

FLÁVIA SÍLVIA CORRÊA TOMAZ

**MODELAGEM DA FUNÇÃO DE INCIDÊNCIA CUMULATIVA NA
PRESENÇA DE RISCOS COMPETITIVOS EM ANÁLISE DE
SOBREVIVÊNCIA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do Título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2017

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

Tomaz, Flávia Sílvia Corrêa, 1982-
T655m Modelagem da função de incidência cumulativa na presença
2017 de riscos competitivos em análise de sobrevivência / Flávia
Sílvia Corrêa Tomaz. – Viçosa, MG, 2017.
xi, 101f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Sebastião Martins Filho.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Análise de sobrevivência (Biometria). 2. Estatística.
I. Universidade Federal de Viçosa. Departamento de Estatística.
Programa de Pós-graduação em Estatística Aplicada e Biometria.
II. Título.

CDD 22 ed. 519.5

FLÁVIA SÍLVIA CORRÊA TOMAZ

**MODELAGEM DA FUNÇÃO DE INCIDÊNCIA CUMULATIVA NA
PRESENÇA DE RISCOS COMPETITIVOS EM ANÁLISE DE
SOBREVIVÊNCIA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Doctor Scientiae*.

APROVADA: 12 de dezembro de 2017.



Cristiane Márcia dos Santos



Leandro Roberto de Macedo



Paulo César Emiliano



Paulo Roberto Cecon



Sebastião Martins Filho
Orientador

**Aos meus pais, Maria Helena e
Exedito, por todo amor
e dedicação.**

AGRADECIMENTOS

Agradeço primeiramente a DEUS por ter me dado forças ao longo desta jornada.

Ao meu orientador, Prof. Dr. Sebastião Martins Filho, pela paciência, dedicação e incentivo durante o desenvolvimento deste trabalho.

Aos meus coorientadores, Prof. Dr. Moysés e Profa. Dra. Ana Carolina, pelo apoio e disponibilidade.

Aos professores Cristiane, Leandro, Paulo Roberto e Paulo Emiliano, membros da banca, pelas valiosas sugestões.

A todos os professores do Programa de Pós-Graduação em Estatística Aplicada e Biometria.

Aos meus pais, Maria Helena e Expedito, pelo amor, carinho e dedicação.

Aos meus irmãos Fernanda e Edevaldo pelas palavras de incentivo.

Ao meu marido, Flávio, pela paciência, companheirismo, amor e compreensão nos momentos em que estive ausente.

À minha grande amiga Simone pelo incentivo, amizade e palavras de apoio ao longo desta caminhada.

À Universidade Federal de Ouro Preto e a todos os colegas de trabalho do Departamento de Ciências Administrativas pelo apoio.

À professora Maria Claudia Feres Monteiro de Castro Souza pelo apoio e a amizade.

Aos colegas de curso, especialmente, a Micherlânia pela amizade.

Aos funcionários do Programa de Pós-Graduação, Júnior e Anita.

À FAPEMIG pelo apoio financeiro.

SUMÁRIO

RESUMO	viii
ABSTRACT	x
INTRODUÇÃO GERAL	1
CAPÍTULO 1	4
1. Referencial Teórico	4
1.1. Análise de Sobrevivência	4
1.2. Riscos competitivos	6
1.2.1. Notação básica	9
1.2.1.1. Função risco específico por causa	9
1.2.1.2. Função de incidência cumulativa	10
1.3. Estimador para função de incidência cumulativa	11
1.4. Modelo de regressão de Cox para o risco específico por causa	12
1.5. Modelo de regressão de Fine e Gray para o risco da subdistribuição	13
1.6. Modelo de regressão binomial	16
1.6.1. Regressão binomial	17
1.6.2. Funções de ligação para a regressão binomial.....	18
1.6.3. Resposta ponderada	19
1.6.4. Equação de estimação ponderada	20
1.7. Simulação de dados com riscos competitivos	20
2. Referências bibliográficas	22
CAPÍTULO 2	27
Efeito da censura sobre a função de incidência cumulativa na presença de riscos competitivos na análise de sobrevivência	27
RESUMO	27
ABSTRACT	28
1. INTRODUÇÃO.....	29
2. MATERIAL E MÉTODOS	30
3. RESULTADOS E DISCUSSÃO	32
4. CONCLUSÕES	38

5. AGRADECIMENTOS.....	38
6. REFERÊNCIAS BIBLIOGRÁFICAS	38

APÊNDICE A: Figura 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.	41
--	----

APÊNDICE B: Figura 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.	42
--	----

APÊNDICE C: Figura 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 porcentagens de censura.	43
---	----

APÊNDICE D: Figura 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.	44
---	----

APÊNDICE E: Figura 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.	45
APÊNDICE F: Tabela 2 - Média e desvio das diferenças das estimativas da incidência 1 obtidas pelo complemento do estimador de Kaplan-Meier e pela função de incidência.	46
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.	50
CAPÍTULO 3	52
Comparação entre métodos de análise de sobrevivência clássica e abordagem baseada em riscos competitivos para análise de eventos competitivos.	52
RESUMO	52
ABSTRACT.....	53
1. INTRODUÇÃO.....	54
2. MATERIAL E MÉTODOS.....	56
3. RESULTADOS E DISCUSSÃO	58
4. CONCLUSÕES.....	67
5. AGRADECIMENTOS.....	67
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	68

APÊNDICE A: Tabela de incidências estimadas para a claudicação ($F_1(t)$) e para a probabilidade de claudicação calculada com base no estimador de Kaplan-Meier ($1-S_1(t)$).	71
APÊNDICE B: Tabela de incidências estimadas para outras injúrias ($F_2(t)$) e para a probabilidade de ocorrer outras injúrias calculada com base no estimador de Kaplan-Meier ($1-S_2(t)$).	72
APÊNDICE C - Código para estimar as funções de incidência cumulativa, para ajustar modelo de Cox para o risco específico por causa e para ajustar modelo de Fine e Gray para o risco de subdistribuição.	73
CAPÍTULO 4	77
Avaliação de modelos de regressão para a função de incidência cumulativa na presença de riscos competitivos.	77
RESUMO	77
ABSTRACT	78
1. INTRODUÇÃO	79
2. MATERIAL E MÉTODOS	80
3. RESULTADOS E DISCUSSÃO	83
4. CONCLUSÕES	93
5. AGRADECIMENTOS	94
6. REFERÊNCIAS BIBLIOGRÁFICAS	94
APÊNDICE A: Código baseado no algoritmo de Haller (2014) para a simulação de dados.	96
APÊNDICE B: Código para a modelagem da função de incidência.	99
CONCLUSÕES GERAIS	101

RESUMO

TOMAZ, Flávia Sílvia Corrêa, D.Sc., Universidade Federal de Viçosa, dezembro de 2017. **Modelagem da função de incidência cumulativa na presença de riscos competitivos em análise de sobrevivência.** Orientador: Sebastião Martins Filho. Coorientadores: Ana Carolina Campana Nascimento e Moysés Nascimento.

Riscos competitivos surgem em situações em que um indivíduo pode falhar devido à várias causas distintas. Na presença de riscos competitivos a estimação e/ou avaliação do efeito de covariáveis sobre a função de incidência cumulativa (subdistribuição) frequentemente é de interesse. Essa função quantifica a probabilidade de um indivíduo experimentar um evento específico, ou seja, falhar devido a uma determinada causa dentre um conjunto de causas de falha. A estimação não paramétrica da função de incidência, por vezes, é obtida por meio do complemento do estimador de Kaplan-Meier, embora esse procedimento não seja adequado e procedimento apropriado para este propósito esteja disponível. No que se refere a modelagem do efeito de covariáveis sobre a função de incidência, abordagens comumente difundidas baseiam-se ou no risco específico por causa ou no risco da subdistribuição. A primeira ignora a presença dos riscos competitivos, enquanto a segunda leva em consideração os riscos competitivos e frequentemente utiliza o modelo de Fine e Gray. Embora existam alternativas ao modelo de Fine e Gray, estas são pouco discutidas. Neste sentido, o objetivo deste trabalho foi avaliar a estimação da função de incidência cumulativa, bem como verificar como a censura e a relação entre proporção de eventos competitivos afetam a estimação dessa função. Ademais objetivou-se avaliar três modelos de regressão para a função de incidência (modelo de regressão com ligação logarítmica, modelo de regressão com ligação logit e modelo de Fine e Gray). Além de um conjunto de dados reais sobre lesões em cavalos foi utilizado também um estudo de simulação. Os resultados encontrados reforçam relatos encontrados na literatura, que apontam a superestimação da função de incidência cumulativa quando a mesma é estimada como complemento do estimador de Kaplan-Meier, bem como a não correspondência entre os efeitos das covariáveis estimados com base no risco específico por causa e o baseado no risco da subdistribuição. Por meio do

estudo de simulação constatou-se que a percentagem de censura bem como a relação entre os eventos competitivos afeta a estimação da função de incidência cumulativa. Verificou-se também, que, em geral, o modelo de regressão com ligação logarítmica mostrou-se uma alternativa ao modelo de Fine e Gray.

ABSTRACT

TOMAZ, Flávia Sílvia Corrêa, D.Sc., Universidade Federal de Viçosa, December, 2017. **Modeling the cumulative incidence function in the presence of competing risks in survival analysis.** Advisor: Sebastião Martins Filho. Co-advisers: Ana Carolina Campana Nascimento and Moysés Nascimento.

Competing risks arise in situations where an individual may fail due to several different causes. In the presence of competing risks the estimation and / or evaluation of the effect of covariates on the cumulative incidence function (subdistribution) is often of interest. This function quantifies the probability of an individual experiencing a specific event, that is, failure due to a cause within a set of causes of failure. The non-parametric estimation of the incidence function is sometimes obtained through the complement of the Kaplan-Meier estimator, although this procedure is not adequate and an appropriate procedure for this purpose is available. With regard to modeling of the effect of covariates on the incidence function, commonly disseminated approaches are based either on the cause-specific hazard or at the subdistribution hazard. The former ignores the presence of competing risks, while the latter takes into account competing risks and often uses the Fine and Gray model. Although there are alternatives to the Fine and Gray model, these are little discussed. In this sense, the objective of this work was to evaluate the estimation of the cumulative incidence function, as well as to verify how the censorship and the relation between the proportion of competitive events affect the estimation of this function. In addition, we evaluated three regression models for the incidence function (regression model with log-link function, regression model with logit-link function and Fine and Gray model). In addition to a set of real data on injuries in horses was also used a simulation study. The results found reinforce reports found in the literature, which point to the overestimation of the cumulative incidence function when it is estimated to complement the Kaplan-Meier estimator, as well as the non match between the effects of covariates estimated based on the cause specific hazard and the of subdistribution hazard. By means of the simulation study it was verified that the percentage of censorship as well as the relation between the competitive events affects the estimation of the function of cumulative incidence. It was also found

that, in general, the regression model with log-link function was an alternative to the Fine and Gray model.

INTRODUÇÃO GERAL

A análise de sobrevivência engloba um conjunto de técnicas utilizadas quando a variável resposta é o tempo até a ocorrência de um determinado evento (KLEINBAUM e KLEIN, 2005). Esse tempo, denominado tempo de falha, pode representar o tempo até a morte, até a cura, recidiva de uma doença, até a recuperação ou qualquer experiência de interesse que possa acontecer a um indivíduo (COLOSIMO e GIOLO, 2006; KLEINBAUM e KLEIN, 2005).

Na análise de sobrevivência clássica atenção é dada a problemas nos quais, para cada indivíduo, só há o registro do tempo até um único evento (CARVALHO *et al.*, 2011). Contudo, um problema habitual na modelagem de tempo de sobrevivência é a ocorrência de falhas devido a várias causas distintas. Neste contexto, a análise de sobrevivência com riscos competitivos apresenta-se como uma extensão da análise de sobrevivência clássica.

Deste modo, riscos competitivos surgem quando é possível observar o tempo até a ocorrência do primeiro evento, que impede que outros aconteçam (CARVALHO *et al.*, 2011) ou alteram fundamentalmente a probabilidade de ocorrência destes outros eventos (GOOLEY *et al.*, 1999).

É comum a ocorrência de eventos competitivos na área médica (BEYERSMANN e SCHEIKE, 2014; KLEIN e BAJORUNAITE, 2004; GOOLEY *et al.*, 1999, REN *et al.*, 2013) em especial em estudos referentes a oncologia e medicina cardiovascular (NOORDZIJ *et al.*, 2013). Nestes e em outros estudos a obtenção da função de incidência cumulativa e/ou estudo do efeito de covariáveis sobre essa função é frequentemente de interesse.

A função de incidência cumulativa, ou subdistribuição, permite quantificar a probabilidade de um indivíduo experimentar um evento pertencente a um conjunto de causas distintas (GESKUS, 2015).

Embora a literatura apresente um estimador não paramétrico adequado para a função de incidência cumulativa, não raro, essa é estimada, incorretamente, através do complemento do estimador de Kaplan-Meier.

No que se refere a modelagem do efeito das covariáveis sobre a função de incidência cumulativa, duas abordagens são frequentemente discutidas: a

modelagem do risco específico por causa e a modelagem do risco da subdistribuição.

A primeira abordagem tem como base o ajuste de um modelo, geralmente, o modelo de Cox para cada tipo de falha e tem como maior crítica o fato de não permitir a avaliação do efeito da covariável diretamente sobre a função de incidência cumulativa.

Já a abordagem baseada no risco da subdistribuição permite avaliar diretamente o efeito de covariáveis sobre a função de incidência cumulativa. Nessa abordagem, ajusta-se um modelo ao risco da subdistribuição, geralmente o modelo de Fine e Gray.

O modelo de Fine e Gray é um modelo tipo Cox para o risco da subdistribuição e é o mais popular devido a sua disponibilidade em software estatístico.

Apesar de, essencialmente, na modelagem de riscos competitivos ser utilizado o modelo de Fine e Gray, existem alternativas a esse modelo, como os modelos de transformação com ligação logarítmica e ligação logit, entretanto, esses são pouco explorados. Todavia o motivo para isso não é claro (GERDS *et al.*, 2012).

Neste sentido, esse trabalho tem como objetivos:

- i) comparar a estimativa não paramétrica da função de incidência cumulativa obtida com o complemento do estimador de Kaplan-Meier e a estimativa não paramétrica considerando o risco competitivo;
- ii) comparar a abordagem baseada na modelagem do risco específico por causa com a abordagem baseada na modelagem do risco da subdistribuição;
- iii) avaliar o efeito da censura sobre a função de incidência cumulativa; e,
- iv) comparar três diferentes modelos para a função de incidência cumulativa, modelo com função de ligação logarítmica (ARR), modelo com função de ligação logit (LRR) e o modelo de Fine e Gray (FG).

Isto posto, o trabalho está estruturado da seguinte forma: no capítulo 1 aborda-se a fundamentação teórica, no capítulo 2, apresenta-se o artigo 1, cujo objetivo é avaliar o efeito da censura sobre a função de incidência, no capítulo

3, tem-se o artigo 2 que compara a abordagem baseada no risco específico por causa com a abordagem baseada no risco da subdistribuição considerando dados de lesões em cavalos e no capítulo 4 finaliza-se com o artigo 3, que faz, por meio de um estudo de simulação, uma comparação de diferentes modelos para o risco da subdistribuição.

CAPÍTULO 1

1. REFERENCIAL TEÓRICO

1.1. Análise de sobrevivência

A análise de sobrevivência é um conjunto de técnicas aplicadas quando a variável resposta é o tempo até a ocorrência de um determinado evento (KLEINBAUM e KLEIN, 2005). Esse tempo é denominado tempo de falha, podendo representar o tempo até a morte de um paciente, até a cura, recidiva de uma doença, a recuperação (por exemplo, volta ao trabalho) ou qualquer experiência de interesse que pode acontecer a um indivíduo (COLOSIMO e GIOLO, 2006; KLEINBAUM e KLEIN, 2005).

Conjuntos de dados de sobrevivência são definidos pelo tempo em que a falha ocorre e frequentemente pelas censuras. Conjunto de covariáveis para cada indivíduo geralmente também é medido (COLOSIMO e GIOLO, 2006).

Dados censurados representam a presença de informações incompletas e parciais, que surgem da perda de acompanhamento de um participante no decorrer do estudo ou a não ocorrência do evento de interesse até o final do experimento (COLOSIMO e GIOLO, 2006; HOUGGARD, 2000). Existem três mecanismos de censura: censura tipo I, censura tipo II e censura tipo III. A censura tipo I é aquela em que o experimento é finalizado após um determinado tempo, enquanto a censura do tipo II estabelece o final do estudo após um número pré-determinado de indivíduos experimentarem o evento de interesse e a do tipo III, do tipo aleatório, mais frequente em ensaios clínicos, caracteriza-se pela retirada do indivíduo do estudo sem a ocorrência da falha, ou quando o participante experimenta um evento diferente do estudado. Todos estes tipos de censura são conhecidos como censura à direita, já que a ocorrência do evento de interesse estará à direita do tempo registrado, ou seja, é maior que o tempo observado. Há também a censura à esquerda e a intervalar. A censura à esquerda ocorre quando o evento de interesse acontece antes do tempo observado, mas não sabemos o tempo exato da ocorrência do desfecho. (CARVALHO *et al.*, 2011; COLOSIMO e GIOLO, 2006; LEE e WANG, 2003). A

censura intervalar é o tipo mais geral, neste caso é sabido que o indivíduo experimentou o evento dentro de um intervalo de tempo. Esse tipo de censura aparece em estudos em que pacientes são acompanhados em visitas periódicas (COLLET, 2003; COLOSIMO e GIOLO, 2006). A censura ainda pode ser classificada em informativa e não informativa. A censura é dita informativa quando a perda de acompanhamento do indivíduo estiver associada ao desfecho em estudo, como exemplo, o abandono do tratamento devido a uma doença decorrente da piora do paciente. Se a perda de informação não está associada ao desfecho, isto é, ocorre por acaso, denomina-se a censura como não informativa (CARVALHO *et al.*, 2011).

Como frequentemente o conjunto de dados de falhas apresenta censura, os métodos estatísticos comuns não são adequados, tornando-se necessário técnicas específicas que acomodem a informação contida nestas observações (COLLET, 2003; COLOSIMO e GIOLO, 2006). Nesse sentido existem duas funções que desempenham papel fundamental na análise de dados de sobrevivência: a função de sobrevivência e a função de taxa de falha ou risco.

A função de sobrevivência, $S(t)$ é definida como a probabilidade de que o tempo de sobrevivência seja maior do que um tempo t , ou seja, $S(t) = P(T \geq t)$, em que T é uma variável aleatória não-negativa, contínua representando o tempo de falha (COLOSIMO e GIOLO, 2006). Essa função representa a probabilidade de uma observação sobreviver ao tempo t . Usando a função de sobrevivência pode-se definir a função acumulada $F(t)$, isto é, $F(t) = 1 - S(t)$, que representa a probabilidade de uma observação não sobreviver ao tempo t .

A função de taxa de falha ou risco é definida como:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (1)$$

Essa função é amplamente utilizada para expressar o risco de um indivíduo em algum tempo t (Collet, 2003), para descrever a distribuição do tempo de vida de pacientes e avaliar como a taxa de falha instantânea muda com o tempo (COLOSIMO e GIOLO, 2006). Outra função importante é a função de taxa de falha acumulada que é dada por:

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

Apesar da função de taxa de falha acumulada não ter uma interpretação prática, ela é útil na avaliação da taxa de falha $\lambda(t)$ (COLOSIMO e GIOLO, 2006).

Visando a estimação da curva de sobrevivência alguns estimadores não paramétricos podem ser encontrados na literatura sendo que o estimador mais utilizado é o estimador de Kaplan-Meier. Para a comparação das curvas de sobrevivência é frequentemente utilizado o teste logrank (COLOSIMO e GIOLO, 2006).

Como o objetivo em muitos estudos é avaliar o efeito de uma ou mais covariáveis sobre a sobrevivência, se somente uma covariável categórica é considerada, como por exemplo gênero ou estágio de uma doença, o efeito dessa covariável pode ser mensurado estimando a curva de sobrevivência para cada nível da covariável e em seguida testar se há diferença significativa entre os níveis da covariável (AALEN *et al.*, 2008).

Contudo, em muitos estudos podem existir várias covariáveis, algumas podendo ser contínuas, o que acarretará na divisão dos dados em estratos de acordo com as covariáveis e o posterior uso de técnicas não paramétricas, como o estimador de Kaplan-Meier e o teste logrank, por exemplo. Neste cenário o uso de modelos de regressão mostra-se adequado.

O modelo de regressão mais utilizado em análise de sobrevivência é o modelo de regressão de Cox (AALEN *et al.*, 2008). O modelo de Cox permite estimar o efeito de covariáveis sem a necessidade de suposições sobre a distribuição dos tempos de sobrevivência. Esse modelo apenas pressupõe a proporcionalidade dos riscos (CARVALHO *et al.*, 2011).

1.2. 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.

Os autores Dignam e Kocherginsky (2008) destacam que até no caso em que indivíduos podem experimentar múltiplos eventos, como por exemplo, uma recaída no mesmo local do tumor primário seguida de uma recaída em órgão distante daquele atingido pelo tumor primário, pode-se designar um evento que ocorre primeiro, criando assim informações com riscos competitivos. Os autores supracitados afirmaram ainda que embora existam métodos para lidar com eventos múltiplos por pacientes, denominados modelos multiestados, a complexidade analítica adicional requerida frequentemente não produz informações diferentes daquelas obtidas pela modelagem via definição de riscos competitivos. Detalhes sobre modelos multiestados podem ser encontrados em Andersen e Keiding (2002), Allignol *et al.* (2011), Arashiro (2008), Beyersmann *et al.* (2012), Hougaard (2000), Machado e Pardinias (2011), Machado *et al.* (2009), Putter *et al.* (2007), Titman (2007).

Diversos autores destacaram a aplicação de métodos de análise de sobrevivência com riscos competitivos na área médica (BEYERSMANN e SCHEIKE, 2014; KLEIN e BAJORUNAITE, 2004; GOOLEY *et al.*, 1999, REN *et al.*, 2013) em especial em estudos referentes a oncologia e medicina cardiovascular (NOORDZIJ *et al.*, 2013). Assane *et al.* (2015) estudaram a incidência de três desfechos distintos para os quais 572 pacientes com hipertensão arterial, acompanhados pelo Programa de Hipertensão Arterial do Hospital Universitário Clementino Fraga Filho da UFRJ estavam em risco: evento cardiovascular não fatal, óbito por causa cardiovascular e óbito por outras causas. Dianatkah *et al.* (2014), avaliaram diferentes modelos para a função de incidência cumulativa no estudo de fatores de risco para doenças cardiovasculares (evento de interesse) e para morte não relacionada a doença cardiovascular. Zhang *et al.* (2013) usaram a estrutura de riscos competitivos

para analisar o efeito de fatores de risco para mortalidade a partir do câncer de mama e para a mortalidade decorrente de outras causas. Gillam *et al.* (2011) usaram metodologia de análise de sobrevivência com riscos competitivos para estudar os fatores associados ao risco de revisão do procedimento de artroplastia na presença do risco competitivo de morte. Logan *et al.* (2006) discutiram modelos de regressão para riscos competitivos para a modelagem de dados de transplante de células tronco.

Na análise de sobrevivência clássica, ou seja, quando somente uma causa de falha é considerada, uma maneira de descrever os dados é por meio da curva de sobrevivência obtida pelo estimador de Kaplan-Meier. No cenário de riscos competitivos a função de distribuição acumulada, $F(t) = 1 - S(t)$, estimada com base no estimador de Kaplan-Meier é superestimada (CARVALHO *et al.*, 2011; HALLER *et al.*, 2013). Isso ocorre porque a estimativa de Kaplan-Meier ignora todos os tipos de eventos diferentes daquele de interesse, considerando-os como censura. Gooley *et al.* (1999) em seu trabalho discutiram a inadequacidade do uso do complemento do estimador de Kaplan-Meier para a probabilidade de falha quando riscos competitivos estão presentes.

Objetivando contornar esse problema, foi proposto a função de incidência acumulada (ou cumulativa) (FIC ou CIF, em inglês, cumulative incidence function), também denominada como subdistribuição. (CARVALHO *et al.*, 2011).

A função de incidência cumulativa permite quantificar a probabilidade de um indivíduo experimentar um evento pertencente a um conjunto de causas distintas e a ela tem sido dado diferentes nomes: risco absoluto, risco bruto, função de incidência cumulativa específica por causa, função de incidência cumulativa bruta, curva de incidência bruta, função de probabilidade bruta, probabilidade real, probabilidade de falha por causa específica, função subdistribuição e incidência cumulativa (GESKUS, 2015).

A FIC juntamente com a função risco específico por causa desempenham um papel importante na análise de sobrevivência com riscos competitivos.

Em riscos competitivos, dois riscos podem ser definidos. Existe um risco que tem uma relação unívoca com a incidência cumulativa, mais comumente denominado de risco da subdistribuição ou subrisco. Este risco representa a fração de indivíduos que experimentarão o evento de interesse em algum tempo, entre os indivíduos que estão em risco ou entre aqueles que experimentaram o

evento competitivo. O segundo tipo de risco é aproximado pelo conceito padrão, isto é, o conceito utilizado na análise de sobrevivência clássica, representando a fração de indivíduos que experimentarão o evento de interesse em algum tempo dentre aqueles que estão em risco. Esse último tipo de risco é mais frequentemente chamado de risco específico por causa e não tem uma relação unívoca com a incidência cumulativa (GESKUS, 2015). A partir desses dois tipos de riscos, duas formas de analisar problemas com eventos competitivos são possíveis, a modelagem do risco específico por causa e a modelagem da função de incidência cumulativa (ou subdistribuição).

1.2.1. Notação básica

Sejam T_i e C_i , respectivamente, o tempo em que o evento ocorre 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). Observa-se n realizações independentes e identicamente distribuídas de $\{\tilde{T}_i, \Delta_i, \Delta_i \varepsilon_i, Z_i, X_i\}$ para $i = 1, \dots, n$ em que $Z_i = (1, Z_{i1}, \dots, Z_{iq})^T$ e $X_i = (X_{i1}, \dots, X_{ip})^T$ são vetores de covariáveis (SCHEIKE e ZHANG, 2011).

O risco específico por causa, a função de incidência cumulativa bem como o risco associado a essa função, denominado como risco da subdistribuição, são algumas quantidades importantes na análise de sobrevivência com riscos competitivos.

