

# EFEITO DA CENSURA SOBRE A FUNÇÃO DE INCIDÊNCIA CUMULATIVA NA PRESENÇA DE RISCOS COMPETITIVOS NA ANÁLISE DE SOBREVIVÊNCIA

Flávia Sílvia Corrêa TOMAZ<sup>1</sup>  
Sebastião MARTINS FILHO<sup>2</sup>

- RESUMO: O objetivo deste trabalho foi avaliar o efeito da censura sobre a função de incidência cumulativa na presença de riscos competitivos. Para tanto, conjuntos de tempos de sobrevivência com riscos competitivos considerando diferentes níveis de censura (0%, 10%, 20%, 30%, 40% e 50%) foram simulados. A simulação foi baseada nos riscos específicos por causa utilizando a distribuição exponencial. Para a simulação dos tempos de censura utilizou-se a distribuição uniforme. Além dos distintos níveis de censura considerou-se diferentes valores para os riscos específicos por causa para os dois eventos competitivos simulados. Comparações entre as curvas de incidência do conjunto de dados sem censura e o conjunto com diferentes porcentagens de censura foram realizadas. Confrontou-se, ainda, as estimativas da função de incidência cumulativa com base na teoria de riscos competitivos com àquela obtida pelo uso do estimador de Kaplan-Meier. Os resultados indicam que as porcentagens de censura bem como relação entre os riscos específicos por causa afetam a função de incidência cumulativa.
- PALAVRAS-CHAVE: Subdistribuição; tempo para eventos; eventos competitivos.

## 1 Introdução

O estudo de tempos de sobrevivência em que a ocorrência de falha decorre de várias causas distintas é denominado na análise de sobrevivência como riscos competitivos (AALEN *et al.*, 2008, CARVALHO *et al.*, 2011; HOSMER *et al.*, 2008).

Os riscos competitivos são uma extensão da análise de sobrevivência clássica na qual um único evento (tipo de falha) é considerado, e é uma metodologia que tem se destacado, especialmente na área médica. Não raro, são encontrados estudos que empregam a avaliação de eventos competitivos em pesquisas médicas (BEYERSMANN e SCHEIKE, 2014; KLEIN e BAJORUNAITE, 2004; GOOLEY *et al.*, 1999), em especial em estudos oncológicos (GOOLEY *et al.*, 1999).

Neste cenário, Zhang *et al.* (2013), usaram a estrutura de riscos competitivos para analisar o efeito de fatores de risco para a mortalidade por câncer de mama e para a mortalidade decorrente de outras causas. Logan *et al.* (2006) modelaram dados de transplante de células tronco. Gillam *et al.* (2011) avaliaram o risco de revisão de

---

<sup>1</sup> Universidade Federal de Ouro Preto - UFOP, Instituto de Ciências Sociais Aplicadas, Departamento de Ciências Administrativas, CEP: 354200-00, Mariana, MG, Brasil. E-mail: [flaviastomaz@yahoo.com.br](mailto:flaviastomaz@yahoo.com.br)

<sup>2</sup> Universidade Federal de Viçosa - UFV, Centro de Ciências Exatas e Tecnológicas, Departamento de Estatística, CEP: 36570-977, Viçosa, MG, Brasil. E-mail: [sebastiaom.ufv@gmail.com](mailto:sebastiaom.ufv@gmail.com)

artroplastia na presença do risco competitivo de morte. Dianatkah *et al.* (2014), estudaram os fatores de risco para doenças cardiovasculares e para morte não relacionada a doença cardiovascular.

Distintos níveis de censura podem ser observados nos trabalhos que aplicavam a metodologia de riscos competitivos para análise dos dados. Ferraz e Moreira-Filho (2017) relatam um nível de censura de, aproximadamente, 48%. Assane *et al.* (2015) descrevem um nível de censura de 75%. No estudo de Dianatkah *et al.* (2014) a porcentagem de censura foi 93,37%. Níveis menores, 43% e 12%, são descritos, respectivamente, por Gillam *et al.* (2011) e Berry *et al.* (2010).

Nestes e em vários trabalhos que fazem uso da abordagem de riscos competitivos um ponto de fundamental importância é a estimativa da função de incidência cumulativa. Essa função permite quantificar a probabilidade de um indivíduo experimentar um evento pertencente a um conjunto de causas de falha distintas (GESKUS, 2015).

Embora existam métodos adequados para se obter a função de incidência cumulativa que estão fundamentados na teoria dos riscos competitivos, com frequência esta função é obtida por meio do complemento do estimador de Kaplan-Meier e interpretada como probabilidade de falha (GOOLEY *et al.*, 1999). Entretanto, diversos estudos têm ressaltado a inadequabilidade deste procedimento (BEUSCART *et al.*, 2012; MARTINEZ-CAMBLOR, 2016; GOOLEY *et al.*, 1999; SHERIF, 2007).

Uma razão para a incorreta obtenção da função de incidência cumulativa é a falta de compreensão da suposição requerida para a interpretabilidade da estimativa obtida por meio do estimador de Kaplan-Meier adicionada a uma falta de entendimento de como a função de incidência cumulativa é calculada (GOOLEY *et al.*, 1999). Ademais, a maneira como o método baseado no estimador de Kaplan-Meier e o método fundamentado na teoria de riscos competitivos lidam com as observações censuradas conduzem a diferentes estimativas ao se empregar as distintas técnicas. Por conseguinte, faz-se necessário o entendimento de como a censura afeta o cálculo da incidência cumulativa (GOOLEY *et al.*, 1999).

Embora existam trabalhos que discutam a inadequabilidade do uso do estimador de Kaplan-Meier e, portanto, do tratamento do evento competitivo como censura não informativa, ainda é limitada as discussões que avaliem, por meio de simulações, como a porcentagem de eventos competitivos influenciam a estimativa da função de incidência cumulativa.

Nesta perspectiva, este artigo tem como objetivo contribuir com as discussões sobre o efeito da censura, bem como o efeito da porcentagem de eventos competitivos, sobre a função de incidência cumulativa.

## **2 Material e métodos**

### **2.1 Riscos competitivos**

Um problema habitual na modelagem do tempo de sobrevivência é a ocorrência de falhas devido a várias causas distintas. Estas diferentes causas de falhas são denominadas eventos competitivos e introduzem os riscos competitivos (AALEN *et al.*, 2008).

Riscos competitivos ou, generalizando, múltiplos modos de falha, refere-se à definição em que o acompanhamento pode ser finalizado devido a mais de um evento (HOSMER *et al.*, 2008). Embora existam distintas causas de falha para os eventos ditos competitivos é

possível observar o tempo até a ocorrência do primeiro evento, que impede que os outros aconteçam (CARVALHO *et al.*, 2011) ou alteram fundamentalmente a probabilidade de ocorrência destes outros eventos (GOOLEY *et al.*, 1999). Assim na estrutura de riscos competitivos é registrado o tempo e o tipo do evento que ocorreu primeiramente.

Na análise de sobrevivência com riscos competitivos duas quantidades de grande importância são a função de incidência cumulativa e a função risco específico por causa.

A função de incidência cumulativa permite quantificar a probabilidade de um indivíduo experimentar um evento pertencente a um conjunto de causas distintas.

O risco específico por causa tem definição aproximado pelo conceito padrão, isto é, o conceito utilizado na análise de sobrevivência clássica, representando a parcela de indivíduos que experimentarão o evento de interesse em algum tempo dentre aqueles que estão em risco. Este risco só terá interpretação quando causas diferentes da considerada podem ser eliminadas.

### 2.1.1 Notação básica

Sejam  $T$  uma variável aleatória contínua representando o tempo para o evento e  $C$  outra variável aleatória contínua e, independente de  $T$ , que representa o tempo de censura. Considere  $T_i$  e  $C_i$ , respectivamente, o tempo em que o evento ocorre (tempo de falha) e o tempo de censura para o  $i$ -ésimo indivíduo;  $\varepsilon_i \in \{1, \dots, k\}$  a causa de falha. Seja  $\tilde{T}_i = \min(T_i, C_i)$  e  $\Delta_i = I(T_i \leq C_i)$ ,  $I(\cdot)$  função indicadora que assume o valor 1 caso uma falha tenha ocorrido e 0 caso contrário (ERIKSSON *et al.*, 2015; GRAW *et al.*, 2009; KUK e VARADHAN, 2013; LI *et al.*, 2015; SARKER *et al.*, 2015; SCHEIKE e ZHANG, 2011; SUN *et al.*, 2006; ZHOU *et al.*, 2011; ZHOU, 2010). As  $n$  realizações independentes e identicamente distribuídas de  $\{\tilde{T}_i, \Delta_i, \Delta_i \varepsilon_i\}$  para  $i = 1, \dots, n$ , formam uma amostra aleatória.

### 2.1.2 Função de incidência cumulativa

Um estimador da função de incidência cumulativa,  $F_i(t)$ ,  $i = 1, 2, \dots, k$ , definida como a probabilidade conjunta,  $F_i(t) = P(T \leq t, \varepsilon = i)$ , em  $\varepsilon$  que representa a causa de falha.

Para estimar a função de incidência cumulativa considera-se  $t_1 < t_2 < \dots < t_r$ , os  $r$  tempos distintos e ordenados de falha,  $d_{ij}$  o número de indivíduos que experimentaram o evento  $i$  no tempo  $t_j$ ,  $n_j$  é o número de indivíduos ainda sob risco no tempo  $t_j$  e  $\hat{S}(t_{j-1})$  a probabilidade geral de sobrevivência, ou seja, sobrevivência a toda e qualquer causa (CARVALHO *et al.*, 2011). Daí, segue que a probabilidade conjunta de um indivíduo estar livre de um evento no tempo imediatamente anterior a  $t_j$  e experimentar um evento do tipo  $i$  no tempo  $t_j$  é dado por (PINTILIE, 2006):

$$\hat{F}_i(t) = \sum_{\forall t, t_j \leq t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1}), \quad (1)$$

a probabilidade geral de sobrevivência,  $\hat{S}(t_{j-1})$ , é obtida pelo uso do estimador de Kaplan-Meier (1958) dado por:

$$\hat{S}(t) = \prod_{j: t_j < t} \left(1 - \frac{d_j}{n_j}\right) \quad (2)$$

em que  $d_j$  é o número de falhas, independentemente do tipo, em  $t_j$  e  $n_j$  é o número de indivíduos sob risco em  $t_j$ .

