pacientes estão em risco: evento cardiovascular não fatal, óbito por causa cardiovascular e óbito por outra causa. As variáveis incluídas no banco estão listadas a seguir.

1. PRONT – Prontuário do paciente.
2. TEMPO – tempo (em meses) desde a data do primeiro exame até a ocorrência de um dos possíveis eventos ou censura.
3. MOTIVO – causa da falha ou evento competitivo ocorrido: evento cardiovascular não fatal, óbito por outra causa e óbito cardiovascular.
4. STATUS – variável indicadora do evento: 0 = censura, 1 = evento cardiovascular não fatal, 2 = óbito por outra causa, 3 = óbito por causa cardiovascular.
5. IDADE – idade (em anos) do paciente ao iniciar o acompanhamento.
6. SEXO – sexo do paciente: 1 = masculino, 2 = feminino.
7. DIABETES – diagnóstico da diabetes: 0 = sem diabetes, 1 = com diabetes.
8. DISLIP – diagnóstico da dislipidemia: 0 = sem dislipidemia, 2 = com dislipidemia.
9. CREATS – creatinina sérica.
10. HARV3 – controle da pressão arterial: 0 = controlada, 1 = não controlada.
11. PAS24H – pressão arterial sistólica de 24 horas.
12. PAD24H – pressão arterial diastólica de 24 horas.

Para este banco de dados, o número de ocorrência dos eventos para cada um dos desfechos ficou distribuído da seguinte forma: Evento cardiovascular não fatal=75, Óbito por outra causa=23, Óbito por causa cardiovascular=44 e Censura = 430.

#### 4.2. Estimando a probabilidade de falha por um evento na presença de riscos competitivos

A figura 1 apresenta as curvas referentes às estimativas de probabilidade de falha por cada um dos desfechos (competitivos entre si) ao longo do tempo. No gráfico (a) estão sobrepostas as curvas da FIA para cada um dos três desfechos definidos. Observe que há maior risco de o paciente vir a sofrer primeiro um evento cardiovascular não fatal antes da sua morte.

Os gráficos (b), (c) e (d) da mesma figura ilustram o que foi amplamente discutido na seção 3, sobre as limitações do estimador clássico de Kaplan-Meier quando utilizado para análise de sobrevivência com riscos competitivos. Podemos observar através dessas curvas que, ao longo do período de acompanhamento, a probabilidade de falha por um evento de interesse estimada utilizando o método tradicional de Kaplan-Meier é maior ou igual à estimativa de probabilidade baseada na FIA. Observe também que a magnitude do viés aumenta à medida que o tempo de acompanhamento dos pacientes aumenta, ou seja, quando a ocorrência de eventos competitivos é maior. A diferença entre as duas estimativas é pequena nos primeiros tempos de acompanhamento, quando a frequência dos eventos competitivos é baixa.



