

UNIVERSIDADE FEDERAL DO PARANÁ

PAULA ALESSANDRA ZEIZER DIMAS

**ANÁLISE DE SOBREVIVÊNCIA NA PRESENÇA DE RISCOS COMPETITIVOS:
UMA APLICAÇÃO EM ENTOMOLOGIA AGRÍCOLA**

CURITIBA

2018

PAULA ALESSANDRA ZEIZER DIMAS

**ANÁLISE DE SOBREVIVÊNCIA NA PRESENÇA DE RISCOS COMPETITIVOS:
UMA APLICAÇÃO EM ENTOMOLOGIA AGRÍCOLA**

Trabalho de Conclusão de Curso apresentado ao Curso de Estatística da Universidade Federal do Paraná como requisito parcial para obtenção do Título de Bacharel em Estatística.

Orientadora: Prof.^a Dr.^a Suely Ruiz Giolo

CURITIBA

2018

AGRADECIMENTOS

Primeiramente, a Deus, por me amar e abençoar tanto.

À minha mãe que me inspira com sua força, determinação e dedicação ao próximo.

Aos melhores avós do mundo.

À professora Dra. Suely Ruiz Giolo por sua paciência, seus ensinamentos e por dedicar seu tempo e conhecimento para a conclusão deste trabalho.

Ao professor Dr. Cesar Augusto Taconeli por sua disponibilidade em ser membro da banca deste trabalho e por me acompanhar desde o Processo Seletivo Estendido.

À Alessandra Benatto e Dra. Maria Cristina Neves de Oliveira (Embrapa-Soja) por terem disponibilizado os dados analisados neste trabalho.

Aos meus amigos, Kenji, Karin, Michele, Henrique, Bruno e Vinicius, que sempre souberam falar as palavras certas nas horas mais difíceis, e por tornarem esta jornada mais divertida.

RESUMO

A análise de sobrevivência é uma área da estatística muito utilizada em diversas áreas do conhecimento. Em medicina, por exemplo, pode ser utilizada em estudos que têm como objetivo analisar o tempo até a ocorrência de um evento de interesse (morte, cura, recidiva etc.), usualmente denominado falha. No contexto de riscos competitivos é possível ter mais de uma causa para a ocorrência do evento. Neste trabalho, foram analisados os dados de um bioensaio na área de entomologia agrícola que teve como objetivo verificar a mortalidade de lagartas *Chrysodeixis includens* alimentadas com uma de três dietas (dieta artificial, folhas de soja ou folhas de algodão) pulverizadas com o vírus *ChinNPV*. A oferta de cada uma das dietas ocorreu após 1 hora, 1 dia, 3 dias ou 8 dias da pulverização do vírus. Morte devido ao vírus e morte por causa desconhecida foram as duas causas de mortalidade consideradas nas análises. As metodologias estatísticas utilizadas foram: o modelo de regressão de Fine-Gray e o modelo de regressão de Cox causa-específica. O primeiro tem como foco a função de incidência acumulada e, o segundo, a função taxa de falha para ambas as causas. Os resultados obtidos a partir dos modelos ajustados aos dados, bem como da análise não paramétrica, foram similares. De modo geral, foi possível observar que o vírus *ChinNPV*, quando pulverizado sobre as plantas de soja e de algodão, produziu mortalidade inferior ao controle positivo (dieta artificial pulverizada com *ChinNPV*). Ademais, comparando as dietas com folhas de algodão e com folhas de soja, a maior redução na mortalidade foi observada quando o vírus *ChinNPV* foi pulverizado sobre as folhas de algodão e ofertado às lagartas 8 dias após a aplicação do vírus. De modo geral, houve uma redução significativa na mortalidade a partir do 3º dia após a pulverização do vírus, indicando que a atividade viral do *ChinNPV* perde sua eficiência com o decorrer do tempo.

Palavras-chave: Modelo de Fine-Gray. Modelo de Cox causa-específica.

SUMÁRIO

1 INTRODUÇÃO	6
2 MATERIAL E MÉTODOS	9
2.1 MATERIAL.....	9
2.1.1 Conjunto de dados.....	9
2.1.2 Recursos Computacionais	12
2.2 MÉTODOS.....	12
2.2.1 Conceitos em Análise de Sobrevida	13
2.2.2 Estimador de Kaplan-Meier.....	13
2.2.3 Modelo de Regressão de Cox.....	14
2.2.4 Método Não-paramétrico no Contexto de Riscos Competitivos.....	16
2.2.5 Modelo de Regressão de Fine-Gray	16
2.2.6 Modelo de regressão de Cox Causa-específica.....	18
3 RESULTADOS E DISCUSSÃO	20
3.1 RESULTADOS DA ANÁLISE NÃO-PARAMÉTRICA	25
3.2 RESULTADOS DO MODELO DE REGRESSÃO DE FINE-GRAY.....	27
3.3 RESULTADOS DO MODELO DE COX CAUSA-ESPECÍFICA.....	31
3.4 COMPARAÇÃO DOS RESULTADOS	34
4 CONCLUSÃO	34
REFERÊNCIAS.....	38
APÊNDICES	40

1 INTRODUÇÃO

A análise de sobrevivência é uma área da estatística responsável pela análise de dados em que a resposta é o tempo até a ocorrência de um evento de interesse, usualmente denominado "falha". Exemplos são: tempo até a morte, tempo até a cura, tempo até a recidiva de uma determinada doença, dentre outros. Uma característica dos dados de sobrevivência é a presença de censuras. Censura é a observação parcial do evento de interesse, o que acontece quando a ocorrência do evento se dá após o término do estudo ou se, por algum motivo, o acompanhamento ou o registro das observações não puderam ser realizados (COLOSIMO; GIOLO, 2006). As censuras podem ser classificadas em: censura à direita (o tempo de ocorrência do evento de interesse está à direita do tempo registrado), censura à esquerda (o tempo de ocorrência do evento de interesse está à esquerda do tempo registrado) e censura intervalar (os registros são realizados em visitas periódicas, tendo-se apenas o conhecimento de que o evento ocorreu em determinado intervalo de tempo). As censuras podem, ainda, ser classificadas em:

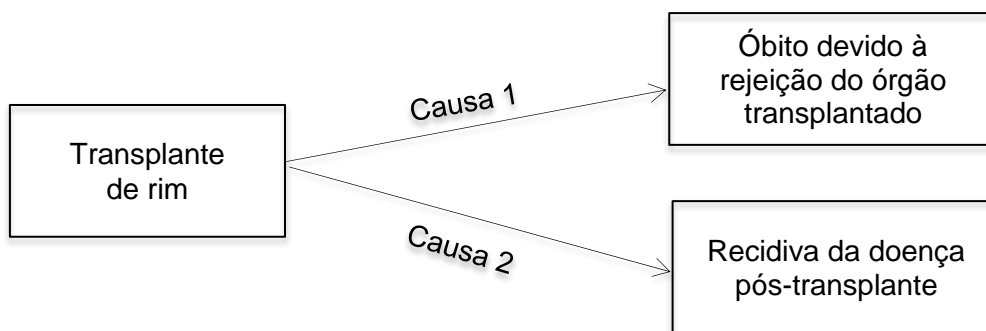
- a) Censuras tipo I: ocorre quando há um tempo predeterminado para a duração do estudo; quando esse tempo termina, os pacientes que não sofreram o evento são considerados censuras.
- b) Censuras tipo II: nesse caso, é estabelecida a proporção de indivíduos que sofrerão a ocorrência do evento de interesse; quando esta quantidade for alcançada o estudo é finalizado, com os indivíduos que não sofreram o evento de interesse sendo classificados como censuras.
- c) Censuras aleatória: ocorre, por exemplo, quando há a necessidade de remover o paciente do estudo por algum motivo ou quando este o abandona.

O tempo até a falha corresponde à variável resposta, geralmente contínua e denotada por T , em que $T > 0$. Há duas funções usuais que caracterizam a variável T , a função de sobrevivência, que fornece a probabilidade de uma observação sobreviver ao tempo t , ou seja, $S(t) = P(T > t)$, e a função taxa de falha, que fornece a taxa instantânea de falha no tempo t , dado que a falha não ocorreu antes deste tempo, sendo definida por

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

Na presença de riscos competitivos, os indivíduos estão sob risco de falha por $k \geq 2$ causas distintas. Por exemplo, para pacientes que se submetem ao transplante de rim, a falha do transplante pode ser caracterizada por um de dois possíveis desfechos (ou causas), o óbito do paciente devido à rejeição do órgão transplantado ou a recidiva da doença pós-transplante, como ilustrado na Figura 1. Nesse caso, há $k = 2$ causas de falha que competem entre si, sendo registrada a que ocorre primeiro. Além disso, é assumido que a ocorrência de uma causa impede a ocorrência da outra.

FIGURA 1 – Representação de duas possíveis causas de falha de transplante de rim



FONTE: O autor (2018).

Em estudos desse tipo, é de interesse analisar o tempo até a falha para cada uma das causas ($\delta = 1, \dots, k$), o que resultará em k curvas de sobrevivência, denotadas por $S_k(t) = P(T > t, \delta = k)$, ou, equivalentemente, em k curvas de incidência acumulada, denotadas por $F_k(t) = P(T \leq t, \delta = k)$.

Quando se tem uma única causa de falha, o estimador de Kaplan-Meier (KAPLAN; MEIER, 1958) é o método não paramétrico mais comum para estimar a probabilidade de sobrevivência, isto é, $S(t) = P(T > t)$. Porém, na presença de riscos competitivos tal estimador não é adequado, pois ignora os riscos que competem entre si. Desse modo, Gray (1988) propôs um método não paramétrico para estimar a função de incidência acumulada, isto é, $F_k(t) = P(T \leq t, \delta = k)$, que fornece a probabilidade de ocorrência do evento para cada causa.

No que se refere aos modelos de regressão no contexto de riscos competitivos, há dois modelos que se destacam na literatura. O primeiro deles, o modelo de regressão de Cox causa-específica, é uma extensão do modelo proposto por Cox (COX, 1972). O segundo modelo, proposto por Fine e Gray (1999), também se baseia no modelo de Cox, explicando a semelhança entre os modelos.

A medicina é uma das principais áreas de aplicação desses modelos. Por exemplo, Ferraz e Moreira Filho (2016) ao estimar a probabilidade de sobrevivência de mulheres com câncer de mama, considerando covariáveis como idade, estadiamento da doença e extensão do tumor, utilizaram os modelos de regressão de Cox causa-específica e o de Fine-Gray. Em outro estudo, Glas et al. (2016) utilizaram os mesmos modelos (de Cox causa-específica e o de Fine-Gray) para a análise de recorrência de câncer de mama em idosas.

Neste trabalho, os métodos citados serão aplicados aos dados de um bioensaio conduzido na área de entomologia agrícola com o objetivo de avaliar a eficácia de um controle biológico de pragas no cultivo de algodão e soja.

2 MATERIAL E MÉTODOS

2.1 MATERIAL

2.1.1 Conjunto de dados

Os dados analisados neste trabalho foram registrados em um bioensaio conduzido na área de entomologia agrícola por Benatto (2018) nas dependências da EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) de Londrina, Paraná, com o objetivo de avaliar a persistência do vírus de poliedrose nuclear da *Chrysodeixis includens* (*ChinNPV*) em folhas de algodão e de soja, mostradas na Figura 2.

FIGURA 2 – Folhas de algodão e de soja utilizadas no bioensaio conduzido com lagartas



FONTE: Marcos Santos/USP Imagens.