### 2.1.3 Função risco específico por causa

A função risco específico por causa para cada tipo de falha  $k$ ,  $\alpha_k(t)$ , descreve o risco instantâneo de falha a partir da causa  $k$ . Assim, essa função representa a probabilidade de um indivíduo falhar devido a causa  $k$  em um pequeno intervalo de tempo, ou seja, representa a taxa de falha instantânea para uma determinada causa (AALEN *et al.*, 2008). Matematicamente representada por:

$$\alpha_k(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{P(t \leq \tilde{T}_i < t + \Delta t, \varepsilon = k | \tilde{T}_i \geq t)}{\Delta t} \right\}. \quad (3)$$

## 2.2 Simulação de dados

Os tempos de sobrevivência foram gerados conforme sugerido por Beyersmann *et al.* (2012) e Beyersmann *et al.* (2009). Essa proposta baseia-se no fato de que o comportamento de um processo de riscos competitivos é completamente determinado pelos riscos específicos por causa (BEYERSMANN *et al.*, 2009).

O algoritmo usado para gerar os dados é descrito abaixo:

- 1) Especifica-se o risco específico por causa para o evento 1 (evento de interesse principal) e em seguida para o evento 2 (evento competitivo), anotados por  $\alpha_{01}$  e  $\alpha_{02}$ ;
- 2) Simula-se tempos de falha  $T$  considerando o risco por todas as causas, ou seja, o risco total, dado por  $\alpha_0 = \alpha_{01} + \alpha_{02}$ ;
- 3) Executa-se um experimento binomial para o tempo de falha  $T$  que define qual evento ocorrerá, com uma probabilidade de  $\alpha_{01}/(\alpha_{01} + \alpha_{02})$  para a ocorrência do evento 1 e a probabilidade complementar para o evento 2;
- 4) Gera-se os tempos de censura.

Os tempos de censura foram gerados a partir de uma distribuição uniforme com valores entre 0 e o tempo de falha máximo do conjunto gerado. Para ajustar a porcentagem de censura foi introduzida censura administrativa de maneira a se obter as porcentagens de censuras desejadas (0%, 10%, 20%, 30%, 40% e 50%).

Na primeira etapa da simulação objetivou-se avaliar o efeito da censura sobre a função de incidência cumulativa. Para tanto gerou-se um conjunto de dados de tamanho 100. Os tempos de falha foram gerados a partir de uma distribuição exponencial. Considerou-se três cenários (cenário 1:  $\alpha_{01}=0,8$  e  $\alpha_{02} = 0,2$ ; cenário 2:  $\alpha_{01}=0,5$  e  $\alpha_{02} = 0,5$ ; cenário 3:  $\alpha_{01}=0,1$

e  $\alpha_{02} = 0,7$ ). Para cada cenário foram consideradas as seguintes porcentagens de censura: 0%, 10%, 20%, 30%, 40% e 50%.

Na segunda etapa do estudo de simulação comparou-se a estimativa da incidência obtida pela equação  $\hat{F}_i(t) = \sum_{\forall t, t_j \leq t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1})$  com a estimativa obtida pelo uso do estimador de Kaplan-Meier, ou seja, estimando a incidência de um evento como o complemento da estimativa de Kaplan-Meier. Para essa finalidade gerou-se conjuntos de tempos de sobrevivência de tamanho 100. Para cada conjunto de dados gerados foi obtido a média das diferenças entre as incidências obtidas pelos dois métodos. O processo foi repetido 1000 vezes e a média das médias das diferenças e o desvio das médias das diferenças foram registrados. Os cenários considerados foram: cenário 1:  $\alpha_{01}=0,9$  e  $\alpha_{02} = 0,1$ ; cenário 2:  $\alpha_{01}=0,8$  e  $\alpha_{02} = 0,1$  e  $0,2$ ; cenário 3:  $\alpha_{01}=0,7$  e  $\alpha_{02} = 0,1; 0,2; 0,3$ ; cenário 4:  $\alpha_{01}=0,6$  e  $\alpha_{02} = 0,1; 0,2; 0,3; 0,4$ ; cenário 5:  $\alpha_{01}=0,5$  e  $\alpha_{02} = 0,1; 0,2; 0,3; 0,4; 0,5$ ; cenário 6:  $\alpha_{01}=0,4$  e  $\alpha_{02} = 0,1; 0,2; 0,3; 0,4; 0,5; 0,6$ ; cenário 7:  $\alpha_{01}=0,3$  e  $\alpha_{02} = 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7$ ; cenário 8:  $\alpha_{01}=0,2$  e  $\alpha_{02} = 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8$ ; cenário 9:  $\alpha_{01}=0,1$  e  $\alpha_{02} = 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9$ . Novamente para cada cenário foram considerados 0%, 10%, 20%, 30%, 40% e 50% de censura.

As funções criadas para a simulação e as análises dos dados foram realizadas no software R (R CORE TEAM, 2018).

### 3 Resultados e discussão

Nas Figuras 1 a 6 pode-se verificar o impacto da censura sobre a função de incidência cumulativa para o evento 1 e 2 considerando diferentes riscos específicos por causa (riscos específicos por causa para o evento 1 e 2, respectivamente iguais a 0,8 e 0,2, nas Figuras 1 e 2; 0,5 e 0,5, nas Figuras 4 e 5 e 0,1 e 0,7, nas Figuras 5 e 6). Percebe-se que o aumento da porcentagem de censura aumenta a diferença entre as curvas das funções de incidências cumulativas. Por conseguinte, quanto maior a porcentagem de censura, em geral, mais distantes as curvas de incidência, considerando as porcentagens de censura de 10%, 20%, 30%, 40% e 50%, estarão da curva de incidência no caso em que não há censuras. Esse afastamento é bem caracterizado quando se considera um alto risco específico para um evento e baixo para o outro (Figuras 1 e 6), em que o afastamento se dá à medida que se aumenta a porcentagem de censura. Nos demais casos ocorre o afastamento sem este rigorosamente acompanhar o aumento da porcentagem da censura. Todavia, observa-se, de maneira geral, o aumento do distanciamento entre as curvas das funções de incidência cumulativas à medida que se aumenta a porcentagem de censura, indicando que a incidência obtida por meio da função de incidência cumulativa no contexto de riscos competitivos, na presença de altas taxas de censura, é um estimador inadequado, o que está de acordo como os resultados encontrados por Martinez-Cambor (2016).

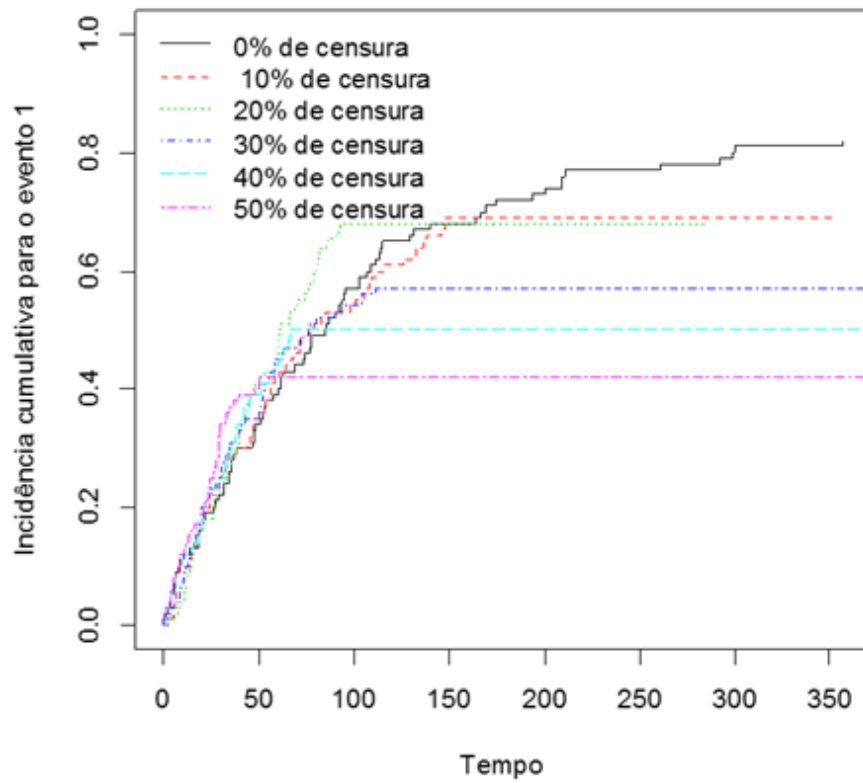


Figura 1 – Estimativa da função de incidência para o evento 1 considerando o risco específico 1 e 2 iguais a 0,8 e 0,2, respectivamente.