1.2.1.1. Função risco específico por causa

A função risco específico por causa para cada tipo de falha k , $\lambda_k(t)$, 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

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

A taxa de risco total é dada pela soma dos riscos específicos por causa, isto é

$$\lambda(t) = \sum_{\varepsilon=1}^k \lambda_{\varepsilon}(t). \quad (4)$$

1.2.1.2. Função de incidência cumulativa (FIC)

A função de incidência cumulativa, $F_{\varepsilon}(t)$, para um evento do tipo ε ($\varepsilon = 1, 2, \dots, k$) é a probabilidade conjunta dada por $F_k(t) = P(\tilde{T} \leq t, \varepsilon = k)$, em que ε representa a causa de falha. Essa função representa a probabilidade de um indivíduo experimentar um evento do tipo k até o tempo t . A função de incidência total é igual a soma das FICs para todos os tipos de eventos (PINTILIE, 2006). Portanto,

$$F(t) = P(\tilde{T}_i \leq t) = \sum_{\varepsilon=1}^k P(\tilde{T} \leq t, \varepsilon = k) = \sum_{\varepsilon=1}^k F_{\varepsilon}(t). \quad (5)$$

A função risco da subdistribuição é definida como

$$\gamma_k(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{P(t < \tilde{T} \leq t + \Delta t, \varepsilon = k | \tilde{T} > t \text{ ou } (\tilde{T} \leq t \text{ e } \varepsilon \neq k))}{\Delta t} \right\} \quad (6)$$

e representa a probabilidade de um indivíduo experimentar um evento k de interesse considerando-se os indivíduos sob risco (aqueles que não falharam) e aqueles que falharam devido ao risco competitivo. A função risco da subdistribuição possui uma relação unívoca com a função de incidência cumulativa permitindo que o efeito de covariáveis sob este risco possa ser traduzido em efeito sobre a função de incidência cumulativa (GESKUS, 2015).

1.3. Estimador para a função de incidência cumulativa.

Na presença de riscos competitivos, a base da estatística descritiva da ocorrência de eventos não é mais a função de sobrevivência, alternativamente, é usada a função de incidência cumulativa para cada evento (HOSMER *et al.*, 2008).

Para a estimar a função de incidência cumulativas considere $0 < t_1 < t_2 < \dots < t_p$ tempos ordenados da ocorrência de qualquer um dos eventos, d_{ij} o número de eventos do tipo i que ocorre no tempo t_j e n_j o número de indivíduos sob risco em t_j . Considere $\hat{S}(t)$ o estimador de Kaplan-Meier da probabilidade de estar livre de eventos no tempo t , isto é, conforme Carvalho *et al.* (2011), a sobrevivência a toda e qualquer causa. A função de incidência cumulativa pode ser obtida pela soma em todos os t_j das probabilidades de se observar um evento k entre os indivíduos que permanecem em risco: ou seja, indivíduos que não experimentaram qualquer um dos eventos possíveis (CARVALHO *et al.*, 2011; PINTILIE, 2006). A probabilidade conjunta de permanecer livre de evento antes do tempo t_j e experimentar um evento do tipo k no tempo t_j , $F_k(t) = P(T \leq t, \varepsilon = k)$, é dada por

$$\hat{F}_k(t) = \sum_{\forall t, t_j \leq t} \hat{\lambda}_{kj} \hat{S}(t_{j-1}) \quad (7)$$

em que $\hat{\lambda}_{kj}$ é risco específico por causa para o evento k no tempo t_j . De forma intuitiva, dado que um indivíduo sob risco ainda não experimentou nenhum evento até o tempo t_{j-1} , este risco pode ser estimado como $\hat{\lambda}_{kj} = \frac{d_{kj}}{n_j}$, segue, portanto,

$$\hat{F}_k(t) = \sum_{\forall t, t_j \leq t} \frac{d_{kj}}{n_j} \hat{S}(t_{j-1}) \quad (8)$$

em que k representa o tipo de falha, d_{kj} o número de indivíduos que experimentaram o evento k no tempo t_j ; n_j o número de indivíduos em risco no

tempo t_j e $\hat{S}(t)$ o estimador de Kaplan-Meier da probabilidade geral de sobrevivência (PINTILIE, 2006).

1.4. Modelo de regressão de Cox para o risco específico por causa

A modelagem dos riscos específicos é uma das maneiras que a literatura apresenta para analisar dados de sobrevivência em que riscos competitivos estão presentes.

A ideia dessa abordagem é especificar algum tipo de modelo de regressão no qual os efeitos das covariáveis são quantificados por meio de parâmetros (GESKUS, 2015). Para cada causa de falha é ajustado um modelo de regressão, para tanto, trata-se os demais eventos como censura informativa (CARVALHO *et al.*, 2011). Além disso, assume-se que as covariáveis têm efeitos específicos por causa, ou seja, efeitos distintos sobre os vários riscos específicos por causa (KLEIN *et al.*, 2014).

Entre os modelos de regressão para o risco específico por causa, o mais popular é o modelo de riscos proporcionais (BAKOYANNIS e TOULOUMI, 2012; GESKUS, 2015; HA *et al.*, 2014, LATOUCHE *et al.*, 2013; MARTINUSSEN e SCHEIKE, 2006; PUTTER *et al.*, 2007).

O modelo de riscos proporcionais supõe que cada risco específico por causa segue o modelo de Cox (COX, 1972). Consequentemente, o efeito das covariáveis sobre o risco específico por causa é descrito por

$$\lambda_k(t|X_i) = \lambda_{k0}(t) \exp(\beta_k^t X_i) \quad (9)$$

em que β_k é um vetor de coeficientes de regressão para a causa k, X_i o vetor de covariáveis e $\lambda_{k0}(t)$ é o risco específico por causa de base para causa k.

A estimação do modelo de Cox para o risco específico por causa para um desfecho k é realizada censurando os indivíduos que experimentaram um evento diferente do evento k. Sendo assim, o procedimento de estimação é o mesmo procedimento utilizado na estimação do modelo de Cox clássico, isto é, em que um único desfecho é considerado (GESKUS, 2015).

Segue, então, que a função de verossimilhança para o modelo de Cox para o risco específico por causa é

$$L(\beta_k) = \prod_{i=1}^k \frac{\exp\{\beta_k^T X_i\}}{\sum_{j \in R(t_i)} \exp\{\beta_k^T X_j\}} \quad (10)$$

em que $R(t_i)$ é o conjunto das observações sob risco no tempo t_i .

Os valores de β_k que maximizam a função de verossimilhança são obtidos pela solução do sistema $U(\beta_k) = 0$ em que

$$U(\beta_k) = \sum_{i=1}^n \left[X_i - \frac{\sum_{j \in R(t_i)} X_j \exp\{\beta_k^T X_j\}}{\sum_{j \in R(t_i)} \exp\{\beta_k^T X_j\}} \right] \quad (11)$$

em que $U(\beta_k)$ é o vetor escore de derivadas de primeira ordem do logaritmo da função de verossimilhança.

1.5. Modelo de regressão de Fine e Gray para o risco da subdistribuição

Na análise de sobrevivência com uma única causa de falha, o modelo de riscos proporcionais de Cox é amplamente utilizado, porém na presença de eventos competitivos esse modelo trata esses eventos competitivos como observações censuradas. Além disso, a função risco específico por causa modelada pelo uso da regressão de Cox não tem uma interpretação direta em termos da função de incidência cumulativa (SCRUCCA *et al.*, 2010), o que tem motivado esforços para a modelagem direta do efeito de covariáveis sobre a função de incidência cumulativa. Nesse sentido Fine e Gray (1999) propuseram um modelo para o risco da subdistribuição da função de incidência cumulativa.

A abordagem direta da função de incidência cumulativa é baseada no risco da subdistribuição de Gray (1988), para a k -ésima causa de falha, que é dado por

$$\gamma_k(t|X) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{P(t < \tilde{T} \leq t + \Delta t, \varepsilon = k | \tilde{T} > t \text{ ou } (\tilde{T} \leq t \text{ e } \varepsilon \neq k))}{\Delta t} \right\} = \frac{-d \log\{1 - F_k(t)\}}{dt} \quad (12)$$

Como há uma relação direta entre a função de incidência cumulativa e a função risco da subdistribuição (ZHANG *et al.*, 2008):

$$F_k(t|X) = 1 - \exp \left[- \int_0^t \gamma_k(u|X) du \right] \quad (13)$$

pode-se, portanto, interpretar o efeito da covariável sobre a função de incidência cumulativa.

O modelo de Fine e Gray estabelece a suposição de proporcionalidade para os riscos da subdistribuição e a expressão para a k-ésima causa de falha pode ser escrita como

$$\gamma_k(t|X) = \gamma_{k0}(t) \exp(\beta_k^T X), \quad (14)$$

em que γ_k é o risco da subdistribuição, $\gamma_{k0}(t)$ é o risco de base da subdistribuição, X é o vetor de covariáveis e β_k é o vetor de coeficientes de regressão. Esse modelo tem semelhanças com o modelo de regressão de Cox e foi motivado como um modelo tipo Cox para o risco da subdistribuição (KLEIN *et al.*, 2014).

A função de verossimilhança parcial para o modelo de riscos proporcionais para a subdistribuição é dada por

$$L(\beta_k) = \prod_{j=1}^r \frac{\exp(\beta_k^T X_j)}{\sum_{i \in R_j} w_{ji} \exp(\beta_k^T X_i)} \quad (15)$$

em que R_j é o conjunto de risco definido como

$R_j = \{i; T_i \geq t \text{ ou } (T_i \leq t \text{ e o indivíduo experimentou um evento de risco competitivo})\}$.

Assim o conjunto de indivíduos sob risco inclui, além dos indivíduos que não sofreram qualquer evento, aqueles que sofreram o evento competitivo, e que até a censura administrativa não saem do conjunto de risco (CARVALHO *et al.*,

2011). Contudo esses indivíduos não contribuem da mesma maneira para a função de verossimilhança sendo necessário a atribuição de peso que muda a cada tempo no qual ocorre o evento de interesse (CARVALHO *et al.*, 2011). Esses pesos são calculados para todos os indivíduos em cada momento t_j no qual ocorre um evento de interesse e são definidos como

$$w_{ji} = \frac{\hat{G}(t_j)}{\hat{G}(\min(t_j, t_i))} \quad (16)$$

em que \hat{G} é o estimador de Kaplan-Meier da distribuição da censura. A distribuição da censura é definida por (T_i, ε_i) , em que T_i é o tempo para o primeiro evento e ε_i é 1 se nenhum evento foi observado e 0 se algum tipo de evento foi observado (PINTILIE, 2006). Cada tempo em que o evento de interesse ocorre é indexado por j . Assim para o indivíduo L o peso será dado por (CARVALHO *et al.*, 2011)

$$w_{ji} = \begin{cases} 1, & \text{se o indivíduo } L \text{ não sofrer evento ou censura em } t_j \\ \frac{\hat{G}(t_j)}{\hat{G}(t_i)}, & \text{se o indivíduo } L \text{ sofrer um evento competitivo em } t_i < t_j \\ 0, & \text{se o indivíduo } L \text{ for retirado do estudo} \end{cases}$$

O valor de β_k que maximiza a função de verossimilhança parcial, $L(\beta_k)$, é obtido resolvendo o sistema $U(\beta_k) = 0$, em que $U(\beta_k)$ é dado por

$$U(\beta_k) = \sum_{j=1}^r \left\{ x_j - \frac{\sum_{i \in R_j} w_{ji} X_i \exp(\beta_k^T X_i)}{\sum_{i \in R_j} w_{ji} \exp(\beta_k^T X_i)} \right\}. \quad (17)$$

$U(\beta_k)$ é o vetor escore de derivadas de primeira ordem do logaritmo da função de verossimilhança.

A flexibilidade e as várias propriedades herdadas do modelo de Cox são vantagens do modelo de Fine e Gray (KLEIN *et al.*, 2014). Nesse sentido para a interpretação dos coeficientes estimados deve-se considerar as propriedades de riscos proporcionais do modelo e, portanto, a interpretação da razão dos

riscos da subdistribuição é semelhante a interpretação da razão dos riscos do modelo de Cox (DIGNAM *et al.*, 2012).

Tomando as razões das taxas do risco da subdistribuição para dois indivíduos i e j , para uma covariável l e mantendo as demais constantes tem-se a razão dos riscos da subdistribuição:

$$\frac{\gamma_i(t)}{\gamma_j(t)} = \exp\{\beta_l(x_{il} - x_{jl})\} = \exp(\beta_l). \quad (18)$$

A razão dos riscos da subdistribuição para covariáveis dicotômicas é a razão do risco de um grupo atual em relação a um grupo de base. Se a covariável é contínua, esta razão refere-se ao efeito do aumento de uma unidade na covariável analisada (LAU *et al.*, 2009; SCRULLA *et al.*, 2010).

1.6. Modelo de regressão binomial

Na modelagem de dados com riscos competitivos, a incidência cumulativa frequentemente é de interesse. A abordagem direta da função de incidência cumulativa é baseada no risco da subdistribuição de Gray (1988), em que o risco da subdistribuição é dado conforme Equação (12).

Sendo que, para a modelagem da função de incidência cumulativa frequentemente tem-se utilizado o modelo de Fine e Gray (1999). O modelo de Fine e Gray assume que a incidência cumulativa para a causa k e sujeito i é dada de acordo com a Equação (13).

Deste modo, quando $\beta_j > 0$ tem-se um aumento na incidência cumulativa e se $\beta_j < 0$ implicará em uma diminuição (SCHEIKE *et al.*, 2008).

Decorrente da flexibilidade do modelo de Fine e Gray, de sua capacidade de herdar propriedades do modelo de Cox, além de sua implementação em softwares estatísticos como o R e o Stata, o modelo de Fine e Gray é o mais popular na modelagem da função de incidência com riscos competitivos.

Contudo pode-se considerar para a modelagem da função de incidência uma classe de modelos da forma

$$h\{F_k(t, z, x)\} = z^T \beta_0(t) + g(x, \beta, t) \quad (19)$$

em que h e g são funções de ligação e $\beta_0(t)$ e β são vetores de coeficientes de regressão (SCHEIKE E ZHANG, 2011) .

Várias funções de ligação podem ser utilizadas, e ao se utilizar a função de ligação complementar log-log obtém-se o modelo proporcional

$$cloglog\{1 - F_k(t, z, x)\} = z^T \beta_0(t) + g(x, \beta, t). \quad (20)$$

O modelo de Fine e Gray é um caso particular do modelo apresentado na Equação (19), que é referido como modelos de transformação (SCHEIKE e ZHANG, 2011).

Além da função de ligação complementar log-log, outras funções têm sido sugeridas na modelagem da incidência cumulativa. Entre elas, a função de ligação logit e a função de ligação logarítmica. Contudo, modelos alternativos ao de Fine e Gray (ligação complementar log-log) existam, são pouco utilizados na prática e a razão para isso não é clara (GERDS *et al.*, 2012).

Klein *et al.* (2014) discutem a modelagem da função de incidência cumulativa considerando funções de ligação logit, logarítmica e complementar log-log empregando a abordagem da regressão binomial.

1.6.1. Regressão binomial

A regressão binomial é uma técnica usada para prever o estado de um evento binário em tempos futuros. O uso desta técnica em estudos envolvendo tempos de eventos se pauta no fato de que, em um instante t , após o início do estudo o status do evento é binário, tomando o valor 1 (um) se o evento ocorre e 0 (zero) caso contrário (KLEIN *et al.*, 2014).

Para a discussão de opções de modelagem, um modelo linear generalizado para o risco absoluto (incidência cumulativa) de um evento ε no tempo t , é considerado

$$F_\varepsilon(t|X) = h(\beta_0(t), \beta(t), X) \quad (21),$$

esse modelo descreve o efeito de variáveis preditores X medidas no tempo de origem por meio de um vetor de coeficientes de regressão dependentes do tempo, $\beta(t)$ e do risco de base $\beta_0(t)$.

1.6.2. Funções de ligação para a regressão binomial

1.6.2.1. Função de ligação logit

A escolha de $h(a, b, x) = \frac{\exp(a+b^T x)}{1+\exp(a+b^T x)}$ define um modelo de regressão no qual os coeficientes são log das razões das probabilidades (chances).

$$OR(t) = \frac{F_\varepsilon(t|X_1 + 1, X_2, \dots, X_p) / (1 - F_\varepsilon(t|X_1 + 1, X_2, \dots, X_p))}{F_\varepsilon(t|X_1, X_2, \dots, X_p) / (1 - F_\varepsilon(t|X_1, X_2, \dots, X_p))},$$

$$OR(t) = \frac{\exp(\beta_0(t) + \beta_1(t)(X_1 + 1) + \beta_2(t)X_2 + \dots + \beta_p(t)X_p)}{\exp(\beta_0(t) + \beta_1(t)X_1 + \beta_2(t)X_2 + \dots + \beta_p(t)X_p)},$$

$$OR(t) = \exp(\beta_1(t)). \quad (22)$$

Se o coeficiente não depende do tempo o modelo é considerado de chances proporcionais.

1.6.2.2. Função de ligação log

A escolha de $h(a, b, x) = \exp(a + b^T x)$ dá origem a um modelo de regressão no qual os coeficientes são razões dos riscos absolutos (funções de incidência).

$$ARR(t) = \frac{F_\varepsilon(t|X_1 + 1, X_2, \dots, X_p)}{F_\varepsilon(t|X_1, X_2, \dots, X_p)},$$

$$ARR(t) = \frac{\exp(\beta_0(t) + \beta_1(t)(X_1 + 1) + \beta_2(t)X_2 + \dots + \beta_p(t)X_p)}{\exp(\beta_0(t) + \beta_1(t)X_1 + \beta_2(t)X_2 + \dots + \beta_p(t)X_p)},$$

$$ARR(t) = \exp(\beta_1(t)). \quad (23)$$

1.6.2.3. Função de ligação complementar log-log

Outra função comumente usada é a função de ligação $h(a, b, x) = 1 - \exp(-\exp(a + b^T x))$ que na presença de risco competitivo produz o modelo de Fine e Gray (Klein *et al.*, 2014).

1.6.3. Resposta ponderada

Considere T_i, C_i e G_i , sendo T_i o tempo do evento, C_i o tempo de censura à direita e G ; a função de sobrevivência para o tempo de censura condicionada às covariáveis para o indivíduo i . Seja $\varepsilon_i \in \{1, \dots, k\}$ o tipo da causa de falha. Assume-se que (T_i, ε_i) são independentes de C_i dado as covariáveis. Denote $G(t|X) = P(C > t|X)$, a probabilidade condicional de uma observação não ser censurada no tempo t . Considere, também, a variável indicadora de censura $\Delta_i = I\{T_i \leq C_i\}$, tal que, $\Delta = 1$ se o tempo do evento é observado e $\Delta = 0$ quando o tempo do evento é censurado. Os dados para cada indivíduo é então $Y_i = (X_i, \tilde{T}_i, \Delta_i \varepsilon_i)$ em que $\tilde{T}_i = \min(T_i, C_i)$.

Seja $N_i(t) = I\{T_i \leq t, \varepsilon_i = 1\}$ o processo de contagem associado a causa 1, que não são observados em todo t . Contudo $\Delta_i N_i(t)$ são computáveis para todo t .

A ideia do processo de estimação é substituir o status do evento $N_\varepsilon(t)$ por uma resposta ponderada $\tilde{N}_\varepsilon(t) = \frac{\Delta N_\varepsilon(t)}{G(T^-|X)}$. Sob suposição de independência $E(\Delta|X, T, G) = G(T^-|X)$ e assim

$$E(\tilde{N}_\varepsilon(t)) = E\left(E\left(\frac{\Delta N_\varepsilon(t)}{G(T^-|X)} \mid T, \varepsilon, X\right)\right) = E(N_\varepsilon(t)|X) = F_\varepsilon(t|X). \quad (24)$$

Tendo em vista que o numerador da resposta ponderada é 0 (zero) quando $T > t$, pode-se substituir o denominador por $G(\min(t, T^-)|X)$.

Para usar a resposta ponderada nas equações de estimação é necessário estimar a distribuição da censura. No caso em que é razoável assumir que o mecanismo de censura independe das covariáveis, o estimador de Kaplan-Meier baseado nos tempos de censura estima consistentemente G (KLEIN *et al.*, 2014).

Para estimar G nas situações em que a censura depende das covariáveis considera-se modelos de regressão para os tempos de censura, tais como o modelo de Cox ou modelo aditivo de Aalen (SCHEIKE *et al.*, 2008).

1.6.4. Equações de estimação ponderada

Considerando-se os dados censurados à direita de n indivíduos independentes, $\beta(t)$ soluciona a equação de estimação $U_n(\beta_0(t), \beta(t), G) = 0$ (KLEIN *et al.*, 2014), sendo

$$U_n(\beta_0(t), \beta(t), \hat{G}) = \sum_{i=1}^n w(t, x_i) \left\{ \frac{\Delta_i N_{\varepsilon, i}(t)}{\hat{G}(T_i|X_i)} - h(\beta_0(t), \beta(t), X_i) \right\} \quad (25)$$

Os pesos $w(t, x_i)$ podem ser p -dimensional e depender da função de ligação.

Conforme Klein *et al.* (2014) as equações de estimação para coeficientes constantes no tempo $\bar{\beta}$ são obtidas em relação a uma sequência de tempos t_1, \dots, t_j e são soluções de

$$\bar{U}_n(\beta_0(t), \bar{\beta}, \hat{G}) = \sum_{i=1}^n \sum_{j=1}^J w(t_j, x_i) \left\{ \frac{\Delta_i N_{\varepsilon, i}(t_j)}{\hat{G}(T_i|X_i)} - h(\beta_0(t_j), \beta, X_i) \right\} \quad (26)$$

1.7. Simulação de dados com riscos competitivos

Estudos de simulação auxiliam na identificação do melhor modelo para uma dada situação, bem como na comparação de métodos relevantes sob

diferentes cenários, para tanto, faz-se necessário um algoritmo correto para a geração de dados (HALLER, 2014).

A simulação de tempos de eventos na presença de riscos competitivos é discutida em Beyersmann *et al.* (2012) e Beyersmann *et al.* (2009) e métodos baseados nos riscos específicos por causa, assim como, baseados em modelos pré-definidos para o risco da subdistribuição, denominado simulação direta, são apresentados.

Na simulação direta gera-se dados de riscos competitivos tais que o risco da subdistribuição da função de interesse (evento 1) siga o modelo dado pela equação

$$\gamma_k(t|X) = \gamma_{k0}(t) \exp(\beta_k^T X) \quad i = 1, \dots, n. \quad (27)$$

De acordo com Beyersmann *et al.* (2012) o algoritmo para simulação dos dados opera em três passos: i) determinação do risco de base para a subdistribuição e para o risco específico por causa; ii) determinação do modelo para os riscos específicos por causa; ii) gerar os dados a partir dos modelos definidos.

Em síntese, uma maneira para gerar os tempos de eventos com riscos competitivos é descrita abaixo:

1) Escolha $\lambda_{10}(t)$ e $\gamma_{10}(t)$, em seguida, calcule $\lambda_{20}(t)$ usando a expressão

$$\lambda_{20}(t) = \gamma_{10}(t) - \lambda_{10}(t) - \frac{d}{dt} \ln \left(\frac{\gamma_{10}(t)}{\lambda_{10}(t)} \right) \quad (28)$$

em $\lambda_{k0}(t)$ é o risco específico de base para a causa de falha k e $\gamma_{10}(t)$ é o risco de base para a subdistribuição.

2) Especificar modelos para a função risco específico por causa 1, $\lambda_1(t|X)$, e para a subdistribuição 1, $\gamma_1(t|X)$, em seguida, determinar o modelo para a função risco específico por causa 2, $\lambda_2(t|X)$, usando a equação

$$\lambda_2(t|X) = \gamma_1(t|X) - \lambda_1(t|X) - \frac{d}{dt} \ln \left(\frac{\gamma_1(t|X)}{\lambda_1(t|X)} \right). \quad (29)$$

3) Gerar os dados a partir dos modelos para o risco específico por causa.

Beyersmann *et al.* (2012) enfatizaram que as funções riscos devem ser diferenciáveis, integráveis e não negativas. Além disso, na presença de risco competitivo, a função de incidência cumulativa para o evento 1, F_1 , relaciona-se ao risco cumulativo da subdistribuição através de $F_1(t) = 1 - \exp(-\Gamma_1(t|X))$, em que, $\Gamma_1(\cdot)$ é o risco cumulativo para a subdistribuição 1, e que converge para $P(\varepsilon = 1)$ quando $t \rightarrow \infty$ (HALLER, 2014).

2. REFERÊNCIAS BIBLIOGRÁFICAS

AALLEN, O. O.; BORGAN, O.; GJESSING, H. K. **Survival and event history analysis: a process point of view**. New York: Springer, 2008. 539 p.

ANDERSEN, P. K.; KEIDING, N. Multi-state models for event history analysis. **Statistical Methods in Medical Research**, v. 11, n. 2, p. 91 – 115, 2002.

ALLIGNOL, A.; SCHUMACHER, M.; BEYERSMANN, J. Empirical transition matrix of multi-state models: the etm package. **Journal of Statistical Software**, v. 38, n. 4, p. 1 - 15, 2011.

ARASHIRO, I. H. I. **Modelos multi-estados markovianos não homogêneos com efeito dinâmico**. 2008. 110 f. Tese (Doutorado em Ciências, Área de concentração: Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2008.

ASSANE, C.; PEREIRA, B. de; MUAUALO, M. A. M. Modelos de riscos competitivos para análise de sobrevivência de pacientes com hipertensão arterial em tratamento no HUCFF/UFRJ. In: **Simpósio Brasileiro de Pesquisa Operacional**, XLVII, 2015, Pernambuco. *Anais ... Pernambuco* [s.n], 2015. p.1225 – 1236.

BAKOYANNIS, G.; TOULOUMI, G. Practical methods for competing risk data: A review. **Statistical Methods in Medical Research**, v. 21, n. 3, p. 257 – 272, 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. (eds). **Handbook of survival analysis**. New York: Taylor & Francis Group, 2014, cap.8, p. 157 – 177.

BEYERSMANN, J.; ALLIGNOL, A.; SCHUMACHER, M. **Competing risks and multistate models with R**. Springer Science & Business Media, 2012. 245 p.

BEYERSMANN J., LATOUCHE A, BUCHHOLZ A., SCHUMACHER M. Simulating competing risks data in survival analysis. **Statistics in Medicine**, v. 28, n. 6, 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. ver.ampl. Rio de Janeiro: Editora Fiocruz, 2011. 432 p.

COLLET, D. **Modelling Survival Data in Medical Research**. 2. ed. London: Chapman and Hall, 2003. 391 p.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. 1. ed., São Paulo: Edgard Blucher, 2006. 370 p.