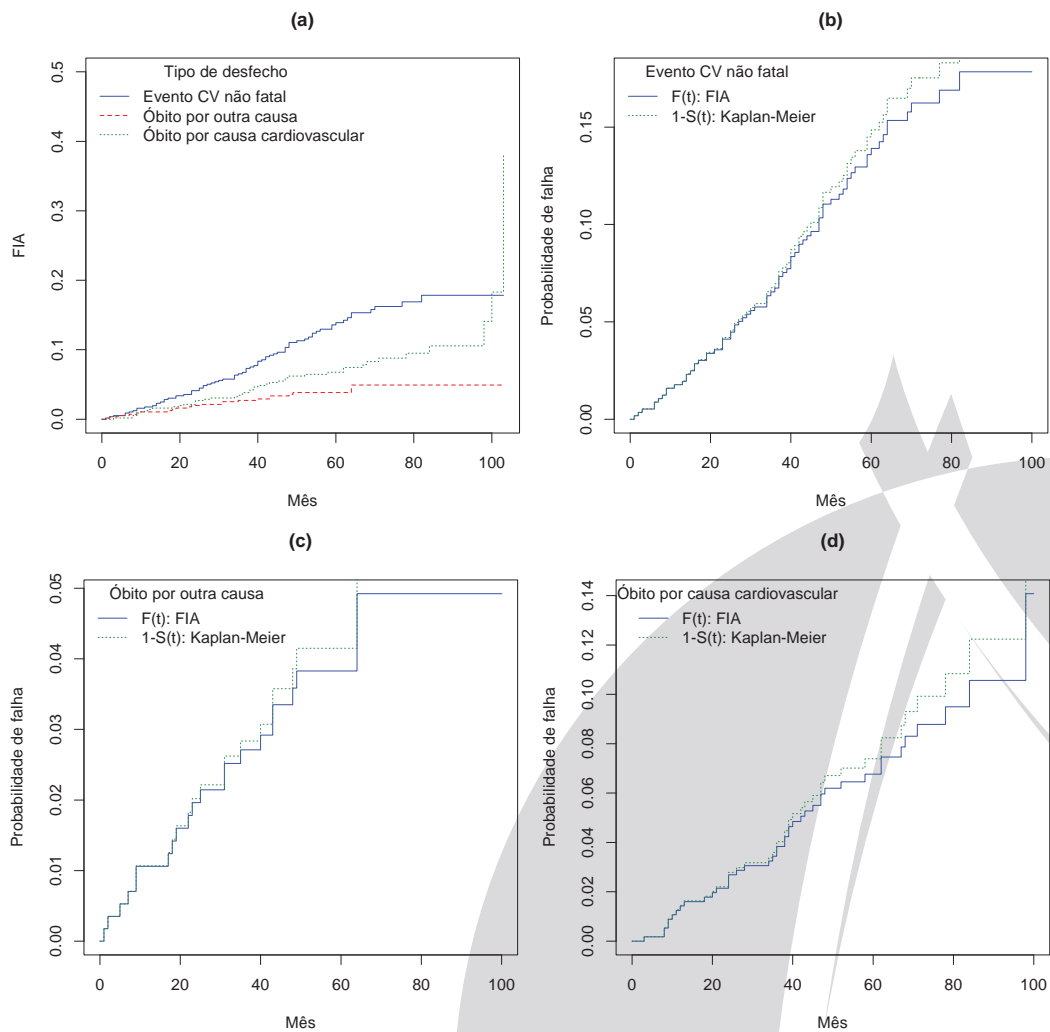


Figura 1: Estimativas de probabilidade de falha por cada evento competitivo. No gráfico (a) estão as curvas de probabilidade de falha por cada um dos desfechos, estimadas através da FIA, e nos gráficos (b), (c) e (d) estão desenhadas as mesmas curvas do (gráfico (a)) e, sobrepostas a essas, as curvas de probabilidade de falha utilizando o método clássico de Kaplan-Meier

### 4.3. Modelando os efeitos das covariáveis para ocorrência dos eventos competitivos

A estimação dos efeitos das covariáveis sobre a ocorrência dos desfechos estudados foi feita com base na função de risco de subdistribuição para cada evento (10) tendo em conta a presença dos eventos competitivos (Fine e Gray, 1999).

Para isso, foi utilizada a função **crprep** da biblioteca **mstate**, que calcula os pesos como descrito na equação 13. A função **crprep** permite reorganizar o banco de dado original como um banco com variáveis mudando no tempo, já que o peso de cada indivíduo vai mudando à medida que ocorrem eventos competitivos. Além disso, a função calcula os pesos considerando cada possível evento como de interesse, sendo os demais considerados competitivos. Ou seja, o novo banco fica com um grande número de linhas, onde cada paciente terá tantas linhas quantos mudanças de peso, repetindo o mesmo procedimento para cada evento de interesse. Para mais detalhes sobre a função **crprep** para análise de riscos competitivos, vide Putter et al. (2007) e Carvalho et al. (2011). Os resultados da modelagem estão resumidos na tabela abaixo.