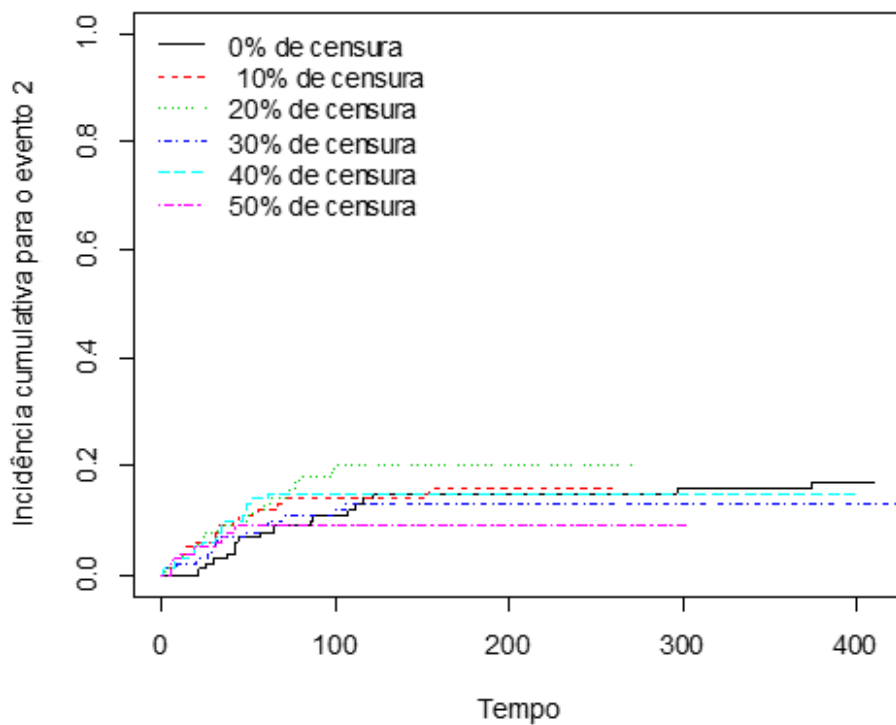


Figura 2 – Estimativa da função de incidência para o evento 2 considerando o risco específico 1 e 2 iguais a 0,8 e 0,2, respectivamente.

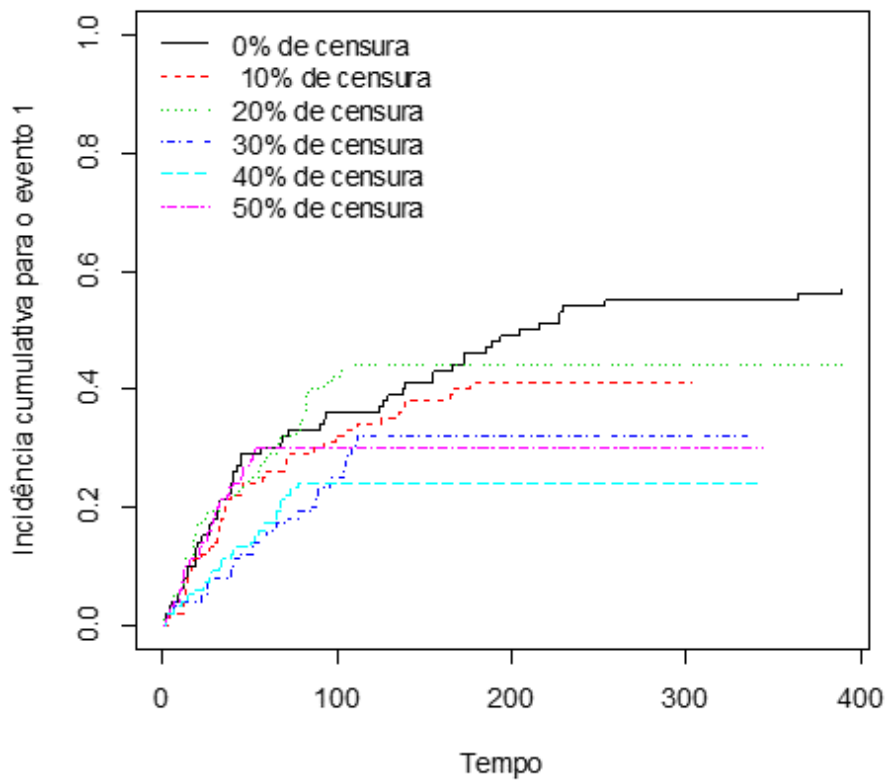


Figura 3 – Estimativa da função de incidência para o evento 1 considerando o risco específico 1 e 2 iguais a 0,5 e 0,5, respectivamente.



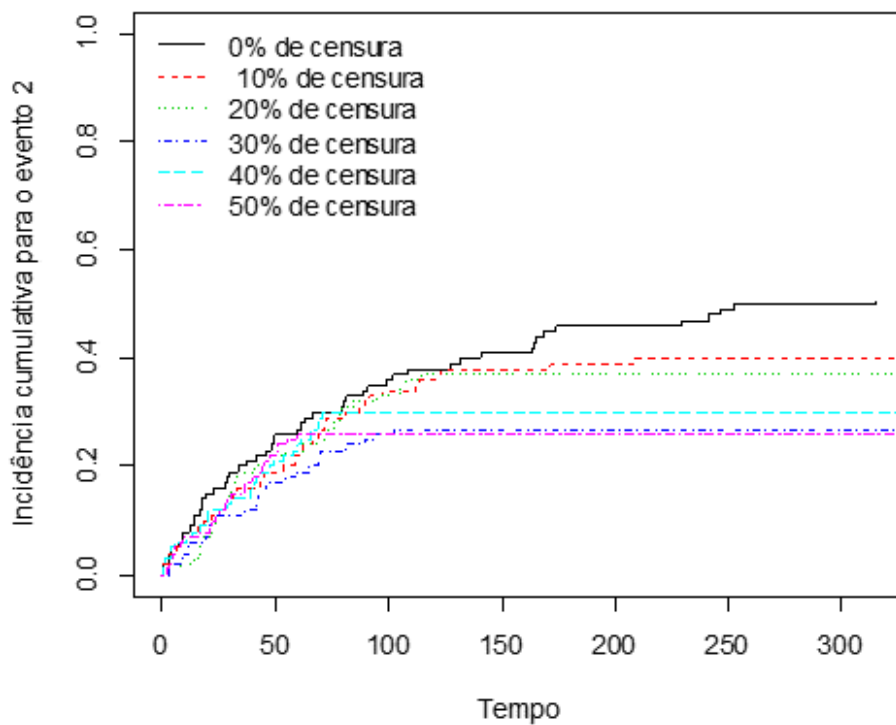


Figura 4 – Estimativa da função de incidência para o evento 2 considerando o risco específico 1 e 2 iguais a 0,5 e 0,5, respectivamente.

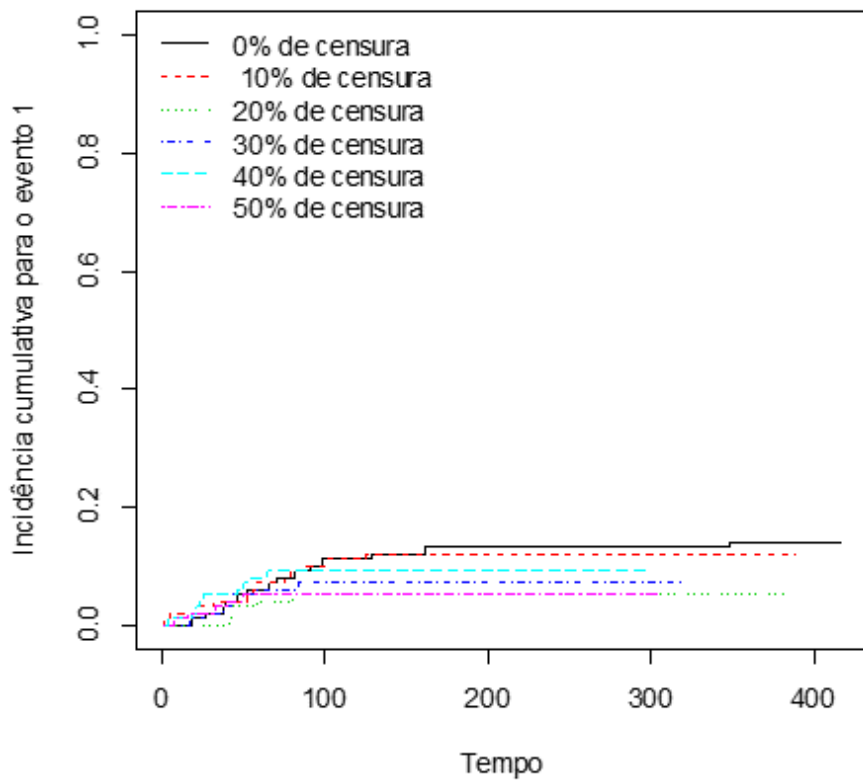


Figura 5 - Estimativa da função de incidência para o evento 1 considerando o risco específico 1 e 2 iguais a 0,1 e 0,7, respectivamente.

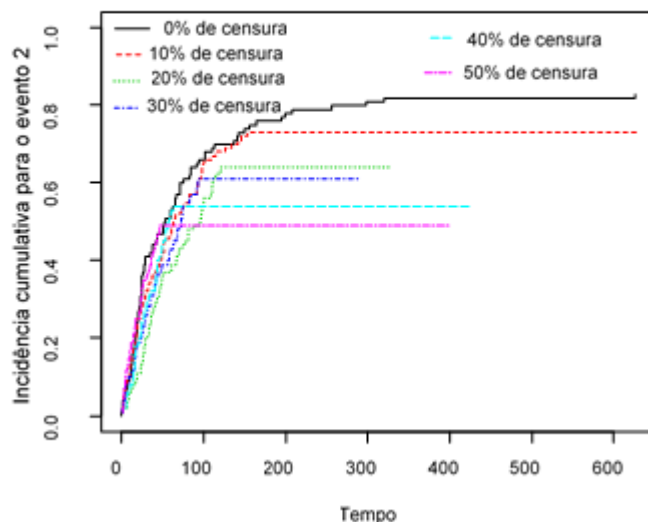


Figura 6 - Estimativa da função de incidência para o evento 2 considerando o risco específico 1 e 2 iguais a 0,1 e 0,7, respectivamente.

A Figura 7 apresenta uma comparação entre as curvas para a função de incidência cumulativa para o evento 1 obtidas pelo uso da equação  $\hat{F}_i(t) = \sum_{v_t, t_j \leq t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1})$  e por meio do complemento do estimador de Kaplan-Meier, considerando dados simulados a partir de um risco específico por causa 1 e 2, respectivamente iguais a 0,8 e 0,2.