COX, D. Regression models and life tables. **Journal of the Royal Statistical Society**, v. 34, n. 2, p. 187 – 220, 1972.

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.

DIGNAM, J. J; KOCHERGINSKY, M. N. Choice and interpretation of statistical tests used when competing risk are present. **Journal of Clinical Oncology**, v. 26, n. 24, p. 4027 – 4034, 2008.

DIGNAM, J. J; ZHANG, Q.; KOCHERGINSKY, M.N. The use and interpretation of competing risks regression models. **Clinical Cancer Research**, v. 18, n. 8, p. 2301 - 2308, 2012.

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.

FINE, J. P.; GRAY, R. J. A proportional hazard model for subdistribution of a competing risk. **Journal of the American Statistical Association**. v. 94, n. 446, p. 496 – 509, 1999.

GERDS, T.A.; SCHEIKE, H.T.; ANDERSEN, P. K. Absolute risk regression on for competing risk: interpretation, link functions, and prediction. **Statistics in Medicine**, v. 31, n. 29, p. 3921-3930, 2012.

GESKUS, R. B. **Data analysis with competing risks and intermediate states**. Boca Raton, Fl: Chapman & Hall/CRC, 2015. 243 p.

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. 18, n. 6, p. 695 – 706, 1999.

GRAY, R. J. A class of k-sample tests for comparing the cumulative incidence of a competing risks. **Annals of Statistics**, v. 16, n. 3, p. 1141 – 1154, 1988.

GRAW, F.; GERDS, T. A.; SCHUMACHER, M. On pseudo-values for regression analysis in competing risks models. **Lifetime Data Anal**, v. 15, p. 241 – 255, 2009.

HA, I. D.; LEE, M.; OH, S.; JEONG, J. H.; SYLVESTER, R. Variable selection in subdistribution hazard frailty models with competing risks data . **Statistics in Medicine**, v. 33, n. 26, p. 4590 – 4604, 2014.

HALLER, B.; SCHMIDT, G.; ULM, K. Applying competing risks regression models: an overview. **Lifetime Data Analysis**, v. 19, n. 1, p. 33 - 58, 2013.

HALLER, B. **The analysis of competing risks data with a focus on estimation of cause-specific and subdistribution hazard ratios from a mixture model**, 2014, 199 p. Dissertation – Fakultät für Mathematik, Informatik und Statistik der Ludwig – Maximilians – Universität München.

HOSMER, D. W.; LEMESHOW, S.; MAY, S. **Regression Modeling of time to event data**. 2 ed. New York: John Wiley & Sons, 2008. 392 p.

HOUGAARD, P. **Analysis of multivariate survival data**. 1.ed., New York: Springer, 2000. 542 p.

KLEIN, J. P.; BAJORUNAITE, R. Inference for competing risks. In: BALAKRISHNAN, C. R.; RAO, R.C. **Handbook of Statistics**. North Holland, 2004, v. 23, cap. 16, p. 291 – 311.

KLEIN, J. P.; HOUWELINGEN, H. C. V.; IBRAHIM, J. G.; SCHEIKE, T. H. (eds). **Handbook of survival analysis**. New York: Taylor & Francis Group, 2014. 632 p.

KLEINBAUM, D. G.; KLEIN, M. **Survival analysis: A self-learning text**. 2.ed, New York: Springer, 2005. 590 p.

KUK, D.; VARADHAN, R. Model selection in competing risk regression. **Statistics in Medicine**, v. 32, n. 18, p. 3077 – 3088, 2013.

LATOUCHE, A.; ALLIGNOL, A.; BEYERSMANN, J. LABOPIN, M.; FINE, J. P. A competing risks analysis should report results on all cause-specific hazard e cumulative incidence function. **Journal of Clinical Epidemiology**, v. 66, n. 6, p. 648 – 653, 2013.

LAU, B.; STEPHEN, R. C.; GANGE, S. Competing risk regression models for epidemiologic data. **American Journal of Epidemiology**, v. 170, n. 2, p. 244 – 259, 2009.

LEE, E. T.; WANG, J. W. **Statistical methods for survival data analysis**. 3. Ed. New Jersey: John Willey & Sons, 2003. 513 p.

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.

MACHADO, L. M.; ALVAREZ, J. de U.; SUAREZ, C. C.; ANDERSEN, P. K. Multi-state models for the analysis of time-to-event-data. **Statistical Methods in Medical Research**, v. 18, n. 2, p. 195 – 222, 2009.

MACHADO, L. M.; PARDINAS, J. R. p3state.msn: Analyzing survival data from an Illness-Death model. **Journal of Statistical Software**, v. 38, n. 3, p. 1 - 18, 2011.

MARTINUSSEN, T.; SCHEIKE, T. H. **Dynamic Regression Models for survival data**. New York: Springer, 2006. 470 p.

NOORDZIJ, M.; LEFFONDRÉ, K.; STRALEN, K. J. V.; ZOCCALI, C. When do we need competing risk methods for survival analysis in nephrology? **Nephrology Dialysis Transplantation**, v. 28, n. 1, p. 2670 - 2677, 2013.

PINTILIE, M. **Competing risk: a practical perspective**. Chichester: John Wiley & Sons, 2006. 224 p.

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, 2007.

REN, Y.; CHANG, C. C. H; ZENAROSA, G. L.; TOMKO, H. E.; DENNELL, D. M.; KANG, H. J., ROBERTS, M. S.; BRYCE, C. L. Gray's Time-varying coefficients model for posttransplant survival of pediatric liver transplant recipients with a diagnosis of cancer. **Computational and Mathematical Methods in Medicine**, v. 2013, 2013.

SARKER, S.; MIN, D. K.; BLACK, T. R.; LIM, H. J. Application of multiplicative and additive hazard models to injury prevention among healthcare workers. **Advances in Research**, v. 5, n. 1, p. 1 – 12, 2015.

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

SCHEIKE, T. H.; ZHANG, M. J. ; GERDS, T. A. Predicting cumulative incidence probability by direct binomial regression. **Biometrika**, v. 95, n. 1, p. 205-220, 2008.

SCRUCCA, L.; SANTUCCI, A.; AVERSA, F. Regression modeling of competing risk using R: an in depth guide for clinicians. **Bone Marrow Transplantation**, v. 45, n. 9, p. 1388 – 1395, 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.

TITMAN, A. C. **Model diagnostics in multi-states models of biological systems**. 2007. 249 f. Thesis (Doctor of Philosophy) – Department of Pure Mathematics and Mathematics Statistics, University of Cambridge, 2007.

ZHANG, M. J.; ZHANG, X.; SCHEIKE, T. H. Modeling cumulative incidence function for competing risks data. **Expert review of clinical pharmacology**, v.1, n. 3, p. 391 – 400, 2008.

ZHANG, S.; IVY, J. S.; WILSON, J. R.; DIEHL, K. M.; YANKASKAS, B. C. Competing risks analysis in mortality estimation for breast cancer pacientes from independent risk groups. **Health Care Management Science**, v. 17, n. 3, p. 259 – 269, 2013.

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.

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

CAPÍTULO 2

Efeito da censura sobre a função de incidência cumulativa na presença de riscos competitivos na análise de sobrevivência.

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.

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.

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

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).

Nesta perspectiva, este artigo tem como objetivo avaliar o efeito da censura sobre a função de incidência cumulativa.

2. MATERIAL E MÉTODOS

2.1. 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 ε que representa a causa de falha é dado por:

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

em que d_{ij} é 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 $\hat{S}(t_{j-1})$ é uma probabilidade geral de sobrevivência, ou seja, sobrevivência a toda e qualquer causa (CARVALHO *et al.*, 2011). Essa probabilidade é obtida pelo uso do estimador de Kaplan-Meier (1958) dado por:

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

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 α_{01} e α_{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 DEVELOPMENT CORE TEAM, 2016).

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 considerada 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-Cambolor (2016).

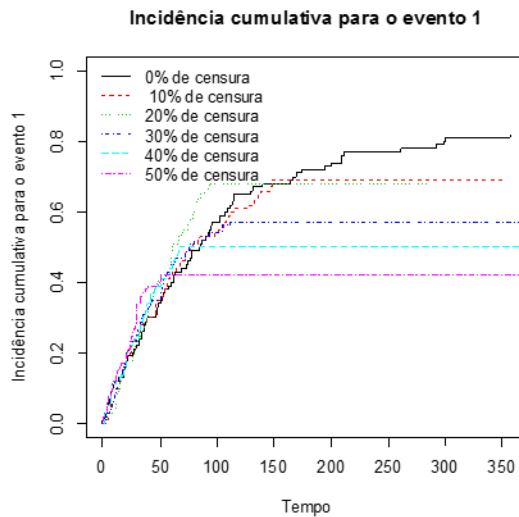


Figura 1 - Estimativa da função de incidência para o evento 1 considerando o risco específico 1 e 2 iguais a 0,8 e 0,2, respectivamente.

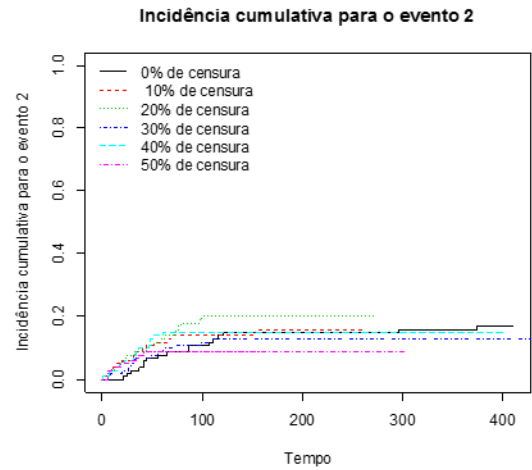


Figura 2 - Estimativa da função de incidência para o evento 2 considerando o risco específico 1 e 2 iguais a 0,8 e 0,2, respectivamente.

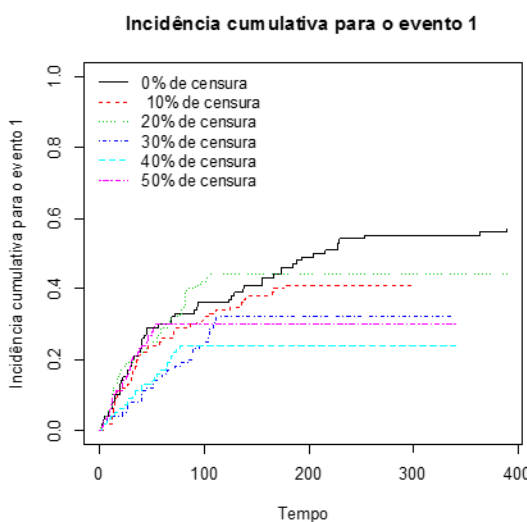


Figura 3 - Estimativa da função de incidência para o evento 1 considerando o risco específico 1 e 2 iguais a 0,5 e 0,5, respectivamente.

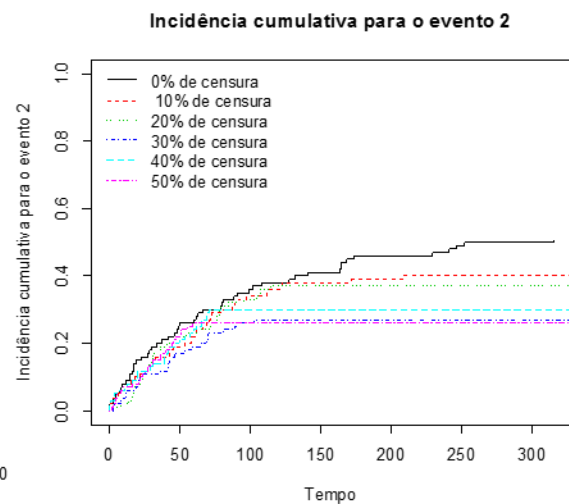


Figura 4 - Estimativa da função de incidência para o evento 2 considerando o risco específico 1 e 2 iguais a 0,5 e 0,5, respectivamente.

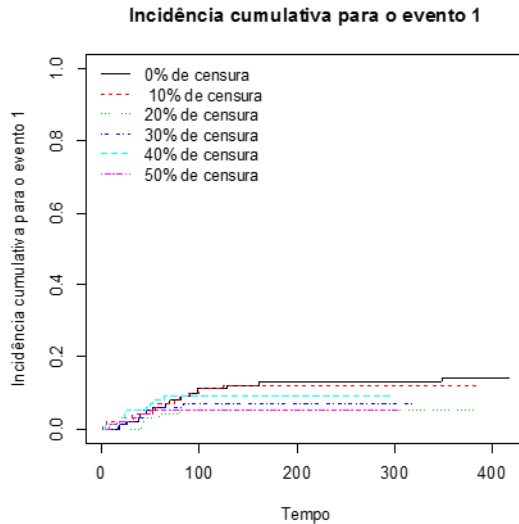


Figura 5 - Estimativa da função de incidência para o evento 1 considerando o risco específico 1 e 2 iguais a 0,1 e 0,7, respectivamente.

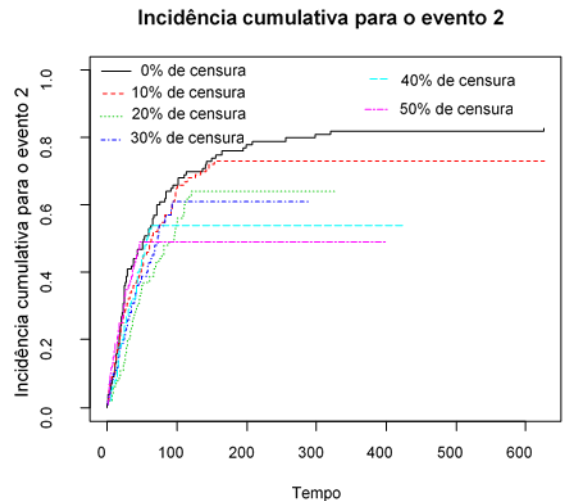


Figura 6 - Estimativa da função de incidência para o evento 2 considerando o risco específico 1 e 2 iguais a 0,1 e 0,7, respectivamente.

A Figura 7 apresenta uma comparação entre as curvas para a função de incidência cumulativa para o evento 1 obtidas pelo uso da Equação $\hat{F}_i(t) = \sum_{\forall t, t_j \leq t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1})$ 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 $\hat{F}_i(t) = \sum_{\forall t, t_j \leq t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1})$ 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).

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 α_{01} e α_{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 α_{01} fixo e aumentando-se o risco específico por causa 2, α_{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).

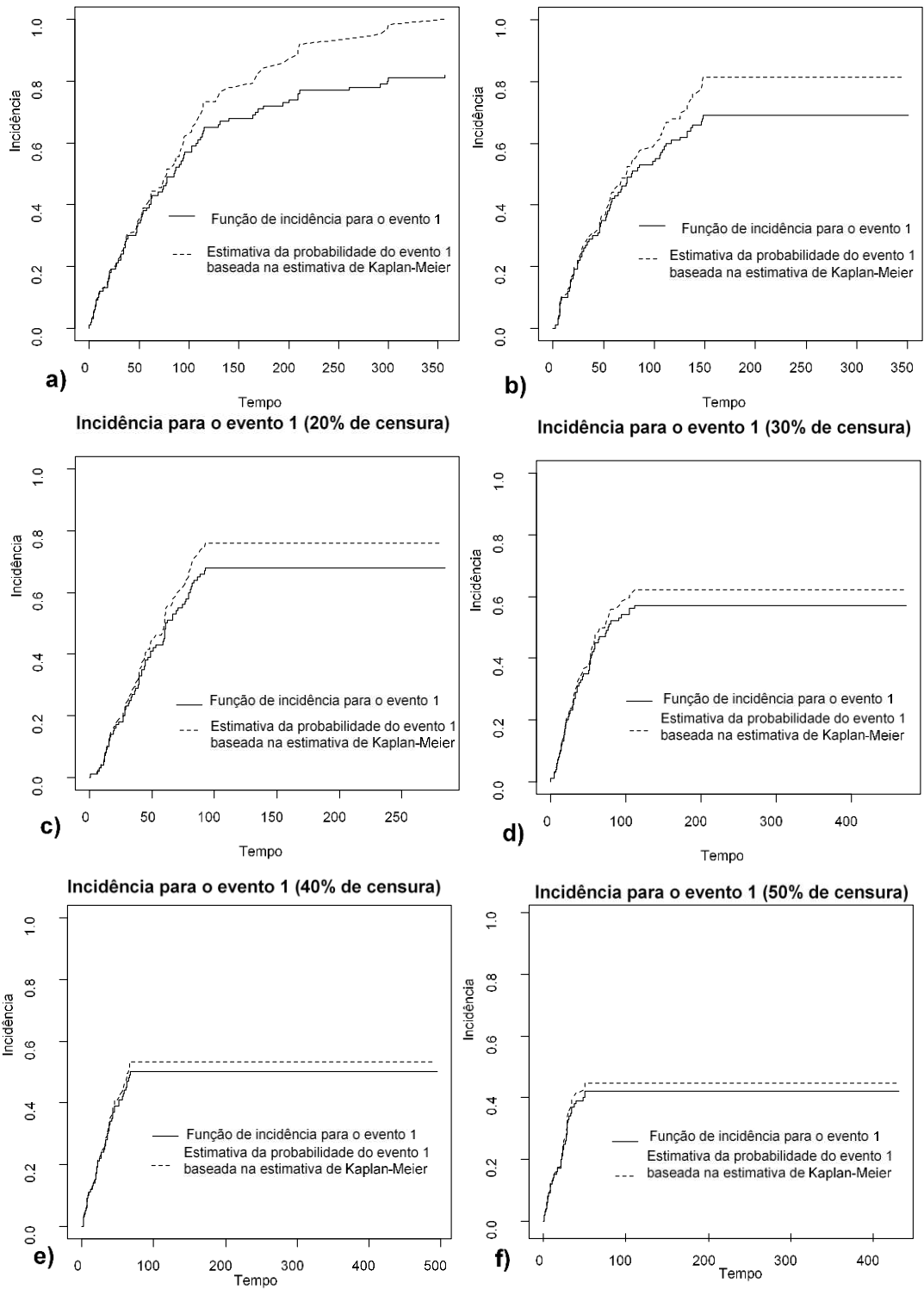


Figura 7 – Curvas de incidência para o evento 1, considerando o risco específico por causa para o evento 1 e para o evento 2 iguais a 0,8 e 0,2, obtidas no contexto de riscos competitivos e pelo estimador de Kaplan-Meier, considerando diferentes porcentagens de censura.

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					
		0%	10%	20%	30%	40%	50%
α_{01}	α_{02}	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)

4. 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 fornece 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 aproxima 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 é tão ruim quanto a obtida pelo uso do estimador de Kaplan-Meier.

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

5. AGRADECIMENTOS

À FAPEMIG pelo apoio financeiro.

6. REFERÊNCIAS BIBLIOGRÁFICAS

AALLEN, O. O.; BORGAN, O.; GJESSING, H. K. **Survival and event history analysis: a process point of view**. New York: Springer, 2008. 539 p.

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.

BEUSCART, J-B.; PAGNIEZ, D., BOULANGER, E.; FOY, C. L. de 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. (eds). **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. 245 p.

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. ver.ampl. Rio de Janeiro: Editora Fiocruz, 2011. 432 p.

DIANATKHAH, M.; RAHGOZAR, M.; TALAEI, M.; KARIMLOUA, M.; SADEGHI, M.; OVEISGHARAM, S.; SARRAFZADEGAN, N. Comparasion of competing risks models baseado on cumulative incidence function in analyzing time to cardiovascular diseases. **ARYA Atheroscler**, v. 10, n. 1, p. 6 – 12, 2014.

GESKUS, R. B. **Data analysis with competing risks and intermediate states**. Boca Raton, Fl: Chapman & Hall/CRC, 2015. 243 p.

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. 18, n. 6, p. 695 – 706, 1999.

HOSMER, D. W.; LEMESHOW, S.; MAY, S. **Regression Modeling of time to event data**. 2 ed. New York: John Wiley & Sons, 2008. 392 p.

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.; BAJORUNAITE, 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.

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. 1161 – 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.

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 DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2016. Disponível em:< <http://www.r-project.org/> >.

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.

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.

ZHANG, S.; IVY, J. S.; WILSON, J. R.; DIEHL, K. M.; YANKASKAS, B. C. Competing risks analysis in mortality estimation for breast cancer pacientes from independent risk groups. **Health Care Management Science**. v. 17, 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.

APÊNDICE A

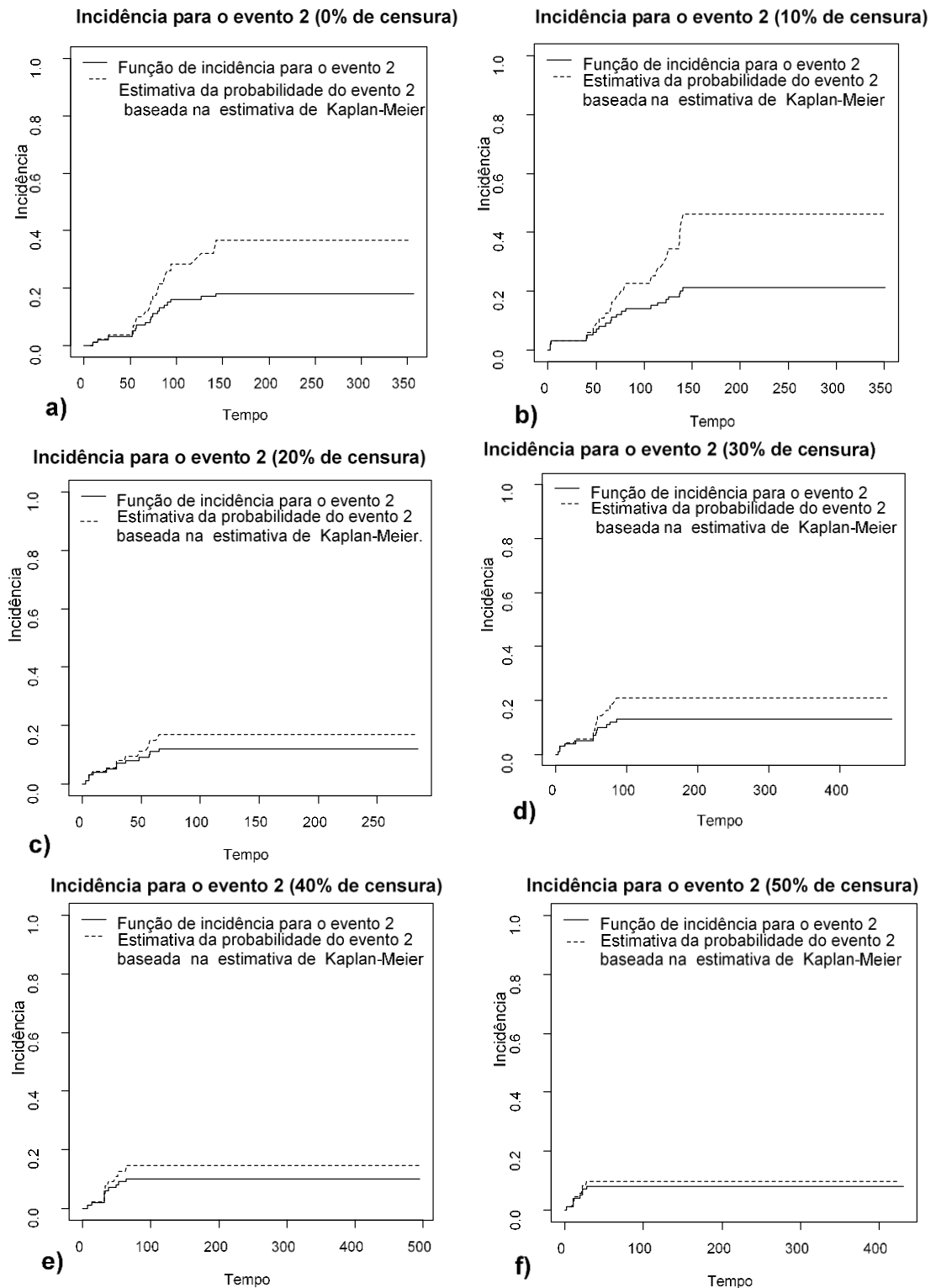


Figura 8 – Curvas de incidência para o evento 2, considerando o risco específico por causa para os eventos 1 e 2 iguais a 0,8 e 0,2, obtidas no contexto de riscos competitivos e pelo estimador de Kaplan-Meier, considerando diferentes porcentagens de censura.

APÊNDICE B

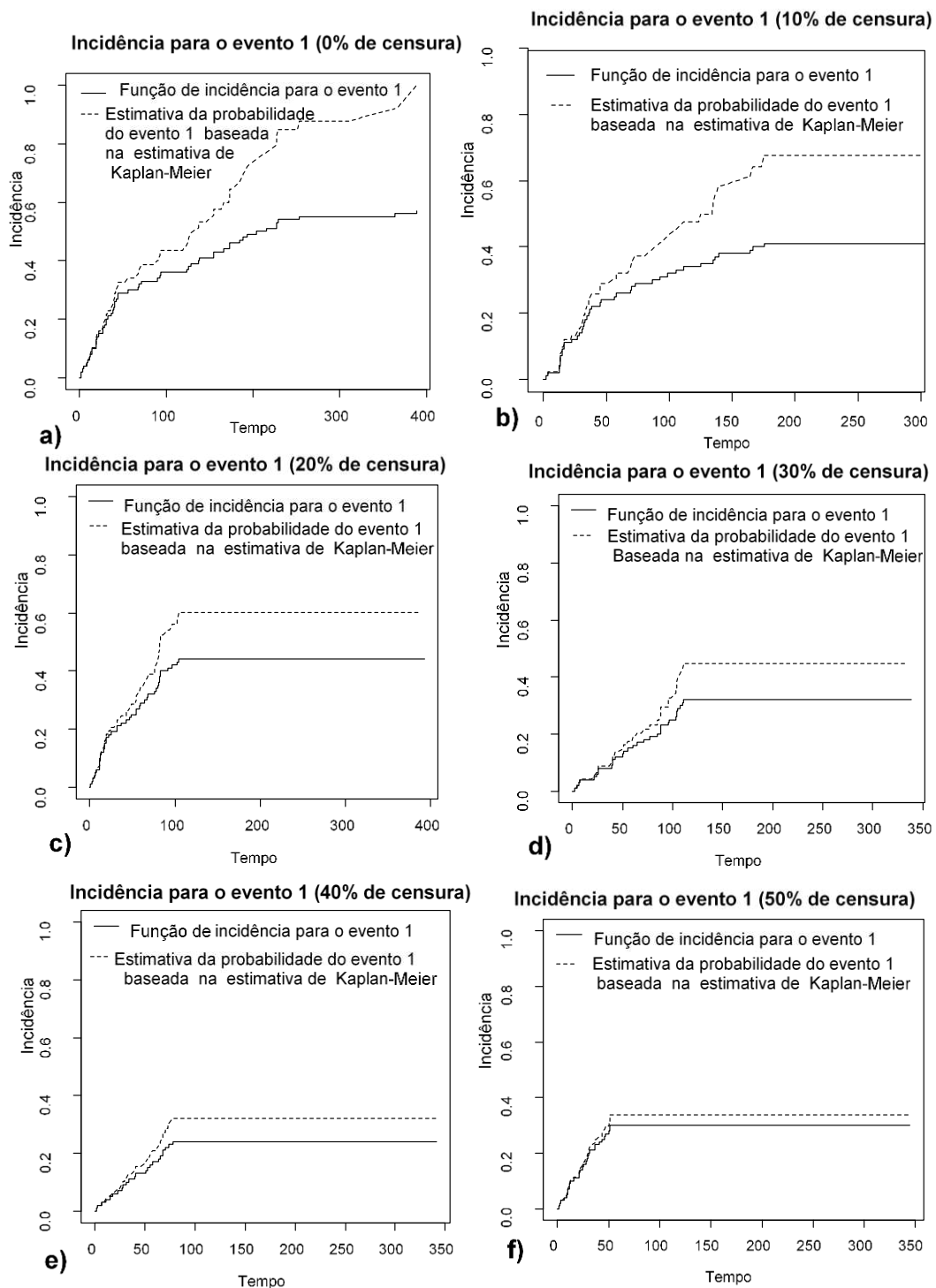


Figura 9 – Curvas de incidência para o evento 1, considerando o risco específico por causa para o evento 1 e para o evento 2 iguais a 0,5 e 0,5, obtidas no contexto de riscos competitivos e pelo estimador de Kaplan-Meier, considerando diferentes porcentagens de censura.