Tabela 1: Modelos de regressão sobre o risco de subdistribuição para cada um dos desfechos

Desfecho	covariável	coef	exp(coef)	se(coef)	z	p
Evento CV não fatal	SEXO2	0.2536	1.289	0.289	0.879	0.3795
	IDADE	0.0008	1.001	0.014	0.057	0.9548
	DIABETES1	0.3827	1.466	0.245	1.561	0.1185
	DISLIP1	0.4364	1.547	0.431	1.012	0.3115
	CREATS	0.0065	1.007	0.002	3.790	0.0001**
	HARV31	0.1815	1.199	0.396	0.462	0.6438
	PAS24H	0.0104	1.010	0.010	1.010	0.3125
	SPAD24H	0.0071	1.007	0.017	0.418	0.6762
Óbito por outra causa	SEXO2	-0.4613	0.6305	0.4730	-0.975	0.3294
	IDADE	0.0335	1.0341	0.0242	1.385	0.1662
	DIABETES1	1.1259	3.0829	0.4611	2.442	0.0146**
	DISLIP1	1.3047	3.6866	1.0329	1.263	0.2065
	CREATS	0.0024	1.0024	0.0032	0.744	0.4569
	HARV31	1.1924	3.2949	0.7669	1.555	0.1200
	PAS24H	-0.0104	0.9897	0.0193	-0.538	0.5906
	PAD24H	0.0012	1.0012	0.0303	0.039	0.9692
Óbito por causa cardiovascular	SEXO2	-0.5976	0.5501	0.3331	-1.794	0.0728*
	IDADE	0.0303	1.0308	0.0176	1.722	0.0850*
	DIABETES1	0.1019	1.1073	0.3170	0.321	0.7479
	DISLIP1	-0.1883	0.8283	0.4350	-0.433	0.6650
	CREATS	0.0025	1.0025	0.0027	0.932	0.3512
	HARV31	-0.0279	0.9724	0.4988	-0.056	0.9553
	PAS24H	0.0241	1.0244	0.0129	1.862	0.0625*
	PAD24H	-0.0191	0.9811	0.0200	-0.953	0.3404

Significância ao nível de: \*10%, \*\* 5%

Observa-se que a Creatinina influencia significativamente no risco de o paciente vir falha por evento cardiovascular não fatal. A cada aumento unitário desta variável, o risco de um paciente vir a sofrer um evento cardiovascular não fatal aumenta em 0.7%. Por outro lado, o diabetes mostrou-se o fator de prognóstico mais importante para ocorrência de óbito por outras causas. Um paciente com diabetes tem risco 3.0829 vezes (208,29%) maior de vir a óbito por outra causa a

cada unidade de tempo do que um paciente sem diabetes. Finalmente, a um nível de 10% de significância, observamos que o sexo do paciente, a idade e pressão arterial sistólica durante 24 horas têm efeitos significativos na observação de óbito cardiovascular. O risco de um paciente de sexo feminino vir a óbito cardiovascular é 45% inferior do que o risco de um paciente de sexo masculino. A pressão arterial sistólica durante 24 horas aumenta o risco (2,4% maior) de o paciente vir ao óbito cardiovascular.

Assim como em modelos clássicos, os pressupostos mais importantes para que o método de Fine e Gray seja adequadamente utilizado na estimação dos efeitos das covariáveis para os eventos competitivos são a proporcionalidade dos riscos e a linearidade em variáveis contínuas (Pintilie, 2006). Assim, a análise dos resíduos de Schoenfeld e os testes de correlação linear do resíduo com o tempo foram efetuados neste trabalho para a validação do modelo proposto. Todos os testes revelaram a proporcionalidade dos riscos em todas as covariáveis do estudo sobre cada um dos desfechos considerados.

## 5. Conclusões

Neste trabalho apresentamos a metodologia para análise de dados de sobrevivência na presença de eventos competitivos. Mostramos que, diante desta situação, o uso de métodos tradicionais como de Kaplan-Meier para calcular a probabilidade de um indivíduo sofrer um evento antes de um determinado tempo ou de modelo de regressão de riscos proporcionais de Cox para estimar o efeito das covariáveis para a ocorrência dos eventos se torna inapropriado, uma vez que estes fornecem estimativas viciadas, não interpretáveis sob ponto de vista clínico. A questão principal que resulta no viés das estimativas é que estes métodos consideram apenas falhas por evento de interesse e tratam como censura não informativa todas as outras falhas causadas por eventos competitivos. Através da aplicação numérica, observamos, por exemplo, que o estimador de Kaplan-Meier superestima a probabilidade de falha por um evento de interesse. A magnitude deste viés tende a aumentar à medida que o tempo de acompanhamento vai crescendo e a ocorrência dos eventos competitivos vai aumentando.