Os gráficos apresentados na Figura 7 mostram que com o aumento da porcentagem de censura tem-se que a incidência estimada pelo uso da equação  $\hat{F}_i(t) = \sum_{v_t, t_j \leq t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1})$  se aproxima da estimativa da incidência obtida inadequadamente pelo uso do estimador de Kaplan-Meier. Essa aproximação das estimativas reforça o resultado anterior que aponta que altas taxas de censura é um complicador para a obtenção da função de incidência, visto que, a estimativa da incidência se aproxima daquela obtida pelo uso do estimador de Kaplan-Meier, que conforme diversos autores não é um estimador adequado para a obtenção da incidência (BAKOYANNIS e TOULOUMI, 2010; MARTINEZ-CAMBLOR, 2016; KIM, 2007; LOGAN *et al.*, 2006; PEPE e MORI, 1993; PUTTER, 2006; SATAGOPAN *et al.*, 2004; SHERIF, 2007; WOLBERS *et al.*, 2014).

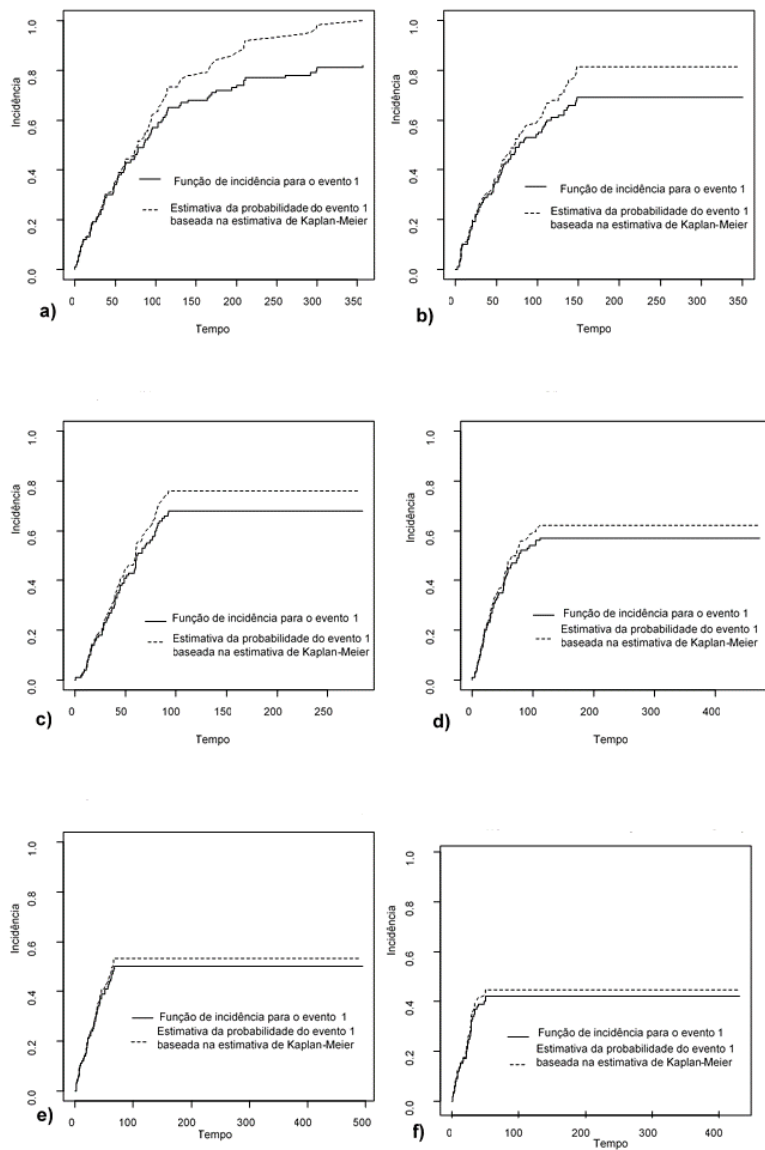


Figure 7 – Curvas de incidência para o evento 1, considerando o risco específico por causa para o evento 1 e para o evento 2 iguais, respectivamente, a 0,8 e 0,2, obtidas por meio da função de incidência no contexto de riscos competitivos e pelo estimador de Kaplan-Meier, considerando diferentes percentagens de censura: a) 0% de censura; b) 10% de censura; c) 20% de censura; d) 30% de censura; e) 40% de censura e f) 50% de censura.

Os gráficos para a incidência do evento 2 considerando riscos específicos por causa para o evento 1 e 2 iguais, respectivamente à 0,8 e 0,2, apresentam comportamento similar (Apêndice A). Resultados análogos, tanto para a incidência para o evento 1 como para o evento 2, foram encontrados considerando outras duas distintas situações: riscos específicos por causa 1 e 2 iguais a 0,5 e 0,5 (Apêndice B e C) e riscos específicos por causa 1 e 2 iguais a 0,7 e 0,1 (Apêndice D e E).

Na Tabela 1 estão apresentados os resultados da diferença média entre as estimativas obtidas pelo uso do complemento do estimador de Kaplan-Meier e as estimativas obtidas com base na teoria de riscos competitivos para alguns valores de  $\alpha_{01}$  e  $\alpha_{02}$ . Uma tabela considerando todos os cenários, ou seja, combinações de riscos específicos por causa consideradas neste estudo, encontra-se no Apêndice F.

Ao se confrontar as estimativas da incidência para o evento 1 obtidas pela equação  $\hat{F}_i(t) = \sum_{\forall t, t_j \leq t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1})$  com as obtidas pelo uso do estimador de Kaplan-Meier, tem-se ainda, que ao se considerar um risco específico por causa 1 fixo e aumentando-se o risco específico por causa 2, tal que  $\alpha_{01} \geq \alpha_{02}$ , ou seja, quando se tem “baixo risco competitivo”, a média das diferenças entre os dois métodos utilizados para o cálculo função de incidência cumulativa aumenta (Tabela 1). No entanto, na situação em que o risco competitivo é alto, ou seja,  $\alpha_{01} < \alpha_{02}$ , para um  $\alpha_{01}$  fixo e aumentando-se o risco específico por causa 2,  $\alpha_{02}$ , tem-se que a diferença entre as estimativas da incidência para o evento diminui (Tabela 1). Isto posto, aponta-se que a equação  $\hat{F}_i(t) = \sum_{\forall t, t_j \leq t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1})$  apresentou pior desempenho quando a ocorrência do evento competitivo é alta (Tabela 1).

Tabela 1 – Média e desvio-padrão das diferenças das estimativas da incidência 1 obtidas pelo complemento do estimador de Kaplan-Meier e pela função de incidência

Risco específico		Porcentagem de censura					
$\alpha_{01}$	$\alpha_{02}$	0%	10%	20%	30%	40%	50%
		Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)
0,8	0,1	0,0299 (0,0079)	0,0277 (0,0079)	0,0265 (0,0245)	0,0207 (0,0062)	0,0159 (0,0054)	0,0122 (0,0042)
0,8	0,2	0,0473 (0,0092)	0,0439 (0,0088)	0,0381 (0,0081)	0,0314 (0,0072)	0,0247 (0,0061)	0,0185 (0,0047)
0,5	0,1	0,0411 (0,0090)	0,0389 (0,0085)	0,0339 (0,0078)	0,0278 (0,0069)	0,0221 (0,0057)	0,0162 (0,0047)
0,5	0,2	0,0621 (0,0091)	0,0572 (0,0088)	0,0492 (0,0080)	0,0409 (0,0071)	0,0314 (0,0060)	0,0230 (0,0049)
0,5	0,3	0,0743 (0,0085)	0,0686 (0,0082)	0,0584 (0,0075)	0,0474 (0,0066)	0,0367 (0,0055)	0,0265 (0,0045)
0,5	0,4	0,0810 (0,0080)	0,0741 (0,0079)	0,0631 (0,0073)	0,0507 (0,0066)	0,0393 (0,0054)	0,0283 (0,0044)
0,5	0,5	0,0850 (0,0074)	0,0769 (0,0073)	0,0647 (0,0070)	0,0516 (0,0063)	0,0399 (0,0052)	0,0288 (0,0043)
0,1	0,1	0,0845 (0,0074)	0,0766 (0,0079)	0,0642 (0,0070)	0,0517 (0,0063)	0,0397 (0,0054)	0,0287 (0,0043)
0,1	0,2	0,0838 (0,0087)	0,0741 (0,0088)	0,0607 (0,0082)	0,0485 (0,0072)	0,0365 (0,0061)	0,0261 (0,0049)
0,1	0,3	0,0743 (0,0101)	0,0652 (0,0099)	0,0530 (0,0094)	0,0421 (0,0081)	0,0317 (0,0065)	0,0226 (0,0054)
0,1	0,4	0,0661 (0,0110)	0,0573 (0,0108)	0,0470 (0,0097)	0,0364 (0,0079)	0,0273 (0,0070)	0,0194 (0,0054)
0,1	0,5	0,0585 (0,0118)	0,0504 (0,0108)	0,0410 (0,0098)	0,0320 (0,0089)	0,0238 (0,0070)	0,0167 (0,0053)
0,1	0,6	0,0525 (0,0119)	0,0449 (0,0115)	0,0362 (0,0099)	0,0284 (0,0084)	0,0210 (0,0066)	0,0150 (0,0053)
0,1	0,7	0,0477 (0,0117)	0,0406 (0,0111)	0,0329 (0,0097)	0,0252 (0,0083)	0,0191 (0,0068)	0,0135 (0,0052)
0,1	0,8	0,0433 (0,0121)	0,0370 (0,0114)	0,0296 (0,0097)	0,0228 (0,0080)	0,0162 (0,0206)	0,0101 (0,0276)
0,1	0,9	0,0401 (0,0115)	0,0341 (0,0109)	0,0265 (0,0245)	0,0205 (0,0171)	0,0146 (0,0246)	0,0091 (0,0276)

## Conclusões

O aumento da porcentagem de censura tem efeito sobre a função de incidência cumulativa. Quanto maior a porcentagem de censura maior o erro na estimativa da função de incidência cumulativa.