APÊNDICE C

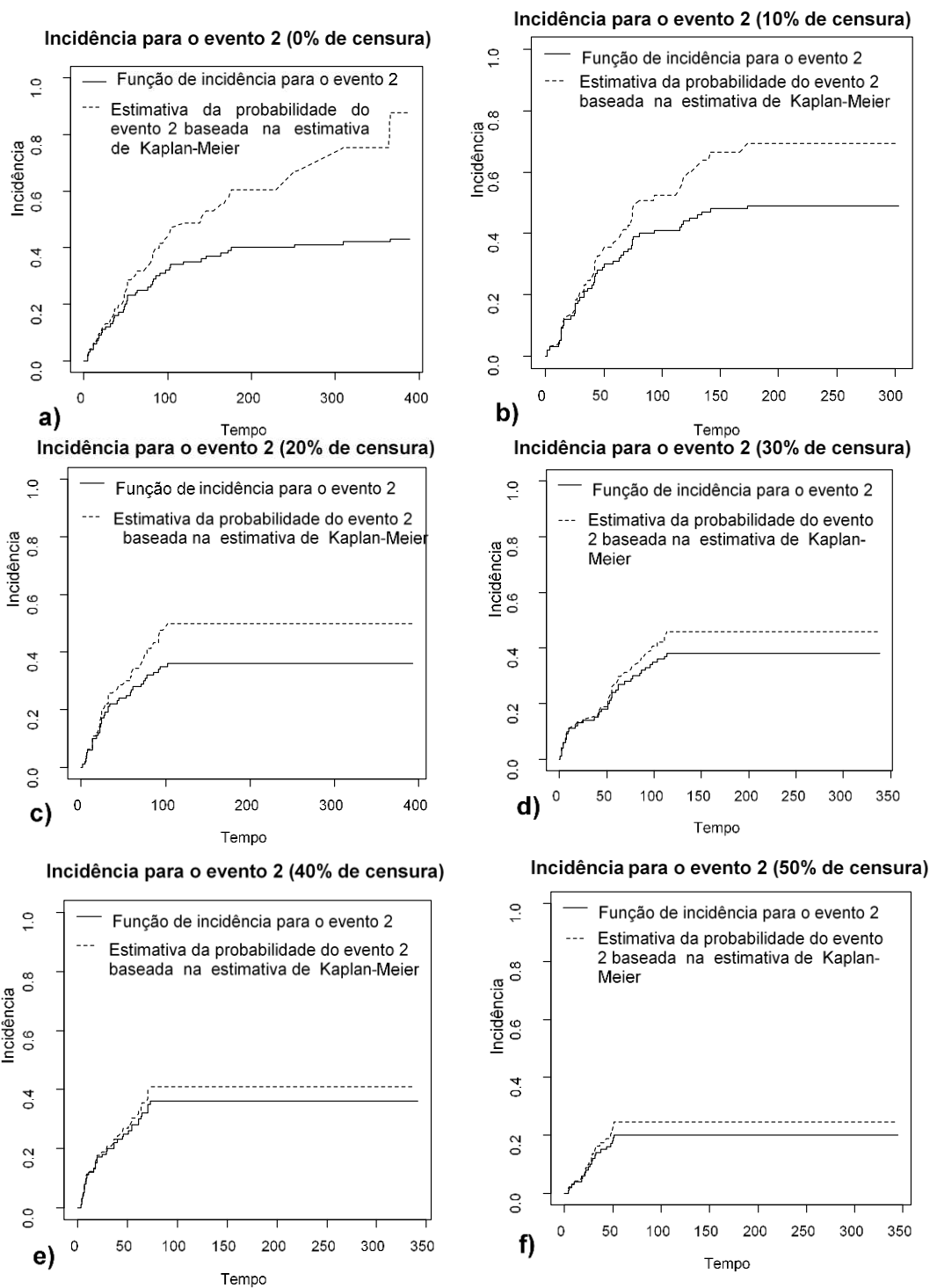


Figura 10 – Curvas de incidência para o evento 2, considerando o risco específico por causa para o evento 1 e para o evento 2 iguais a 0,5 e 0,5, obtidas no contexto de riscos competitivos e pelo estimador de Kaplan-Meier, considerando diferentes porcentagens de censura.

APÊNDICE D

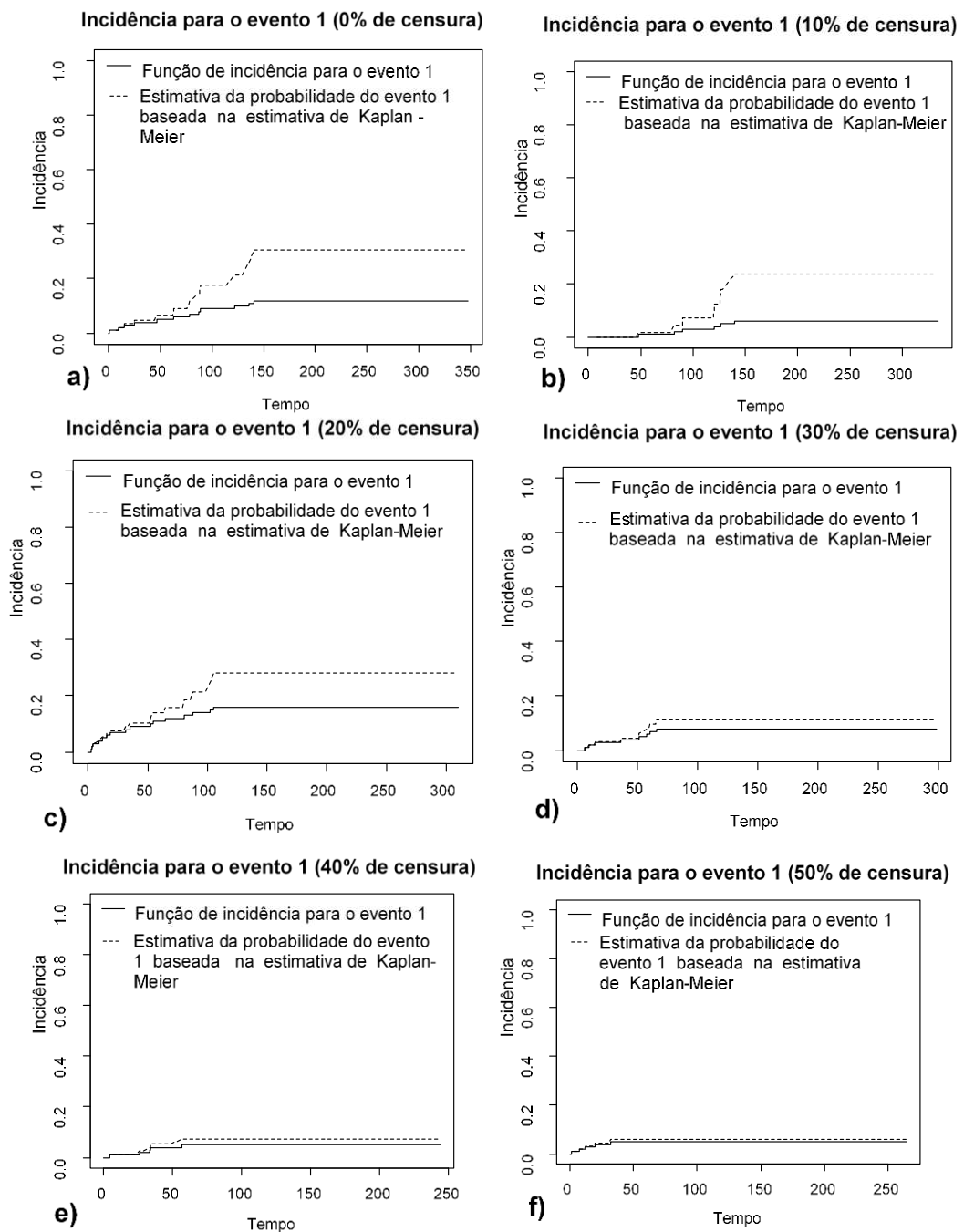


Figura 11 – Curvas de incidência, considerando o risco específico por causa para o evento 1 e para o evento 2 iguais, respectivamente, a 0,1 e 0,7, obtidas no contexto de riscos competitivos e pelo estimador de Kaplan-Meier, considerando diferentes porcentagens de censura.

APÊNDICE E

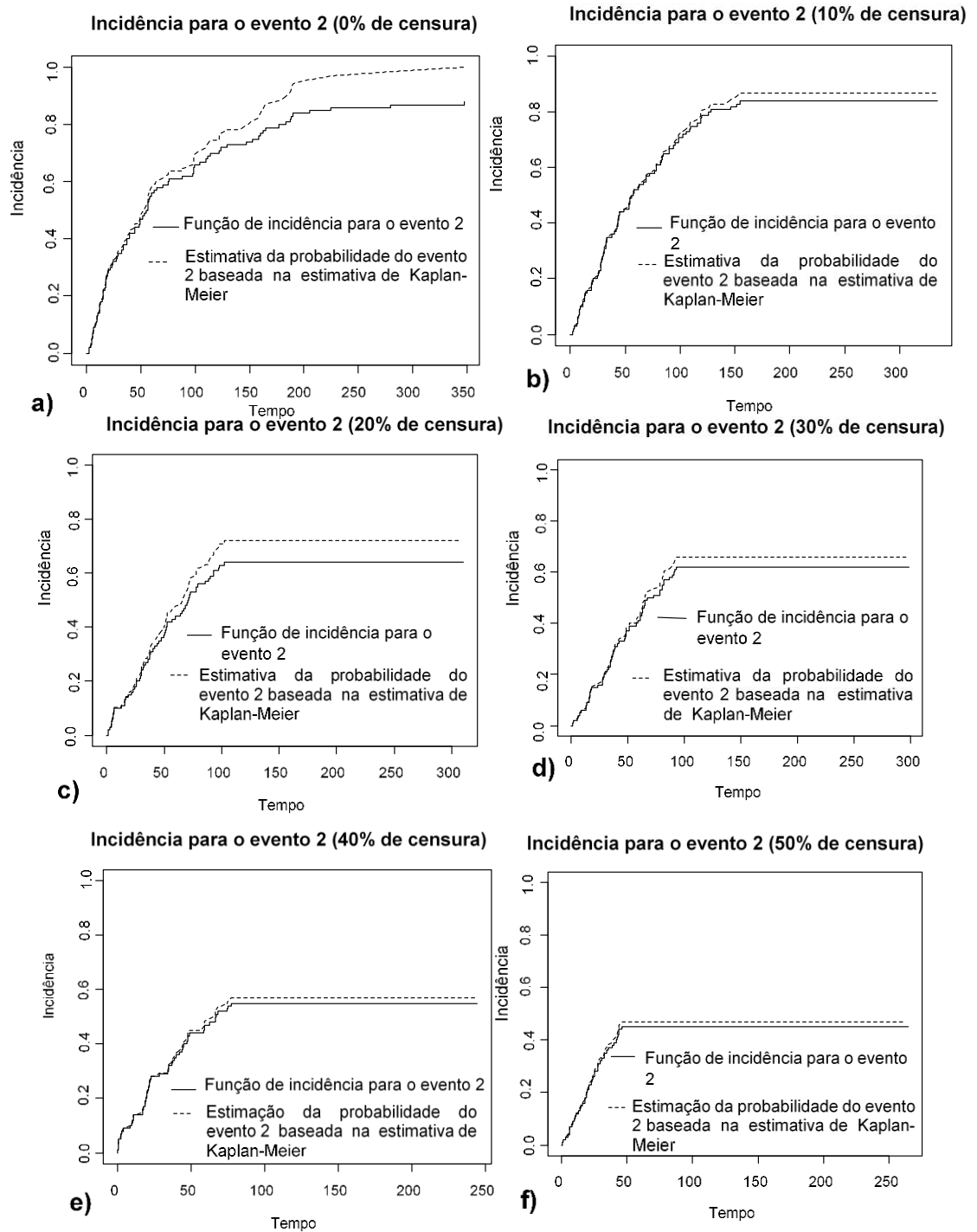


Figura 12 – Curvas de incidência, considerando o risco específico por causa para o evento 1 e para o evento 2 iguais, a 0,1 e 0,7, obtidas no contexto de riscos competitivos e pelo estimador de Kaplan-Meier, considerando diferentes porcentagens de censura

APÊNDICE 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.

(Continua)

Riscos específicos por causa		Porcentagem de censura					
α_{01}	α_{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 - Média e desvio-padrão das diferenças das estimativas da incidência 1 obtidas pelo complemento do estimador de Kaplan-Meier e pela função de incidência.

(Continua)

Riscos específicos por causa		Porcentagem de censura					
α_{01}	α_{02}	0%	10%	20%	30%	40%	50%
		Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)	Média (Desvio)
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)
0,4	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 - Média e desvio-padrão das diferenças das estimativas da incidência 1 obtidas pelo complemento do estimador de Kaplan-Meier e pela função de incidência.

(Continua)

Riscos específicos por causa		Porcentagem de censura					
		0%	10%	20%	30%	40%	50%
α_{01}	α_{02}	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 - Média e desvio-padrão das diferenças das estimativas da incidência 1 obtidas pelo complemento do estimador de Kaplan-Meier e pela função de incidência.

(Conclusão)