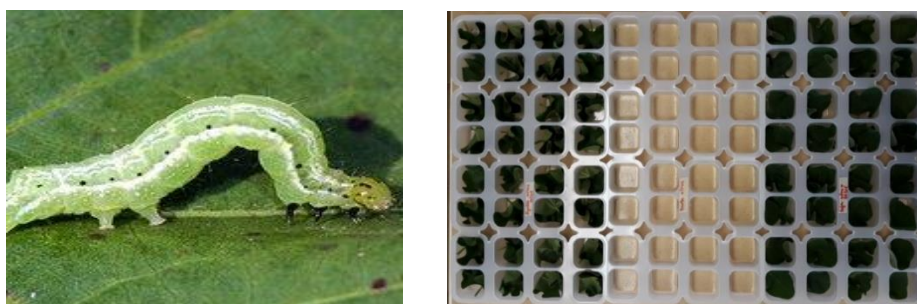
A soja e o algodão foram cultivados em vasos em ambiente controlado, sendo suas folhas utilizadas entre 30 e 40 dias após a semeadura. O vírus entomopatogênico *ChinNPV* foi obtido a partir de lagartas infectadas, o qual foi multiplicado por inoculação (inserção do vírus em lagartas). As 2.435 lagartas sadias de *Chrysodeixis includens* utilizadas no estudo foram produzidas em laboratório. Os 6 tratamentos (dietas) utilizados no estudo foram:

- Tratamento 1 - dieta artificial de Greene, Leppa e Dickerson (1976),
- Tratamento 2 - folhas de algodão,
- Tratamento 3 - folhas de soja,
- Tratamento 4 - dieta artificial pulverizada com o vírus *ChinNPV*,
- Tratamento 5 - folhas de algodão pulverizadas com o vírus *ChinNPV*,
- Tratamento 6 - folhas de soja pulverizadas com o vírus *ChinNPV*.

Os três primeiros tratamentos (T1, T2 e T3), que não receberam a aplicação do vírus *ChinNPV*, são usualmente denominados controles negativos. Nos outros três tratamentos (T4, T5 e T6), a pulverização do vírus *ChinNPV* foi realizada sobre a dieta artificial, bem como sobre as folhas de algodão e de soja. Após 1 hora da pulverização, um 1º grupo de lagartas recebeu a dieta artificial com vírus, um 2º grupo recebeu a dieta de folhas de algodão com vírus, e um 3º grupo recebeu a dieta de folhas de soja com vírus. Após 1 dia da pulverização, um 4º grupo de lagartas recebeu a dieta artificial com vírus, um 5º grupo a dieta de folhas de algodão com vírus e um 6º grupo a dieta de folha de soja com vírus. Procedimento similar foi realizado com outros 3 grupos de lagartas após 3 dias da pulverização e, ainda, com outros 3 grupos após 8 dias da pulverização, totalizando 12 grupos que receberam dietas com vírus, bem como outros 12 grupos que receberam dietas sem o vírus. Foram realizadas, ainda, três repetições desse esquema, resultando em $24 \times 3 = 72$ grupos (2.435 lagartas).

Os grupos, em cada uma das três repetições, foram compostos de 30 a 40 lagartas sadias de 3º instar ¹ (grupos = combinações das dietas e das datas após a pulverização do vírus), sendo tais lagartas dispostas separadamente em bandejas com células individuais. Cada bandeja (mostrada na Figura 3) foi devidamente vedada para que não houvesse contato algum entre as lagartas. Cuidados também foram tomados para que não houvesse contato entre as lagartas que receberam as dietas com vírus e as que receberam as dietas sem vírus e, também, entre as lagartas que receberam dietas distintas.

FIGURA 3 – Lagarta *Chrysodeixis includens* e bandeja utilizada no bioensaio



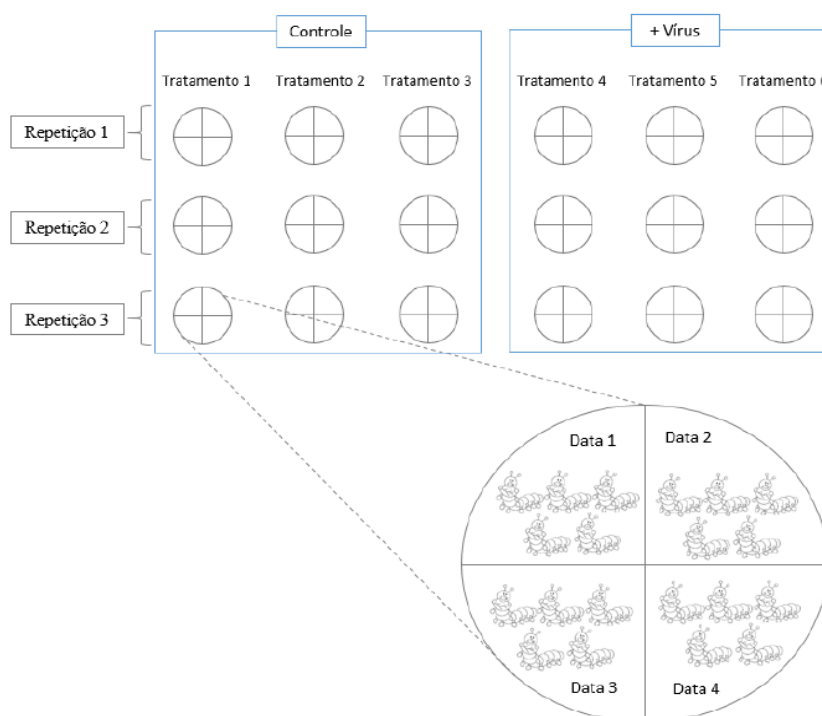
FONTE: Embrapa Soja de Londrina.

¹ Instar: estágio entre as mudas, desde a eclosão da larva até a última fase larval (3º instar: entre 2 e 3 dias de vida).

Cada lagarta foi acompanhada diariamente por 12 dias, sendo que no decorrer dos dias foi possível observar um de três desfechos: morte devido ao vírus, morte por causa desconhecida, ou não ocorrência de morte. As lagartas com o desfecho de morte devido ao vírus apresentam uma característica bem peculiar, a mudança no volume do intestino e da coloração do mesmo, possibilitando a identificação desse tipo de desfecho.

Ao final das três repetições, as informações registradas para cada uma das 2.435 lagartas, divididas em 72 grupos de acordo com as combinações das dietas e datas mencionadas, foram: o tratamento, a data de oferta da dieta ou das folhas, o tempo (em dias) até a ocorrência da morte e a causa da morte. Os tempos (em dias) das lagartas que atingiram o estágio pré-pupa (estágio que antecede o estágio de transformação da lagarta em borboleta) foram considerados nas análises como tempos censurados, uma vez que a não ocorrência de morte caracteriza a presença de censuras. A Figura 4 ilustra o esquema do delineamento do bioensaio realizado e o Quadro 1 apresenta a lista de variáveis disponíveis no conjunto de dados do bioensaio descrito.

FIGURA 4 – Esquema do bioensaio conduzido com lagartas *Chrysodeixis includens*



FONTE: O autor (2018).

QUADRO 1 – Lista de variáveis dos dados do bioensaio na área de entomologia agrícola

Variável	Descrição
Data de oferta da dieta	D1 = oferta da dieta 1 hora após a aplicação do vírus D2 = oferta da dieta 1 dia após a aplicação do vírus D3 = oferta da dieta 3 dias após a aplicação do vírus D4 = oferta da dieta 8 dias após a aplicação do vírus
Tratamento (Dieta)	T1 = Dieta artificial T2 = Folhas de algodão T3 = Folhas de soja T4 = Dieta artificial com vírus <i>ChinNPV</i> T5 = Folhas de algodão com vírus <i>ChinNPV</i> T6 = Folhas de soja com vírus <i>ChinNPV</i>
Sub-amostra	Identificação da lagarta (de 1 a 38)
Repetição	Repetição do experimento (1 a 3)
Tempo	Tempo, em dias, até a morte da lagarta
Causa	0 = Censura 1 = Morte devido ao vírus <i>ChinNPV</i> 2 = Morte por causa desconhecida

FONTE: O autor (2018).

2.1.2 Recursos Computacionais

Para a realização das análises foi utilizado o *software* R, versão 3.2.4 (R CORE TEAM, 2017). O pacote *cmprsk* (GRAY, 2015) foi utilizado para o ajuste do modelo de regressão de Fine-Gray e o pacote *riskRegression* para o ajuste do modelo de regressão de Cox causa-específica (OZENNE et al., 2018).

2.2 MÉTODOS

Nesta seção, são apresentados alguns conceitos na área de análise de sobrevivência, bem como o estimador não-paramétrico de Kaplan-Meier e o modelo de Cox. Na sequência, são apresentadas as metodologias estatísticas consideradas para a análise dos dados descritos. São elas: (a) abordagem não-paramétrica no contexto de riscos competitivos, que serviu como ferramenta exploratória para o ajuste de modelos de regressão, (b) o modelo de regressão de Fine-Gray, e (c) o modelo de regressão de Cox causa-específica.

2.2.1 Conceitos em Análise de Sobrevivência

A análise de sobrevivência compreende um conjunto de métodos que permite analisar dados cuja variável resposta é o tempo até a ocorrência de um evento. Uma característica desse tipo de dados, como mencionado previamente, é a presença de censuras. As principais funções de interesse em análise de sobrevivência são: (a) a função de sobrevivência, $S(t) = P(T > t)$, que fornece a probabilidade de uma observação sobreviver ao tempo t , (b) a função de distribuição acumulada, $F(t) = 1 - S(t)$, que fornece a probabilidade de falha, isto é, de ocorrência do evento até o tempo t , e (c) a função taxa instantânea de falha $\lambda(t)$, dada por

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

em que $\lambda(t)\Delta t$ pode ser vista como a probabilidade de um indivíduo falhar em um intervalo de tempo muito pequeno, dado que ele sobreviveu ao tempo t .

A função (1) apresenta uma relação direta com a função de sobrevivência, uma vez que $\lambda(t) = -\frac{d}{dt}[\ln(S(t))]$. Tem-se, ainda, a função taxa de falha acumulada, dada por $\Lambda(t) = \int_0^t \lambda(u)du = -\ln[S(t)]$, e a função densidade de probabilidade, definida como a probabilidade de falha no intervalo $[t, t + \Delta t)$, com $\Delta t \rightarrow 0$, isto é,

$$f(t) = \lim_{\Delta t \rightarrow 0} P(t \leq T < t + \Delta t),$$

tal que $F(t) = \int_0^t f(u)du$ corresponde à função de distribuição acumulada.

No contexto não paramétrico, estimação da função de sobrevivência é geralmente realizada pelo estimador de Kaplan-Meier ou, ainda, pelo estimador de Nelson-Aalen (COLOSIMO; GIOLO, 2006).

2.2.2 Estimador de Kaplan-Meier

O estimador não-paramétrico de Kaplan-Meier (KAPLAN; MEIER, 1958) fornece estimativas da probabilidade de sobrevivência, as quais podem ser utilizadas para representações gráficas das curvas de sobrevivência. Duas propriedades desse estimador são: consistência e normalidade assintótica.