O estimador para a função de incidência cumulativa dado por  $\hat{F}_i(t) = \sum_{\forall t, t_j \leq t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1})$  não forneceu bons resultados quando se tem altas porcentagens de censura.

À medida que se aumenta a porcentagem de censura, a estimativa da incidência cumulativa, obtida por meio da função de incidência cumulativa no contexto de riscos competitivos, se aproximou da incidência cumulativa obtida pelo uso do estimador de Kaplan-Meier. Em decorrência disso, tem-se que, para altas taxas de censura, a incidência cumulativa obtida por meio da função de incidência foi tão ruim quanto a obtida pelo uso do estimador de Kaplan-Meier.

A relação entre os riscos específicos por causa afetou a função de incidência cumulativa.

## Agradecimentos

À FAPEMIG pelo apoio financeiro e aos revisores e editores pelas sugestões.

TOMAZ, F. S. C., MARTINS FILHO, S. Effect of censorship on a cumulative incidence function in the presence of competing risks in the survival analysis. *Rev. Bras. Biom.* Lavras, v.37, n.2, p.149-177, 2019.

- **ABSTRACT:** *The objective of this work was to evaluate the effect of censorship on the estimation of cumulative incidence in the presence of competing risks. For this, sets of survival times with competing risks considering different levels of censorship (0%, 10%, 20%, 30%, 40% and 50%) were simulated. The simulation was based on the cause-specific hazard and the exponential distribution was used. For the simulation of the censorship times, the uniform distribution was used. In addition to the different levels of censorship, different values were considered for the cause-specific hazard in the two simulated competing events. Comparisons between the incidence curves of the uncensored data set and the set with different percentages of censorship were performed. We also compared the estimates of the cumulative incidence function based on the theory of competing risks with that obtained by using the Kaplan-Meier estimator. The results indicate that censorship percentages as well as the relation between cause-specific hazard affect the cumulative incidence.*
- **KEYWORDS:** *Subdistribution; time-to-event; competing events.*

## Referências

- AALEN, O. O.; BORGAN, O.; GJESSING, H. K. *Survival and event history analysis: a process point of view*. New York: Springer, 2008. 539p.
- ASSANE, C. C.; PEREIRA, B.B., MUAUALO, M. A. Modelos e riscos competitivos para análise de sobrevivência de pacientes com hipertensão arterial em tratamento no HUCFF/UFRJ. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 47, 20015, Porto de Galinhas, *Anais ...* 1226 – 1236.
- BAKOYANNIS, G.; TOULOUMI, G. Practical methods for competing risks data: a review. *Statistical Methods in Medical Research*, v.21, n.3, p.257-272, 2010.
- BERRY, S. D.; NGO, L.; SAMELSON, E. J.; KIEL, D. P. Competing risk of death: an important consideration in studies of older adults. *Journal of the American Geriatrics Society*, v.58, n.4, p.783-787, 2010.
- BEUSCART, J-B.; PAGNIEZ, D., BOULANGER, E.; FOY, C. L. S.; SALLERON, J.; FRIMAT, L.; DUHAMEL, A. Overestimation of the probability of death on peritoneal dialysis by Kaplan-Meier method: advantages of a competing risks approach. *BMC Nephrology*, v.13, n.31, p.1471 – 2369, 2012.
- BEYERSMANN, J.; SCHEIKE, T. H. Classical regression models for competing risks. In: KLEIN, J. P.; HOUWELINGEN, H. C. V.; IBRAHIM, J. G.; SCHEIKE, T. H. (ed.). *Handbook of survival analysis*. New York: Taylor & Francis Group, 2014, cap. 8, p.157 – 177.
- BEYERSMANN, J.; SCHUMACHER, M.; ALLIGNOL, A. *Competing risk and multistate models with R*. New York, 2012. 245p.
- BEYERSMANN, J.; LATOUCHE, A.; BUCHHOLZ, A.; SCHUMACHER, M. Simulating competing risk data in survival analysis. *Statistics in Medicine*, v.28, n.5, p.956 – 971, 2009.
- CARVALHO, M. S.; ANDREOZZI, V. L.; CODEÇO, C. T.; CAMPOS, D. P.; SHIMAKURA, S. E. *Análise de sobrevivência: teoria e aplicações em saúde*. 2.ed. Rio de Janeiro: Editora Fiocruz, 2011. 432p.
- DIANATKHAH, M.; RAHGOZAR, M.; TALAEI, M.; KARIMLOUA, M.; SADEGHI, M.; OVEISGHARAM, S.; SARRAFZADEGAN, N. Comparison of competing risks models based on cumulative incidence function in analyzing time to cardiovascular diseases. *ARYA Atheroscler*, v.10, n.1, p.6–12, 2014.
- ERIKSSON, F.; LI, J.; SCHEIKE, T.; ZHANG, M. J. The proportional odds cumulative incidence model for competing risks. *Biometrics*, v.71, n.3, p.687–695, 2015.
- FERRAZ, R. O.; MOREIRA-FILHO, D. C. Análise de sobrevivência de mulheres com câncer de mama: modelos de riscos competitivos. *Ciência & Saúde Coletiva*, v.22, p.3743-3754, 2017.
- GESKUS, R. B. *Data analysis with competing risks and intermediate states*. Boca Raton: Chapman & Hall/CRC, 2015. 243p.



- GILLAM, M. H.; SALTER, A.; RYAN, P.; GRAVES, S. E. Different competing risks models applied to data from the Australian Orthopaedic Association National joint Replacement Registry. *Acta Orthopaedica*, v.82, n.5, p.513–520, 2011.
- GOOLEY, T. A.; WENDY, L.; CROWLEY, J.; STORER, B. E. Estimation of Failure probabilities in the presence of competing risk: New representations of old estimators. *Statistics in Medicine*. v.8, n.6, p.695–706, 1999.
- GRAW, F.; GERDS, T. A.; SCHUMACHER, M. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Anal*, v.15, p.24 –255, 2009.
- HOSMER, D. W.; LEMESHOW, S.; MAY, S. *Regression modeling of time to event data*. 2.ed. New York: John Wiley & Sons, 2008. 392p.
- KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, v.53, n.282, p.457-481, 1958.
- KIM, H. T. Cumulative incidence in competing risks data e competing risks regression analysis. *Clinical Cancer Research*, v.13, n.2, p.559–565, 2007.
- KLEIN, J. P.; BAJORUNAITIS, R. Inference for competing risks. In: BALAKRISHNAN, C. R.; RAO, R.C. *Handbook of Statistics*. North Holland, v.23, cap.16, p.291–311, 2004.
- KUK, D.; VARADHAN, R. Model selection in competing risk regression. *Statistics in Medicine*, v.32, n.18, p.3077–3088, 2013.
- LI, J.; SCHEIKE, T. H.; ZHANG, M-J. Checking Fine and Gray subdistributions hazard model with cumulative sums of residuals. *Lifetime Data Anal*, v.21, n.2, p.197–217, 2015.
- LOGAN, B. R.; ZHANG, M. J.; KLEIN, J. P. Regression models for hazard rates versus cumulative incidence probabilities in hematopoietic cell transplantation data. *Biology of Blood and Marrow transplantation*, v.12, n.1, p.107–112, 2006.
- MARTINEZ-CAMBLOR, P. Kaplan-Meier estimator in competing risk contexts. *Hacettepe Journal of Mathematics and Statistics*, v.45, n.4, p.116 –1173, 2016.
- PEPE, M. S.; MORI, M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure times data? *Statistics in Medicine*, v.12, n.8, p.737-751, 1993.
- PINTILIE, M. *Competing risk: a practical perspective*. Chichester: John Wiley & Sons, 2006. 224p.
- PUTTER, H.; FIOCCO, M.; GESKUS, R.B. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, v.26, n.11, p.2389–2430, 2006.
- R CORE TEAM. *R: A language and environment for statistical computing*. Vienna, Austria: R Fundation for Statistical Computing, 2018. Disponível em:< <http://www.r-project.org/> >.
- SARKER, S.; MIN, D. K.; BLACK, T. R.; LIM, H. J. Application of multiplicative and additive hazard models to injury prevention among health care workers. *Advances in Research*, v.5, n.1, p. 8–12, 2015.
- SATAGOPAN, J. M.; BEN-PORAT, L.; BERWICK, M.; ROBSON, M.; KUTLER, D.; AUERBACH, A. D. A note on competing risks in survival analysis. *British Journal of Cancer*, v.91, n.7, p.1229–1235, 2004.

SCHEIKE, T. H.; ZHANG, M. J. Analyzing competing risk data using the R timereg package. *Journal of Statistical Software*, v.38, n.2, p.1-15, 2011.

SHERIF, B. N. *A comparison of Kaplan-Meier e cumulative incidence estimate in the presence or absence of competing risks in breast cancer data*. 2007, 30 f. Thesis (Master of Science) – School of Public Health, University of Pittsburgh, 2007.

ZHOU, B.; LATOUCHE, A.; ROCHA, V.; FINE, J. Competing risks regression for stratified data. *Biometrics*, v.67, n.2, p.661–670, 2011.

ZHOU, B. *Contributions to competing risks regression*. 2010. 93 p. Thesis (Doctor of Philosophy) – Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, 2010.

SUN, L.; LIU, J.; SUN, J.; ZHANG, M. J. Modeling the subdistribution of a competing risks. *Statistica Sinica*, v.16, n.4, p.1367–1385, 2006.

ZHANG, S.; IVY, J. S.; WILSON, J. R.; DIEHL, K. M.; YANKASKAS, B. C. Competing risks analysis in mortality estimation for breast cancer patients from independent risk groups. *Health care management Science*, v.7, n.3, p.259 – 269, 2013.