Riscos específicos por causa		Porcentagem de censura					
		0%	10%	20%	30%	40%	50%
α_{01}	α_{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$surv; incidencekm
estimatecif<-timepoints(cif, times= km$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)
}

```


CAPÍTULO 3:

Comparação entre métodos de análise de sobrevivência clássica e riscos competitivos para análise de eventos competitivos.

RESUMO

O objetivo deste trabalho foi confrontar duas distintas abordagens na análise de dados de sobrevivência na presença de eventos competitivos. A primeira abordagem foca-se na análise de sobrevivência padrão, mais especificadamente na obtenção da incidência cumulativa por meio do uso do complemento do estimador de Kaplan-Meier, e na modelagem do efeito das covariáveis pelo ajuste do modelo de riscos proporcionais de Cox. Nessa abordagem os eventos competitivos são tratados como censura. A segunda abordagem, denominada riscos competitivos, enfatiza a obtenção da incidência cumulativa e modelagem do efeito das covariáveis tendo como base, respectivamente, a função de incidência cumulativa e o modelo de Fine e Gray. Para a ilustração e comparação destas duas metodologias utilizou-se dados referentes a lesões em equinos de corrida. Os eventos considerados neste estudo foram lesões devido à claudicação (evento principal) e lesões decorrente de outras injúrias (evento competitivo). Os resultados encontrados indicam que a estimativa da incidência para cada um dos eventos, quando se utiliza o estimador de Kaplan-Meier, é superestimada. Além disso, a modelagem dos efeitos das covariáveis sobre o risco específico, ajustado pelo modelo de Cox, não corresponde ao efeito sobre a incidência deste evento ajustado pelo modelo de Fine e Gray.

Palavras-chave: Kaplan-Meier, Modelo de Cox, Modelo de Fine e Gray.

ABSTRACT

The objective of this work was to compare two different approaches in the analysis of survival data in the presence of competing risks. The first approach focuses on standard survival analysis, more specifically on the cumulative incidence estimation through the use of the Kaplan-Meier estimator complement, and on modeling the covariates effect by fitting the Cox proportional hazards model. In this approach, competing events are treated as censorship. The second approach, called competitive risks, emphasizes the cumulative incidence and modeling of the effect of covariates based on the cumulative incidence function and the Fine and Gray model, respectively. For the illustration and comparison of these two methodologies, we used data on injuries in equines. The events considered in this study were injuries due to lameness (main event) and other injuries (competing event). The results indicate that the estimation of the incidence for each of the events, when using the Kaplan-Meier estimator, is overestimated. In addition, the modeling of covariates effects on specific risk, adjusted by the Cox model, does not correspond to the effect on the incidence of this event adjusted by the Fine and Gray model.

Keywords: Kaplan-Meier, Cox model, Fine and Gray model

1. INTRODUÇÃO

Um problema clássico na modelagem de tempos de sobrevivência é a ocorrência de falhas devido a vários eventos distintos. Como exemplo, em estudos oncológicos, pode-se ter interesse em modelar o efeito de fatores de riscos para óbito devido a dois eventos distintos: mortalidade devido ao câncer ou mortalidade decorrente de outras causas (ZHANG *et al.*, 2013). Ou, ainda, estudar diferentes desfechos em pacientes com hipertensão arterial: evento cardiovascular não fatal, óbito por causa cardiovascular e óbito por outras causas (ASSANE *et al.*, 2015).

Esses diferentes tipos de falhas são denominados eventos competitivos e introduzem os riscos competitivos (AALEN *et al.*, 2008). Embora existam distintos tipos de falhas 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 tipos de eventos (GOOLEY *et al.*, 1999).

Riscos competitivos podem também estar presentes em estudos de diferentes tipos de eventos não fatais em que o interesse está no evento que ocorre primeiro. A título de ilustração, ao se avaliar diferentes desfechos: diagnóstico de câncer, diagnóstico de doença cardíaca ou diagnóstico de morte; um diagnóstico de câncer antes da constatação de doença do coração ou morte impossibilita que os dois últimos ocorram primeiro, sendo assim cada um desses serve como um evento competitivo (AUSTIN *et al.*, 2016).

Assim, de maneira geral, a estrutura de riscos competitivos pode ser usada se diferentes tipos de eventos são estudados e o foco está no tempo e no tipo do primeiro evento (WOLBERS *et al.*, 2014, BEYERSMANN *et al.*, 2012).

Na análise de riscos competitivos, assim como na análise de sobrevivência padrão a estimação da incidência bem como a mensuração dos efeitos das covariáveis frequentemente são de interesse (KIM, 2007).

A incidência cumulativa do k -ésimo evento, $P(T \leq t, \varepsilon = k)$, em que $\varepsilon = k$ é o tipo de falha, representa a probabilidade desse evento ocorrer. A estimativa desta quantidade na análise de sobrevivência sem a presença de riscos competitivos, ou seja, na análise de sobrevivência padrão, é obtida por meio do complemento da função de sobrevivência, isto é, por $1 - \hat{S}(t)$, em que $\hat{S}(t)$ é

obtido pelo uso do estimador de Kaplan-Meier. Entretanto, na presença de eventos competitivos, estudos têm apontado a inadequabilidade deste estimador para obter a incidência cumulativa para uma causa específica (ANDERSEN *et al.*, 2012; KIM, 2007; GOOLEY *et al.*, 1999).

Na presença de riscos competitivos a incidência baseada no estimador de Kaplan-Meier é superestimada (CARVALHO *et al.*, 2011), e isto ocorre porque este estimador ignora todos os eventos distintos daquele de interesse, considerando-os como censura.

Objetivando contornar esse problema foi proposta a função de incidência cumulativa, também denominada, subdistribuição, que estima a probabilidade de ocorrência de um evento de interesse sem desconsiderar o evento competitivo.

No que se refere à modelagem do efeito das covariáveis, em muitas aplicações os riscos competitivos têm sido ignorados e o modelo de regressão de Cox, usado na análise de sobrevivência padrão, tem sido ajustado (WOLBERS *et al.*, 2009). Deste modo, o modelo de Cox modela o risco específico por causa para o k-ésimo evento, $\lambda_k(t)$, dado por:

$$\lambda_k(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{P(t \leq T < t + \Delta t, \varepsilon = k | T \geq t)}{\Delta t} \right\}$$

e que representa a probabilidade de um indivíduo falhar em um pequeno intervalo de tempo.

Nessa modelagem os riscos competitivos são ignorados e, portanto, o efeito da covariável reflete o efeito “puro”, ou melhor, o efeito quando os outros tipos de eventos não existem (PINTILIE, 2006). No entanto, o modelo de Cox pode também conduzir a erros nas estimativas quando os riscos competitivos estão presentes (BERRY, 2010).

Além disso, o efeito de covariáveis sobre um evento, quando se utiliza o modelo de Cox, pode ser diferente do efeito da covariável sobre a incidência do referido evento quando se considera o risco competitivo (KIM, 2007). Como o modelo de Cox não tem uma interpretação direta em termos de função de incidência (SCRUCCA *et al.*, 2010), Fine e Gray (1999) propuseram um modelo para o risco da função de incidência cumulativa. Esse modelo permite avaliar diretamente o efeito das covariáveis sobre a função de incidência cumulativa.

Assim, o objetivo foi comparar a abordagem baseada na análise de sobrevivência padrão com a abordagem baseada em riscos competitivos na estimação da incidência cumulativa, bem como comparar o modelo de regressão de Cox padrão e o modelo de Fine e Gray na modelagem de dados sobre lesões em equinos de corridas.

2. MATERIAL E MÉTODOS

Os dados utilizados neste trabalho referem-se a 158 equinos da raça Puro Sangue Inglês (PSI) que permaneceram no Centro de Treinamento em Pedro do Rio, Petrópolis, RJ, no período de janeiro de 2001 a dezembro de 2003.

A variável resposta é o tempo até a primeira falha e o tipo da primeira falha. O evento de interesse principal é a falha decorrente da claudicação enquanto o evento competitivo é a falha devido a outras injúrias.

Dos 158 animais, 90 (57%) experimentaram o evento de interesse principal (claudicação), 27 (17%) experimentaram o evento competitivo (outras injúrias) e 41 (25,9%) até o final do estudo não haviam experimentado nenhum dos eventos, ou seja, foram censurados.

A probabilidade total estimada da ocorrência de um evento, independentemente do tipo de lesão, com base no estimador de Kaplan-Meier, foi obtida usando a expressão $\hat{F}(t) = 1 - \hat{S}(t)$, em que $\hat{S}(t)$ é a sobrevivência livre de evento fornecida pelo estimador de Kaplan-Meier (1958) definido como:

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

em que d_j é o número de falhas em t_j , $j = 1, 2, \dots, r$, r tempos distintos, e n_j é o número de indivíduos sob risco em t_j (GESKUS, 2015; CARVALHO *et al.*, 2011; COLOSIMO e GIOLO, 2006).

Alternativamente, estimou-se a probabilidade total de ocorrência de uma lesão por meio da estimação da função de incidência, dada pela soma das incidências para cada tipo de evento. A função de incidência total é dada pela expressão:

$$F(t) = P(T \leq t) = \sum_{k=1}^2 P(T \leq t, \varepsilon = k) = F_1(t) + F_2(t), \quad (2)$$

em que k é o tipo de evento, $k = 1, 2$, sendo que $k = 1$ refere-se ao evento de interesse principal e $k = 2$ ao evento competitivo. $F_1(t)$ e $F_2(t)$ são respectivamente, as probabilidades de claudicação e lesão devido à outras injúrias e são estimadas por

$$\hat{F}_k(t) = \sum_{\forall j, t_j < t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1}), \quad (3)$$

com $k=1, 2$, d_{ij} representando o número de animais que experimentaram o evento k no tempo t_j , n_j o número de animais em risco no tempo t_j e $\hat{S}(t_{j-1})$ a sobrevivência livre de eventos no tempo t_{j-1} , obtida conforme descrito em (1) (PINTILIE, 2006).

Estimou-se a incidência de cada evento, $\hat{F}_k(t) = P(T \leq t, \varepsilon = k)$, de duas formas distintas. Na primeira, $\hat{F}_k(t)$ foi obtida como o complemento do estimador de Kaplan-Meier para o evento k , ou seja, $\hat{F}_k(t) = 1 - KM_k$, em que KM_k é o estimador de Kaplan-Meier para o k -ésimo evento, dado pela expressão (1), obtido ao se considerar o evento distinto do k como censura. A segunda forma de obtenção da probabilidade de experimentar o k -ésimo evento, $\hat{F}_k(t) = P(T \leq t, \varepsilon = k)$, baseou-se na abordagem da função de incidência quando riscos competitivos estão presentes, conforme apresentado na expressão (3).

Em seguida dois diferentes modelos de regressão foram ajustados: o modelo de Cox para o risco específico por causa e o modelo de Fine e Gray para o risco da subdistribuição.

O modelo semiparamétrico de riscos proporcionais de Cox ajustado para o risco específico por causa para cada tipo de evento é dado por:

$$\lambda_k(t|Z) = \lambda_{k,0} \exp(\beta_k^T Z_i), \quad (4)$$

sendo β_k o vetor $p \times 1$ de coeficientes de regressão, Z_i o vetor $p \times 1$ de covariáveis e $\lambda_{k,0}$ a função de risco de base para a causa k , não negativa e não especificada (GESKUS, 2015; CROWDER, 2001).

O modelo de riscos proporcionais para o risco da subdistribuição, proposto por Fine e Gray (1999) pode ser escrito como:

$$\gamma_k(t|Z) = \gamma_{k0} \exp(\beta_k^T Z_i) \quad (5),$$

com β_k o vetor $p \times 1$ de coeficientes de regressão, Z_i o vetor $p \times 1$ de covariáveis e $\gamma_{k,0}$ a função de risco de base da subdistribuição para a causa k , não negativa e não especificada.

As covariáveis que tiveram seus efeitos avaliados foram: sexo (1 para fêmea e 2 para macho), idade (meses); peso, variável categórica, definida como 1 se peso do animal é inferior a 480 kg e 2 se o peso do animal é maior ou igual a 480 kg, estações de treino.

As análises dos dados foram realizadas no software R (R DEVELOPMENT CORE TEAM, 2016), utilizando as funções *survfit*, e *coxph* do pacote *survival* e as funções *cuminc* e *crr* do pacote *cmprsk*.

3. RESULTADOS E DISCUSSÃO

Dos 158 cavalos do estudo, 62 (39,2%) eram fêmeas e 96 (60,8%) machos. A idade dos animais variou entre 25 meses a 75 meses, sendo que 120 animais (75,9%) tinham entre 25 meses a 36 meses; 38 (24,1%) tinham mais de 36 meses. A idade média foi de 32,6 meses. 67 animais (42,4%) pesam menos de 480 kg e 91 (57,6%) pesam 480 kg ou mais. O número de estações de treino variou de 1 a 6 estações com valor mediano igual a 2. 90 (57%) animais experimentaram o evento principal (claudicação com ou sem a presença de outras injúrias), 27 (17,1%) o evento competitivo (outras injúrias) e 41 (25,9%) foram censurados.

As funções de incidência cumulativa (FIC) para a claudicação e para outras injúrias estão apresentadas na Figura 1 juntamente com a incidência total, isto é, incidência de qualquer lesão independentemente do tipo.

A FIC total pode ser estimada de duas maneiras distintas, sendo a primeira por meio do uso do estimador de Kaplan-Meier do tempo do primeiro evento, ignorando o tipo do evento. Alternativamente, pode-se estimar esta quantidade como a soma das FICs para os diferentes eventos (WOLBERS *et al.*, 2014). A incidência cumulativa total, $1 - \hat{S}(t)$, estimada pelo complemento da função de Kaplan-Meier para a sobrevivência livre de injúrias é mostrada na Figura 1. Essa incidência é igual a soma das incidências cumulativas para a claudicação e outros tipos de injúrias (Tabela 1). Essa análise do evento de interesse como um único tipo de evento de falha é correta, porém limitada para avaliar várias questões de pesquisa importante (KIM, 2007). Nessa direção, a análise da combinação de diferentes causas de falhas e a posterior análise como um único evento responde às questões relacionadas as diferenças entre grupos e/ou tratamentos considerando a combinação dos tipos de evento, entretanto não permite comparar as incidências cumulativas para os diferentes tipos de eventos.

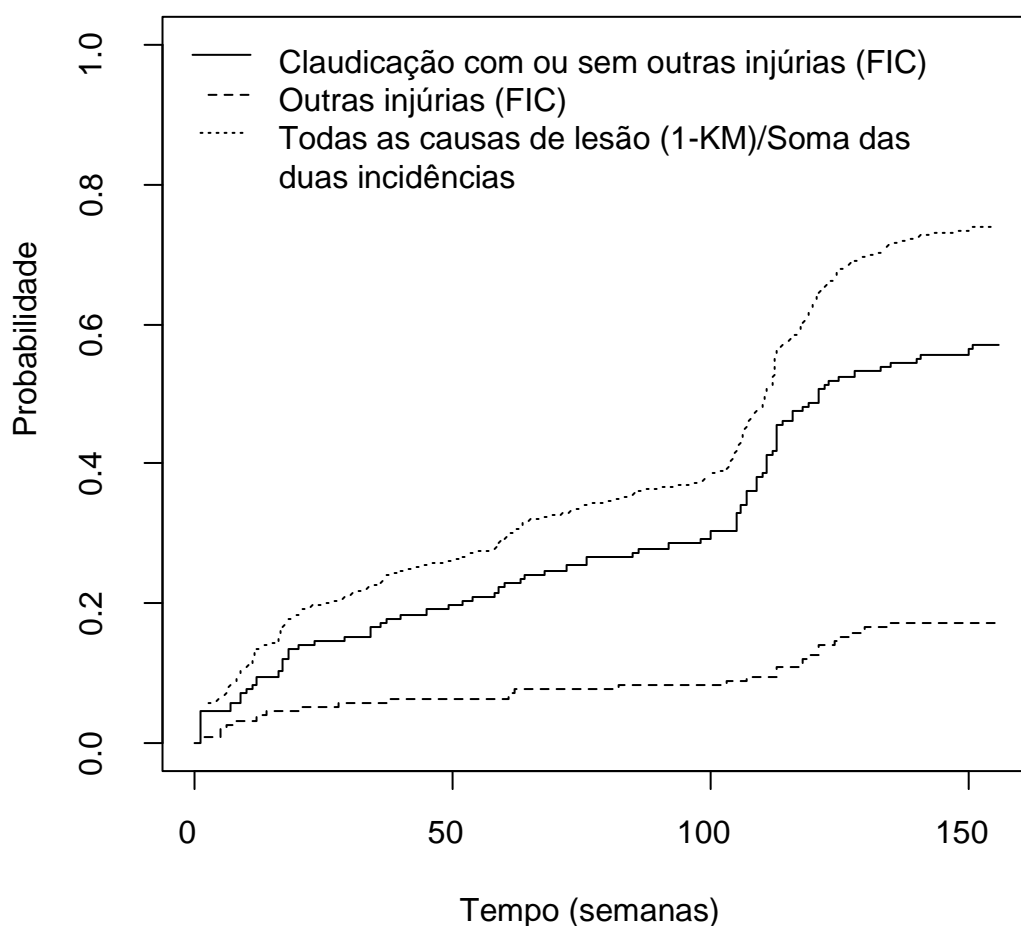


Figura 1 - Função de incidência cumulativa para a claudicação, função de incidência cumulativa para outras injúrias e o estimador de Kaplan-Meier (KM) para a incidência de todos os tipos de lesões (soma das duas funções de incidência).

Tabela 1 - Incidências estimadas para a claudicação ($\hat{F}_1(t)$), para outras injúrias ($\hat{F}_2(t)$), para a incidência total ($\hat{F}(t)$) e para a incidência calculada com base no estimador de Kaplan-Meier ($1 - \hat{S}(t)$)

(Continua)

Tempo	$\hat{F}_1(t)$	$\hat{F}_2(t)$	$\hat{F}(t)$ = $\hat{F}_1(t) + \hat{F}_2(t)$	$1 - \hat{S}(t)$	Tempo	$\hat{F}_1(t)$	$\hat{F}_2(t)$	$\hat{F}(t)$ = $\hat{F}_1(t) + \hat{F}_2(t)$	$1 - \hat{S}(t)$
1	0,0443	0,0063	0,0506	0,0506	82	0,2658	0,0823	0,3481	0,3481
5	0,0443	0,0190	0,0633	0,0633	85	0,2722	0,0823	0,3544	0,3544
6	0,0443	0,0253	0,0696	0,0696	86	0,2785	0,0823	0,3608	0,3608
7	0,0570	0,0253	0,0823	0,0823	92	0,2848	0,0823	0,3671	0,3671
8	0,0570	0,0316	0,0886	0,0886	98	0,2911	0,0823	0,3734	0,3734
9	0,0696	0,0316	0,1013	0,1013	100	0,3038	0,0823	0,3861	0,3861
10	0,0759	0,0316	0,1076	0,1076	103	0,3038	0,0886	0,3924	0,3924
11	0,0823	0,0316	0,1139	0,1139	105	0,3291	0,0886	0,4177	0,4177
12	0,0949	0,0380	0,1329	0,1329	106	0,3418	0,0886	0,4304	0,4304
14	0,0949	0,0443	0,1392	0,1392	107	0,3608	0,0949	0,4557	0,4557
16	0,1013	0,0443	0,1456	0,1456	109	0,3797	0,0949	0,4747	0,4747
17	0,1203	0,0443	0,1646	0,1646	110	0,3861	0,0949	0,4810	0,4810
18	0,1329	0,0443	0,1772	0,1772	111	0,4114	0,0949	0,5063	0,5063
20	0,1392	0,0443	0,1835	0,1835	112	0,4177	0,0949	0,5127	0,5127
21	0,1392	0,0506	0,1899	0,1899	113	0,4557	0,1076	0,5633	0,5633
23	0,1456	0,0506	0,1962	0,1962	114	0,4620	0,1076	0,5696	0,5696
28	0,1456	0,0570	0,2025	0,2025	116	0,4747	0,1076	0,5823	0,5823
29	0,1519	0,0570	0,2089	0,2089	117	0,4747	0,1139	0,5886	0,5886
34	0,1646	0,0570	0,2215	0,2215	118	0,4810	0,1203	0,6013	0,6013
36	0,1709	0,0570	0,2278	0,2278	119	0,4873	0,1266	0,6139	0,6139
37	0,1772	0,0633	0,2405	0,2405	121	0,5063	0,1392	0,6456	0,6456
40	0,1835	0,0633	0,2468	0,2468	122	0,5127	0,1392	0,6519	0,6519
45	0,1899	0,0633	0,2532	0,2532	123	0,5190	0,1392	0,6582	0,6582

Tabela 1 - Incidências estimadas para a claudicação ($\hat{F}_1(t)$), para outras injúrias ($\hat{F}_2(t)$), para a incidência total ($\hat{F}(t)$) e para a incidência calculada com base no estimador de Kaplan-Meier ($1 - \hat{S}(t)$)

(Conclusão)

Tempo	$\hat{F}_1(t)$	$\hat{F}_2(t)$	$\hat{F}(t)$ = $\hat{F}_1(t) + \hat{F}_2(t)$	$1 - \hat{S}(t)$	Tempo	$\hat{F}_1(t)$	$\hat{F}_2(t)$	$\hat{F}(t)$ = $\hat{F}_1(t) + \hat{F}_2(t)$	$1 - \hat{S}(t)$
52	0,2025	0,0633	0,2658	0,2658	125	0,5253	0,1519	0,6772	0,6772
54	0,2089	0,0633	0,2722	0,2722	127	0,5253	0,1582	0,6835	0,6835
58	0,2152	0,0633	0,2785	0,2785	128	0,5316	0,1582	0,6899	0,6899
59	0,2215	0,0633	0,2848	0,2848	130	0,5316	0,1646	0,6962	0,6962
60	0,2278	0,0633	0,2911	0,2911	133	0,5380	0,1646	0,7025	0,7025
61	0,2278	0,0696	0,2975	0,2975	135	0,5443	0,1709	0,7152	0,7152
62	0,2278	0,0759	0,3038	0,3038	140	0,5506	0,1709	0,7215	0,7215
63	0,2342	0,0759	0,3101	0,3101	141	0,5570	0,1709	0,7278	0,7278
64	0,2405	0,0759	0,3165	0,3165	150	0,5633	0,1709	0,7342	0,7342
68	0,2468	0,0759	0,3228	0,3228	151	0,5696	0,1709	0,7405	0,7405
72	0,2532	0,0759	0,3291	0,3291	156	0,5696	0,1709	0,7405	0,7405
76	0,2658	0,0759	0,3418	0,3418					

A soma das estimativas obtidas por Kaplan-Meier para claudicação e para outras injúrias é maior do que a estimativa da incidência total, obtida pela soma das duas funções de incidência (Figura 2). Essas curvas ilustram a estimativa incorreta da incidência total ao se usar soma dos estimadores de Kaplan-Meier dos eventos distintos para se estimar essa função.

A superestimação da FIC quando essa é estimada como a soma das estimativas de Kaplan-Meier para os distintos eventos é relatada por Austin *et al.* (2016) em seu estudo sobre causas de morte (causa cardiovascular e causa não cardiovascular) em pacientes com doença cardiovascular.

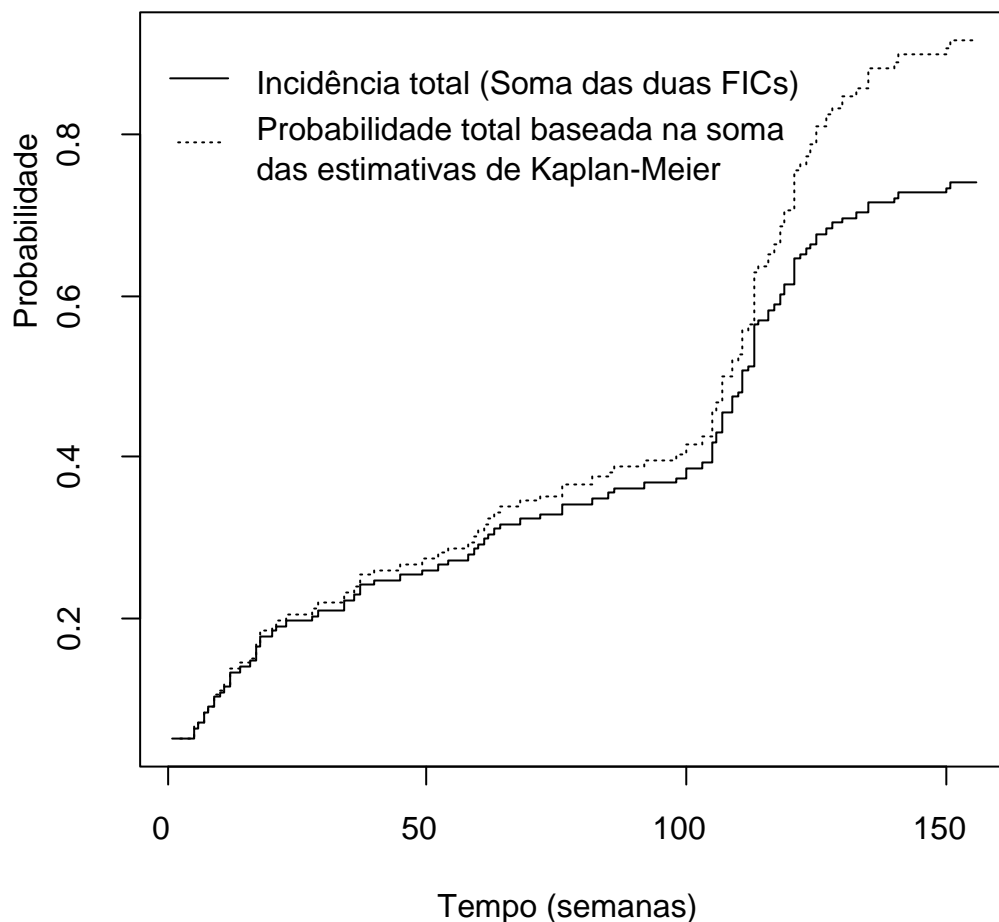


Figura 2 – Função de incidência cumulativa total e incidência calculada pela soma das estimativas da probabilidade de claudicação e de outras injúrias obtidas com base no estimador de Kaplan-Meier.

Nas Figuras 3 e 4 estão apresentadas, respectivamente, as probabilidades de incidência para claudicação e outras injúrias, obtidas por meio do complemento do estimador de Kaplan-Meier e também as estimativas obtidas com base na função de incidência para cada um dos eventos.

As curvas de probabilidade obtidas com base no estimador de Kaplan-Meier para ambos eventos são iguais ou superiores às estimativas obtidas a partir da função de incidência cumulativa. Por conseguinte, também se tem a superestimação das incidências para cada tipo de evento, sendo que o viés aumenta com o tempo de acompanhamento do animal.

Assane *et al.* (2015) em seu trabalho também destacaram limitações do estimador clássico de Kaplan-Meier quando o mesmo foi aplicado para estimação da incidência de diferentes tipos de eventos (evento cardiovascular não fatal, óbito por causa cardiovascular, óbito por outra causa) em pacientes com pressão arterial. Conforme os autores supracitados, a probabilidade de falha para um evento de interesse estimada por meio do estimador de Kaplan-Meier é igual ou maior à probabilidade estimada com base na função de incidência cumulativa. Esses autores, assim como Berry *et al.* (2010), destacaram ainda que a magnitude do viés aumenta com o aumento do tempo de acompanhamento do paciente o que está de acordo como os resultados encontrados neste trabalho.

O viés é consequência do tratamento dos eventos competitivos como censuras e da estimação da incidência do evento de interesse com base no estimador de Kaplan-Meier. Ao se estimar a incidência usando este procedimento viola-se a suposição subjacente ao estimador de Kaplan-Meier: a independência entre o tempo do evento e a distribuição da censura (PUTTER *et al.*, 2007).

No cálculo utilizado para o estimador de Kaplan-Meier ao se censurar o tempo de evento de um indivíduo está se indicando que o indivíduo ainda não experimentou o evento de interesse até o tempo de censura, mas que poderá experimentar esse evento em um tempo posterior, não observável (WOLBERS *et al.*, 2014). Consequentemente, indivíduos censurados podem ser representados por aqueles que permanecem sob observação, deste modo, o risco para indivíduos censurados é igual ao daqueles que permaneceram sob observação (PUTTER *et al.*, 2007).

Contudo, na estrutura de riscos competitivos, sabe-se que indivíduos que experimentam um evento competitivo nunca experimentaram o evento de interesse principal como primeiro evento (WOLBERS *et al.*, 2014). Uma vez que os indivíduos que não falharam devido ao evento de interesse são tratados como se ainda pudessem experimentar esse tipo de evento, ou seja, são censurados, assim, o estimador de Kaplan-Meier acaba por superestimar a probabilidade de falha (PUTTER *et al.*, 2007).

O impacto do incorreto tratamento dos eventos competitivos como eventos censurados têm importância prática, visto que, em geral, quanto maior a porcentagem de eventos competitivos maior o viés (AUSTIN *et al.*; 2016, WOLBERS *et al.*, 2014).

Os resultados deste trabalho apontam que as estimativas das funções de incidências cumulativas para os dois eventos considerados, claudicação e outras injúrias, com base no estimador de Kaplan-Meier, isto é, ignorando a estrutura de risco competitivo, são enviesadas, o que corrobora com os resultados encontrados na literatura (ANDERSEN *et al.*, 2012; KIM, 2007; PUTTER, 2007; SHERIF, 2007; LOGAN *et al.*, 2006).

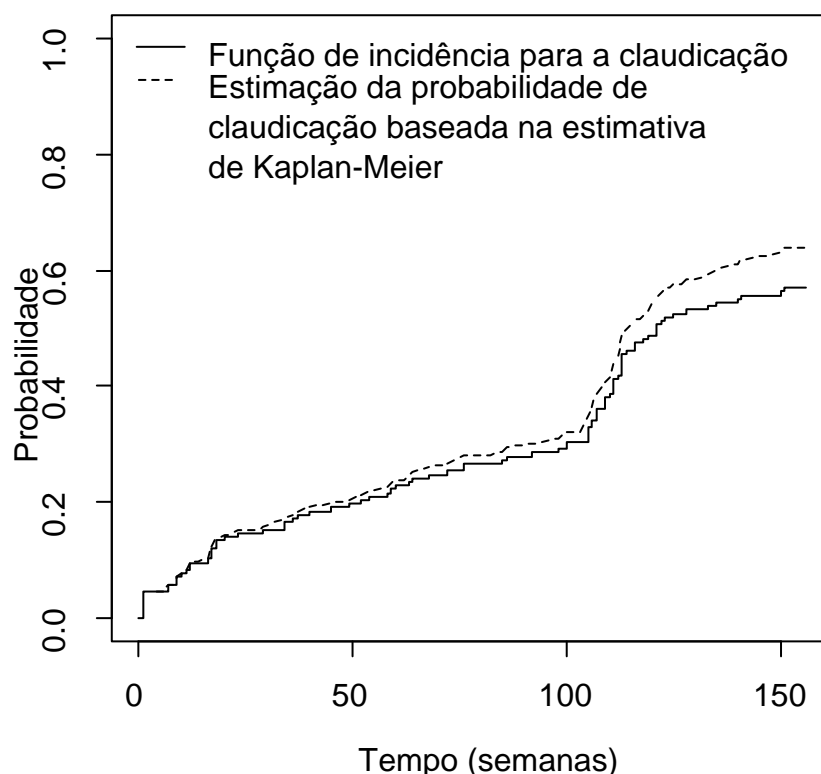


Figura 3 – Função de incidência cumulativa para claudicação e probabilidade de claudicação obtida com base nas estimativas de Kaplan-Meier.

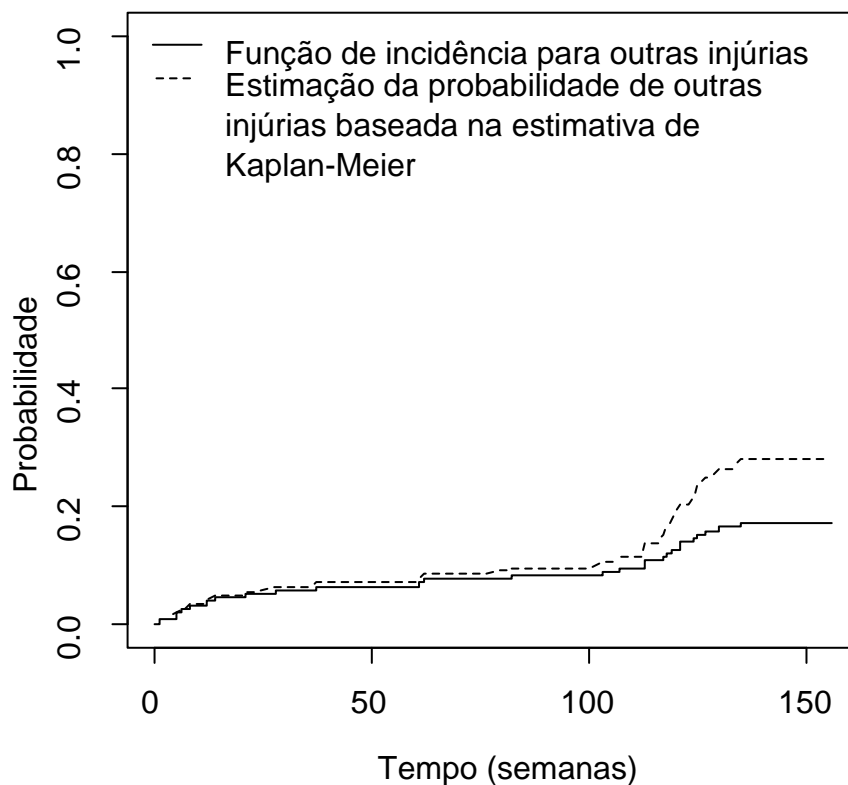


Figura 4 – Função de incidência cumulativa para outras injúrias e probabilidade de outras injúrias obtida com base nas estimativas de Kaplan-Meier.

Os resultados dos modelos de regressão para o risco específico por causa e para a o risco da subdistribuição são apresentados na Tabela 2.

No que se refere ao evento outras injúrias, nenhuma das covariáveis estudadas afeta significativamente a incidência e o risco específico por causa.

Os resultados mostram ainda que o com aumento da idade diminui-se significativamente a probabilidade de incidência de claudicação (RSR = 0,9479; valor-p= 0,0146) e o risco específico por causa (RR = 0,9485, valor-p=0,0062) para a claudicação, ou seja, o risco de claudicação quando não é considerado o risco do evento competitivo (“outras injúrias”).

Estudos descritivos sobre lesões musculoesqueléticas relacionadas ao treino relatam que o índice de lesões varia entre os grupos etários, com os mais altos níveis relatados em animais de dois anos de idade (LINDNER e DINGERKUS, 1993). Contudo aconselha-se precaução na interpretação dos resultados destes estudos, já que podem ser influenciados por fatores de

confusão, como a exposição ao treinamento e as corridas que podem influenciar a relação entre idade e o risco de lesão (COGGER *et al.*, 2006).

A covariável estações aumentou significativamente a incidência (RSR = 1,3868, valor-p=0,0006) e o risco específico por causa (RR = 1,4565, valor-p < 0,0001) para a claudicação e teve nenhum efeito significativo para outras injúrias.

Embora, em geral, quando riscos competitivos estão presentes, o efeito das covariáveis nos riscos específicos e no risco da subdistribuição serão diferentes e em direção oposta. Entretanto, em situação em que há um efeito que diminui (ou aumenta) o risco específico por causa do evento de interesse e não tem nenhum efeito significativo sobre o risco específico por causa do evento competitivo, ocorrerá uma diminuição (ou aumento) na função de incidência para o evento de interesse (BEYERSMANN *et al.*, 2012). Esse comportamento foi verificado para as covariáveis idade e estações.

Tabela 2 – Estimativas obtidas para os modelos de regressão de Fine e Gray para a função de incidência cumulativa para o evento claudicação e para o evento outras injúrias e estimativas obtidas para o modelo de regressão de Cox para o risco específico de claudicação e para o risco específico para outras injúrias.

Modelo de Fine e Gray para o risco da subdistribuição			Modelo de Cox para o risco específico por causa	
Evento: claudicação				
Covariáveis	Valor-p	RSR (I.C 95%)	Valor-p	RR (I.C 95%)
Sexo	0,4236 ^{ns}	1,2114 [0,7553; 1,9378]	0,6732 ^{ns}	1,1052 [0,6943; 1,7593]
Idade	0,0146*	0,9479 [0,9081; 0,9895]	0,0062*	0,9485 [0,9133; 0,9851]
Peso	0,3523 ^{ns}	1,2408 [0,7875; 1,9549]	0,2932 ^{ns}	1,2736 [0,8133; 1,9992]
Estações	0,0006*	1,3868 [1,1501; 1,6723]	<0,0001*	1,4564 [1,2261; 1,7299]
Evento: outras injúrias				
Covariáveis	Valor-p	RSR (I.C 95%)	Valor-p	RR (I.C 95%)
Sexo	0,3902 ^{ns}	0,6951 [0,3033; 1,5934]	0,3350 ^{ns}	0,6706 [0,2976; 1,5109]
Idade	0,5982 ^{ns}	1,0133 [0,9648; 1,0642]	0,9856 ^{ns}	1,0004 [0,9560; 1,0469]
Peso	0,7514 ^{ns}	1,1411 [0,5042; 2,5823]	0,5148 ^{ns}	1,3091 [0,5821; 2,9438]
Estações	0,5571 ^{ns}	1,1113 [0,7813; 1,5808]	0,0519 ^{ns}	1,3888 [0,9973; 1,9340]

RR: razão dos riscos para a causa específica. RSR: razão dos subriscos (riscos da subdistribuição). I.C 95%: Intervalo de 95% de confiança para a razão dos riscos e para a razão dos subriscos. * significativo a 5%. ns: não significativo.

Em relação às diferenças na significância do efeito de covariáveis sobre o risco específico por causa, Bimali e He (2015) em seu estudo sobre a associação entre a obesidade e o risco de câncer encontraram um risco específico por causa para o câncer estatisticamente maior para a população obesa. Entretanto a incidência de câncer não foi estatisticamente maior na população obesa. Conforme os autores, um risco específico por causa estatisticamente maior não significa alta taxa de incidência deste evento. Deste modo, a influência de covariáveis sobre o risco específico por causa pode ser diferente da influência sobre a probabilidade de ocorrência de um evento (GILLAM *et al.*, 2011).

Embora, em geral, a relação entre as covariáveis e o risco específico por causa não possa ser traduzida em uma relação direta com o risco da subdistribuição (ANDERSEN *et al.*, 2012), observou-se que o aumento e a diminuição do risco específico por causa foi trasladado para o risco da subdistribuição, explicação para este fato se pauta no argumento de Beyersmann *et al.* (2012) supracitado.

4. CONCLUSÕES

As incidências para os eventos claudicação e outras injúrias obtidas com base no estimador de Kaplan-Meier superestimam as funções de incidências cumulativa para os respectivos eventos.

A incidência total obtida pela soma das incidências para cada um dos eventos, derivadas a partir do estimador de Kaplan-Meier é viesada.

As covariáveis avaliadas afetam da mesma maneira, aumentando ou diminuindo, o risco da subdistribuição, ou seja, a função de incidência cumulativa, e o risco específico por causa para o evento claudicação. Isto decorre do fato destas covariáveis não terem efeito sobre o evento competitivo.

A abordagem baseada na análise de sobrevivência padrão quando eventos competitivos estão presentes é inadequada, alternativa mais apropriada baseia-se na abordagem dos riscos competitivos.

5. AGRADECIMENTOS

À FAPEMIG pelo apoio financeiro.

6. REFERÊNCIAS BIBLIOGRÁFICAS

AALLEN, O. O.; BORGAN, O.; GJESSING, H. K. **Survival and event history analysis: a process point of view**. New York: Springer Science & Business Media, 2008. 539 p.

ANDERSEN, P. K.; GESKUS, R.B.; WITTE, T. de; PUTTER, H. Competing risks in epidemiology: possibilities and pitfalls. **International Journal of Epidemiology**, v. 41, n. 3, p. 861 - 870, 2012.

ASSANE, C.; PEREIRA, B. de; MUAUALO, M. A. M. Modelos de riscos competitivos para análise de sobrevivência de pacientes com hipertensão arterial em tratamento no HUCFF/UFRJ. In: **Simpósio Brasileiro de Pesquisa Operacional**, XLVII, 2015, Pernambuco. *Anais ... Pernambuco* [s.n], 2015. p. 1225 – 1236.

AUSTIN, P. C.; LEE, D. S.; FINE, J. P. Introduction to the analysis of survival data in the presence of competing risks. **Circulation**, v. 133, n. 6, p. 601-609, 2016.

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.

BEYERSMANN, J.; SCHUMACHER, M.; ALLIGNOL, A. **Competing risk and multistate models with R**. New York: Springer Science & Business Media, 2012. 245 p.

BIMALI, M.; HE, J. Association between Obesity and Cancer: An Analysis Using the Competing Risk Regression Approach. **Advances in Epidemiology**, v. 2015, 2015.

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. ver.ampl. Rio de Janeiro: Editora Fiocruz, 2011. 432 p.

COGGER, N.; EVANS, D.; PERKINS, N.; HODGSON, D.; REID, S. **Risk factors for injuries in thoroughbred racehorses**. Kingston: Rural Industries Research and Development Corporation, 2006, 54p. Relatório.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. 1. ed., São Paulo: Edgard Blucher, 2006. 370 p.

CROWDER, M. J. **Classical competing risks**. New York: CRC Press, 2001, 186 p.

FINE, J. P.; GRAY, R. J. A proportional hazard model for subdistribution of a competing risk. **Journal of the American Statistical Association**, v. 94, n. 446, p. 496-509, 1999.

GESKUS, R. B. **Data analysis with competing risks and intermediate states**. Boca Raton, FL: Chapman & Hall/CRC, 2015. 243 p.

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. 18, n. 6, p. 695 – 706, 1999.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observation. **Journal of the Statistical Association**, v. 53, n. 282, p. 457-481, 1958.

KIM, H. T. Cumulative incidence in competing risks data and competing risks regression analysis. **Clinical Cancer Research**, v. 13, n. 2, p. 559-565, 2007.

LINDNER, A.; DINGERKUS, A. Incidence of training failure among thoroughbred horses at Cologne. **Germany Preventive Veterinary Medicine**, v. 16, n.2, 1993, p. 85-94, 1993.

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.

PINTILIE, M. **Competing risk: a practical perspective**. Chichester: John Wiley & Sons, 2006. 224 p.

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, 2007.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2016. Disponível em:< <http://www.r-project.org/> >.

SCRUCCA, L. SANTUCCI, A.; AVERSA, F. Regression modeling of competing risk using R: an in-depth guide for clinicians. **Bone Marrow Transplantation**, v. 45, n. 9, p. 1388 – 1395, 2010.

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

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

WOLBERS, M.; KOLLER, M. T.; WITTEMAN, J. C. M.; STEYERBERG, E. W. Prognostic models with competing risks: methods and application to coronary risk prediction. **Epidemiology**, v. 20, n. 4, p. 555-561, 2009.

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. 17, n. 3, p. 259 – 269, 2013.

APÊNDICE A: Tabela de incidências estimadas para a claudicação ($\hat{F}_1(t)$) e para a probabilidade de claudicação calculada com base no estimador de Kaplan-Meier ($1 - \hat{S}_1(t)$)

Tempo	$\hat{F}_1(t)$	$1 - \hat{S}_1(t)$	Tempo	$\hat{F}_1(t)$	$1 - \hat{S}_1(t)$	Tempo	$\hat{F}_1(t)$	$1 - \hat{S}_1(t)$
1	0,0443	0,0443	52	0,2025	0,2107	111	0,4114	0,4423
5	0,0443	0,0443	54	0,2089	0,2175	112	0,4177	0,4494
6	0,0443	0,0443	58	0,2152	0,2243	113	0,4557	0,4923
7	0,0570	0,0573	59	0,2215	0,2311	114	0,4620	0,4997
8	0,0570	0,0573	60	0,2278	0,2380	116	0,4747	0,5144
9	0,0696	0,0704	61	0,2278	0,2380	117	0,4747	0,5144
10	0,0759	0,0769	62	0,2278	0,2380	118	0,4810	0,5219
11	0,0823	0,0835	63	0,2342	0,2449	119	0,4873	0,5295
12	0,0949	0,0966	64	0,2405	0,2518	121	0,5063	0,5526
14	0,0949	0,0966	68	0,2468	0,2587	122	0,5127	0,5606
16	0,1013	0,1032	72	0,2532	0,2657	123	0,5190	0,5686
17	0,1203	0,1232	76	0,2658	0,2795	124	0,5190	0,5686
18	0,1329	0,1364	82	0,2658	0,2795	125	0,5253	0,5767
20	0,1392	0,1431	85	0,2722	0,2865	127	0,5253	0,5767
21	0,1392	0,1431	86	0,2785	0,2935	128	0,5316	0,5852
23	0,1456	0,1498	92	0,2848	0,3005	130	0,5316	0,5852
28	0,1456	0,1498	98	0,2911	0,3075	133	0,5380	0,5938
29	0,1519	0,1565	100	0,3038	0,3215	135	0,5443	0,6025
34	0,1646	0,1700	103	0,3038	0,3215	140	0,5506	0,6113
36	0,1709	0,1768	105	0,3291	0,3498	141	0,5570	0,6202
37	0,1772	0,1835	106	0,3418	0,3639	150	0,5633	0,6290
40	0,1835	0,1903	107	0,3608	0,3851	151	0,5696	0,6378
45	0,1899	0,1971	109	0,3797	0,4065	156	0,5696	0,6378
49	0,1962	0,2039	110	0,3861	0,4137			

APÊNDICE B: Tabela de incidências estimadas para outras injúrias ($\hat{F}_2(t)$) e para a probabilidade de ocorrer outras injúrias calculada com base no estimador de Kaplan-Meier ($1 - \hat{S}_2(t)$)

Tempo	$\hat{F}_2(t)$	$1 - \hat{S}_2(t)$	Tempo	$\hat{F}_2(t)$	$1 - \hat{S}_2(t)$	Tempo	$\hat{F}_2(t)$	$1 - \hat{S}_2(t)$
1	0,0063	0,0063	52	0,0633	0,0694	111	0,0949	0,1141
5	0,0190	0,0196	54	0,0633	0,0694	112	0,0949	0,1141
6	0,0253	0,0262	58	0,0633	0,0694	113	0,1076	0,1371
7	0,0253	0,0262	59	0,0633	0,0694	114	0,1076	0,1371
8	0,0316	0,0329	60	0,0633	0,0694	116	0,1076	0,1371
9	0,0316	0,0329	61	0,0696	0,0777	117	0,1139	0,1501
10	0,0316	0,0329	62	0,0759	0,0860	118	0,1203	0,1632
11	0,0316	0,0329	63	0,0759	0,0860	119	0,1266	0,1765
12	0,0380	0,0398	64	0,0759	0,0860	121	0,1392	0,2035
14	0,0443	0,0468	68	0,0759	0,0860	122	0,1392	0,2035
16	0,0443	0,0468	72	0,0759	0,0860	123	0,1392	0,2035
17	0,0443	0,0468	76	0,0759	0,0860	124	0,1456	0,2182
18	0,0443	0,0468	82	0,0823	0,0948	125	0,1519	0,2330
20	0,0443	0,0468	85	0,0823	0,0948	127	0,1582	0,2480
21	0,0506	0,0542	86	0,0823	0,0948	128	0,1582	0,2480
23	0,0506	0,0542	92	0,0823	0,0948	130	0,1646	0,2634
28	0,0570	0,0617	98	0,0823	0,0948	133	0,1646	0,2634
29	0,0570	0,0617	100	0,0823	0,0948	135	0,1709	0,2791
34	0,0570	0,0617	103	0,0886	0,1041	140	0,1709	0,2791
36	0,0570	0,0617	105	0,0886	0,1041	141	0,1709	0,2791
37	0,0633	0,0694	106	0,0886	0,1041	150	0,1709	0,2791
40	0,0633	0,0694	107	0,0949	0,1141	151	0,1709	0,2791
45	0,0633	0,0694	109	0,0949	0,1141	156	0,1709	0,2791
49	0,0633	0,0694	110	0,0949	0,1141			

APÊNDICE C - Código para estimar as funções de incidência cumulativa, para ajustar modelo de Cox para o risco específico por causa e para ajustar modelo de Fine e Gray para o risco de subdistribuição.