Supondo que t_1, t_2, \dots, t_r são os tempos distintos de falha, d_j o número de eventos que ocorre em cada tempo t_j , e n_j o número de indivíduos sob risco no tempo imediatamente anterior a t_j , o estimador de Kaplan-Meier é dado por

$$\hat{S}(t) = \prod_{t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right),$$

sendo sua variância estimada pela fórmula de Greenwood, dada por

$$\widehat{Var}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}.$$

Como $\hat{S}(t)$, para t fixo, segue distribuição assintótica Normal, um intervalo aproximado de $100(1 - \alpha)\%$ de confiança para $S(t)$ pode ser obtido por

$$[\hat{S}(t)] \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{S}(t))},$$

em que $\alpha/2$ denota o $\alpha/2$ -quantil da distribuição Normal padrão. Como para valores extremos de t , tal intervalo pode produzir limite inferior negativo ou limite superior maior do que 1, Kalbfleisch e Prentice (2002) sugeriram a obtenção de um intervalo de confiança para $S(t)$ o qual tem como base a transformação $\hat{U}(t) = \log[-\log(\hat{S}(t))]$, de modo que o intervalo fica dado por

$$[\hat{S}(t)]^{\exp\{\pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{U}(t))}\}},$$

sendo a variância assintótica de $\hat{U}(t)$ dada por

$$\widehat{Var}(\hat{U}(t)) = \frac{\sum_{t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}}{[\log(\hat{S}(t))]^2}.$$

2.2.3 Modelo de Regressão de Cox

O modelo de regressão de Cox (COX, 1972) é um dos mais utilizados na literatura para a análise de dados de sobrevivência. A expressão geral do modelo é dada pela taxa de falha da seguinte forma

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

em que $\mathbf{x} = (x_1, x_2, \dots, x_p)$ é o vetor de valores observados para as p covariáveis, $\boldsymbol{\beta}$ é o vetor de parâmetros associados às covariáveis e $\lambda_0(t)$ é a função taxa de falha de base.

O modelo de Cox é considerado multiplicativo semi-paramétrico por resultar da multiplicação de dois componentes: um paramétrico, denotado por $g(\mathbf{x}'\boldsymbol{\beta}) = \exp(\mathbf{x}'\boldsymbol{\beta})$, e outro não-paramétrico, $\lambda_0(t)$, o que o torna um modelo bastante flexível. Também é denominado modelo de taxas de falha proporcionais devido à razão de taxas de falha de dois indivíduos quaisquer ser assumida constante para qualquer tempo t . Ou seja, considerando os indivíduos i e j , tem-se que a razão

$$\frac{\lambda(t | \mathbf{x}_i)}{\lambda(t | \mathbf{x}_j)} = \exp(\mathbf{x}'_i \boldsymbol{\beta} - \mathbf{x}'_j \boldsymbol{\beta}),$$

não depende do tempo.

Para a estimação do vetor $\boldsymbol{\beta}$ utiliza-se o método da máxima verossimi-lhança, baseado na maximização da função de verossimilhança parcial dada por

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left(\frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{\sum_{j \in R(t_i)} \exp\{\mathbf{x}'_j \boldsymbol{\beta}\}} \right)^{\delta_i},$$

em que δ_i é o indicador de falha (1 se falha e 0 se censura) e $R(t_i)$ é o conjunto de indivíduos sob risco em t_i .

Para estimar o componente $\Lambda_0(t)$ e, em consequência $\lambda_0(t)$, é utilizado o estimador de Nelson-Aalen-Breslow dado por

$$\hat{\Lambda}_0(t) = \sum_{t_j < t} \frac{d_j}{\sum_{j \in R(t_j)} \exp(\mathbf{x}'_j \hat{\boldsymbol{\beta}})},$$

em que d_j é o número de eventos que ocorreram até o tempo t_j e $R(t_j)$ o conjunto de indivíduos sob risco no tempo t_j .

O modelo de Cox serviu de base para modelos no contexto de riscos competitivos, dentre eles, o modelo de regressão de Fine-Gray e o modelo de Cox causa-específica, descritos em mais detalhes nas Seções 2.2.5 e 2.2.6.

2.2.4 Método Não-paramétrico no Contexto de Riscos Competitivos

Em análise de sobrevivência, o estimador de Kaplan-Meier é muito utilizado quando há apenas uma causa para o evento de interesse. Contudo, no contexto de riscos competitivos, as estimativas obtidas por meio desse estimador podem ser viesadas, pois ao obtê-las para cada causa k as ocorrências devido às demais causas são consideradas como censuras (PINTILIE, 2006).

Pensando nisso, foi proposto um estimador empírico da função de incidência acumulada associada à cada causa k (isto é, da probabilidade de um indivíduo sofrer o evento devido à causa k até o tempo t). Esse estimador é dado da seguinte forma

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

em que $\hat{S}(t_{j-1})$ é o estimador de Kaplan-Meier considerando todas as falhas como sendo da mesma causa, d_{kj} é o número de eventos ocorridos até o tempo t_j devido à causa k e n_j é o número de indivíduos sob risco em t_j .

Para testar se duas ou mais curvas obtidas por meio do estimador $\hat{F}_k(t)$ diferem, Gray (1988) propôs um teste, conhecido como teste de Gray, que permite comparar as funções de incidência causa-específica entre s categorias ($s \geq 2$) de uma dada covariável (PINTILIE, 2006).

2.2.5 Modelo de Regressão de Fine-Gray

Fine e Gray (1999) propuseram um modelo baseado na função de incidência acumulada para cada uma das causas de falha separadamente (comumente chamada de subdistribuição), que estima a função $F_k(t | \mathbf{x})$ para cada causa, isto é,

$$F_k(t | \mathbf{x}) = P(T \leq t, \delta = k | \mathbf{x}),$$

fornecendo a probabilidade de falha pela causa k condicional às covariáveis.

Sob este modelo, estima-se apenas a subdistribuição para a principal causa de interesse, na presença das demais causas. Portanto, considerando que a principal causa de interesse seja a causa 1, tem-se que a taxa de falha para essa causa fica definida por

$$\lambda_1(t | \mathbf{x}) = -\frac{d}{dt} \ln[1 - F_1(t | \mathbf{x})],$$

e a função de incidência acumulada por

$$F_1(t | \mathbf{x}) = 1 - \exp\{-\Lambda_{01}(t) \exp(\mathbf{x}'\boldsymbol{\beta}_1)\},$$

em que $\Lambda_{01}(t)$ é a função taxa de falha de base, $\boldsymbol{\beta}_1$ é o vetor de coeficientes associado à causa 1, e \mathbf{x} é o vetor de valores observados das covariáveis.

A estimação de $\Lambda_{01}(t)$ pode ser feita pelo estimador de Nelson-Aalen Breslow e a do vetor de parâmetros $\boldsymbol{\beta}_1$ pela maximização da seguinte função pseudo-verossimilhança:

$$L(\boldsymbol{\beta}_1) = \prod_{i=1}^n \left(\frac{\exp(\boldsymbol{\beta}_1' \mathbf{x}_i)}{\sum_{j \in R_i} w_{ij} \exp(\boldsymbol{\beta}_1' \mathbf{x}_j)} \right)^{\delta_i = 1},$$

em que R_i é o conjunto de indivíduos sob risco em t_i , que contém os indivíduos que ainda estão sob observação e aqueles que experimentaram a falha por uma causa competindo antes de t_i . Os pesos w_{ij} são calculados da seguinte forma: os indivíduos que não experimentaram nenhum evento de interesse antes do tempo t_i recebem o peso $w_{ij} = 1$. Já os indivíduos que experimentaram eventos competitivos antes de t_i recebem um peso w_{ij} que diminui com o tempo. Os pesos são calculados da seguinte forma:

$$w_{ij} = \frac{\hat{G}(t_i)}{\hat{G}(\min(t_i, t_j))},$$

em que $\hat{G}(t)$ é a estimativa de Kaplan-Meier da função de sobrevivência da distribuição de censura, que é a probabilidade cumulativa de um indivíduo sob risco no tempo t (GUO; SO, 2018).

Quanto à suposição de taxas de falha proporcionais, esta pode ser avaliada, tal qual no modelo de Cox, pelos resíduos de Schoenfeld (COLOSIMO; GIOLO, 2006). Para que a suposição esteja atendida, o gráfico desses resíduos versus o tempo, para cada um dos parâmetros do modelo, deve apresentar inclinação próxima de zero.

2.2.6 Modelo de regressão de Cox Causa-específica

No contexto de riscos competitivos, tem-se duas ou mais causas de falha que competem entre si (causas concorrentes), registrando-se o tempo até a ocorrência da causa que ocorrer primeiro. O modelo de Cox no contexto de riscos competitivos, denominado modelo de Cox causa-específica, considera que a função taxa de falha para cada causa ($\delta = 1, \dots, k$) é dada por

$$\lambda_k(t | \mathbf{x}) = \lambda_{0k}(t) \exp(\mathbf{x}' \boldsymbol{\beta}_k), \quad (2)$$

em que $\lambda_{0k}(t)$ é a taxa de falha de base associada à causa k , $\boldsymbol{\beta}_k$ é o vetor de parâmetros desconhecidos e \mathbf{x} é o vetor de valores observados das covariáveis (DIGNAM et al., 2012). Assim, para $k = 1$, $\lambda_1(t)$ fornece a taxa instantânea de falha devido à causa 1, condicionado à sobrevivência até o tempo anterior a t .

Apesar de o modelo de Fine-Gray e o modelo de Cox causa-específica serem baseados no modelo de regressão de Cox, sendo parecidos em suas estruturas e com ambos assumindo a suposição de taxa de falhas proporcionais, o modelo de Cox causa-específica, quando comparado ao de Fine-Gray, não fornece a mesma informação. O procedimento de estimação associado ao modelo de Cox causa-específica fornece estimativas para cada uma das k causas simultaneamente, enquanto o do modelo de Fine-Gray apenas para a causa de principal interesse (LATOUCHE et al., 2013).

Nota-se que o modelo em (2) assume que os parâmetros são diferentes entre as causas e que as covariáveis são consideradas invariantes no tempo. Sob esse modelo, a taxa de falha acumulada, para cada causa k , é dada da seguinte forma

$$\Lambda_k(t | \mathbf{x}) = \Lambda_{0k}(t) \exp(\mathbf{x}' \boldsymbol{\beta}_k),$$

em que $\Lambda_{0k}(t)$ corresponde à função taxa de falha acumulada de base causa-específica, que pode ser estimada por meio do estimador de Nelson-Aalen Breslow (HANLEY, 2008), e $\boldsymbol{\beta}_k$ corresponde ao vetor de parâmetros associado à causa k , estimado pelo método da máxima verossimilhança parcial modificado (GUO; SO, 2018), em que a função de verossimilhança é dada por

$$L(\boldsymbol{\beta}) = \prod_i \left(\frac{\exp(\boldsymbol{\beta}_k' \mathbf{x}_i)}{\sum_{j \in R_i} \exp(\boldsymbol{\beta}_k' \mathbf{x}_j)} \right)^{\delta_i=k}$$

em que R_i é o conjunto de indivíduos sob risco no tempo t_i que contém os indivíduos que não falharam por qualquer uma das causas ou que não foram censurados antes de t_i . A causa de falha é indicada por δ_i .

Sob o modelo de Cox causa-específica, a função de incidência acumulada para cada uma das k causas também pode ser estimada por

$$\hat{F}_k(t | \mathbf{x}) = \int_0^t \hat{S}(s^- | \mathbf{x}) d\hat{\Lambda}_k(s | \mathbf{x}),$$

tal que $\sum_i^K \hat{F}_k(t | \mathbf{x}) = 1 - \hat{S}(t | \mathbf{x})$ e em que s^- denota o instante imediatamente anterior a s .

No que se refere à suposição de taxas de falha proporcionais, ela também pode ser avaliada por meio dos resíduos de Schoenfeld (COLOSIMO; GIOLO, 2006).