A regressão sobre a função de risco de subdistribuição permitiu-nos estimar os efeitos das covariáveis para cada um dos desfechos considerando a presença dos eventos competitivos. Um resultado importante desta análise é que o diagnóstico de diabetes revela um impacto significativo na progressão da morte por causa não cardiovascular. Por outro lado, a pressão arterial sistólica aumenta o risco de o paciente vir ao óbito cardiovascular. O modelo proposto mostrou-se adequado aos dados disponíveis na medida em que satisfaz as suposições fundamentais da proporcionalidade dos riscos e linearidades em variáveis.

Como trabalho futuro, propomos a análise dos dados de pacientes com hipertensão arterial como uma estrutura de riscos semi-competitivos. No caso, o objetivo seria estudar o tempo até a ocorrência de dois eventos principais, considerando o desfecho “evento cardiovascular não fatal” como um evento intermediário e óbito (seja qual for a causa) como evento terminal. Assim, teríamos um par de variáveis observadas no mesmo paciente e seria razoável considerar a existência de uma estrutura de dependência entre essas duas variáveis, identificando uma distribuição conjunta dos tempos de falhas por meio de uma função cópula.

## Agradecimentos

Os autores são gratos pelo apoio financeiro recebido pelo CNPq-Conselho Nacional de Desenvolvimento Científico e Tecnológico e ANP-Agência Nacional do Petróleo, Gas Natural e Biocombustíveis. Ao professor Gil F. Salles do Programa de Hipertensão Arterial do HUCFF/UFRJ pelo fornecimento do banco de dados e importantes sugestões.

## Referências

- Berry, S.D., Ngo, L., Samelson e E.J., Kiel, D.P.** (2010), Competing Risk of Death: An Important Consideration in Studies of Older Adults. *J American Geriatrics Society*, 58, 4, 783-787.
- Carvalho, M.S, Andreozzi, V.L., Codeço, C.T., et al.** *Análise de Sobrevivência: teoria e aplicações em saúde*. 2.ed, Fiocruz, Rio de Janeiro, 2011.
- Collett, D.** *Modelling Survival Data in Medical Research*. Chapman Hall, 1994.
- Cox, D.R.** (1972), Regression models and life-tables. *Journal of the Royal Statistical Society*, 34, 2, 187-220.
- Fine, J.P.e Gray, R.J.** (1999), A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94, 446, 496-509.
- Gijbels, I.** (2010), Censored data. *Computational Statistics*, 2, 2, 178-188.
- Gooley, T.A., Leisenring, W., Crowley, J. e Storer, B.E.** (1999), Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*, 18, 6, 695-706.
- Kalbfleisch, J. D. e Prentice, R. L.** *The Statistical Analysis of Failure Time Data*. Wiley, New York, 1980.
- Kaplan, E. L. e Meier, P.** (1958), Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457-481.
- Kim, H.T.** (2007), Cumulative Incidence in Competing Risks and Competing Risks Regression Analysis. *Clinical Cancer Research*, 13, 2, 559-565.
- Lim, H.J., Zhang, X., Dyck, R. e Osgood, N.** (2010), Methods of Competing Risks Analysis of End-Stage Renal Disease and Mortality among People with Diabetes. *BioMed Central*, 10, 97, 1-26.
- Pintilie, M.** *Competing risks: A practical perspective*. 1.ed. Wiley, New York, 2006.
- Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., et al.** (1978), The analysis of failure times in the presence of competing risks. *Biometrics*, 34, 4, 541-554.
- Putter, H., Fiocco, M. e Geskus, R. B.** (2007), Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, 26, 11, 2389-2430.