WOLBERS, M. KOLLER, M. T.; STEL, V. S.; SCHAER, B.; JAGER, K., J.; LEFFONDRÉ, K.; HEINZE, G. Competing risks analyses: objectives e approaches. *European Heart Journal*, v.35, n.42, p.2936–2941, 2014.

Recebido em 10.04.2018

Aprovado após revisão em 16.07.2018

## APÊNDICE A

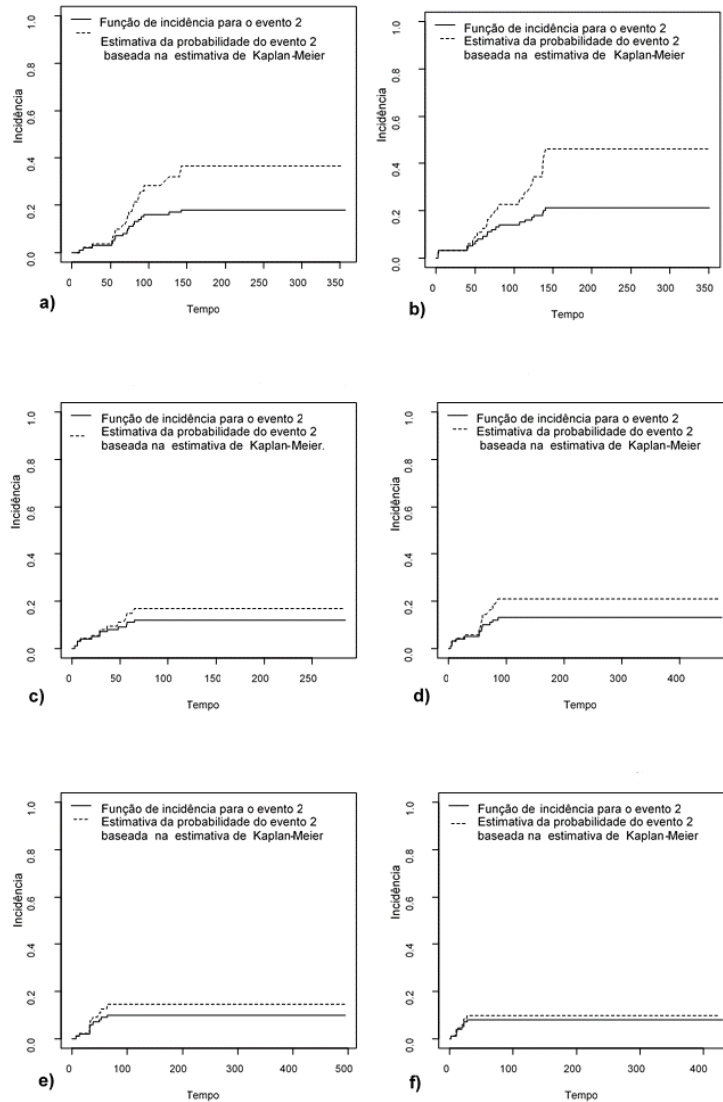


Figure 8 – Curvas de incidência para o evento 2, considerando o risco específico por causa para o evento 1 e para o evento 2 iguais, respectivamente, a 0,8 e 0,2, obtidas por meio da função de incidência no contexto de riscos competitivos e pelo complemento do estimador de Kaplan-Meier, considerando diferentes porcentagens de censura: a) 0% de censura; b) 10% de censura; c) 20% de censura; d) 30% de censura; e) 40% de censura e f) 50% de censura.

## APÊNDICE B

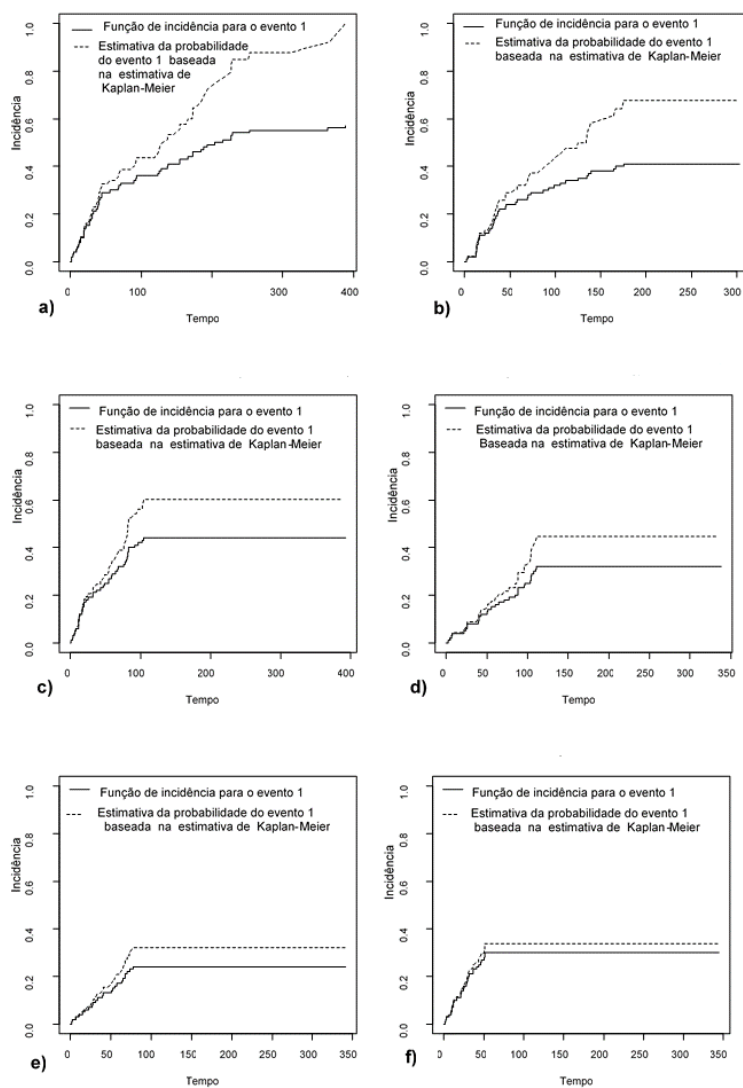


Figure 9 – Curvas de incidência para o evento 1, considerando o risco específico por causa para o evento 1 e para o evento 2 iguais, respectivamente, a 0,5 e 0,5, obtidas por meio da função de incidência no contexto de riscos competitivos e pelo complemento do estimador de Kaplan-Meier, considerando diferentes porcentagens de censura: a) 0% de censura; b) 10% de censura; c) 20% de censura; d) 30% de censura; e) 40% de censura e f) 50% de censura.

## APÊNDICE C

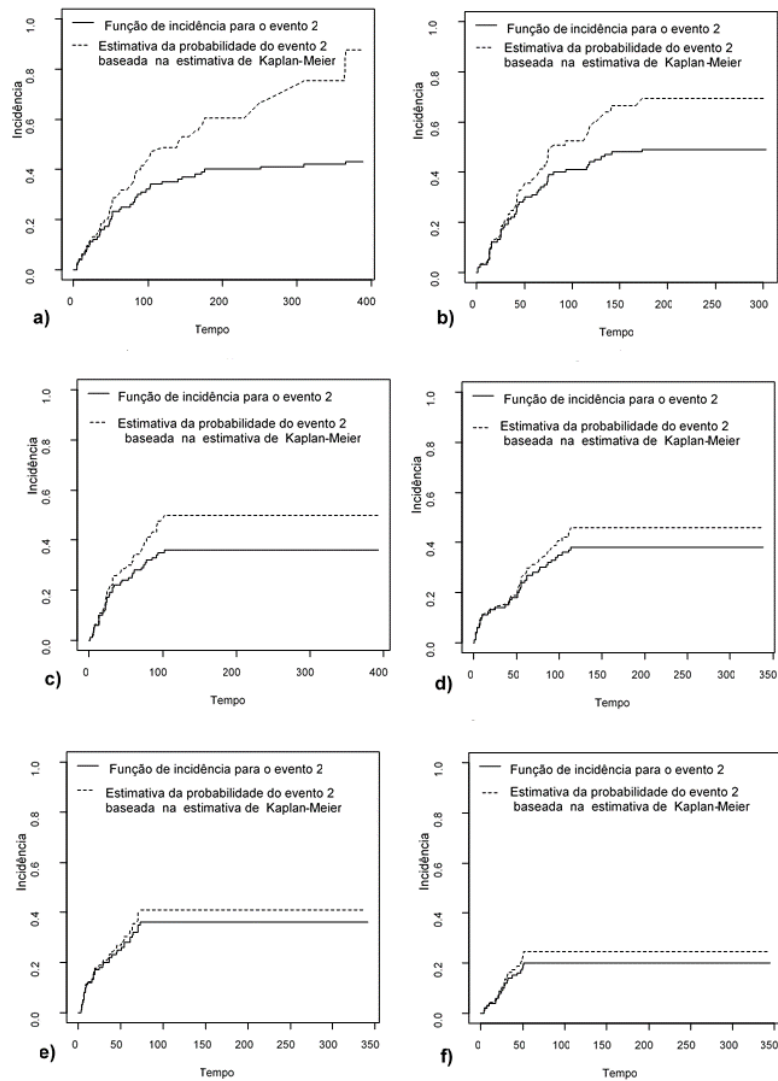


Figure 10 – Curvas de incidência para o evento 2, considerando o risco específico por causa para o evento 1 e para o evento 2 iguais, respectivamente, a 0,5 e 0,5, obtidas por meio da função de incidência no contexto de riscos competitivos e pelo complemento do estimador de Kaplan-Meier, considerando diferentes percentagens de censura: a) 0% de censura; b) 10% de censura; c) 20% de censura; d) 30% de censura; e) 40% de censura e f) 50% de censura.