Os resultados da análise dos dados descritos na Seção 2.1, obtidos a partir da análise exploratória, do estimador não-paramétrico apresentado na Seção 2.2.4, bem como do ajuste dos dois modelos descritos nas Seções 2.2.5 e 2.2.6, são discutidos a seguir. O nível de significância considerado em todos os testes e procedimentos de seleção de covariáveis foi de 5%.

3 RESULTADOS E DISCUSSÃO

Para a análise inicial dos dados foram consideradas as 2 causas de falha competindo entre si, com a causa 1 correspondendo à morte devido à aplicação do vírus e a causa 2 à morte por causa desconhecida. A data após a aplicação do vírus e o tratamento (dieta) ofertado à lagarta foram as covariáveis utilizadas nas análises. Apenas os tratamentos pulverizados com o vírus *ChinNPV* (tratamentos T4, T5 e T6) foram considerados, tendo em vista que nos tratamentos T1, T2 e T3 não foram registradas mortes devido à causa 1.

Ao todo, 2.435 lagartas foram utilizadas no bioensaio, porém, ao considerar apenas os tratamentos T4, T5 e T6, o total de lagartas recebendo tais tratamentos foi de 1.217. Portanto, foram considerados nas análises 36 grupos (12 grupos em cada uma das 3 repetições), totalizando 1.217 lagartas, sendo que 375 sobreviveram, 783 morreram devido à causa 1 (ingestão do vírus) e apenas 59 morreram por causa desconhecida. O número de lagartas para as 12 combinações de tratamentos e datas (somando as frequências das 3 repetições), se deu conforme apresentado na Tabela 1. Os tratamentos foram identificados como T4, T5 e T6 e as datas como D1, D2, D3 e D4, conforme descrito no Quadro 1.

TABELA 1 – Frequências de lagartas de acordo com cada combinação de tratamento e data após a pulverização do vírus, considerando a soma das três repetições

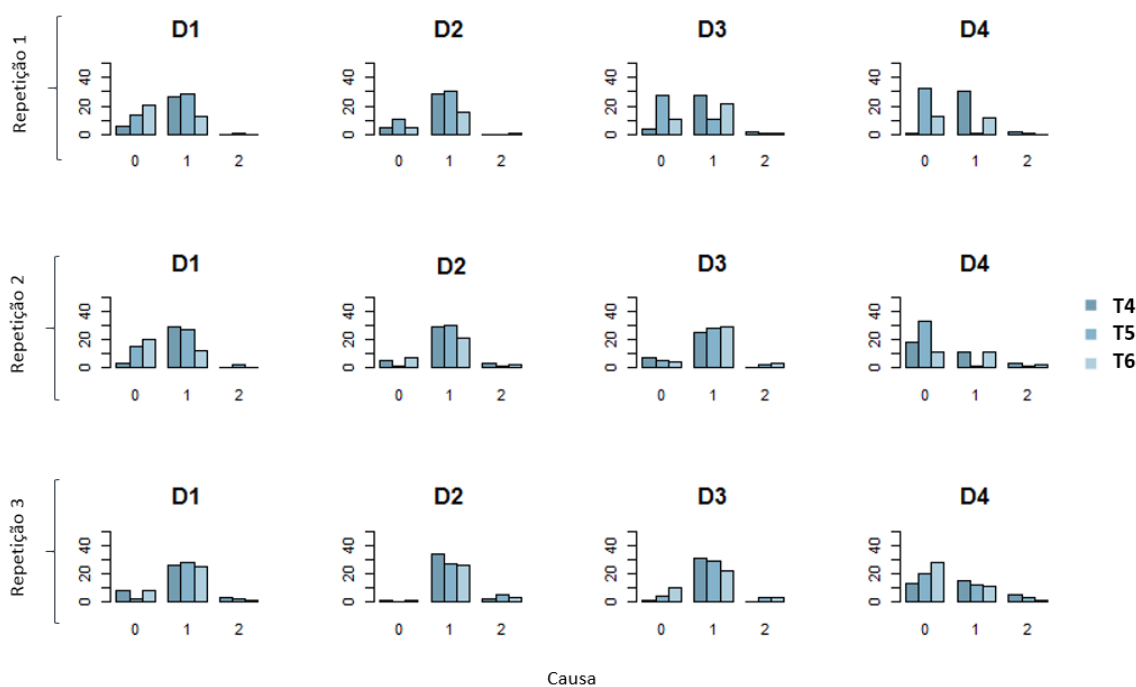
Tratamentos	Censura				Morte pelo vírus				Morte por causa desconhecida			
	D1	D2	D3	D4	D1	D2	D3	D4	D1	D2	D3	D4
T4	17	11	12	32	81	91	83	56	3	5	2	10
T5	31	12	36	85	83	87	68	14	5	6	6	5
T6	49	13	25	52	50	63	73	34	1	6	7	3

FONTE: O autor (2018).

Para verificar se há consistência entre as três repetições realizadas no bioensaio, foi realizada uma breve análise descritiva dos dados. A Figura 5, que mostra as frequências das observações de acordo com as covariáveis tratamento e data, nos permite avaliar se há o mesmo padrão ao comparar os gráficos entre as 3 repetições, ou seja, se as colunas de frequências, quando comparadas para cada data e tratamento, não mudam de comportamento entre as repetições. Ao comparar os gráficos da data D1 entre as repetições, nota-se que a distribuição dos dados é

diferente na repetição três, tanto para censura (causa = 0) quanto para a causa 1 (morte pelo vírus). Para a data D3, a repetição 1 difere das demais. Na data D4, a repetição 3 é a que difere. Também é possível perceber que a causa 2 (morte por causa desconhecida) ocorre com menos frequência quando comparada com a causa 1 e com a censura (causa = 0).

FIGURA 5 – Frequência de lagartas de acordo com as datas e tratamentos em cada repetição



FONTE: O autor (2018).

Por outro lado, a Figura 6, que mostra os gráficos do tempo até a ocorrência do evento de acordo com as duas causas para cada observação (lagarta), nos permite verificar se há diferenças significativas de ocorrência das duas causas entre as repetições. Por exemplo, ao comparar os gráficos da data D2 e tratamento T5 para as 3 repetições, nota-se que o comportamento de mortes devido às duas causas difere na repetição 1, quando comparado com as outras duas repetições; na repetição 1 não há mortes pela causa 2, além disso, há mais ocorrência de censuras na primeira repetição.

Além da Figura 6, foram analisadas as curvas de mortalidade acumulada, estimadas via análise não-paramétrica, para complementar as suspeitas de diferenças entre as repetições (Figura 7). Ao se observar a combinação data D2 e tratamento T5 na Figura 7, nota-se que há uma diferença significativa entre as curvas de mortalidade,

quando são comparadas as repetições 1 com a 2, e a 1 com a 3. Além disso, utilizando o teste não-paramétrico de Gray, a conclusão foi a mesma, resultando na rejeição da hipótese de igualdade entre as duas curvas (Tabela 2). Para a combinação data D2 e tratamento T5, os resultados na Tabela 2 mostra um valor p igual a 0,0007 ao testar as curvas associadas às repetições 1 e 2 para a causa 1 e, também, um valor p igual a 0,0139 ao comparar as repetições 1 e 3 para esta mesma causa.

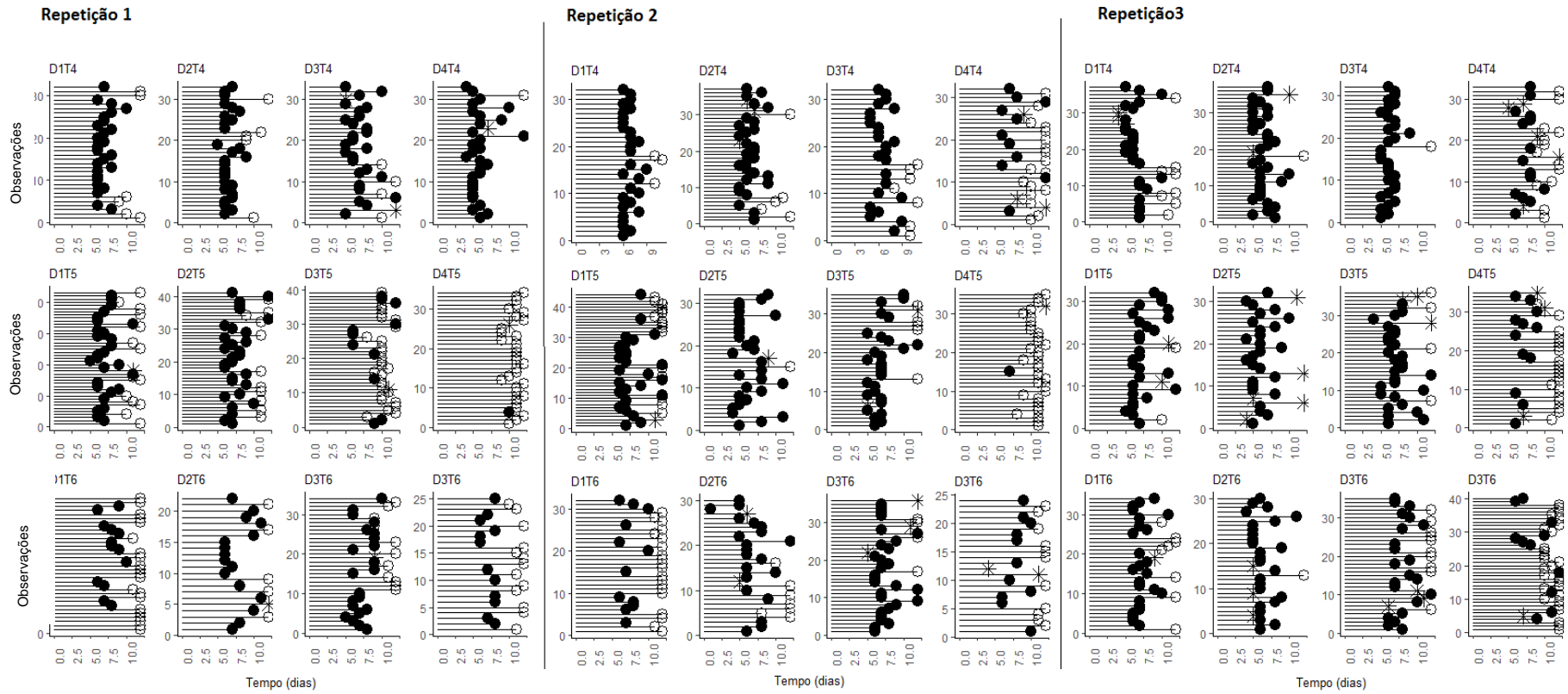
TABELA 2 – Teste de Gray utilizado para comparação das curvas de mortalidade entre as três repetições considerando as combinações entre datas e tratamentos

Data	Comparação entre as repetições	Causa de morte	Tratamento 4		Tratamento 5		Tratamento 6	
			Estatística	Valor p	Estatística	Valor p	Estatística	Valor p
D1	1 e 2	1	1.29430	0.25520	0.70270	0.40190	0.00190	0.96570
		2	--	--	0.12280	0.72610	--	--
	1 e 3	1	0.14800	0.70040	3.10880	0.07790	10.31590	0.00130
		2	2.67250	0.10210	0.66150	0.41600	0.98510	0.32100
	2 e 3	1	1.68540	0.19420	7.71000	0.00550	8.99430	0.00270
		2	2.67250	0.10210	0.18190	0.66980	0.92630	0.33580
D2	1 e 2	1	0.22200	0.63750	11.36470	0.00070	0.29020	0.59010
		2	2.76800	0.09620	1.22750	0.26790	0.17230	0.67810
	1 e 3	1	0.96170	0.32680	6.05440	0.01390	4.57290	0.03250
		2	1.35670	0.24410	4.61610	0.03170	0.49330	0.48250
	2 e 3	1	2.39040	0.12210	0.42800	0.51300	2.81670	0.09330
		2	0.24690	0.61930	2.69510	0.10070	0.22480	0.63540
D3	1 e 2	1	0.07651	0.78209	21.58510	0.00000	1.32988	0.24883
		2	0.96970	0.32476	0.00304	0.95603	0.60190	0.43786
	1 e 3	1	7.11434	0.00765	20.86569	0.00000	0.41385	0.52002
		2	2.41561	0.12013	0.33609	0.56210	0.80546	0.36947
	2 e 3	1	6.47409	0.01095	0.03909	0.84327	3.97498	0.04618
		2			0.17674	0.67419	0.00228	0.96196
D4	1 e 2	1	31.82537	0.00000	0.00011	0.99182	0.19449	0.65920
		2	0.32747	0.56715	0.00215	0.96299	2.02474	0.15476
	1 e 3	1	25.71843	0.00000	11.14242	0.00084	3.22033	0.07273
		2	1.76090	0.18451	0.94878	0.33003	0.62396	0.42958
	2 e 3	1	0.80322	0.37013	11.23059	0.00080	2.16763	0.14094
		2	0.71587	0.39750	0.51902	0.47126	0.91029	0.34004

FONTE: O autor (2018).

