

Aline D'Avila Pereira

**Modelos Bayesianos de Análise de
Sobrevivência aplicados a dados de
Neoplasia Trofoblástica Gestacional**

Niterói - RJ, Brasil

12 de dezembro de 2022

Aline D'Avila Pereira

**Modelos Bayesianos de Análise de
Sobrevivência aplicados a dados de
Neoplasia Trofoblástica Gestacional**

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em
Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof. Dr. Luis Guillermo Coca Velarde

Co-Orientador(a): Profa. Dra. Fernanda Freitas Oliveira Cardoso

Niterói - RJ, Brasil

12 de dezembro de 2022

Aline D'Avila Pereira

**Modelos Bayesianos de Análise de
Sobrevivência aplicados a dados de
Neoplasia Trofoblástica Gestacional**

Monografia de Projeto Final de Graduação sob o título “*Modelos Bayesianos de Análise de Sobrevivência aplicados a dados de Neoplasia Trofoblástica Gestacional*”, defendida por Aline D'Avila Pereira e aprovada em 12 de dezembro de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Prof. Dr. Luis Guillermo Coca Velarde
Departamento de Estatística – UFF

Prof. Dr. Jony Arrais Pinto Junior
Departamento de Estatística – UFF

Prof. Dr. Antônio Rodrigues Braga Neto
Maternidade Escola da UFRJ – UFRJ

Niterói, 12 de dezembro de 2022

Ficha catalográfica automática - SDC/BIME
Gerada com informações fornecidas pelo autor

P436m Pereira, Aline D'Avila
Modelos Bayesianos de Análise de Sobrevida aplicados a dados de Neoplasia Trofoblástica Gestacional / Aline D'Avila Pereira. - 2022.
63 f.

Orientador: Luis Guillermo Coca Velarde.
Coorientador: Fernanda Freitas Oliveira Cardoso.
Trabalho de Conclusão de Curso (graduação)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2022.

1. Análise de Sobrevida. 2. Letalidade. 3. Doença Trofoblástica Gestacional. 4. Modelos Bayesianos. 5. Produção intelectual. I. Velarde, Luis Guillermo Coca, orientador. II. Cardoso, Fernanda Freitas Oliveira, coorientadora. III. Universidade Federal Fluminense. Instituto de Matemática e Estatística. IV. Título.

CDD - XXX

Resumo

A Doença Trofoblástica Gestacional (DTG) constitui grupo de tumores relacionados à gestação e originários do tecido trofoblástico placentário, cujo marcador biológico-hormonal é o hCG (gonadotrofina coriônica humana). A verdadeira etiologia dessa anomalia ainda não está clara, no entanto a origem da DTG parece ser por gametogênese e fertilização anormais, além da transformação maligna do tecido trofoblástico. A DTG é classificada em dois grupos de importância clínica: mola hidatiforme e neoplasia trofoblástica gestacional (NTG). Esta última é a forma maligna da doença e inclui a mola invasora, o coriocarcinoma, o tumor trofoblástico do sítio placentário (TTSP) e o tumor trofoblástico epitelióide (TTE). Trata-se de patologia rara, que ainda demanda estudos para melhor compreendê-la, em especial com relação ao tempo até o óbito, uma variável que apresenta censura, uma subárea para análise de sobrevivência cuja definição está relacionada ao fato de não existir uma data de ocorrência do desfecho. Esse parâmetro pode possuir inúmeras distribuições, dentre elas a distribuição de Weibull, a qual foi abordada no presente estudo. Assim, o objetivo do presente estudo é avaliar a letalidade NTG, utilizando o modelo bayesiano de análise de sobrevivência para estimar o tempo até o óbito.

As características coletadas foram analisadas através da técnica de análise de sobrevivência com enfoque Bayesiano para estimar o tempo até o óbito. Assim, utilizou-se como variáveis explicativas: classificação quanto à dosagem de hCG < 100 mil (UI/L), histologia, origem, escore prognóstico OMS, metástase, classificação do escore prognóstico, tratamento no centro de referência (CR), tratamento inicial e intervalo entre término da gravidez e início do tratamento. Sendo a sobrevida (em dias) a variável de interesse. Além disso, avaliou-se a curva de sobrevida a partir da utilização do estimador de Kaplan-Meier. Para a obtenção das curvas de sobrevida utilizou-se o software RStudio e para o modelo Bayesiano o OpenBugs. Vale destacar que para definir a *priori* dos parâmetros da Weibull, utilizou-se uma *priori* não informativa para β e r .