```
cavalo<-read.csv("cavalos.csv",sep=";")
cavalo
```

```
# tempo denota o tempo até a ocorrência do primeiro evento em semanas.
#status é uma variável indicadora do tipo de evento:
#1: claudicação com ou sem a presença de outras injúrias.
# 2: outras injúrias.
# 0: observação censurada: o animal não sofreu lesões.
```

```
cavalo$claudicacao<-ifelse(cavalo$status==1, 1,0) #Cria uma variável que
#denota a ocorrência de claudicação.
cavalo$outras.injurias<-ifelse(cavalo$status==2,1,0) #Cria uma variável que
#denota a #ocorrência de outros tipos de injúrias
cavalo$evento<- ifelse(cavalo$status>0,1,0) #Cria uma variável que indica a
#ocorrência de um evento de qualquer tipo.
cavalo
attach(cavalo)
```

```
library(survival)
library(cmprsk)
```

```
#####
#Figura 1: Gráfico das funções de incidência cumulativas para a claudicação,
#outras injúrias e lesões por todas as causas.
#####
```

```
cif1<-cuminc(cavalo$tempo,cavalo$status, cencode=0)
plot(cif1$"1 1"$time,cif1$"1 1"$est, type="l", ylim=c(0,1),xlab="Tempo
(semanas)", ylab="Probabilidade")
lines(cif1$"1 2"$time,cif1$"1 2"$est, type="l", ylim=c(0,1),lty=2)
```

```
km.combinado<-survfit(Surv(tempo,evento)~1)
lines(km.combinado$time,1-km.combinado$surv, type="l",lty=3)
```

```
legend("topleft", legend=c("Claudicação com ou sem outras injúrias (FIC)",
"Outras injúrias (FIC)",
"Todas as causas de lesão (1-KM)/Soma das duas incidências"),
lty=c(1,2,3),bty="n")
```

```
#####
#Tabela 1 - Incidências estimadas para a claudicação, (F1(t)), outras injúrias,
#F2(t), e #incidência total (F(t)), calculada pela soma das incidências
#específicas e a incidência #total (F(t))calculada com base no estimador de
```

```

#Kaplan-Meier (1-#S(t)).
#####

estimativa1.km<-1-km.combinado$surv ; estimativa1.km
estimativa1.cif<-timepoints(cif1, times=km.combinado$time); estimativa1.cif
estimativa1.cif$est
cif.total<- estimativa1.cif$est[1,]+estimativa1.cif$est[2,]; cif.total
cbind(km.combinado$time,
estimativa1.cif$est[1,],estimativa1.cif$est[2,],cif.total,estimativa1.km)

```

```

#####
# Figura 2: Gráfico da função de incidência cumulativa total e estimativas da
#probabilidade da incidência total baseada nas estimativas de Kaplan-Meier
#para a claudicação e outras injúrias
#####

```

```

km.claudicacao<-survfit(Surv(cavalo$tempo, claudicacao)~1); km.claudicacao
km.outras<-survfit(Surv(cavalo$tempo, outras.injurias)~1); km.outras

soma.km<-(1-km.claudicacao$surv)+(1-km.outras$surv); soma.km
#soma a probabilidade de claudicação com a probabilidade de outras injúrias
#baseada no estimador de Kaplan-Meier.

```

```

cif.total<- estimativa1.cif$est[1,]+estimativa1.cif$est[2,]

```

```

tempo1<-km.claudicacao$time
plot(tempo1,soma.km, type='s',xlab="Tempo
(semanas)",ylab="Probabilidade",lty=3)

```

```

lines(tempo1, cif.total,type='s',lty=1)

```

```

legend('topleft',legend=c("Incidência total (Soma das duas FICs)",
"Probabilidade total baseada na soma das estimativas
de Kaplan-Meier"), lty=c(1,3),bty='n')

```

```

#####
# Figura 3: Gráfico da função de incidência cumulativa para a claudicação e
#estimativas de Kaplan-Meier para o evento claudicação
#####

```

```

plot(cif1$"1 1"$time,cif1$"1 1"$est, type="l", ylim=c(0,1),xlab="Tempo
(semanas)", ylab="Probabilidade")

```

```

km.claudicacao<-survfit(Surv(cavalo$tempo, claudicacao)~1)

```

```

lines(km.claudicacao$time,1-km.claudicacao$surv, type="l",lty=2)

```

```

legend("topleft", legend=c("Função de incidência para a claudicação",
"Estimação da probabilidade de claudicação
baseada na estimativa de Kaplan-Meier"), lty=c(1,2),bty="n")

```

```
#####
#Figura 4: Gráfico da função de incidência cumulativa para outras injúrias e
#estimativas de Kaplan-Meier para o evento outras injúrias
#####
```

```
plot(cif1$"1 2"$time,cif1$"1 2"$est, type="l", ylim=c(0,1),xlab="Tempo
(semanas)", ylab="Probabilidade")
km.outras<-survfit(Surv(cavalo$tempo, outras.injurias)~1)
lines(km.outras$time,1-km.outras$surv, type="l",lty=2)
```

```
legend("topleft", legend=c("Função de incidência para outras injúrias",
"Estimação da probabilidade de outras injúrias baseada na estimativa de
Kaplan-Meier"), lty=c(1,2),bty="n")
```

```
#####
#Modelos de regressão para o risco da subdistribuição e modelos de regressão
#para o risco específico por causa.
#####
```

```
cov<-cbind(cavalo$sexo,cavalo$idade,cavalo$peso,cavalo$estacoes);cov
```

```
crr.1<-crr(cavalo$tempo,cavalo$status,cov,failcode=1,cencode=0) #Modelo de
#regressão de Fine e Gray para claudicação.
crr.1
```

```
crr.2<-crr(cavalo$tempo,cavalo$status,cov,failcode=2,cencode=0) #Modelo de
#regressão de Fine e Gray para outras injúrias.
crr.2
```

```
cox.1<-coxph(Surv(cavalo$tempo,
claudicacao)~cavalo$sexo+cavalo$idade+cavalo$peso+cavalo$estacoes)
#Modelo de regressão de Cox para o risco específico de claudicação.
cox.1
```

```
cox.2<-coxph(Surv(cavalo$tempo,
outras.injurias)~cavalo$sexo+cavalo$idade+cavalo$peso+cavalo$estacoes)
#Modelo de regressão de Cox para o risco específico para outras injúrias.
cox.2
```

```
fine.1<- summary(crr.1, digits=4); fine.1
fine.2<-summary(crr.2, digits=4); fine.2
coxph.1<- summary(cox.1); coxph.1
coxph.2<-summary(cox.2); coxph.2
round(fine.1$coef,4)
round(fine.1$conf.int,4)
round(fine.2$coef,4)
round(fine.2$conf.int,4)
round(coxph.1$coefficients,4)
```



```
round(coxph.1$conf.int,4)
round(coxph.2$coefficients,4)
round(coxph.2$conf.int,4)
```

```
#####
#Tabela (Apêndice A) - Incidências estimadas para a claudicação (F1(t)) e
#estimativas da probabilidade de claudicação baseada no estimador de
#Kaplan-Meier.
#####
```

```
estimativa2.km<-1-km.claudicao$surv; estimativa2.km
estimativa2.cif<-timepoints(cif1, times=km.claudicao$time); estimativa2.cif
estimativa2.cif$est #fornece a estimativa da incidência da claudicação
cbind(km.claudicao$time, estimativa2.cif$est[1,], estimativa2.km)
```

```
#####
#Tabela (Apêndice B) - Incidências estimadas para outras injúrias (F2(t)) e
#estimativas da probabilidade de ocorrência de outras injúrias baseada no
#estimador de Kaplan-Meier.
#####
```