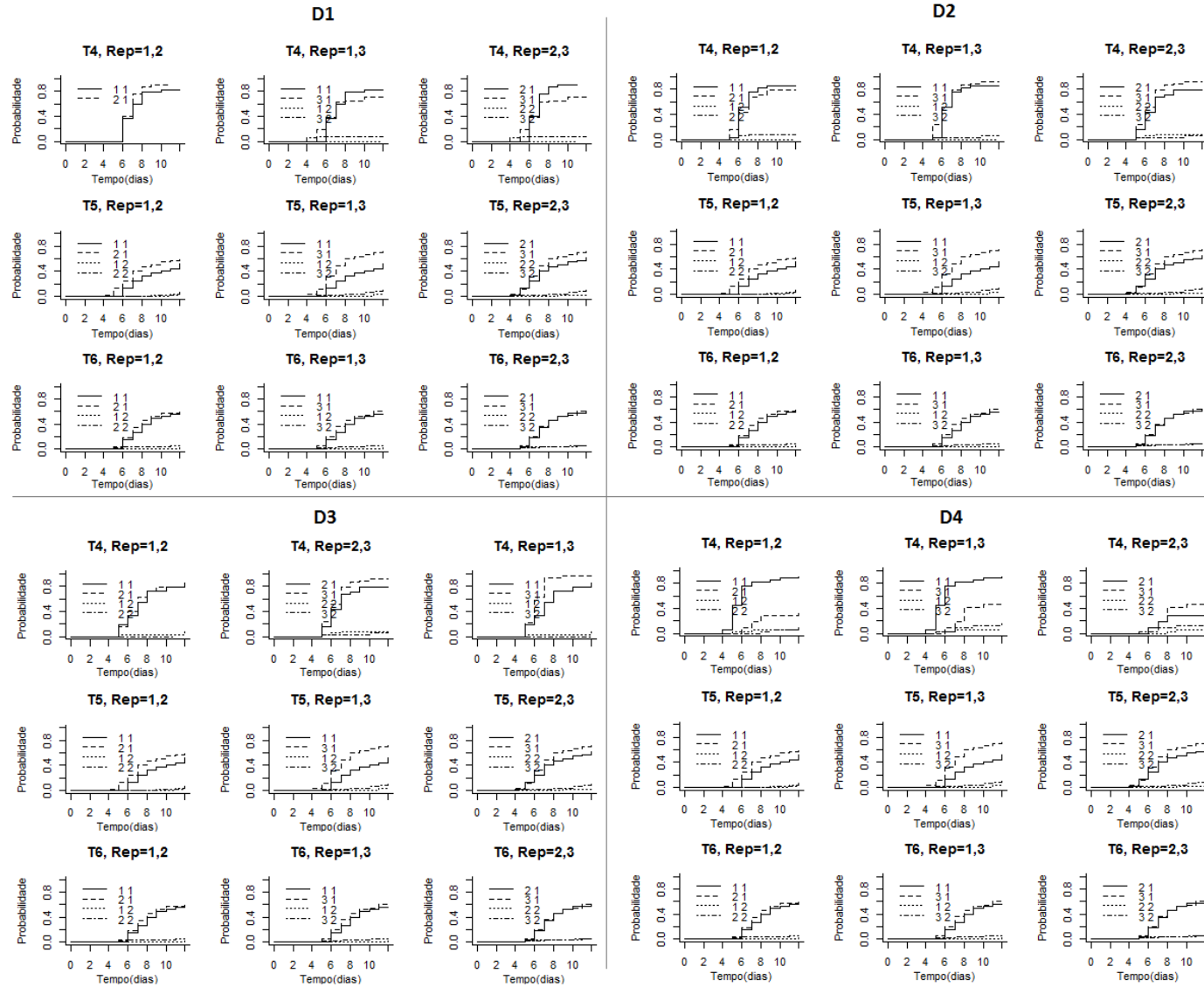
Como alguns grupos de lagartas mostraram inconsistências entre as repetições (valor p significativo no teste de Gray, mostrados na Tabela 2, eles foram excluídos das análises. Os grupos excluídos foram: T4D3 – repetição 3, T4D4 – repetição 1, T5D1 – repetição 3, T5D2 – repetição 1, T5D3 – repetição 1, T5D4 – repetição 3, T6D1 – repetição 3, e T6D2 – repetição 3, totalizando 276 lagartas excluídas.

FIGURA 6 – Tempo até a ocorrência do evento para cada lagarta em seu respectivo grupo de repetição, tratamento e data de aplicação do vírus



FONTE: O autor (2018).

FIGURA 7 – Curvas de mortalidade acumulada para as causas 1 e 2 comparando as repetições por data e tratamento



FONTE: O autor (2018).

3.1 RESULTADOS DA ANÁLISE NÃO-PARAMÉTRICA

Após remoção das 276 lagartas, prosseguiu-se a análise com 941 lagartas, que somando as 3 repetições, resultou nas frequências apresentadas na Tabela 3.

TABELA 3 – Frequências de lagartas de acordo com cada combinação de tratamento e data após a pulverização do vírus, considerando a soma das três repetições

Tratamentos	Censura				Morte pelo vírus				Morte por causa desconhecida			
	D1	D2	D3	D4	D1	D2	D3	D4	D1	D2	D3	D4
T4	17	11	11	31	81	91	52	26	3	5	2	8
T5	29	1	9	65	55	57	57	2	3	6	5	2
T6	41	12	25	52	25	37	73	34	0	3	7	3

FONTE: O autor (2018).

Foram realizados testes de comparação das curvas de mortalidade acumuladas entre as datas fixando o tratamento, mostrados na Tabela 4, bem como entre os tratamentos fixando a data, mostrados na Tabela 5. Para a causa 1, nota-se, a partir da Tabela 4, que a mortalidade para o tratamento T4 diferiu entre as datas D1 e D4 e, também, entre as datas D2 e D4, e D3 e D4. Já para o tratamento T5, a mortalidade diferiu entre todas as datas e, para o tratamento 6, apenas não diferiu entre as datas D1 e D4, e D2 e D3.

TABELA 4 – Teste de Gray para comparação entre datas fixando o tratamento

Tratamento	Data	Causa 1		Causa 2	
		Estatística	Valor p	Estatística	Valor p
T4	D1,D2	2.5444	0.1107	0.4220	0.5159
	D1,D3	2.0322	0.1540	0.0965	0.7560
	D1,D4	9.9248	0.0016	3.9767	0.0461
	D2,D3	0.0092	0.9236	1.0183	0.3129
	D2,D4	19.8946	0.0000	1.9438	0.1632
	D3,D4	19.3484	0.0000	4.6079	0.0318
T5	D1,D2	27.3152	0.0000	1.7761	0.1826
	D1,D3	6.5591	0.0104	0.7018	0.4022
	D1,D4	48.5109	0.0000	0.6994	0.4030
	D2,D3	6.6668	0.0098	0.2532	0.6148
	D2,D4	114.9311	0.0000	0.3762	0.5397
	D3,D4	19.3484	0.0000	4.6079	0.0318
T6	D1,D2	15.0221	0.0001	3.8408	0.0500
	D1,D3	8.7548	0.0031	4.0614	0.0439
	D1,D4	1.66e-5	0.9968	2.4063	0.1208
	D2,D3	1.5462	0.2137	9.6e-5	0.9922
	D2,D4	18.1154	0.0000	0.3272	0.5673
	D3,D4	10.8463	0.0010	0.4837	0.4867

FONTE: O autor (2018).

Para a causa 1, nota-se, a partir da Tabela 5, que a mortalidade para as datas D1 e D4 diferiu entre os 3 tratamentos. Já para a datas D2 e D3, diferiu somente entre os tratamentos T4 e T6, e T5 e T6.

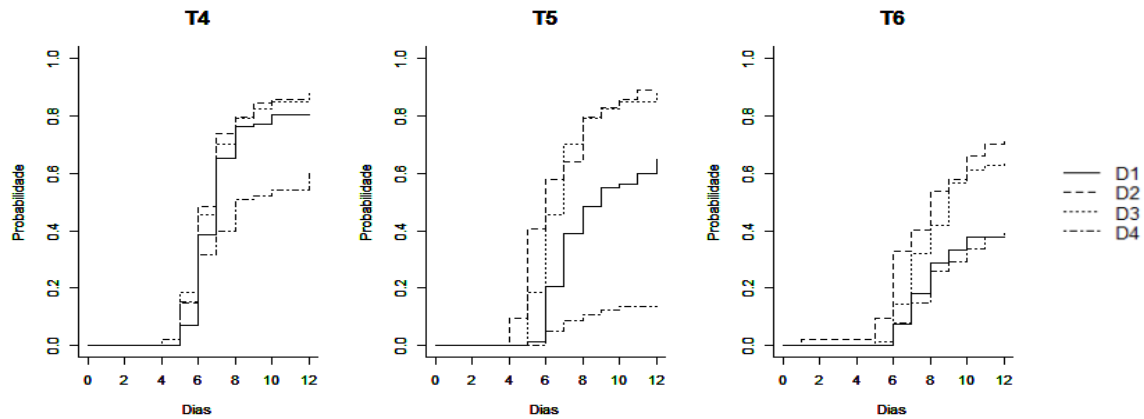
TABELA 5 – Teste de Gray para comparação entre tratamentos fixando a data

Data	Tratamento	Causa 1		Causa 2	
		Estatística	Valor <i>p</i>	Estatística	Valor <i>p</i>
D1	T4,T5	12.3669	0.0004	0.0107	0.9176
	T4,T6	36.2082	0.0000	1.9842	0.1590
	T5,T6	10.4049	0.0013	2.5713	0.1088
D2	T4,T5	1.5132	0.2187	0.6146	0.4331
	T4,T6	7.9489	0.0048	0.0115	0.9147
	T5,T6	10.5466	0.0012	0.4030	0.5255
D3	T4,T5	4.2091	0.0402	0.9029	0.3420
	T4,T6	21.6404	0.0000	0.8334	0.3613
	T5,T6	7.2208	0.0072	0.0303	0.8618
D4	T4,T5	43.4188	0.0000	1.7283	0.1886
	T4,T6	10.7893	0.0010	3.5569	0.0593
	T5,T6	14.0489	0.0002	0.3920	0.5312

FONTE: O autor (2018).