## APÊNDICE D

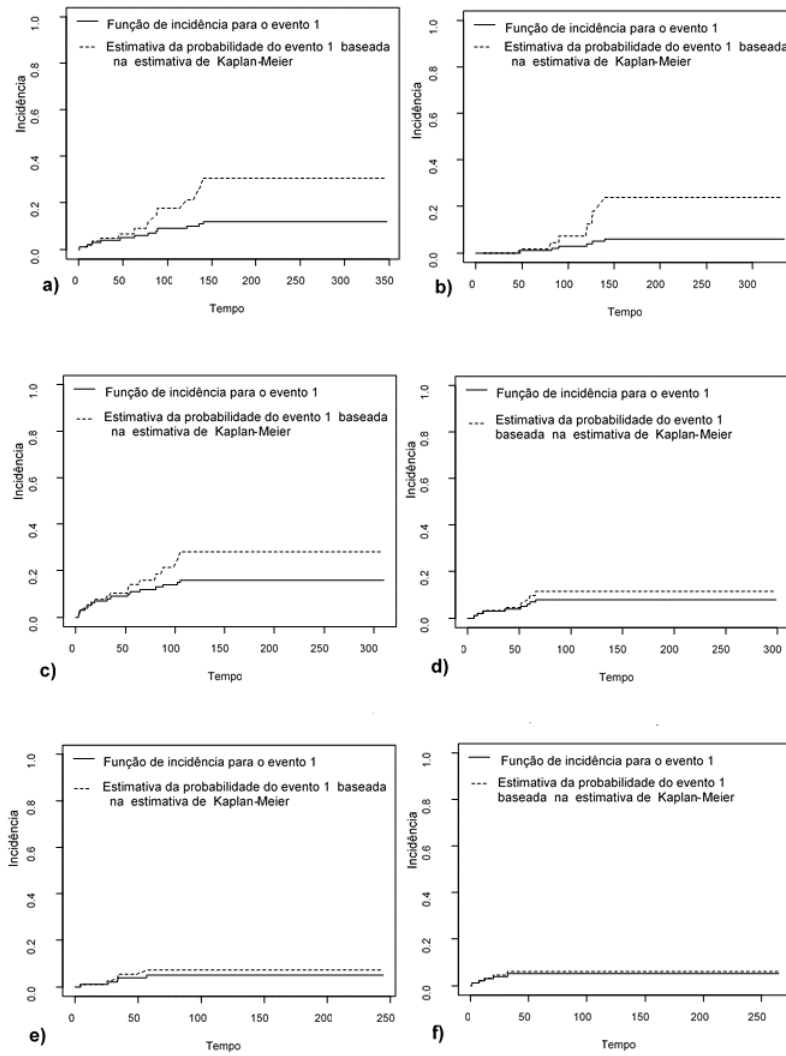


Figure 11 – Curvas de incidência, considerando o risco específico por causa para o evento 1 e para o evento 2 iguais, respectivamente, a 0,1 e 0,7, obtidas por meio da função de incidência no contexto de riscos competitivos e pelo complemento do estimador de Kaplan-Meier, considerando diferentes porcentagens de censura: a) 0% de censura; b) 10% de censura; c) 20% de censura; d) 30% de censura; e) 40% de censura e f) 50% de censura.

## APÊNDICE E

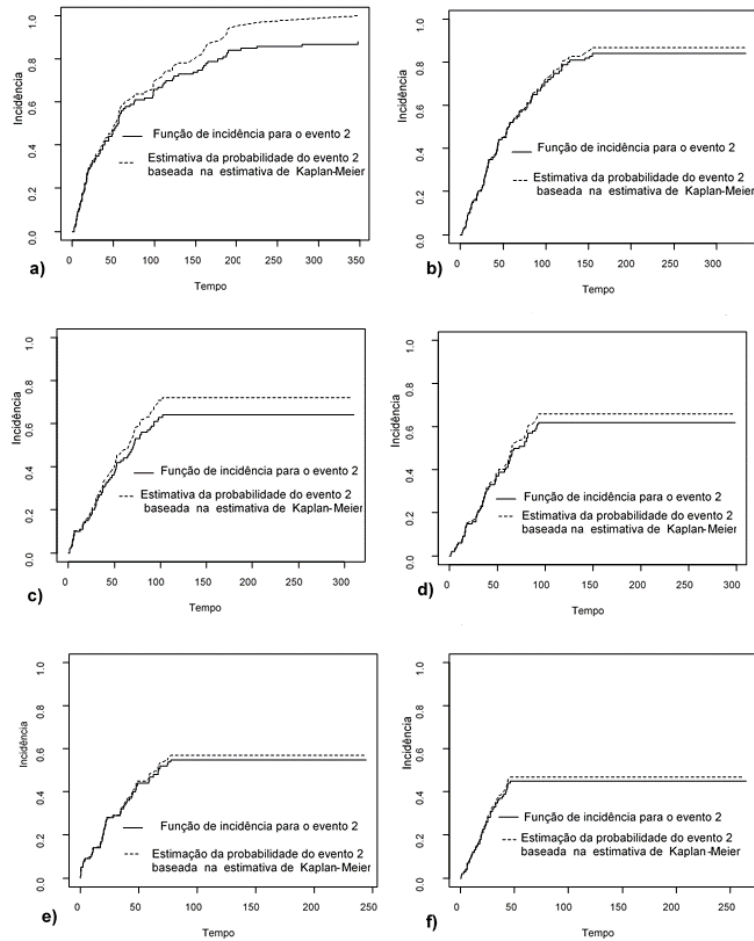


Figure 12 – Curvas de incidência, considerando o risco específico por causa para o evento 1 e para o evento 2 iguais, respectivamente, a 0,1 e 0,7, obtidas por meio da função de incidência no contexto de riscos competitivos e pelo complemento do estimador de Kaplan-Meier, considerando diferentes porcentagens de censura: a) 0% de censura; b) 10% de censura; c) 20% de censura; d) 30% de censura; e) 40% de censura e f) 50% de censura.

ANEXO F

Tabela 2 - Média e desvio-padrão das diferenças das estimativas da incidência 1 obtidas pelo complemento do estimador de Kaplan-Meier e pela função de incidência

Riscos específicos por causa		Porcentagem de censura					
$\alpha_{01}$	$\alpha_{02}$	0%	10%	20%	30%	40%	50%
		Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)
0,9	0,1	0,0276 (0,0075)	0,0257 (0,0073)	0,0224 (0,0067)	0,0190 (0,0060)	0,0148 (0,0050)	0,0111 (0,0042)
0,8	0,1	0,0299 (0,0079)	0,0277 (0,0079)	0,0265 (0,0245)	0,0207 (0,0062)	0,0159 (0,0054)	0,0122 (0,0042)
0,8	0,2	0,0473 (0,0092)	0,0439 (0,0088)	0,0381 (0,0081)	0,0314 (0,0072)	0,0247 (0,0061)	0,0185 (0,0047)
0,7	0,1	0,0334 (0,0082)	0,0309 (0,0081)	0,0271 (0,0073)	0,0226 (0,0064)	0,0178 (0,0053)	0,0133 (0,0044)
0,7	0,2	0,0513 (0,0092)	0,0478 (0,0087)	0,0414 (0,0079)	0,0343 (0,0072)	0,0267 (0,0059)	0,0199 (0,0048)
0,7	0,3	0,0647 (0,0096)	0,0598 (0,0092)	0,0515 (0,0083)	0,0418 (0,0070)	0,0327 (0,0060)	0,0239 (0,0047)
0,6	0,1	0,0366 (0,0082)	0,0341 (0,0084)	0,0298 (0,0077)	0,0250 (0,0068)	0,0195 (0,0058)	0,0147 (0,0045)
0,6	0,2	0,0564 (0,0093)	0,0525 (0,0091)	0,0454 (0,0082)	0,0370 (0,0074)	0,0291 (0,0061)	0,0214 (0,0049)
0,6	0,3	0,0692 (0,0092)	0,0636 (0,0092)	0,0546 (0,0084)	0,0450 (0,0067)	0,0346 (0,0059)	0,0252 (0,0047)
0,6	0,4	0,0770 (0,0083)	0,0704 (0,0081)	0,0601 (0,0074)	0,0487 (0,0069)	0,0377 (0,0056)	0,0272 (0,0044)
0,5	0,1	0,0411 (0,0090)	0,0389 (0,0085)	0,0339 (0,0078)	0,0278 (0,0069)	0,0221 (0,0057)	0,0162 (0,0047)
0,5	0,2	0,0621 (0,0091)	0,0572 (0,0088)	0,0492 (0,0080)	0,0409 (0,0071)	0,0314 (0,0060)	0,0230 (0,0049)
0,5	0,3	0,0743 (0,0085)	0,0686 (0,0082)	0,0584 (0,0075)	0,0474 (0,0066)	0,0367 (0,0055)	0,0265 (0,0045)
0,5	0,4	0,0810 (0,0080)	0,0741 (0,0079)	0,0631 (0,0073)	0,0507 (0,0066)	0,0393 (0,0054)	0,0283 (0,0044)



Tabela 2 (continuação) - Média e desvio-padrão das diferenças das estimativas da incidência 1 obtidas pelo complemento do estimador de Kaplan-Meier e pela função de incidência