```
estimativa3.km<-1-km.outras$surv; estimativa3.km
estimativa3.cif<-timepoints(cif1, times=km.outras$time); estimativa3.cif
estimativa3.cif$est #fornece a estimativa da incidência da outras injúrias
cbind(km.outras$time, estimativa3.cif$est[2,], estimativa3.km)
```

CAPÍTULO 4

Avaliação de modelos de regressão para a função de incidência cumulativa na presença de riscos competitivos.

RESUMO

O objetivo deste trabalho foi comparar modelos de regressão para a função de incidência baseados na abordagem da regressão binomial, o modelo de regressão com função de ligação logarítmica (ARR) e o modelo de regressão com função de ligação logit (LRR), com o modelo de regressão de Fine e Gray (FG). Para verificar o desempenho dos três modelos supracitados, realizou-se, inicialmente, um estudo de simulação. Neste estudo foram consideradas duas causas de falhas, diferentes níveis de censuras e diferentes proporções de ocorrência do evento de interesse (evento 1) e do evento competitivo (evento 2). Posteriormente, procedeu-se a comparação dos três modelos por meio da análise de um conjunto de dados reais referente à lesões em 158 equinos. As covariáveis consideradas foram sexo, idade inicial, peso e o número de estações de treino. Os resultados do estudo de simulação apontam que o modelo de Fine e Gray apresentou o melhor desempenho, entretanto, o modelo ARR mostrou-se uma alternativa ao modelo anterior. Porém, em geral, o modelo LRR pareceu não ser adequado para a modelagem de dados de riscos competitivos. No que se refere a análise de dados reais os modelos considerados foram concordantes na identificação de covariáveis significativamente associadas à probabilidade de claudicação. Ademais, as probabilidades previstas para a claudicação por esses três modelos foram próximas, indicando que tanto o modelo com função de ligação logarítmica como o modelo com função de ligação logit são boas alternativas para a modelagem da incidência de claudicação. Entretanto, na modelagem do evento “outras injúrias” os modelos considerados foram divergentes na identificação das covariáveis significativas e as probabilidades previstas por esses modelos foram bem distintas.

Palavras-chave: Análise de sobrevivência, Eventos competitivos, Modelos de regressão para a subdistribuição.

ABSTRACT

The objective of this study was to compare the regression models for the incidence function based on the direct binomial regression model, regression model with logarithmic-link function (ARR), and regression model with logit-link function (LRR), Fine and Gray regression model (FG). In order to verify the performance of the three models above, a simulation study was carried out initially. In this study two causes of failure, different levels of censorship and different proportions of occurrence of the event of interest (event 1) and of the competing event (event 2) were considered. Subsequently, the three models were compared by of the analysis of a set of real data regarding the injuries in 158 equines. The covariates considered were gender, initial age, weight and the number of training stations. The results of the simulation study show that the Fine and Gray model presented the best performance, however, the ARR model was an alternative to the previous model. However, in general, the LRR model seems not to be suitable for modeling competing risk data. Regarding analysis of real data, the models considered were concordant in the identification of covariates significantly associated with the probability of lameness. In addition, the predicted probabilities for lameness by these three models were close, indicating that both the model with log-link function and the model with logit-link function are good alternatives for modeling the incidence for lameness. However, in the modeling of the event "other injuries" the models considered were divergent in the identification of significant covariates and the predicted probabilities by these models were quite different.

Keywords: Survival analysis, Competing events, Regression models for subdistribution.

1. INTRODUÇÃO

Os riscos competitivos surgem quando um indivíduo pode experimentar mais de um evento, isto é, pode falhar por mais de uma causa, sendo que a ocorrência de um evento impede a ocorrência de outros ou altera fundamentalmente a probabilidade de ocorrência dos outros tipos de eventos (CARVALHO *et al.*, 2011; GOOLEY *et al.*, 1999).

Em análise de sobrevivência, na presença de riscos competitivos, uma quantidade de interesse é a função de incidência cumulativa, também denominada de subdistribuição (GESKUS, 2015). Essa função permite quantificar a probabilidade de um indivíduo experimentar um evento pertencente a um conjunto de causas distintas de falha.

Para a modelagem da função de incidência cumulativa frequentemente é usado o modelo de Fine e Gray (FINE e GRAY, 1999). Esse modelo é muito popular devido a sua similaridade com o modelo de Cox (COX, 1972) e da implementação desta metodologia em softwares estatísticos como o R e Stata.

Entretanto alguns autores (KLEIN *et al.*, 2014; SCHEIKE e ZHANG, 2011; SCHEIKE *et al.*, 2008) discutem modelos alternativos ao modelo de Fine e Gray, denominados como modelos de transformação, e cuja estimação é pautada na abordagem da regressão binomial. Neste enfoque funções de ligação tais como ligação logarítmica, ligação logit e ligação complementar log-log podem ser utilizadas. O modelo de Fine e Gray corresponde ao uso da função de ligação complementar log-log no modelo de transformação (GRAW *et al.*, 2009).

Embora existam modelos alternativos ao modelo de regressão de Fine e Gray, estes são pouco utilizados na prática, todavia a razão para isto não é clara (GERDS *et al.*, 2012).

Neste sentido, o objetivo deste trabalho é avaliar o desempenho do modelo de regressão para subdistribuição com função de ligação logarítmica (ARR), o modelo de regressão com função de ligação logit (LRR) e o modelo de Fine e Gray (FG) na modelagem do risco da subdistribuição.

2. MATERIAL E MÉTODOS

2.1. Simulação de dados

Os dados foram simulados segundo o algoritmo apresentado por Haller (2014) com algumas alterações. Considerou-se, na simulação, duas causas de falha, assim como feito por Leoce (2016) e Fine e Gray (1999). Para a causa de falha principal (evento 1) considerou-se três covariáveis $X = (X_1, X_2, X_3)$, tais que, $X_1 \sim$ Uniforme $[0; 3]$, $X_2 \sim$ Bernoulli $(0,5)$ e $X_3 \sim$ Normal $(0; 1)$ e os valores dos parâmetros iguais a $\beta = (\beta_1, \beta_2, \beta_3) = (\ln(1,15); \ln(0,9); \ln 2)$ conforme Haller (2014).

Avaliou-se seis níveis de censura (0%, 10%, 20%, 30%, 40% e 50%) e diferentes proporções de ocorrência do evento 1 e do evento 2.

O processo de simulação foi repetido 1000 vezes e cada conjunto de dados simulado continha 100 observações.

Para cada cenário, ou seja, combinação do nível de censura e proporções dos eventos 1 e 2 foram estimados os coeficientes de regressão para as três covariáveis citadas acima usando o modelo de regressão de Fine e Gray, modelo de regressão com a função de ligação logarítmica (ARR) e o modelo de regressão com função de ligação logit (LRR). Em seguida, obteve-se a média e o desvio dos coeficientes de regressão estimados, $\hat{E}(\hat{\beta})$ e $S(\hat{\beta})$, a partir das 1000 réplicas.

2.2. Dados sobre lesões em equinos

Para a comparação dos modelos de regressão na presença de riscos competitivos, considerou-se dados sobre lesões em equinos de corrida. Neste conjunto de dados, 158 animais foram acompanhados, sendo que 90 experimentaram a claudicação, 27 outras injúrias e 41 foram censurados.

Foram utilizados como covariáveis para o evento principal (claudicação) e para o evento competitivo (outras injúrias): o sexo, a idade inicial, o peso categorizado (1 para peso inferior a 480 kg e 2 para peso igual ou superior a 480 kg) e o número de estações de treino.

Para cada causa de falhas foram ajustados os modelos de regressão para o risco absoluto (função de incidência) considerando a função de ligação logarítmica e a função de ligação logit, e também o modelo de Fine e Gray, o mais comumente utilizado na modelagem de riscos competitivos.

Para a avaliação dos modelos de regressão alternativos (modelo de regressão com a função de ligação logarítmica e modelo de regressão com a função de ligação logit), comparou-se graficamente a incidência cumulativa predita de cada um dos modelos com a predita pelo modelo de regressão de Fine e Gray, conforme sugerido por Gerds *et al.*(2012).

2.3. Métodos

A regressão de Fine e Gray (1999) modela o risco da subdistribuição (GRAY, 1988) dado por

$$\gamma_k(t|X) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{P(t < \tilde{T} \leq t + \Delta t, \varepsilon = k | \tilde{T} > t \text{ ou } (\tilde{T} \leq t \text{ e } \varepsilon \neq k))}{\Delta t} \right\} = \frac{-d \log\{1 - F_k(t)\}}{dt} \quad (1)$$

tal que

$$\gamma_k(t|X) = \gamma_{k0}(t) \exp(\beta_k^T X), \quad (2)$$

em que γ_k é o risco da subdistribuição, $\gamma_{k0}(t)$ é o risco de base da subdistribuição, X é o vetor de covariáveis e β_k é o vetor de coeficientes de regressão. Assume-se, neste modelo, que a incidência cumulativa para a causa k (BIMALI e HE, 2015; HALLER *et al.*, 2013; SCHEIKE *et al.*, 2008; SUN *et al.*, 2006; ZHANG *et al.*, 2008) é dada por:

$$F_k(t, X_i) = 1 - \exp\{-\Lambda_0(t) \exp(\beta_k^T X_i)\}. \quad (3)$$

A função de verossimilhança parcial é dada por (PINTILIE, 2006):

$$L(\beta_k) = \prod_{j=1}^r \frac{\exp(\beta_k^T X_j)}{\sum_{i \in R_j} w_{ji} \exp(\beta_k^T X_i)} \quad (4)$$

em que R_j é o conjunto de risco definido como:

$R_j = \{i; T_i \geq t \text{ ou } (T_i \leq t \text{ e o indivíduo experimentou um evento de risco competitivo})\}$.

Assim o conjunto de indivíduos sob risco inclui, além dos indivíduos que não sofreram qualquer evento, aqueles que sofreram o evento competitivo, e que até a censura administrativa não saem do conjunto de risco (CARVALHO *et al.*, 2011).

Os w_{ji} são pesos que indicam a contribuição de cada indivíduo para a função de verossimilhança, definidos como:

$$w_{ji} = \frac{\hat{G}(t_j)}{\hat{G}(\min(t_j, t_i))} \quad (5)$$

em que \hat{G} é o estimador de Kaplan-Meier da distribuição da censura.

Para o ajuste dos modelos de regressão à função de incidência, usando as funções de ligação logarítmica (log) e ligação logit, baseou-se na abordagem de regressão binomial (KLEIN *et al.*, 2014; SCHEIKE e ZHANG, 2011; SCHEIKE *et al.*, 2008). Para tanto considerou-se um modelo linear generalizado para o risco absoluto de um evento:

$$F_\varepsilon(t) = E(N_\varepsilon(t)|X) = h(\beta_0(t), \beta(t), X). \quad (6)$$

O modelo descreve o efeito do vetor de variáveis preditoras X por meio de um vetor de coeficientes de regressão $\beta(t)$, uma função de ligação e um intercepto $\beta_0(t)$ (KLEIN *et al.*, 2014).

Algumas funções de ligação utilizadas são a logarítmica, logit e complementar log-log. A última, fornece o modelo de Fine e Gray (GERDS *et al.*, 2012).

No processo de estimação trabalha-se com a resposta ponderada de um processo de contagem, $\tilde{N}_\varepsilon(t) = \frac{\Delta N_\varepsilon(t)}{G(T^-|X)}$ sendo que

$$E(\tilde{N}_\varepsilon(t)) = E\left(E\left(\frac{\Delta N_\varepsilon(t)}{G(T^-|X)} \mid T, \varepsilon, X\right)\right) = E(N_\varepsilon(t)|X) = F_\varepsilon(t|X). \quad (7)$$

em que ε representa o tipo de falha $\varepsilon \in \{1, \dots, k\}$, $G(t|X) = P(C > t|X)$ denota a probabilidade de uma observação não ser censurada, $\Delta_i = I\{T_i \leq C_i\}$, variável indicadora de censura e N_ε o processo de contagem associado a causa ε .

O valor de $\beta(t)$ soluciona a equação $U_n(\beta_0(t), \beta(t), G) = 0$ (KLEIN *et al.*, 2014) sendo

$$U_n(\beta_0(t), \beta(t), \hat{G}) = \sum_{i=1}^n w(t, x_i) \left\{ \frac{\Delta_i N_{\varepsilon,i}(t)}{\hat{G}(T_i|X_i)} - h(\beta_0(t), \beta(t), X_i) \right\}. \quad (8)$$

Para a implementação dos modelos no software R (R DEVELOPMENT CORE TEAM, 2016) foram utilizados os pacotes *cmprsk* e *riskRegression*.

3. RESULTADOS E DISCUSSÃO

Inicialmente, são mostrados os resultados obtidos por meio do estudo de simulação usado para avaliar o desempenho dos modelos de Fine e Gray, modelo de regressão com função de ligação log (ARR) e o modelo de regressão com função de ligação logit (LRR).

A partir da Tabela 1 pode-se observar que quando as proporções de ocorrência dos eventos estão próximas, as estimativas obtidas para quaisquer modelos têm menor desvio padrão independentemente do nível de censura considerado.

Em geral, o modelo de Fine e Gray apresentou coeficientes de mesma direção e magnitude próxima a dos verdadeiros valores para os coeficientes das covariáveis $\beta = (\beta_1, \beta_2, \beta_3) = (\ln(1,15); \ln(0,9); \ln 2)$, o que é esperado tendo em vista que os dados foram gerados a partir de um modelo de riscos proporcionais para subdistribuição (Tabela 1).

O modelo de transformação utilizando a função de ligação logarítmica (ARR) apresentou um bom desempenho quando a proporção de eventos 1 e 2 eram próximas, contudo, quando o número de eventos 2 era demasiadamente maior do que a do evento 1 os coeficientes estimados, em geral, apresentam

magnitudes consideravelmente diferentes dos verdadeiros valores dos coeficientes das covariáveis e maiores desvios padrão o que ocasiona pouca precisão das estimativas obtidas (Tabela 1).

O modelo de transformação utilizando a função de ligação logit (LRR) apresentou o pior desempenho, tendo em vista que suas estimativas, na maior parte dos casos, apresentam maiores diferenças em magnitude em relação aos verdadeiros valores e também maiores desvios padrão. Menores distorções entre os coeficientes estimados e os verdadeiros valores são notados quando a proporção de eventos do tipo 1 e evento do tipo 2 são próximas (Tabela 1).

A taxa de censura, embora afete a estimação dos coeficientes, parece não ser determinante no modelo de Fine e Gray. Para este, a proporção de eventos 1 e 2 parece afetar mais intensamente o desempenho do modelo (Tabela 1).

No que diz respeito ao modelo ARR, este apresenta um bom desempenho desde que a proporção do evento 2 não seja muito maior do que a do evento do tipo 1.

Entretanto, o modelo LRR apresenta o pior desempenho, todavia, este mau desempenho é atenuado quando as proporções entre os eventos estão próximas. Além disso quanto maior a ocorrência do evento 2 em relação ao evento 1, pior o desempenho deste modelo (Tabela 1).

Tabela 1 – Estimativa da média e desvio padrão dos coeficientes estimados pelos modelos de regressão de Fine e Gray (FG), modelo de regressão com ligação log (ARR) e modelo de regressão com função de ligação logit (LRR).

(Continua)

Censura	Evento 1	Evento 2		Modelo FG $\hat{E}(\hat{\beta}) (s(\hat{\beta}))$	Modelo ARR $\hat{E}(\hat{\beta}) (s(\hat{\beta}))$	Modelo LRR $\hat{E}(\hat{\beta}) (s(\hat{\beta}))$
0%	25%	75%	$\hat{\beta}_1$	0,1548 (0,2475)	0,1648 (0,4707)	1,4426 (36,2961)
			$\hat{\beta}_2$	- 0,1198 (0,2288)	- 0,1224 (0,2634)	- 1,5170 (39,8716)
			$\hat{\beta}_3$	0,7354 (0,5897)	0,7236 (0,6377)	2,8732 (60,9997)
0%	39%	61%	$\hat{\beta}_1$	0,1504 (0,1941)	0,1191 (0,1760)	0,1855 (0,2645)
			$\hat{\beta}_2$	-0,1225(0,1645)	- 0,1011 (0,1534)	- 0,1588 (0,2331)
			$\hat{\beta}_3$	0,7216 (0,3475)	0,6167 (0,3369)	0,8938 (0,4774)
0%	52%	48%	$\hat{\beta}_1$	0,1455 (0,1736)	0,1033 (0,1403)	0,1894 (0,2520)
			$\hat{\beta}_2$	- 0,1110 (0,1475)	- 0,0780 (0,1186)	- 0,1444 (0,2203)
			$\hat{\beta}_3$	0,7438 (0,2970)	0,5773 (0,2650)	0,9639 (0,4249)
10%	22%	68%	$\hat{\beta}_1$	0,1327 (0,2634)	0,2818 (4,0182)	1,4795 (27,0439)
			$\hat{\beta}_2$	- 0,1042 (0,2257)	0,3330(14,1412)	- 1,4661(33,9421)
			$\hat{\beta}_3$	0,7550 (0,4961)	0,8711 (2,7675)	3,0858 (44,4196)
10%	35%	55%	$\hat{\beta}_1$	0,1492 (0,2001)	0,1296 (0,2021)	0,1898 (0,2905)
			$\hat{\beta}_2$	-0, 1113 (0,1826)	- 0,0921 (0,1749)	- 0,1387 (0,2587)
			$\hat{\beta}_3$	0,7278 (0,3841)	0,6464 (0,3905)	0,8899 (0,5339)
10%	45%	45%	$\hat{\beta}_1$	0,1508 (0,1814)	0,1245 (0,1545)	0,1960 (0,2512)
			$\hat{\beta}_2$	- 0,1087 (0,1591)	- 0,0850 (0,1455)	- 0,1437 (0,2357)
			$\hat{\beta}_3$	0,6627 (0,3320)	0,5298 (0,3019)	0,8212 (0,4590)
20%	19%	61%	$\hat{\beta}_1$	0,1405 (0,2927)	0,1793 (4,1684)	- 1,8321 (49,7141)
			$\hat{\beta}_2$	- 0,1013 (0,2492)	- 0,1495 (3,1494)	- 0,4670 (23,5995)
			$\hat{\beta}_3$	0,7286 (0,5241)	0,6774 (3,4581)	2,1428 (28,9232)
20%	30%	50%	$\hat{\beta}_1$	0,1515 (0,2220)	0,1362 (0,3209)	0,2310 (1,1585)
			$\hat{\beta}_2$	- 0,1122 (0,1922)	-0, 1010 (0,2094)	- 0,1456 (0,3182)
			$\hat{\beta}_3$	0,7117 (0,3894)	0,6617 (0,4269)	0,8930 (0,8231)
20%	39%	41%	$\hat{\beta}_1$	0,1383 (0,1965)	0,1153 (0,1828)	0,1780 (0,2796)
			$\hat{\beta}_2$	- 0,1188 (0,1761)	- 0,0915(0,1594)	- 0,1478 (0,2495)
			$\hat{\beta}_3$	0,7557 (0,3475)	0,6198 (0,3361)	0,8994 (0,4845)
30%	17%	53%	$\hat{\beta}_1$	0,1574 (0,3174)	1,3212(24,7612)	9,3986 (153,5484)
			$\hat{\beta}_2$	- 0,1206 (0,2691)	- 0,1451(3,3196)	- 0,1724 (56,9233)
			$\hat{\beta}_3$	0,7210 (0,7466)	0,7046 (4,9842)	3,5510 (50,9841)
30%	26%	44%	$\hat{\beta}_1$	0,1410 (0,2420)	0,1492 (0,7999)	0,9339 (14,8970)
			$\hat{\beta}_2$	- 0,0911 (0,2187)	- 0,0680(1,0724)	- 0,3496 (19,1360)
			$\hat{\beta}_3$	0,6997 (0,4535)	0,7380 (1,2488)	1,6857 (14,981)
30%	34%	36%	$\hat{\beta}_1$	0,1418 (0,2041)	0,1220 (0,2231)	0,1794 (0,2901)
			$\hat{\beta}_2$	- 0,1139 (0,1892)	- 0,0926(0,1831)	- 0,1380 (0,2627)
			$\hat{\beta}_3$	0,7129 (0,3952)	0,6424 (0,4061)	0,8735 (0,5413)

Tabela 1 – Estimativa da média e desvio padrão dos coeficientes estimados pelos modelos de regressão de Fine e Gray (FG), modelo de regressão com ligação log (ARR) e modelo de regressão com função de ligação logit (LRR).

							(Conclusão)
Censura	Evento 1	Evento 2		Modelo FG $\hat{E}(\hat{\beta}) (s(\hat{\beta}))$	Modelo ARR $\hat{E}(\hat{\beta}) (s(\hat{\beta}))$	Modelo LRR $\hat{E}(\hat{\beta}) (s(\hat{\beta}))$	
40%	14%	46%	$\hat{\beta}_1$	0,1649 (0,3518)	1,2212(21,0361)	8,9754 (95,0077)	
			$\hat{\beta}_2$	- 0,1112 (0,3065)	- 0,2567(6,3734)	- 7,4325 (90,3249)	
			$\hat{\beta}_3$	0,7995 (0,9086)	1,3444 (5,6092)	5,6854 (70,1424)	
40%	22%	38%	$\hat{\beta}_1$	0,1490 (0,2808)	0,1470 (0,6053)	0,2219 (35,7589)	
			$\hat{\beta}_2$	- 0,1101 (0,2387)	- 0,1122(0,4477)	- 1,3901 (28,8228)	
			$\hat{\beta}_3$	0,773 (0,6297)	0,8040 (0,8857)	1,3676 (24,7535)	
40%	28%	32%	$\hat{\beta}_1$	0,1331 (0,2225)	0,1161 (0,2228)	0,1621 (0,3098)	
			$\hat{\beta}_2$	- 0,1093 (0,1969)	- 0,1078(0,2242)	- 0,1780 (0,5941)	
			$\hat{\beta}_3$	0,7082 (0,4527)	0,6700 (0,4723)	0,8638 (0,7267)	
50%	12%	38%	$\hat{\beta}_1$	0,1690 (0,4089)	2,0613(35,0018)	19,4230 (244,1380)	
			$\hat{\beta}_2$	- 0,1148 (0,3400)	2,0904(78,4404)	- 5,8386 (114,2733)	
			$\hat{\beta}_3$	0,9655 (1,5934)	0,3549(32,2145)	6,1028 (83,9648)	
50%	18%	32%	$\hat{\beta}_1$	0,1506 (0,3007)	0,2702 (4,7125)	1,5534 (21,4979)	
			$\hat{\beta}_2$	- 0,1129 (0,2599)	- 0,1075(3,9076)	- 1,3069 (38,2990)	
			$\hat{\beta}_3$	0,7270 (0,6769)	0,8017 (3,9309)	0,4759 (40,9200)	
50%	23%	27%	$\hat{\beta}_1$	0,1507 (0,2596)	0,1677 (0,4326)	0,5086 (7,8850)	
			$\hat{\beta}_2$	- 0,1123 (0,2169)	- 0,1092(0,3316)	- 0,2084 (5,2021)	
			$\hat{\beta}_3$	0,7365 (0,4926)	0,7649 (0,7542)	1,2837 (9,0147)	

A seguir são mostrados os resultados da análise dos dados sobre lesões em equinos. Na Tabela 2 é apresentado o resultado obtido pelo uso de diferentes modelos de regressão para incidência cumulativa, com efeito das covariáveis sobre o evento claudicação. Os três modelos de regressão para a subdistribuição do risco de claudicação indicaram que as covariáveis idade e estações são estatisticamente significativas. O modelo de regressão com a função de ligação logarítmica (ARR) apresentou o menor erro padrão para todos os coeficientes de regressão estimados.

A variável idade, independentemente do modelo usado, reduz significativamente a probabilidade de claudicação, enquanto que a variável estações aumenta significativamente a probabilidade de claudicação.

Contudo a interpretação do efeito da variável, ou seja, exponencial do coeficiente estimado, $\exp(\text{coef})$, dependerá do modelo considerado. Assim, no modelo de regressão para a incidência cumulativa utilizando a função de ligação logarítmica, o $\exp(\text{coef})$ corresponde a razões log transformadas dos riscos

absolutos (incidências cumulativas) para dois indivíduos que têm os mesmos valores de covariáveis com exceção da covariável considerada (KLEIN *et al.*, 2014). Já quando se considera o modelo com a função de ligação logit, o exp (coef) corresponde a razão das chances, ou seja, razão entre a probabilidade de um indivíduo experimentar o evento e a probabilidade de não experimentar o evento. Conforme destacado por Gerds *et al.* (2012) a probabilidade complementar de ocorrer um evento de interesse, ou seja, de não experimentar o evento em questão, inclui tanto a possibilidade de não experimentar um evento de interesse como a de experimentar um evento competitivo.

Dessa forma, considerando o modelo com a função de ligação logarítmica, a cada 1 mês de aumento na idade implica em uma diminuição significativa na probabilidade de claudicação ($\exp(\text{coef}) = 0,9711$, IC[95%] = [0,9438; 0,9993], valor-p=0,0443). Ao se analisar o modelo logístico constata-se um decréscimo estatisticamente significativo e maior do que o anterior na probabilidade de claudicação ($\exp(\text{coef}) = 0,9437$, IC [95%]= [0,8966; 0,9933], valor – p=0,0266), portanto a chance de experimentar claudicação é aproximadamente 6% menor do que a de não experimentar claudicação neste modelo.

No que se refere ao modelo de Fine e Gray tem-se uma redução no risco da subdistribuição ($\exp(\text{coef}) = 0,9477$, IC [95%] = [0,908; 0,989], valor-p = 0,0146) e conseqüentemente uma redução na probabilidade de claudicação.

Nota-se que a magnitude do efeito da covariável idade foram próximos, especialmente para os modelos com função de ligação logit e o modelo de Fine e Gray.

Avaliando o efeito da covariável estações, tem-se que no modelo com a função de ligação logarítmica, a cada aumento de 1 estação ocorrerá um aumento de aproximadamente 26% na probabilidade de claudicação. Para o modelo com a função de ligação logit, pode-se observar um aumento de aproximadamente 82% na chance de animal experimentar um evento claudicação em relação a de não experimentar. Analisando o risco da subdistribuição ($\exp(\text{coef}) = 1,3868$, IC [95%] = [1,150; 1,672], valor-p = 0,0006), tem-se um aumento aproximado de 39% no risco de claudicação.

Tabela 2 - Resultados do ajuste de diferentes modelos de riscos competitivos para o evento claudicação: modelo de regressão com funções de ligação log e logit e o ajuste do modelo de Fine e Gray.

Modelo de regressão com ligação logarítmica					
covariável	coef.	exp(coef)	se (coef)	IC_95%	valor-p
sexo	0,0700	1,0725	0,1710	[0,7670; 1,4995]	0,6824 ^{ns}
idade	- 0,0293	0,9711	0,0146	[0,9438; 0,9993]	0,0443*
peso	0,1799	1,1971	0,1656	[0,8652; 1,6562]	0,2776 ^{ns}
estações	0,2341	1,2638	0,0480	[1,1503; 1,3885]	<0,0001*

Modelo de regressão com ligação logit					
covariável	coef.	exp(coef)	se (coef)	IC_95%	valor-p
sexo	0,0640	1,0661	0,3384	[0,5492; 2,0693]	0,8500 ^{ns}
idade	- 0,0580	0,9437	0,0261	[0,8966; 0,9933]	0,0266*
peso	0,3081	1,3609	0,3356	[0,7049; 2,6271]	0,3586 ^{ns}
estações	0,5968	1,8163	0,1434	[1,3712; 2,4058]	< 0,0001*

Modelo de regressão de Fine e Gray					
covariável	coef.	exp(coef)	se (coef)	IC_95%	valor-p
sexo	0,1918	1,2114	0,2397	[0,757; 1,938]	0,4236 ^{ns}
idade	- 0,0535	0,9477	0,0219	[0,908; 0,989]	0,0146*
peso	0,2157	1,2407	0,2319	[0,788; 1,955]	0,3523 ^{ns}
estações	0,3270	1,3868	0,0955	[1,150; 1,672]	0,0006*

coef. – coeficiente de regressão estimado; se (coef) – erro padrão do coeficiente de regressão; IC_95% - intervalo de confiança com 95% de confiança, ns – não significativo a 5%; * significativo a 5%.

A Tabela 3 mostra o efeito das covariáveis sobre a incidência cumulativa para o evento outras injúrias. No modelo de Fine e Gray todas as covariáveis foram não significativas. Contudo, o modelo de regressão com a função de ligação logarítmica apresentou duas covariáveis significativas (idade e estações) enquanto que no modelo com a função de ligação logit somente a covariável idade foi significativa.

No modelo com a função ligação logarítmica, a covariável idade aumentou significativamente a probabilidade de um animal experimentar “outras injúrias” em aproximadamente 8% enquanto a variável estações aumentou em aproximadamente 76% (Tabela 3).

Para o modelo com função de ligação logit, o aumento acarretado pela covariável idade na chance de um animal sofrer claudicação foi de aproximadamente (5%) (Tabela 3).

Tabela 3 - Resultados do ajuste de diferentes modelos de riscos competitivos para o evento outras injúrias: modelo de regressão com funções de ligação log e logit e o ajuste do modelo de Fine e Gray.

Modelo de regressão com função de ligação logarítmica					
covariável	coef.	exp(coef)	se (coef)	IC_95%	valor-p
sexo	-1,2597	0,2837	0,7297	[0,0678; 1,1859]	0,0843 ^{ns}
idade	0,0781	1,0812	0,0159	[1,0481; 1,1153]	<0,0001 [*]
peso	1,1348	3,1107	0,7074	[0,7775; 12,4449]	0,1087 ^{ns}
estações	0,5627	1,7553	0,1474	[1,3149; 2,3433]	0,0001 [*]

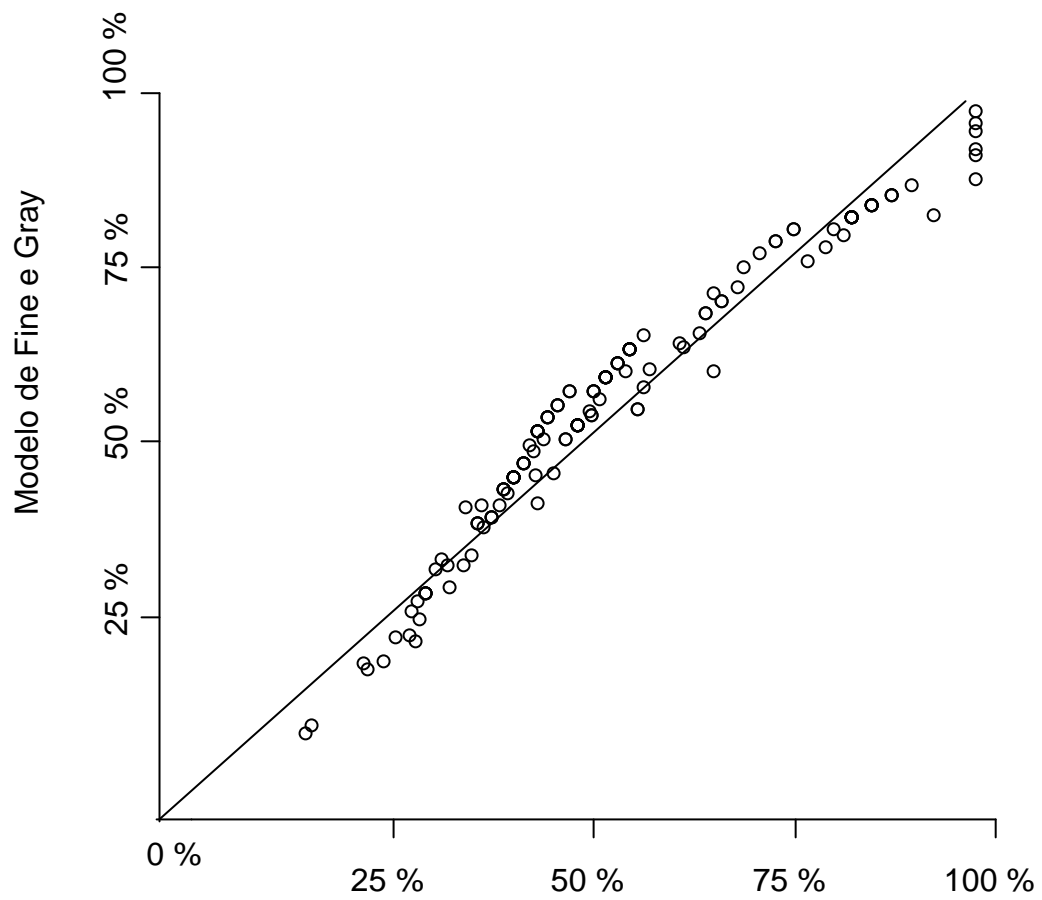
Modelo de regressão com função de ligação logit					
covariável	coef.	exp(coef)	se (coef)	IC_95%	valor-p
sexo	- 1,0627	0,3455	0,6157	[0,1034; 1,1550]	0,0844 ^{ns}
idade	0,0514	1,0528	0,0255	[1,0014; 1,1068]	0,0441 [*]
peso	0,2799	1,3229	0,5694	[0,4334; 4,0381]	0,6231 ^{ns}
estações	0,4222	1,5253	0,2451	[0,9434; 2,4660]	0,0850 ^{ns}

Modelo de regressão de Fine e Gray					
covariável	coef	exp(coef)	se (coef)	IC_95%	valor-p
sexo	- 0,3636	0,6951	0,423	[0,303; 1,59]	0,390 ^{ns}
idade	0,0132	1,0133	0,025	[0,965; 1,06]	0,598 ^{ns}
peso	0,1320	1,1411	0,417	[0,504; 2,58]	0,751 ^{ns}
estações	0,1055	1,1113	0,180	[0,781; 1,58]	0,557 ^{ns}

coef. – coeficiente de regressão estimado; se (coef) – erro padrão do coeficiente de regressão; IC_95% - intervalo de confiança com 95% de confiança, ns – não significativo a 5%; * significativo a 5%.

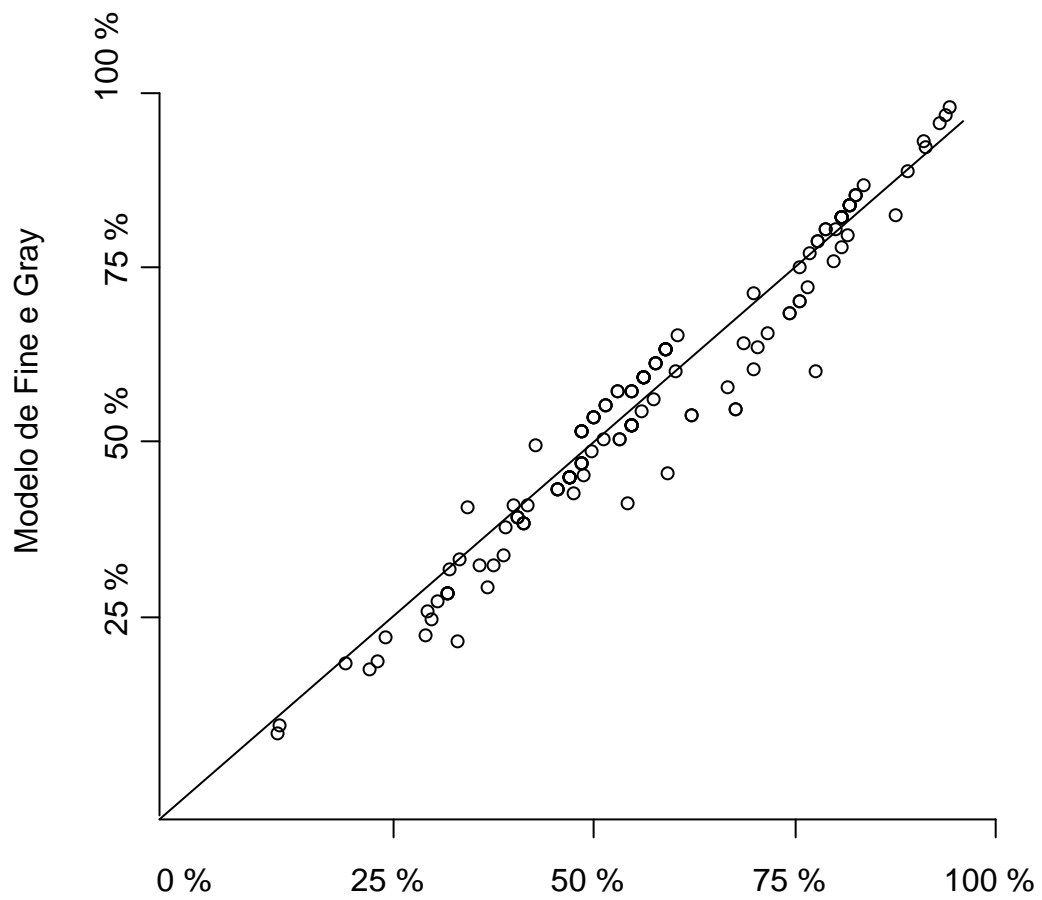
Para avaliar os modelos alternativos (modelo com função de ligação logarítmica e o modelo com função de ligação logit) em relação ao modelo de Fine e Gray utilizou-se análise gráfica. As Figuras 1 e 2 apresentam a comparação da incidência cumulativa estimada pelos modelos alternativos (função de ligação logarítmica e função de ligação logit) em relação a predita pelo modelo de Fine e Gray. Essas figuras não mostram grandes diferenças entre as probabilidades preditas para a claudicação pelos modelos de regressão com função de ligação logarítmica e função de ligação logit em relação ao modelo de Fine e Gray.

Entretanto, para a probabilidade de outras injúrias os modelos considerados apresentam consideráveis diferenças nas probabilidades preditas em relação ao modelo de Fine e Gray (Figuras 3 e 4).



Modelo para o risco absoluto (função de ligação log)

Figura 1 – Comparação entre as probabilidades cumulativas previstas para a claudicação em animais durante 156 semanas, considerando o modelo de regressão para o risco absoluto considerando a função de ligação log (eixo x) e o modelo de Fine e Gray (eixo y).



Modelo para o risco absoluto (função de ligação logit)

Figura 2 – Comparação entre as probabilidades cumulativas previstas para a claudicação em animais durante 156 semanas, considerando o modelo de regressão para o risco absoluto considerando a função de ligação logit (eixo x) e o modelo de Fine e Gray (eixo y).

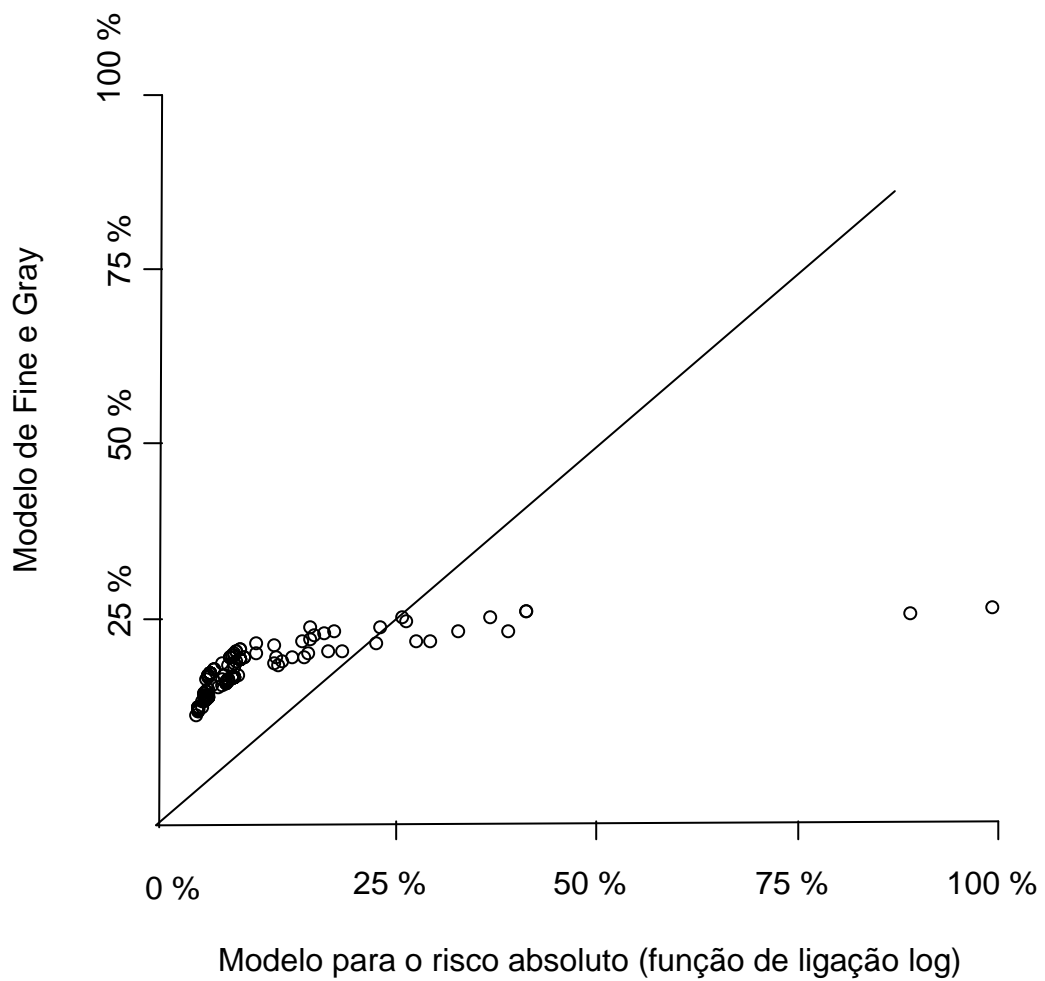


Figura 3 – Comparação entre as probabilidades cumulativas previstas para a outras injúrias em animais durante 156 semanas, considerando o modelo de regressão para o risco absoluto considerando a função de ligação log (eixo x) e o modelo de Fine e Gray (eixo y).

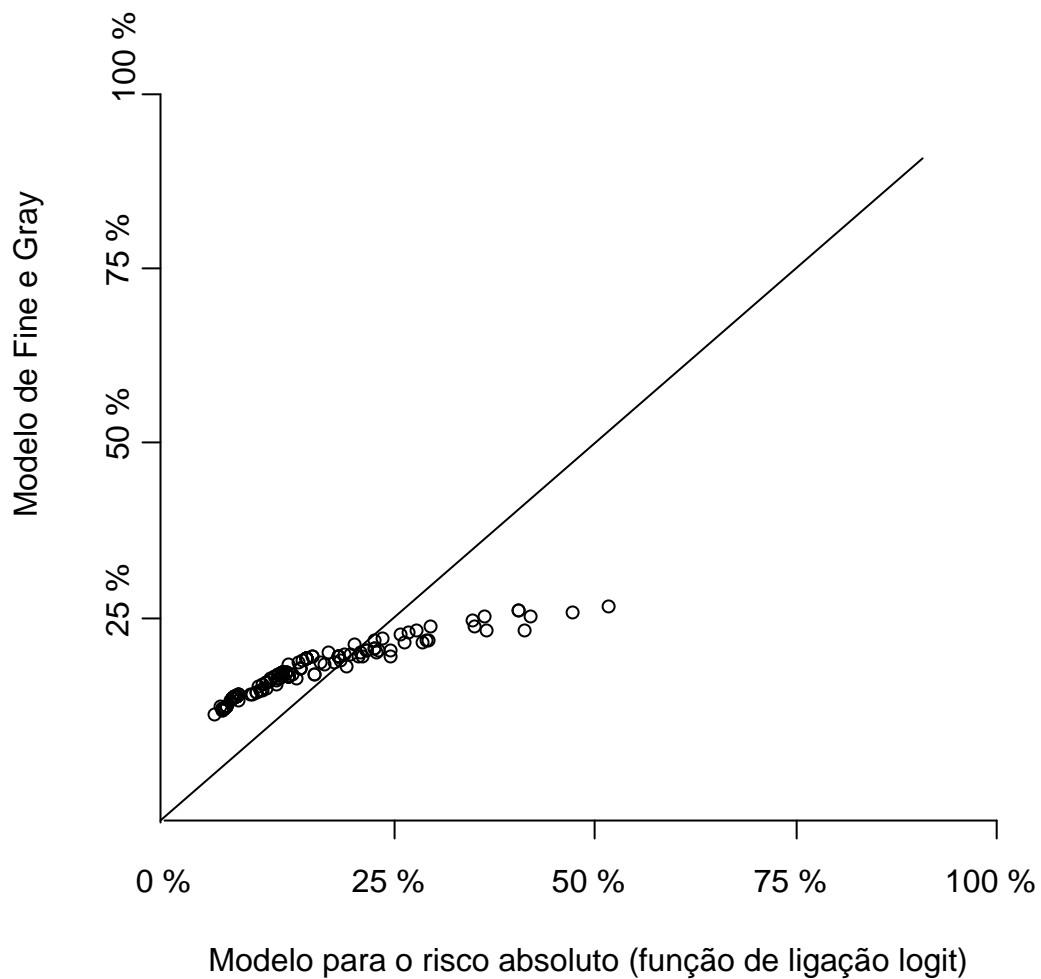


Figura 4 – Comparação entre as probabilidades cumulativas previstas para a outras injúrias em animais durante 156 semanas, considerando o modelo de regressão para o risco absoluto considerando a função de ligação logit (eixo x) e o modelo de Fine e Gray (eixo y).

4. CONCLUSÕES

Com base no estudo de simulação tem-se que o modelo de regressão de Fine e Gray apresentou melhor desempenho do que os modelos de regressão com função de ligação logarítmica (ARR) e o modelo com função de ligação com ligação logit (LRR).

O modelo ARR mostrou-se adequado quando não há grandes diferenças entre as proporções entre os eventos considerados.

O modelo LRR, em geral, apresentou-se inadequado para a modelagem de riscos competitivos.

Considerando os dados sobre lesões em equinos, os modelos considerados neste trabalho, modelo de regressão para a função de incidência cumulativa utilizando as funções de ligação logarítmica e a função de ligação logit, bem como o modelo de Fine e Gray permitiram identificar como fatores significativamente associados à probabilidade de claudicação as covariáveis idade e estações.

Contudo para a incidência de outras injúrias os modelos foram divergentes. O modelo com a função de ligação logarítmica identificou como significativa as covariáveis idade e estações, o modelo utilizando a função de ligação logit identificou somente a covariável idade como significativa. Enquanto que no modelo de Fine e Gray todas as covariáveis estudadas foram não significativas.

5. AGRADECIMENTOS

À FAPEMIG pelo apoio financeiro.

6. REFERÊNCIAS BIBLIOGRÁFICAS

BIMALI, M.; HE, J. Association between obesity and câncer: An analysis using the competing risk regression approach. **Advances in Epidemiology**, v. 2015, p. 1 – 7, 2015.

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. ver.ampl. Rio de Janeiro: Editora Fiocruz, 2011. 432 p.

COX, D. Regression models and life tables. **Journal of the Royal Statistical Society**, v. 34, n.2, p. 187 – 220, 1972.

FINE, J. P.; GRAY, R. J. A proportional hazard model for subdistribution of a competing risk. **Journal of the American Statistical Association**, v. 94, n. 446, p.496 – 509, 1999.

GERDS, T. A.; SCHEIKE, H. T.; ANDERSEN, P. K. Absolute risk regression on for competing risk: interpretation, link functions, and prediction. **Statistics in Medicine**, v. 31, n. 29, p. 3921-3930, 2012.

GESKUS, R. B. **Data analysis with competing risks and intermediate states**. Boca Raton, Fl: Chapman & Hall/CRC, 2015. 243p.

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. 18, n. 6, p. 695 – 706, 1999.

GRAY, R.J. A class of k-sample tests for comparing the cumulative incidence of a competing risks. **Annals of Statistics**, v. 16, n. 3, p. 1141 – 1154, 1988.

GRAW, F.; GERDS, T. A.; SCHUMACHER, M. On pseudo – values for regression analysis in competing risks models. **Lifetime Data Analysis**, v. 15, n. 2, p. 241 - 255, 2009.

HALLER, B. **The analysis of competing risks data with a focus on estimation of cause-specific and subdistribution hazard ratios from a mixture model**, 2014, 199 p. Dissertation – Fakultät für Mathematik, Informatik und Statistik der Ludwig – Maximilians – Universität München.

HALLER, B.; SCHMIDT, ULM. K. Applying competing risks regression models: an overview. **Lifetime Data Analysis**, v. 19, n. 1, p. 33 - 58, 2013.

KLEIN, J. P.; HOUWELINGEN, H. C. V.; IBRAHIM, J. G.; SCHEIKE, T. H. (eds). **Handbook of survival analysis**. New York: Taylor & Francis Group, 2014. 632 p.

LEOCE, N. M. **Prognostic Modeling in the Presence of Competing Risks: An Application to Cardiovascular and Cancer Mortality in Breast Cancer Survivors**. 2016. 184 f. Thesis (doctorate) - Public Health, Mailman School Of Public Health, Columbia University, New York.

PINTILIE, M. **Competing risk: a practical perspective**. Chichester: John Wiley & Sons, 2006. 224 p.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2016. Disponível em:< <http://www.r-project.org/> >.

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

SCHEIKE, T. H.; ZHANG, M. J. ; GERDS, T. A. Predicting cumulative incidence probability by direct binomial regression. **Biometrika**, v. 95, n. 1, p. 205-220, 2008.

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

ZHANG, M. J.; ZHANG, X.; SCHEIKE, T. H. Modeling cumulative incidence function for competing risks data. **Expert review of clinical pharmacology**, v.1, n. 3, p. 391 – 400, 2008.

APÊNDICE A: Código baseado no algoritmo de Haller (2014) para a simulação de dados (considerando o cenário mostrado na primeira linha da Tabela 1).