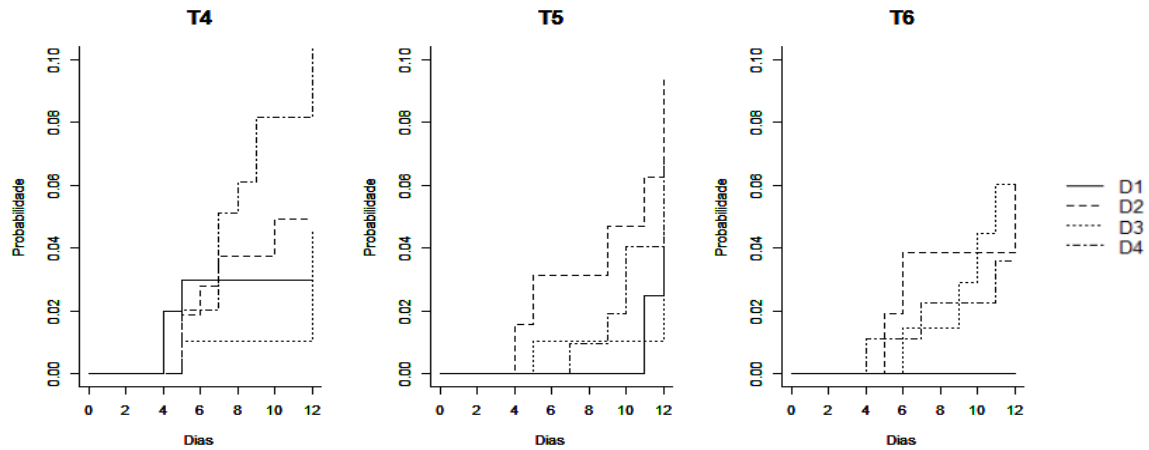
As curvas de mortalidade acumulada para cada tratamento, considerando as quatro datas e as causas de morte 1 e 2, podem ser observadas nas Figuras 8 e 9. Focando nos resultados para a causa 1, pode-se notar, a partir da Figura 8, que as curvas de mortalidade das lagartas que receberam o tratamento T4 com 1 dia (data D2) e 3 dias após a aplicação do vírus (data D3) estão muito próximas. Esse comportamento se repete para os demais tratamentos, sendo que para o tratamento T6 a probabilidade de morte foi menor para essas duas datas (D2 e D3) do que para os tratamentos T4 e T5. Ainda, para os três tratamentos, as curvas cresceram de forma menos acentuada para a data D4 (8 dias após a aplicação do vírus), indicando probabilidade de morte menor para esta data, em particular para o tratamento T5. Ademais, para o tratamento T6, as curvas correspondentes às datas D1 e D4 (1 hora e 8 dias após a aplicação do vírus, respectivamente) estão bem próximas, o que não se repete para os outros dois tratamentos.

FIGURA 8 – Gráficos das curvas de mortalidade acumulada por data e tratamento para a causa 1



FONTE: O autor (2018).

FIGURA 9 - Gráficos das curvas de mortalidade acumulada por data e tratamento para a causa 2



FONTE: O autor (2018).

3.2 RESULTADOS DO MODELO DE REGRESSÃO DE FINE-GRAY

O modelo de regressão de Fine-Gray ajustado aos dados, considerando como causa principal a morte devido ao vírus e as covariáveis data e tratamento, bem como a interação entre elas, que teve de ser mantida no modelo por apresentar efeito significativo, apresenta 11 parâmetros. Na Tabela 6 tem-se a expressão do modelo ajustado para cada combinação de data (D1, D2, D3 e D4) e tratamento (T4, T5 e T6).

TABELA 6 – Expressão do modelo de Fine-Gray ajustado aos dados, considerando as combinações entre data e tratamento para a causa 1

Data e tratamento	Expressão do modelo
D1T4	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}$
D1T5	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}\exp\{-0,617\}$
D1T6	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}\exp\{-1,332\}$
D2T4	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}\exp\{0,226\}$
D2T5	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}\exp\{0,226 - 0.617 + 0.784\}$
D2T6	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}\exp\{0,226 - 1.332 + 0.759\}$
D3T4	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}\exp\{-0.029\}$
D3T5	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}\exp\{-0.029 - 0.617 + 0.500\}$
D3T6	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}\exp\{-0.029 - 1.332 + 0.871\}$
D4T4	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}\exp\{-1.163\}$
D4T5	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}\exp\{-1.163 - 0.617 - 2.206\}$
D4T6	$1 - \exp\{-\hat{\Lambda}_{01}(t)\}\exp\{-1.163 - 1.332 + 1.999\}$

FONTE: O autor (2018).

As estimativas obtidas para a taxa de falha de base $\hat{\Lambda}_{01}(t)$ para cada tempo de falha t (com t variando de 1 a 12 dias) estão dispostas na Tabela 7.

TABELA 7 - Estimativas das taxas de falha de base associadas à causa 1 em cada tempo de falha

Tempo	1	4	5	6	7	8	9	10	11	12
$\hat{\Lambda}_{01}(t)$	0.002	0.012	0.115	0.408	0.789	1.137	1.137	1.489	1.594	1.746

FONTE: O autor (2018).

O comportamento das curvas de mortalidade obtidas por meio do ajuste do modelo de Fine-Gray, mostradas na Figura 10, são similares aos da análise não paramétrica. Porém, algumas observações adicionais podem ser feitas. Para o tratamento T4 (dieta artificial + vírus), observa-se que, exceto para a data D4, as curvas são próximas umas das outras. Entretanto, para o tratamento T5, observa-se que todas as curvas diferiram entre si, enquanto para o tratamento T6 as curvas associadas às datas D1 e D4 (D1 = 1 hora e D4 = 8 dias) voltaram a ficar próximas.

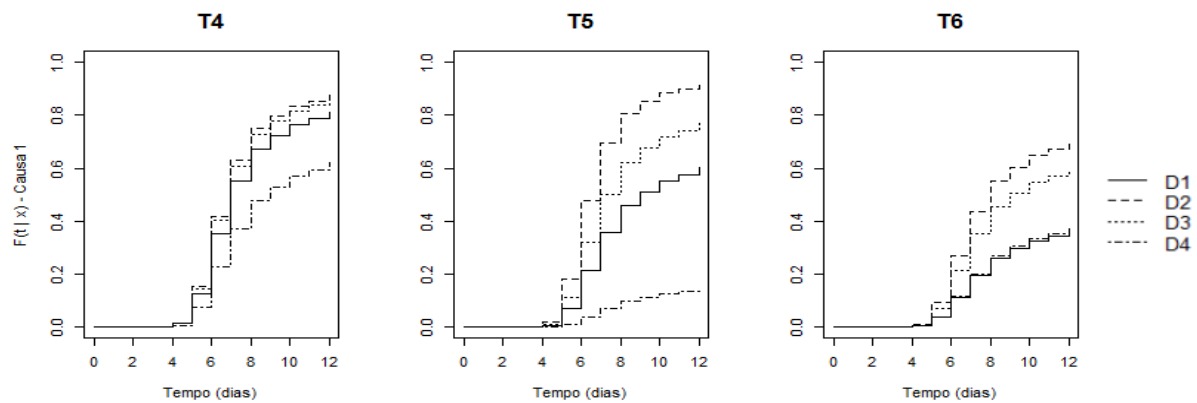
A Tabela 8 mostra as estimativas dos parâmetros e dos resultados do teste Z associados ao modelo ajustado. As categorias de referência consideradas foram: data D1 (1 hora) e tratamento T4 (dieta artificial + vírus).

TABELA 8 – Estimativas dos parâmetros e teste Z associados ao modelo de regressão de Fine-Gray

	β	$\exp(\beta)$	$se(\beta)$	z	Valor p	
Data 2	0.226	1.254	0.135	1.677	9.3e-02	
Data 3	-0.029	0.971	0.158	-0.186	8.5e-01	
Data 4	-1.163	0.313	0.207	-5.626	1.8e-03	***
Tratamento 5	-0.617	0.540	0.157	-3.920	8.9e-05	***
Tratamento 6	-1.332	0.264	0.216	-6.167	7.0e-10	***
Data 2: Tratamento 5	0.784	2.190	0.229	3.422	6.2e-04	***
Data 3: Tratamento 5	0.500	1.649	0.231	2.168	3.0e-01	
Data 4: Tratamento 5	-2.206	0.110	0.742	-2.973	2.9e-04	***
Data 2: Tratamento 6	0.759	2.136	0.277	2.740	6.1e-03	**
Data 3: Tratamento 6	0.871	2.390	0.266	3.270	1.1e-02	
Data 4: Tratamento 6	1.999	3.320	0.323	3.175	2.0e-02	*

FONTE: O autor (2018).

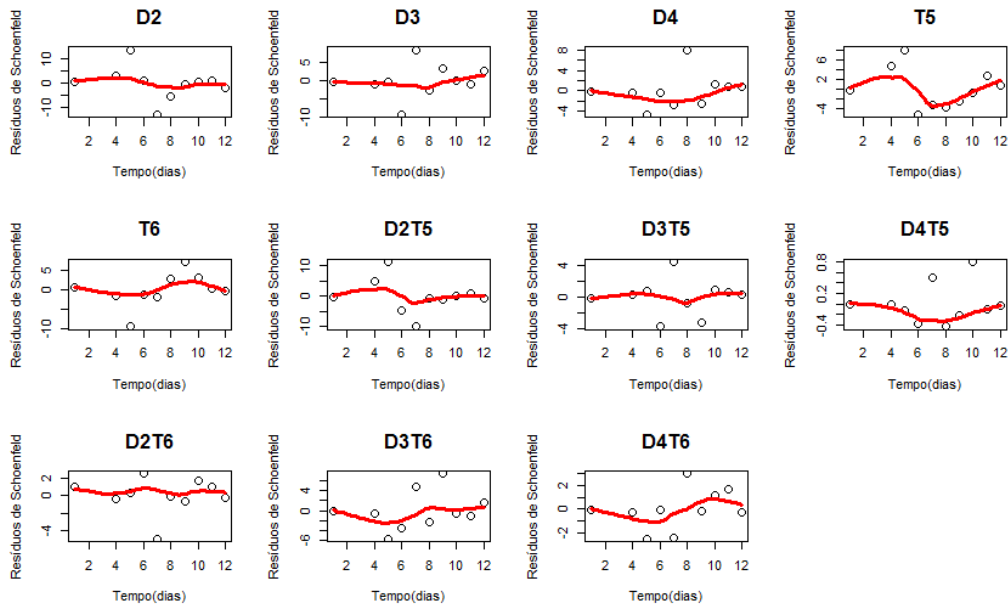
FIGURA 10 – Curvas de mortalidade associadas ao modelo de Fine-Gray por data e tratamento



FONTE: O autor (2018).

Os resíduos de Schoenfeld versus o tempo, úteis para avaliar o pressuposto de taxa de falhas proporcionais, apresentaram as curvas (em vermelho) com inclinação próxima de zero, não evidenciando violação preocupante quanto ao pressuposto mencionado (Figura 11). Sendo assim, o modelo foi utilizado para obtenção dos riscos mostrados na Tabela 9. Com base nessa tabela, tem-se as seguintes interpretações: o risco de uma lagarta morrer devido ao tratamento T4, ao recebê-lo 1 dia após a pulverização (data D2), foi de 1,2411. Este mesmo risco, decorrente do tratamento T4, quando ofertado 3 dias após a pulverização (data D3), foi de 1,17, e quando ofertado 8 dias após a pulverização (data D4) foi de 0,5845.