Foram acompanhadas 2181 pacientes, desde o primeiro registro em um Centro de Referência até o óbito ou remissão, sendo que dessas 89 foram a óbito devido à NTG, o que representa 4,08% de alcance do desfecho. Utilizando o estimador de Kaplan-Meier, observou-se diferenças nas curvas de sobrevida de todas as variáveis analisadas. Já em relação à avaliação do modelo Bayesiano, observou-se que as variáveis sem histologia, escore OMS ≤ 6 e tratamento inicial único contribuem para aumentar o tempo até o óbito.

Com isso, notou-se que ajustar modelos com o enfoque Bayesiano permite analisar problemas altamente complexos, de forma eficiente e de fácil interpretação, sendo possível calcular a probabilidade do efeito positivo.

Palavras-chave: Análise de Sobrevivência. Letalidade. Epidemiologia. Gravidez. Neoplasia Trofoblástica. Doença Trofoblástica Gestacional. Modelos Bayesianos.

Dedicatória

Dedico este trabalho aos meus pais, Ivan e Marcia, pelo carinho, amor, ensinamentos, educação e apoio.

Agradecimentos

Em primeiro lugar agradeço a Deus, pois tudo que acontece na minha vida é por vontade Dele.

Aos meus pais, Marcia Pereira e Ivan Pereira, e aos meus irmãos, André Pereira e Adriano Pereira, pelo amor, apoio e incentivo nas horas boas e ruins. À Bia, minha sobrinha, que com o tempo foi se tornando uma grande amiga, confidente e ainda joga The Sims 4 comigo nas horas de folga e diversão. Amo muito vocês.

Ao Prof. Luis Guillermo Coca Velarde por me orientar no quesito estatístico e por ter me dado a oportunidade de ajudar as pessoas a passarem na sua matéria. Obrigada também pela sua atenção e por ter me motivado a amar a estatística e seguir nessa nova faculdade. Não tenho palavras para agradecer. Escolhi esse curso por você e espero que você fique orgulhoso de mim.

A minha coorientadora Fernanda Cardoso, por todo cuidado com a pesquisa, pela coleta dos dados e por me ajudar na escrita e confortar com palavras. Agradeço por todo carinho.

Aos professores do curso de estatística da UFF. Vou levar vocês pra sempre no meu coração. Em especial: Jony, Jessica, Ana Maria, Rafael, Yasmin e Ludmilla. Muito obrigada por todos os ensinamentos e cuidado com os alunos.

Aos amigos que ganhei ao longo de toda a faculdade de Estatística. Em especial, aos "amigos da Gabi", Carol e Leandro que me ouviram e me ajudaram a passar nas matérias mais difíceis.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 12
1.1	Doença Trofoblástica Gestacional	p. 12
1.2	Doença Trofoblástica Gestacional e Dados Censurados	p. 14
1.3	Objetivos	p. 16
2	Materiais e Métodos	p. 17
2.1	Desenho e tipo do estudo	p. 17
2.2	Análise de sobrevivência	p. 18
2.2.1	Definições básicas	p. 19
2.2.2	Funções básicas	p. 21
2.2.2.1	Função densidade de probabilidade	p. 21
2.2.2.2	Função de sobrevivência e função de risco	p. 22
2.2.3	Análise exploratória	p. 25
2.2.3.1	Estimador produto de Kaplan-Meier	p. 26
2.2.4	Modelo paramétrico de sobrevivência	p. 29
2.2.4.1	Distribuição Weibull	p. 29
2.2.5	Enfoque Bayesiano	p. 31
2.2.5.1	Aplicação	p. 31
2.2.5.2	Teorema de Bayes	p. 31

2.2.5.3	Distribuição <i>a priori</i> e <i>a posteriori</i>	p. 32
2.2.5.4	Cadeia de Markov	p. 34
2.2.5.5	Seleção do modelo	p. 36
3	Resultados	p. 37
3.1	Análise descritiva	p. 37
3.2	Modelo de regressão sob enfoque Bayesiano	p. 44
4	Conclusão	p. 50
	Referências	p. 52
5	Apêndice A - Tabelas com a função de sobrevivência das variáveis analisadas	p. 53
6	Apêndice B - <i>Script</i> do modelo implementado no <i>OpenBugs</i>	p. 61
7	Apêndice C - Modelos Final	p. 63