```

# Required libraries
library(numDeriv) # Calcula as derivadas
library(cmprsk) # Estima o modelo de Fine e Gray
library(riskRegression) #Ajusta diferentes modelos para o risco absoluto.

# Define o risco de base para o evento de interesse (evento1)
gamma1_0 <- function(t) 0.001 * exp(-0.001*t/log(1.3))
lambda1_0 <- function(t) 0.001

# Calcula o risco específico por causa de base para o evento competitivo
lambda2_0 <- function(t) gamma1_0(t) - lambda1_0(t) -
grad(gamma1_0,t)/gamma1_0(t) + grad(lambda1_0,t)/lambda1_0(t)
#Define os coeficientes de regressão
xi <- c(log(1.15),log(0.9),log(2))

#Define os riscos como funções do tempo, os coeficientes de regressão e
valores das covariáveis.
gamma1_x <- function(t,xi,X) gamma1_0(t)*exp(sum(xi*X))
lambda1_x <- function(t,X) lambda1_0(t)*exp(xi[1]*X[1]*exp(-0.0005*t) +
xi[2]*X[2]*exp(-0.0005*t) + xi[3]*X[3]*exp(-0.0005*t))

# Calcula o risco específico por causa para o evento competitivo

lambda2_x <- function(t,xi,X) gamma1_x(t,xi,X) - lambda1_x(t,X)-
grad(gamma1_x,t,xi=xi,X=X)/gamma1_x(t,xi,X) +
grad(lambda1_x,t,X=X)/lambda1_x(t,X)
Runs<- 1000 #número de simulações
n<-100 #número de indivíduos em cada simulação
# Matriz para guardar os resultados
CRR <- matrix(nrow=Runs,ncol=3)
ARR <- matrix(nrow=Runs,ncol=3)
LRR <- matrix(nrow=Runs,ncol=3)
matriz.status<-matrix(nrow=Runs,ncol=n)
# Loop para repetir a simulação
for(RR in 1:Runs)
{
#Amostragem aleatória das covariáveis a partir de distribuições definidas.
# X2 é restrita a valores entre 0 e 10.
X1 <- runif(n,0,3)
X2 <- pmax(pmin(rnorm(n,5,1),10),0)
X3 <- sample(0:1,n,repl=T)
XX <- cbind(X1,X2,X3)
# Vetor para salvar os resultados dentro do loop
Evstat <- rep(0,n)
Evertime <- c()
# Determina o tempo do evento e o tipo do evento para cada indivíduo

```

```

for(j in 1:n)
{
  Ti <- 0
  while(Evstat[j]==0)
  {
    Ti <- Ti+1
    # Probabilidade de um evento pra o indivíduo j no tempo Ti
    Prob <- lambda1_x(t=Ti,X=XX[j,]) +
    lambda2_x(t=Ti,xi=xi,X=XX[j,])
    # Determina se algum evento ocorreu em Ti
    Event <- sample(0:1,1,prob=c(1-Prob,Prob))
    # se um evento ocorreu, determina o tipo do evento
    if(Event==1)
      Evstat[j] <- sample(1:2,1,prob=c(lambda1_x(t=Ti,X=XX[j,]),
      lambda2_x(t=Ti,xi=xi,X=XX[j,])))
  }
  Evertime[j] <- Ti
}
#Inserir censura administrativa
perc.cens=0
cens.adm<-as.vector(quantile(Evertime, 1-perc.cens))
stat<-c(Evertime<=cens.adm)*Evstat
# Transforma os dados gerados em um dataframe
dat <- data.frame(Time=Evertime,Stat=stat,X1=X1,X2=X2,X3=X3)
#Matriz de status (0, 1,2) para cada simulação
matriz.status[RR,]<- dat$Stat

# Ajusta o modelo de Fine e Gray e salva o resultado dos coeficientes

CRR[RR,] <- crr(dat$Time,dat$Stat,
cov1=cbind(dat$X1,dat$X2,dat$X3))$coef
# Calcula a média dos coeficientes do modelo FG
MEAN.FG <- apply(CRR,2,mean,na.rm=T)
#Calcula o desvio padrão dos coeficientes do modelo FG
SD.FG <- apply(CRR,2,sd,na.rm=T)
#Imprime a média em cada simulação

#Ajusta o modelo para o risco absoluto ligação log (ARR)
Time<-dat$Time
Stat<-dat$Stat
ARR[RR,] <-
riskRegression(Hist(Time,Stat)~X1+X2+X3,data=dat,cause=1,link='relative')$ti
meConstantEffects$coef

# Calcula a média dos coeficientes do modelo ARR
MEAN.ARR <- apply(ARR,2,mean,na.rm=T)
#Calcula o desvio padrão dos coeficientes do modelo ARR
SD.ARR <- apply(ARR,2,sd,na.rm=T)

#Ajusta o modelo logit para o risco absoluto (LRR)

```

```

LRR[RR,] <-
riskRegression(Hist(Time,Stat)~X1+X2+X3,data=dat,cause=1,link='logistic')$tim
eConstantEffects$coef #salva os coeficientes

# Calcula a média dos coeficientes do modelo LRR
MEAN.LRR <- apply(LRR,2,mean,na.rm=T)
#Calcula o desvio padrão dos coeficientes do modelo LRR
SD.LRR <- apply(LRR,2,sd,na.rm=T)
}

#Mostra a média final dos coeficientes do modelo FG
print(paste("Média coef. FG: ",round(MEAN.FG[1], 4)," / ",
round(MEAN.FG[2],4)," / ",round(MEAN.FG[3], 4)))

#Mostra o desvio padrão dos coeficientes
print(paste("Desvio padrão coef. FG ",round(SD.FG[1], 4)," / ",
round(SD.FG[2],4)," / ",round(SD.FG[3], 4)))

#Mostra a média final dos coeficientes do modelo ARR
print(paste("Média coef. ARR: ",round(MEAN.ARR[1], 4)," / ",
round(MEAN.ARR[2],4)," / ",round(MEAN.ARR[3], 4)))

#Mostra o desvio padrão dos coeficientes
print(paste("Desvio padrão coef. ARR ",round(SD.ARR[1], 4)," / ",
round(SD.ARR[2],4)," / ",round(SD.ARR[3], 4)))

#Mostra a média final dos coeficientes do modelo LRR
print(paste("Média coef. LRR: ",round(MEAN.LRR[1], 4)," / ",
round(MEAN.LRR[2],4)," / ",round(MEAN.LRR[3], 4)))

#Mostra o desvio padrão dos coeficientes
print(paste("Desvio padrão coef. LRR ",round(SD.LRR[1], 4)," / ",
round(SD.LRR[2],4)," / ",round(SD.LRR[3], 4)))

#Calcula a proporção de ocorrência de cada evento (0, 1, 2) em RR
simulações

d<-RR*n
round(table(matriz.status)/d,2)

```

APÊNDICE B: Código para a modelagem da função de incidência.

```
library(survival)
library(riskRegression)
library(pec) ## para estimar o desempenho preditivo (comparação gráfica)
library(cmprsk) ## para comparar com o modelo de Fine e Gray
head (cavalo)
## distribuição da censura

coxph(Surv(cavalo$tempo,cavalo$status==0)~sexo+idade+peso+estacoes,data
=cavalo)#a distribuição da censura independe das covariáveis

#####
#Usando as covariáveis sexo,idade,peso,estacoes
#####
##PARA CLAUDICAÇÃO
## Regressão para o risco absoluto
arr.claudicao <- ARR(Hist(tempo,status)~sexo+idade+peso
+estacoes,data=cavalo,cause=1) #dist. Censura = KM
summary(arr.claudicao, digits=4)

## Regressão risco logístico
lrr.claudicao <- LRR(Hist(tempo,status)~ sexo+idade+peso
+estacoes,data=cavalo,cause=1)
summary(lrr.claudicao,digits=4)

## Fine-Gray regression model
fg.claudicao <- FGR(Hist(tempo,status)~ sexo+idade+peso
+estacoes,data=cavalo,cause=1)
summary(fg.claudicao, digits=4)

#PARA OUTRAS INJÚRIAS
## Regressão para o risco absoluto –
arr.outras <- ARR(Hist(tempo,status)~sexo+idade+peso
+estacoes,data=cavalo,cause=2) #dist. Censura = KM
summary(arr.outras, digits=4)

## Regressão risco logístico
lrr.outras <- LRR(Hist(tempo,status)~ sexo+idade+peso
+estacoes,data=cavalo,cause=2)
summary(lrr.outras, digits=4)

## Fine-Gray regression model
fg.outras <- FGR(Hist(tempo,status)~ sexo+idade+peso
+estacoes,data=cavalo,cause=2)
summary(fg.outras, digits=4)

#Comando para comparar predições em horizonte de tempo
#Para claudicação
```



```

#Compara ARR (relativo) com Fine e Gray
#horizonte de predição (intervalo de tempo)=156 semanas
plot(predictEventProb(arr.claudicao,times=156,newdata=cavalo),predictEvent
Prob(fg.claudicao,times=156,newdata=cavalo),xlim=c(0,1),ylim=c(0,1),
xlab="Modelo para o risco absoluto (função de ligação log)",ylab="Modelo de
Fine e Gray", axes=FALSE)
prodlim:::PercentAxis(1,at=seq(0,1,.25))
prodlim:::PercentAxis(2,at=seq(0,1,.25))
abline(a=0,b=1)

```

```

#Compara LRR (logístico) com Fine e Gray
#horizonte de predição (intervalo de tempo)=156 semanas
plot(predictEventProb(lrr.claudicao,times=156,newdata=cavalo),predictEvent
Prob(fg.claudicao,times=156,newdata=cavalo),xlim=c(0,1),ylim=c(0,1),
xlab="Modelo para o risco absoluto (função de ligação logit)",ylab="Modelo de
Fine e Gray", axes=FALSE)
prodlim:::PercentAxis(1,at=seq(0,1,.25))
prodlim:::PercentAxis(2,at=seq(0,1,.25))
abline(a=0,b=1)

```

#Para outras injúrias

```

#horizonte de predição (intervalo de tempo)=156 semanas
plot(predictEventProb(arr.outras,times=156,newdata=cavalo),predictEventProb(
fg.outras,times=156,newdata=cavalo), xlim=c(0,1),ylim=c(0,1), xlab="Modelo
para o risco absoluto (função de ligação log)",ylab="Modelo de Fine e Gray",
axes=FALSE)
prodlim:::PercentAxis(1,at=seq(0,1,.25))
prodlim:::PercentAxis(2,at=seq(0,1,.25))
abline(a=0,b=1)

```

```

#Compara LRR (logístico) com Fine e Gray
#horizonte de predição (intervalo de tempo)=156 semanas
plot(predictEventProb(lrr.outras,times=156,newdata=cavalo),predictEventProb(f
g.outras,times=156,newdata=cavalo), xlim=c(0,1),ylim=c(0,1), xlab="Modelo
para o risco absoluto (função de ligação logit)",ylab="Modelo de Fine e Gray",
axes=FALSE)
prodlim:::PercentAxis(1,at=seq(0,1,.25))
prodlim:::PercentAxis(2,at=seq(0,1,.25))
abline(a=0,b=1)

```

CONCLUSÕES GERAIS

Por meio dos resultados desta pesquisa pôde-se verificar que:

- o aumento da porcentagem de censura bem como a proporção dos eventos principal e competitivo afetam a estimação não paramétrica da função de incidência cumulativa;
- a estimativa da incidência baseada no estimador de Kaplan-Meier superestima a função de incidência cumulativa;
- embora, na análise de dados reais realizada, as covariáveis afetem de maneira semelhante o risco específico por causa e o risco da subdistribuição, aumentando-os ou diminuindo-os, a abordagem para análise de riscos competitivos baseada no risco específico por causa e a baseada no risco da subdistribuição não são equivalentes; e,
- ao se comparar três modelos de regressão para a função de incidência cumulativa (modelo de regressão de Fine e Gray, modelo de regressão com função de ligação logarítmica e o modelo de regressão com função de ligação logit) observou-se que o modelo de Fine e Gray apresentou o melhor desempenho.