FIGURA 11 – Resíduos de Schoenfeld versus tempo para cada coeficiente do modelo de Fine-Gray



FONTE: O autor (2018).

TABELA 9 – Riscos obtidos a partir do modelo de Fine-Gray

Tratamento	Data	Risco	IC (2.5%)	IC (97.5%)
T4	D1	1.0000	--	--
	D2	1.2411	0.963	1.599
	D3	1.1700	0.900	1.522
	D4	0.5845	0.423	0.808
T5	D1	0.5516	0.409	0.744
	D2	1.4814	1.073	2.045
	D3	0.8746	0.592	1.291
	D4	0.0934	0.048	0.182
T6	D1	0.2728	0.180	0.414
	D2	0.7196	0.512	1.012
	D3	0.5461	0.364	0.819
	D4	0.2822	0.168	0.475

FONTE: O autor (2018).

Considerando, agora, a razão de riscos, mostradas na Tabela 10, tem-se as seguintes interpretações:

- Na data 1, o risco de morte devido ao tratamento T4 foi 1,813 vezes o risco de morte devido ao tratamento T5 e 3,666 vezes o risco de morte devido ao tratamento T6. Ainda, o risco de morte devido ao tratamento T5 foi 2,022 vezes o risco de morte devido ao tratamento T6.
- Na data 2, o risco de morte devido ao tratamento T5 foi 1,194 vezes o risco de morte devido ao tratamento T4 e 2,059 vezes o risco de morte devido ao

tratamento T6. Ademais, o risco de morte devido ao tratamento T4 foi 1,725 vezes o risco de morte devido ao tratamento T6.

- (c) Na data 3, o risco de morte devido ao tratamento T4 foi 1,338 vezes o risco de morte devido ao tratamento T5 e 2,146 vezes o risco de morte devido ao tratamento 6. Ainda, o risco de morte devido ao tratamento T5 foi 1,602 vezes o risco de morte devido ao tratamento T6.
- (d) Na data 4, o risco de morte devido ao tratamento T4 foi 6,259 vezes o risco de morte devido ao tratamento T5, e 2,071 vezes o risco de morte devido ao tratamento T6. Por fim, o risco de morte devido ao tratamento T6 foi 3,022 vezes o risco de morte devido ao tratamento T5.

TABELA 10 – Razão de riscos comparando os tratamentos em cada data

Data	Tratamentos	Razão de riscos	IC (2.5%)	IC (97.5%)
D1	T4xT5	1.813	1.469	2.266
	T4xT6	3.666	3.006	4.568
	T5xT6	2.022	1.911	2.111
D2	T5xT4	1.194	1.183	1.205
	T4xT6	1.725	1.676	1.778
	T5xT6	2.059	2.027	2.093
D3	T4xT5	1.338	1.296	1.378
	T4xT6	2.143	2.141	2.146
	T5xT6	1.602	1.580	1.624
D4	T4xT5	6.259	4.622	7.462
	T4xT6	2.071	1.998	2.133
	T6xT5	3.022	2.379	3.555

FONTE: O autor (2018).

3.3 RESULTADOS DO MODELO DE COX CAUSA-ESPECÍFICA

Para o modelo de Cox causa-específica, foram observados resultados similares aos obtidos para o modelo de Fine-Gray. Utilizando o pacote *riskRegression*, foi possível ajustar modelos com covariáveis distintas para as causas 1 e 2. Para a causa 1, foram observados efeitos significativos para as covariáveis data e tratamento, bem como para a interação entre elas. Já para a causa 2, apenas as covariáveis “data” e “tratamento na presença de data” apresentaram efeito significativo (a interação não). Estimativas dos parâmetros do modelo ajustado para a causa 1 se encontram na Tabela 11 e para a causa 2 na Tabela 12.

TABELA 11 – Estimativas dos parâmetros do modelo de Cox causa-específica para a causa 1

	β	$\exp(\beta)$	$se(\beta)$	z	$Valor\ p$	
Data 2	0.256	1.292	0.153	1.674	0.094072	.
Data 3	-0.058	0.943	0.177	-0.329	0.742221	
Data 4	-1.287	0.276	0.226	-5.696	1.2e-08	***
Tratamento 5	-0.706	0.493	0.175	-4.030	5.6e-05	***
Tratamento 6	-1.463	0.231	0.230	-6.366	1.9e-10	***
Data 2: Tratamento 5	0.831	2.297	0.243	3.419	0.000628	***
Data 3: Tratamento 5	0.580	1.787	0.260	2.236	0.009383	*
Data 4: Tratamento 5	-2.164	0.115	0.754	-2.869	0.025370	**
Data 2: Tratamento 6	0.781	2.184	0.301	2.598	0.001133	**
Data 3: Tratamento 6	0.951	2.589	0.292	3.255	0.004113	**
Data 4: Tratamento 6	1.319	3.743	0.347	0.347	0.000144	***

FONTE: O autor (2018).

TABELA 12 – Estimativas dos parâmetros do modelo de Cox causa-específica para a causa 2

	β	$\exp(\beta)$	$se(\beta)$	z	$Valor\ p$	
Data 2	1.562	4.769	0.498	3.137	0.00171	**
Data 3	1.247	3.482	0.490	2.545	0.01093	*
Data 4	0.674	1.963	0.495	1.361	0.17357	
Tratamento 5	-0.233	0.792	0.353	-0.660	0.50898	
Tratamento 6	-0.888	0.411	0.378	-2.350	0.01876	*

FONTE: O autor (2018).

Os riscos, obtidos a partir do modelo de Cox causa-específica, apresentados na Tabela 13 para a causa 1, retorna as seguintes conclusões: o risco de óbito devido ao tratamento T4, considerando a data D2, foi de 1,319, enquanto para a data D3 foi de 1,176 e, para a data D4, foi de 0,604, e assim por diante.

TABELA 13 – Riscos obtidos a partir do modelo de Cox causa-específica (causa 1)

Tratamento	Data	Riscos	IC (2.5%)	IC (97.5%)
T4	D1	1.0000	--	--
	D2	1.3190	0.9472	1.7246
	D3	1.1760	0.8696	1.6041
	D4	0.6040	0.3778	0.7479
T5	D1	0.5050	0.3584	0.7125
	D2	1.4699	1.0249	2.1081
	D3	0.8416	0.6840	1.0354
	D4	0.0798	0.0440	0.1448
T6	D1	0.2230	0.1530	0.3766
	D2	0.6670	0.3180	1.3990
	D3	0.4968	0.2590	0.9529
	D4	0.2473	0.1087	0.5626

FONTE: O autor (2018).

No que diz respeito às razões de riscos, mostradas na Tabela 14, tem-se, comparando os tratamentos na data D1, que: o risco de óbito devido ao tratamento T4 foi 1,98 vezes o risco de óbito devido ao tratamento T5, e 4,484 vezes o risco de óbito devido ao tratamento T6. Ainda, o risco de óbito devido ao tratamento T5 foi 2,265 vezes o risco de óbito devido ao tratamento T6.

Para a data D2, o risco de óbito devido ao tratamento T5 foi 1,114 vezes o risco de óbito devido ao tratamento T4, e 2,204 vezes o risco de óbito devido ao tratamento T6. Ademais, o risco de óbito devido ao tratamento T4 foi 1,977 vezes o risco de óbito devido ao tratamento T6. Já para a data D3, o risco de óbito devido ao tratamento T4 foi 1,397 vezes o risco de óbito devido ao tratamento T5, e 2,367 vezes o risco de óbito devido ao tratamento T6. Ainda, o risco de óbito devido ao tratamento T5 foi 1,694 vezes o risco de óbito devido ao tratamento T6. Finalmente, para a data D4, o risco de óbito devido ao tratamento T4 foi 7,566 vezes o risco de óbito devido ao tratamento T5, e 2,442 vezes o risco de óbito devido ao tratamento T6. Por fim, o risco de óbito devido ao tratamento T6 foi 3,098 vezes o do tratamento T5.

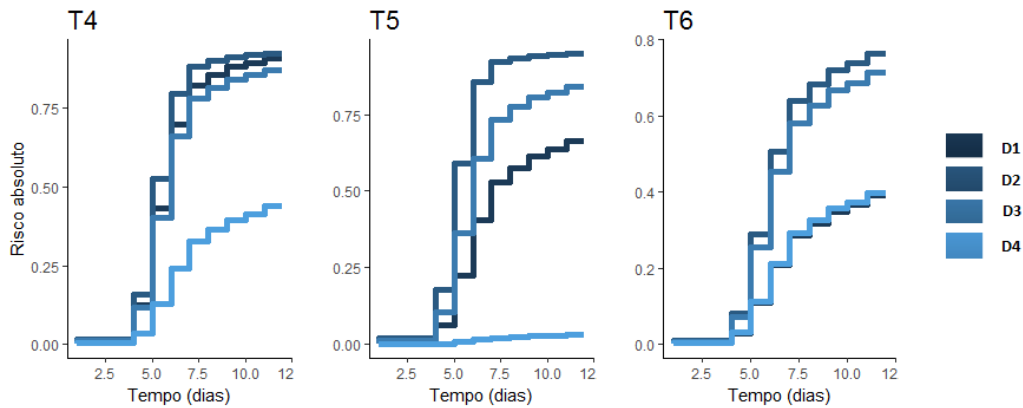
TABELA 14 – Razão de riscos comparando os tratamentos em cada data ($k = 1$)

Data	Tratamentos	Razão de riscos	IC (2.5%)	IC (97.5%)
D1	T4xT5	1.980	1.636	2.433
	T4xT6	4.484	3.824	5.386
	T5xT6	2.265	2.174	2.354
D2	T5xT4	1.114	1.103	1.125
	T4xT6	1.977	1.928	2.03
	T5xT6	2.204	2.172	2.238
D3	T4xT5	1.397	1.355	1.437
	T4xT6	2.367	2.365	2.368
	T5xT6	1.694	1.672	1.716
D4	T4xT5	7.566	5.929	8.769
	T4xT6	2.442	2.369	2.504
	T6xT5	3.098	2.455	3.631

FONTE: O autor (2018).

A partir da Figura 12, nota-se, para as datas D1, D2 e D3, que as curvas são similares para o tratamento T4, o que não ocorre para os outros dois tratamentos. Para o tratamento T6, as curvas associadas às datas D1 e D4 se aproximam, sendo quase iguais após o tempo 5.

FIGURA 12 – Curvas de mortalidade acumulada associadas ao modelo de Cox causa-específica (causa 1) por data e tratamento



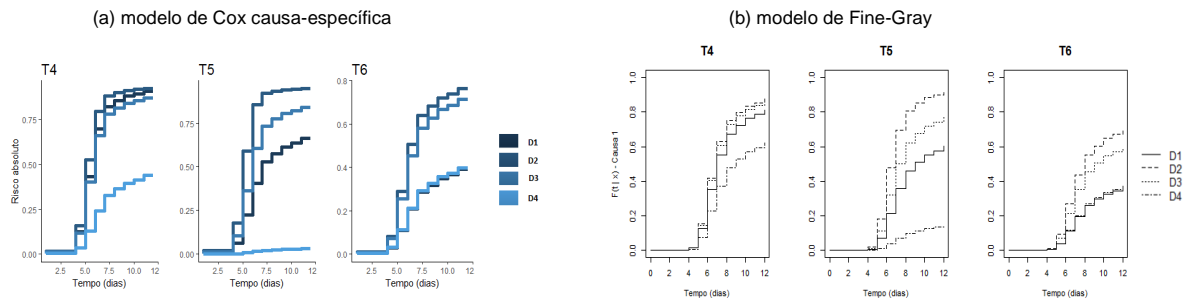
FONTE: O autor (2018).