Lista de Figuras

1	Distribuição dos dados entre os óbitos e não óbitos	p. 40
2	Função de sobrevivência estimada pelo método de Kaplan-Meier	p. 41
3	Função de sobrevivência estimada pelo método de Kaplan-Meier das variáveis avaliadas	p. 42
4	Monitoramento da Convergência dos Parâmetros do Modelo Completo .	p. 47
5	Densidade a <i>posteriori</i> dos parâmetros do Modelo Completo	p. 48

Lista de Tabelas

1	Variáveis clínicas, histológicas e medicamentosas obtidas das pacientes que participaram do presente estudo	p. 18
2	Características clínicas das pacientes estudadas	p. 38
3	Efeito das variáveis no tempo até o óbito, considerando o modelo simples	p. 46
4	Estimativas do parâmetros do Modelo Completo	p. 48
5	Avaliação da qualidade do modelo completo e dos modelos com apenas uma variável explicativa	p. 49
6	Função de sobrevivência estimada pelo método de Kaplan-Meier	p. 53
7	Função de sobrevivência estimada pelo método de Kaplan-Meier conforme hCG	p. 53
8	Função de sobrevivência estimada pelo método de Kaplan-Meier conforme histologia	p. 54
9	Função de sobrevivência estimada pelo método de Kaplan-Meier conforme origem da NTG	p. 55
10	Função de sobrevivência estimada pelo método de Kaplan-Meier conforme estadiamento	p. 56
11	Função de sobrevivência estimada pelo método de Kaplan-Meier conforme presença ou não de metástase	p. 57
12	Função de sobrevivência estimada pelo método de Kaplan-Meier conforme escore prognóstico OMS	p. 58
13	Função de sobrevivência estimada pelo método de Kaplan-Meier conforme local de tratamento	p. 59
14	Função de sobrevivência estimada pelo método de Kaplan-Meier conforme tratamento inicial utilizado	p. 60

1 Introdução

1.1 Doença Trofoblástica Gestacional

A Doença Trofoblástica Gestacional (DTG) constitui grupo de tumores relacionados à gestação e originários do tecido trofoblástico placentário, cujo marcador biológico-hormonal é o hCG (gonadotrofina coriônica humana). Apesar de ser uma patologia rara, por vezes pouco conhecida pela população, a DTG começou a ser estudada em meados do século XX. Em 1959, foi inaugurado o primeiro Centro de Referência (CR) para o tratamento das doenças do trofoblasto, permitindo o início do desenvolvimento de protocolos específicos, que viriam a ser adotados em outros CR pelo país posteriormente. No ano de 2017, já havia 39 CR no Brasil. No ano de 2016, foi criada a Comissão Nacional Especializada em DTG da Federação Brasileira das Associações de Ginecologia e Obstetrícia (FEBRASGO, 2017). No entanto, vale destacar que no Brasil, não há registro nacional desta patologia, diferente do que acontece em outros países, e apesar do crescimento, da visibilidade e da importância da DTG (CARDOSO, 2017).

A etiologia da DTG não está clara, no entanto, segundo BERKOWITZ (2009) e CARDOSO (2017), a origem da DGT pode ter relação com a gametogênese e fertilização anormais, podendo estar associado também a transformações malignas do tecido trofoblástico. Apresenta-se na forma benigna e maligna. Na benigna, tem-se a mola hidatiforme (MH), que pode ser subdividida em mola hidatiforme completa (MHC) ou parcial (MHP). Já na forma maligna, temos a neoplasia trofoblástica gestacional (NTG), que inclui a mola invasora, o coriocarcinoma, o tumor trofoblástico do sítio placentário e o tumor trofoblástico epitelióide (CARDOSO, 2017). CARDOSO (2017) e BRAGA et al. (2018) destacam que a estimativa da incidência da DTG é de 1 a cada 200 a 400 gestações e que 7 a 25% dos seus casos são malignos, taxas superiores ao que é encontrado em outros países. No entanto, apesar de apresentar um risco para a vida e fertilidade das mulheres, a NTG possui alta taxa de cura, sendo importante destacar o diagnóstico precoce e o tratamento adequado e oportuno em CR .