Riscos específicos por causa		Porcentagem de censura					
		0%	10%	20%	30%	40%	50%
$\alpha_{01}$	$\alpha_{02}$	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)
0,4	0,1	0,0480 (0,0094)	0,0442 (0,0095)	0,0384 (0,0086)	0,0318 (0,0074)	0,0249 (0,0062)	0,0188 (0,0046)
0,4	0,2	0,0689 (0,0092)	0,0639 (0,0086)	0,0549 (0,0077)	0,0444 (0,0068)	0,0349 (0,0057)	0,0254 (0,0046)
0,4	0,3	0,0797 (0,0082)	0,0723 (0,0081)	0,0616 (0,0075)	0,0496 (0,0064)	0,0385 (0,0055)	0,0279 (0,0042)
0,4	0,4	0,0844 (0,0081)	0,0763 (0,0077)	0,0646 (0,0073)	0,0520 (0,0064)	0,0396 (0,0055)	0,0287 (0,0042)
0,4	0,5	0,0865 (0,0074)	0,0778 (0,0076)	0,0654 (0,0071)	0,0524 (0,0066)	0,0397 (0,0053)	0,0283 (0,0045)
0,4	0,6	0,0860 (0,0077)	0,0768 (0,0078)	0,0640 (0,0073)	0,0515 (0,0064)	0,0386 (0,0055)	0,0280 (0,0046)
0,3	0,1	0,0564 (0,0093)	0,0525 (0,0087)	0,0450 (0,0082)	0,0376 (0,0073)	0,0290 (0,0059)	0,0212 (0,0048)
0,3	0,2	0,0766 (0,0083)	0,0700 (0,0084)	0,0597 (0,0076)	0,0490 (0,0065)	0,0375 (0,0056)	0,0271 (0,0044)
0,3	0,3	0,0842 (0,0080)	0,0766 (0,0076)	0,0645 (0,0072)	0,0522 (0,0060)	0,0398 (0,0054)	0,0285 (0,0043)
0,3	0,4	0,0864 (0,0077)	0,0774 (0,0077)	0,0649 (0,0073)	0,0517 (0,0066)	0,0395 (0,0055)	0,0282 (0,0045)
0,3	0,5	0,0851 (0,0077)	0,0763 (0,0082)	0,0638 (0,0077)	0,0505 (0,0066)	0,0383 (0,0058)	0,0273 (0,0046)
0,3	0,6	0,0830 (0,0085)	0,0736 (0,0086)	0,0611 (0,0081)	0,0488 (0,0071)	0,0368 (0,0061)	0,0258 (0,0051)

Tabela 2 (continuação) - Média e desvio-padrão das diferenças das estimativas da incidência 1 obtidas pelo complemento do estimador de Kaplan-Meier e pela função de incidência

Riscos específicos por causa		Porcentagem de censura					
$\alpha_{01}$	$\alpha_{02}$	0%	10%	20%	30%	40%	50%
		Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)
0,3	0,7	0,0809 (0,0092)	0,0714 (0,0093)	0,0588 (0,0087)	0,0462 (0,0077)	0,0349 (0,0063)	0,0246 (0,0050)
0,2	0,1	0,0697 (0,0087)	0,0638 (0,0084)	0,0546 (0,0076)	0,0447 (0,0068)	0,0346 (0,0059)	0,0251 (0,0047)
0,2	0,2	0,0845 (0,0073)	0,0765 (0,0072)	0,0647 (0,0073)	0,0520 (0,0065)	0,0400 (0,0051)	0,0285 (0,0044)
0,2	0,3	0,0859 (0,0078)	0,0774 (0,0077)	0,0645 (0,0074)	0,0516 (0,0066)	0,0393 (0,0056)	0,0277 (0,0046)
0,2	0,4	0,0831 (0,0089)	0,0735 (0,0086)	0,0611 (0,0084)	0,0484 (0,0072)	0,0366 (0,0061)	0,0257 (0,0048)
0,2	0,5	0,0796 (0,0094)	0,0697 (0,0091)	0,0576 (0,0086)	0,0452 (0,0074)	0,0341 (0,0064)	0,0243 (0,0051)
0,2	0,6	0,0752 (0,0104)	0,0650 (0,0099)	0,0536 (0,0092)	0,0422 (0,0078)	0,0319 (0,0069)	0,0225 (0,0053)
0,2	0,7	0,0697 (0,0087)	0,0612 (0,0105)	0,0501 (0,0094)	0,0391 (0,0082)	0,0293 (0,0067)	0,0207 (0,0053)
0,2	0,8	0,0845 (0,0073)	0,0571 (0,0106)	0,0468 (0,0097)	0,0366 (0,0083)	0,0270 (0,0069)	0,0189 (0,0056)
0,1	0,1	0,0845 (0,0074)	0,0766 (0,0079)	0,0642 (0,0070)	0,0517 (0,0063)	0,0397 (0,0054)	0,0287 (0,0043)
0,1	0,2	0,0838 (0,0087)	0,0741 (0,0088)	0,0607 (0,0082)	0,0485 (0,0072)	0,0365 (0,0061)	0,0261 (0,0049)
0,1	0,3	0,0743 (0,0101)	0,0652 (0,0099)	0,0530 (0,0094)	0,0421 (0,0081)	0,0317 (0,0065)	0,0226 (0,0054)
0,1	0,4	0,0661 (0,0110)	0,0573 (0,0108)	0,0470 (0,0097)	0,0364 (0,0079)	0,0273 (0,0070)	0,0194 (0,0054)

Tabela 2 (continuação) - Média e desvio-padrão das diferenças das estimativas da incidência 1 obtidas pelo complemento do estimador de Kaplan-Meier e pela função de incidência

Riscos específicos por causa		Porcentagem de censura					
		0%	10%	20%	30%	40%	50%
$\alpha_{01}$	$\alpha_{02}$	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)
0,1	0,5	0,0585 (0,0118)	0,0504 (0,0108)	0,0410 (0,0098)	0,0320 (0,0089)	0,0238 (0,0070)	0,0167 (0,0053)
0,1	0,6	0,0525 (0,0119)	0,0449 (0,0115)	0,0362 (0,0099)	0,0284 (0,0084)	0,0210 (0,0066)	0,0150 (0,0053)
0,1	0,7	0,0477 (0,0117)	0,0406 (0,0111)	0,0329 (0,0097)	0,0252 (0,0083)	0,0191 (0,0068)	0,0135 (0,0052)
0,1	0,8	0,0435 (0,0120)	0,0370 (0,0114)	0,0296 (0,0097)	0,0228 (0,0080)	0,0162 (0,0206)	0,0101 (0,0276)
0,1	0,9	0,0401 (0,0115)	0,0341 (0,0109)	0,0265 (0,0245)	0,0205 (0,0171)	0,0146 (0,0246)	0,0091 (0,0276)

## APÊNDICE G

Código da função usada para simular os dados de sobrevivência com riscos competitivos e da função que compara a estimativa da incidência obtida pelo uso da abordagem de riscos competitivos com a obtida pelo uso do complemento do estimador de Kaplan-Meier.

```

simula.data<- function(n,alfa.01,alfa.02, lambda, perc.cens) {
#####
#Simula tempos de sobrevivência para eventos competitivos
#n- tamanho da amostra a ser simulada
# alfa.01 –risco específico por causa 1
# alfa.02 – risco específico por causa 2
#alfa= alfa.01 + alfa.02 – risco específico por todas as causas.
#lambda – parâmetro lambda para a distribuição exponencial. #E(X)=1/lambda.
#perc.cens – porcentagem de censura nos dados. Deve ser um valor entre 0 #e 1.
#####
T<-rexp(n, lambda) # gera tempos exponenciais com média 1/lambda
p=alfa.01/(alfa.01 + alfa.02)
f.cause<-rbinom(n,size=1,prob=p)
f.cause<-ifelse(f.cause==0,2,1) # codifica as causas de falha: 1- falha 1, #2-falha 2
cens.time<-runif(n, 0, round(max (T)))
obs.time<-pmin(T, cens.time)
obs.cause<-c(T <= cens.time)*f.cause
cens.adm<- as.vector(quantile (obs.time,1-perc.cens))
status<-c(obs.time <= cens.adm)*f.cause # gera dados com p% de censura;
data.simul<-data.frame(obs.time,status,event=ifelse(status>0,1,0),
event.1=ifelse(status==1, 1,0), event.2=ifelse(status ==2,1,0) )
return(data.simul)
}

diference<-function(n,alfa.01,alfa.02, lambda, perc.cens, event.type)
{
#####Calcula a
média das diferenças entre as estimativas da incidência com #base no estimador de
Kaplan-Meier (1 - km) e a baseada na função de #incidência no contexto de riscos
competitivos para o evento especificado #em “event.type”.
#####

library(survival)
library(cmprsk)
data.simul<- simula.data (n, alfa.01,alfa.02, lambda, perc.cens)
cif<- cuminc(data.simul[,1],data.simul[,2], cencode=0); #cif
ifelse(event.type ==1, i<-c(4),i<-5)
km<-survfit(Surv(data.simul[,1],data.simul[,i])~1); #km

```

```

incidencekm<-1- km$urv; incidencekm
estimatecif<-timepoints(cif, times= km$time); #estimatecif
diference<- (incidencekm- as.vector(estimatecif$est[event.type,])); #diference
mean<-mean(diference, na.rm=TRUE)
sd<-sd(diference, na.rm = TRUE)
result<-data.frame(mean,sd)
return(result)
}

summary.diference<-function(n,alfa.01, alfa.02, lambda, perc.cens, event.type, N){
#####
#Calcula a média das médias e o desvio padrão das diferenças encontradas #nas N
simulações.
#n- tamanho da amostra a ser simulada
#alfa.01 – risco específico por causa para o evento 1
#alfa.02 – risco específico por causa para o evento 2
#lambda – parâmetro lambda para a distribuição exponencial. #E(X)=1/lambda.
#perc.cens – porcentagem de censura nos dados. Deve ser um valor entre 0 #e 1.
#event.type– 1 se o evento é de interesse e 2 se o evento é o competitivo
#####
vector.mean<-matrix (rep(0,N), nrow=N,ncol=1)
vector.sd<-matrix (rep(0,N), nrow=N,ncol=1)
for (j in 1:N){
d<-diference(n,alfa.01,alfa.02, lambda, perc.cens, event.type)
vector.mean[j,1]<-d$mean
vector.sd[j,1]<-d$sd
}
mean.mean<-mean(vector.mean, na.rm=TRUE)
sd.mean<-sd(vector.mean, na.rm=TRUE)
est<- data.frame(mean.mean, sd.mean)
return(est)
}

```