Na literatura, o diagnóstico dos modelos é geralmente realizado verificando a suposição de proporcionalidade das taxas de falha. Porém, outros gráficos (resíduos martingal e deviance) poderiam também ser úteis para auxiliar a verificar a qualidade de ajuste dos modelos. Esses resíduos, contudo, não estão disponíveis nas saídas dos pacotes utilizados nesse trabalho.

3.4 COMPARAÇÃO DOS RESULTADOS

Ao comparar as curvas de mortalidade obtidas a partir dos dois modelos de regressão (Figura 13), nota-se que elas apresentam comportamento similar. No geral, as probabilidades de morte estimadas para o tratamento T6 diminuíram em relação ao tratamento T4. Além disso, com o aumento do tempo entre a aplicação do vírus e a oferta das dietas para as lagartas, menor foram as probabilidades estimadas.

FIGURA 13 - Comparação das curvas de mortalidade obtidas a partir dos dois modelos ajustados



FONTE: O autor (2018).

As razões de riscos calculadas a partir dos dois modelos foram próximas, como pode ser observado na Tabela 15. Contudo, nota-se que as razões de risco obtidas a partir do modelo de Cox causa-específica foram, em sua maioria, maiores do que as obtidas a partir do modelo de Fine-Gray.

TABELA 15 – Comparação entre as razões de risco dos dois modelos ajustados aos dados

Data	Tratamentos	Modelo de Fine-Gray			Modelo de Cox causa-específica		
		Razão de riscos	IC (2.5%)	IC (97.5%)	Razão de riscos	IC (2.5%)	IC (97.5%)
D1	T4xT5	1.813	1.469	2.266	1.980	1.636	2.433
	T4xT6	3.666	3.006	4.568	4.484	3.824	5.386
	T5xT6	2.022	1.911	2.111	2.265	2.174	2.354
D2	T5xT4	1.194	1.183	1.205	1.114	1.103	1.125
	T4xT6	1.725	1.676	1.778	1.977	1.928	2.03
	T5xT6	2.059	2.027	2.093	2.204	2.172	2.238
D3	T4xT5	1.338	1.296	1.378	1.397	1.355	1.437
	T4xT6	2.143	2.141	2.146	2.367	2.365	2.368
	T5xT6	1.602	1.580	1.624	1.694	1.672	1.716
D4	T4xT5	6.259	4.622	7.462	7.566	5.929	8.769
	T4xT6	2.071	1.998	2.133	2.442	2.369	2.504
	T6xT5	3.022	2.379	3.555	3.098	2.455	3.631

FONTE: O autor (2018).

4 CONCLUSÃO

Os dados analisados neste trabalho possibilitaram o ajuste e a comparação de dois modelos muito utilizados em análise de sobrevivência no contexto de riscos competitivos, o modelo de Fine-Gray e o de Cox causa-específica.

Inicialmente, em um contexto não paramétrico, foram obtidas as curvas de mortalidade de acordo com as combinações das covariáveis data e tratamento. A partir das curvas estimadas e dos resultados do teste de Gray, foi possível concluir que as lagartas que receberam os tratamentos T4, T5 ou T6 após 1 e 3 dias da aplicação do vírus (datas D2 e D3, respectivamente), apresentaram probabilidade maior de vir a óbito. Outro padrão observado nos três tratamentos, foi o de que as lagartas que receberam T4, T5 ou T6 após 8 dias da aplicação do vírus, apresentaram probabilidade menor de vir a óbito.

No contexto dos modelos de regressão, notou-se a partir dos coeficientes estimados que todas as estimativas com valor negativo no modelo de Fine-Gray também apresentaram estimativas negativas no modelo de Cox causa-específica. Além disso, as estimativas foram muito próximas umas das outras. Ao comparar as curvas de mortalidade, notou-se, para o tratamento T4 (dieta artificial + vírus), que elas não apresentaram muita diferença (comportamento similar), sendo que a probabilidade de as lagartas virem a óbito foi maior para aquelas que receberam a dieta após 1 dia da aplicação do vírus (data D2), seguido das que receberam com 3 dias, 1 hora e 8 dias após a aplicação do vírus (datas D3, D1 e D4, respectivamente). Já para o tratamento T5, ocorreu uma mudança na curva associada à data D3 (3 dias); nesta data as probabilidades estimadas diminuíram em relação ao tratamento T4. Para o tratamento T6, não ocorreu muita alteração entre os dois modelos, mas comparado aos tratamentos T4 e T5 houve um decréscimo na probabilidade de óbito. Além disso, a probabilidade de óbito foi menor entre as lagartas que receberam as folhas de soja com vírus (T6), comparado com as que receberam as folhas de algodão com vírus (T5) ou a dieta artificial com o vírus (T4). Também houve menos discrepâncias entre as curvas de mortalidade associadas ao tratamento T4 do que entre as curvas associadas aos tratamentos T5 e T6.

Os riscos mostrados nas Tabelas 9 e 13 foram similares. Porém, notou-se, para o tratamento T4, que os riscos de óbito foram maiores para o modelo de Cox causa-

específica. Já para os demais tratamentos, os riscos de óbito foram maiores para o modelo de Fine-Gray. Em geral, as razões de riscos obtidas a partir do modelo de Cox causa-específica foram maiores do que as obtidas a partir do modelo de Fine-Gray, porém, os resultados foram muito parecidos.

De modo geral, foi possível observar que o vírus *ChinNPV*, quando aplicado sobre plantas de soja e de algodão, produziu mortalidade inferior ao controle positivo (dieta artificial + *ChinNPV*). A maior redução na mortalidade foi observada para o tratamento T5 (folhas de algodão com vírus *ChinNPV*) quando este foi ofertado às lagartas 8 dias após a aplicação do vírus. Houve, ainda, uma redução significativa na mortalidade a partir do 3º dia após a aplicação do vírus, indicando que a atividade viral do *ChinNPV* perde sua eficiência com o decorrer do tempo.

REFERÊNCIAS

- BENATTO, A. **Aumento da eficiência dos vírus ChinNPV e HzNPV com agentes químicos que realçam sua atividade inseticida**. Tese (Doutorado), Universidade Federal do Paraná. Setor de Ciências Biológicas, 2018.
- COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. São Paulo: Edgard Blücher, 2006, 392 p.
- COX, D. R. Regression Models and Life-Tables. **Journal of the Royal Statistical Society**. London, v. 34, n. 2, p. 187-220, 1972.
- DIGNAM, J. J.; ZHANG, Q.; KOCHERGINSKY, M. The use and interpretation of competing risks regression models. **Clinical Cancer Research, Denville**, v. 18, n. 8, p. 2301-2308; 2012.
- FINE, J. P., GRAY, R. J. A proportional hazards model for the subdistribution of a competing risk. **Journal of the American Statistical Association**. v. 94, n. 446, p. 496-509, 1999.
- FERRAZ, R. O.; MOREIRA FILHO, D. C. Survival analysis of women with breast cancer: competing risk models. **Ciência & Saúde Coletiva**, v. 22, n. 11, 2017.
- GLAS, N. A.; KIDERLEN, M.; VANDENBROUCKE, J. P.; CRAEN, A. J. M.; PORTIELJE, J. E. A.; VAN DE VELDE, C. J. H.; LIEFERS, G.; BASTIAANNET, E.; LE CESSIE, S. Performing Survival Analyses in the Presence of Competing Risks: A Clinical Example in Older Breast Cancer Patients. **JNCI: Journal of the National Cancer Institute**, v. 108, i. 5, 2016.
- GRAY, R. J. **Subdistribution analysis of competing risks**. Package cmprsk, 2015. Disponível em <<https://cran.r-project.org/web/packages/cmprsk/cmprsk.pdf>>.
- GRAY, R. J. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. **The Annals of Statistics**, v. 16, n. 3, 1988.
- GREENE, G. L.; LEPPLA, N. C.; DICKERSON, W. A., Velvetbean caterpillar: a rearing procedure and artificial medium. **Journal of Economic Entomology**, v. 69, p. 487-488, 1976.
- GUO, C.; SO, Y. Cause-specific analysis of competing risks using the PHREG procedure. **SAS Institute Inc**. Paper SAS2159, 2018.
- HANLEY, J. A. The Breslow estimator of the nonparametric baseline survivor function in Cox's regression model: some heuristics. **Epidemiology**, Cambridge, v. 19, n. 1, p. 101 - 102, 2008.
- KALBFLEISCH, J.D.; PRENTICE, R. L. **The statistical analysis of failure time data**. New York: John Wiley & Sons, 2002. 462p.

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

LATOUCHE, A., ALLIGNOL, A., BEYERSMANN, J., LABOPIN, M., and FINE, J. P. A Competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. **Journal of Clinical Epidemiology**, v. 66, n. 6, p. 648 - 653, 2013.

OZENNE, B.; SORENSEN, A. L.; SCHEIKE, T.; TORP-PEDERSEN, C.; GERDS, T. A. riskRegression: predicting the risk of an event using Cox regression models. **The R Journal**. v. 9, n. 2, 2017.

PINTILIE, M. **Competing risks: a practical perspective**. Chichester: Wiley, 2006.

APÊNDICES

A1 – Formato de algumas linhas dos dados das lagartas

data	tratamento	obs	repeticao	tempo	causa
1	4	1	1	12	0
1	4	2	1	10	0
1	4	3	1	8	1
1	4	4	1	6	1
1	4	5	1	9	0
1	4	6	1	10	0
1	4	7	1	6	1
1	4	8	1	6	1
1	4	9	1	8	1
1	4	10	1	6	1
1	4	11	1	7	1
1	4	12	1	8	1
1	4	13	1	6	1
1	4	14	1	6	1
1	4	15	1	7	2
1	4	16	1	6	1
...					

Observação:

causa = 0 (censura)
causa = 1 (morte devido ao vírus)
causa = 2 (morte por causa desconhecida)

A2 – Comandos utilizados para ajuste dos modelos de regressão

Ajuste do modelo de Fine-Gray

```
fit04 <- coxph(Surv(tempo, as.numeric(cens)) ~
factor(data)+factor(tratamento)+factor(tratamento)*factor(data),
              data=df, x=T)
Y04 = model.matrix(fit04)
fit14 <- crr(ftime=df$tempo, fstatus=df$causa==1, cov1=Y04)
summary(fit14)
```

Ajuste do modelo de Cox causa-específica

```
fit27 <- CSC(formula
=list(Hist(tempo,causa)~tratamento+data+tratamento*data,
      Hist(tempo,causa)~tratamento+data),
data=df)

# Taxa de falha de base
H0<-matrix(0,10,2)
for(i in 1:10){
  H0[i,1]<- fit14$uftime[i]
  H0[i,2]<-sum(fit14$bfitj[1:i])}

# Resíduos de Schoenfeld
names(fit14$coef) <- c("D2", "D3", "D4",
                      "T5", "T6",
                      "D2T5", "D3T5", "D4T5", "D2T6", "D3T6", "D4T6")

par(mfrow=c(2,6))
for(j in 1:ncol(fit14$res)){
  scatter.smooth(fit14$uft, fit14$res[,j],
                 lpars=list(col='red', lwd=3),
                 mains=names(fit14$coef)[j],
                 xlab='Tempo(dias)',
                 ylab='Resíduos de Schoenfeld')}
```