A DTG possui como marcador biológico o hormônio gonadotrofina coriônica humana (hCG) que é produzido durante a gravidez, e que pode apresentar valores bastante elevados nesta doença. Com níveis hormonais mais altos, acima de 200.000 UI/L, podemos ter maior chance de desenvolvimento de formas malignas da doença (RENGANATHAN, 2016). Se após o término da gravidez, o tecido trofoblástico ainda for detectado, através de níveis de hCG em ascensão ou platô, também podemos estar diante de doença maligna, a NTG. Nesse sentido, é importante realizar acompanhamento, sendo necessário mensurar os níveis de hCG no sangue a cada uma ou duas semanas, até a obtenção de três resultados consecutivos negativos, ou seja, inferiores a 5 UI/L. Logo após esse período, é preciso avaliar exames mensais por mais 6 meses. Como o hCG é um hormônio produzido durante a gestação, é imprescindível que a mulher não inicie uma nova gravidez no período de acompanhamento (CARDOSO, 2017; FEBRASGO, 2017).

Em relação ao processo de tratamento e cura, a evolução da medicina tem proporcionado um maior índice de cura, visto a utilização de tecnologias de diagnóstico como o ultrassom, que também contribuiu para o tratamento, cura e preservação da capacidade reprodutiva. No entanto, é possível notar que os casos de morte ocorrem quando há um diagnóstico ou um esvaziamento uterino tardio, assim, destacando, mais uma vez, a importância de um diagnóstico precoce (FEBRASGO, 2017).

A NTG, forma maligna da DTG, é classificada através da distribuição anatômica da neoplasia e sistema de escore de risco apresentado pela Organização Mundial da Saúde. Se o escore foi menor ou igual a 6, é considerada uma doença de baixo-risco para resistência a monoquimioterapia, podendo ser tratada com agente único inicialmente; caso o escore de risco seja maior que 6, é considerada doença de alto-risco para resistência a monoquimioterapia, sendo necessário o tratamento com múltiplos agentes quimioterápicos desde a abordagem terapêutica inicial. Logo, a indicação do tratamento quimioterápico adequado é dependente da classificação da NTG (CARDOSO, 2017). CARDOSO (2017) destaca ainda que seria importante avaliar outros fatores associados com o tempo até o óbito em mulheres com NTG.

Dessa forma, o método de Análise de Sobrevida é muito utilizado na área de epidemiologia para identificar fatores que influenciam o tempo até o óbito. Com o conhecimento mais amplo sobre os fatores, pode-se prevenir e/ou tratar a doença e, assim, contribuir para a melhora da qualidade de vida dos pacientes. Esse método será introduzido e aprofundado na Seção 2.1

1.2 Doença Trofoblástica Gestacional e Dados Censurados

Os estudos na área da saúde podem, muitas das vezes, envolver covariáveis relacionadas com o tempo de sobrevivência. Assim, com o aprofundamento das técnicas estatísticas, a Análise de Sobrevivência cresceu muito no final do século XX. Esse tipo de método também pode ser chamado de Análise de sobrevivência e foi definida por KLEINBAUM (1996) como um conjunto de procedimentos estatísticos utilizado para analisar dados que a variável resposta é o tempo de ocorrência de um evento de interesse, denominado falha ou desfecho. Além disso, CARVALHO et al. (2011) eluciu que esse método consiste em uma classe de modelos quantitativos estocásticos utilizado para avaliar quesitos e características até a ocorrência de um evento em específico.

Além da área da saúde, essa análise pode ser utilizada em outras áreas como, por exemplo na engenharia. Nesse campo, a Análise de Sobrevivência é chamada de Confiabilidade e, nesse contexto, é avaliado o tempo até um equipamento apresentar falha (EBELING, 1997). Também pode ser utilizada nas ciências sociais, criminologia, demografia etc.

Na epidemiologia, a Análise de Sobrevivência como supracitado avalia o tempo até a ocorrência de um evento que pode ser a recidiva de um tumor, o risco de ocorrência de um recém-nascido vir a óbito após o nascimento. Assim, essa perspectiva, avaliando esses cenários, tem-se um fator importante a ser avaliado que é a presença de censura, uma subárea para análise de sobrevivência cuja definição está relacionada ao fato de não existir uma data de ocorrência do desfecho, ou seja, quando há observações parciais dos dados visto que o evento de interesse não ocorreu no período estudado (COLOSIMO; GIOLO, 2014). Assim, visando a sua análise exploratória, pode-se utilizar o estimador de Kaplan-Meier. Vale destacar, nesse sentido, que a presença de dados censurados é a principal diferença desse tipo de análise de regressão e planejamento de experimentos. Dessa maneira, na presença de dados censurados, outras abordagens podem ser utilizadas como o modelo clássico de regressão de riscos proporcionais de Cox ou o modelo Bayesiano como definido por COLOSIMO e GIOLO (2014) e PAULINO (2003).

Com essa abordagem clássica de dados censurados, CARVALHO (2019) avaliou modelos de Análise de Sobrevivência aplicados a dados de Neoplasia Trofoblástica Gestacional de mulheres de Porto Alegre e observou que menos de 5% das pacientes apresentaram histologia por MOLA invasora ou tumor trofoblástico do sítio placentário, gravidez ante-

cedente ectópica ou a termo/pré-termo e estadiamento FIGO I ou IV. Como resultado da análise frequentista, notou-se que a dosagem de hCG (UI/L) pré-tratamento, histologia NTG, estadiamento FIGO, escore prognóstico, intervalo entre o término da gestação e início do tratamento (em semanas) e quimiorresistência são variáveis significativas para modelar o risco de óbito. No entanto, a idade da paciente, a presença de gestação antecedente e o local de tratamento não foram variáveis significativas.

Outro estudo, realizado em 2020, utilizando os dados da presente pesquisa, utilizou modelos semiparamétricos de risco proporcional de Cox construídos utilizando cada variável com valor de $p < 0,2$ para construir um modelo múltiplo contendo apenas variáveis com influência significativa no risco relativo (RR) de morte por NTG, incluindo intervalos de confiança de 95% (IC 95%). Nesse sentido, o modelo foi construído utilizando uma estratégia “forward”, inserindo variáveis com menores valores de p para o teste de Wald avaliando a significância do efeito do fator, uma por vez. Observaram que apesar das variáveis concentração sérica de gonadotrofina coriônica humana pré-tratamento, tipo de gravidez antecedente, tempo entre o fim da gravidez anterior e o tratamento, estadiamento FIGO, local de metástase e tratamento inicial terem apresentado significância no RR, o efeito dessas variáveis foi absorvido por outras variáveis ao longo da construção do modelo ajustado pelo método “forward” e, portanto, não incluído no modelo final de riscos proporcionais de Cox. Assim, observaram que as variáveis significativas que compuseram o modelo final foram a histologia, presença de metástase, tratamento inicial, escore OMS e tratamento no centro de referência (CR) (FREITAS et al., 2020). Diferente desse estudo, a atual pesquisa visa modelar o tempo até o óbito, avaliando quais variáveis teriam associação significativa, ou seja, aquela que contribui para o aumento ou diminuição do tempo até o óbito sob um enfoque Bayesiano.

Como supracitado, pelas abordagens clássicas algumas covariáveis significativas não foram incluídas no modelo final. Nesse sentido, será utilizada uma abordagem Bayesiana para a estimação dos parâmetros do modelo de sobrevivência, aprofundada na Seção 2.2, para o estudo da letalidade da neoplasia trofoblástica gestacional apresentada nesta Seção. Esse tipo de estudo promove uma avaliação das variáveis, sendo o tempo até o óbito a variável resposta. Assim, o presente estudo apresenta uma importância na literatura epidemiológica.

Vale destacar, que será utilizada uma abordagem paramétrica para estimar as funções básicas da análise de sobrevivência. Assim, uma vantagem dessa abordagem é assumir-se que o tempo T até o evento segue uma distribuição conhecida de probabilidade, no caso

do presente estudo, a distribuição Weibull, e a partir disso, os parâmetros correspondentes são estimados. Assim, diferente do modelo semiparamétrico de riscos proporcionais que visa estimar o efeito de covariáveis, com a abordagem paramétrica é possível estimar os parâmetros da distribuição do tempo de sobrevivência.

No modelo de regressão paramétrico, uma distribuição é escolhida para a variável resposta e, a partir disso, identifica-se um dos modelos de regressão, tais como: linear, Poisson, exponencial, logístico, entre outros. Na análise de sobrevivência, a ideia é a mesma, entretanto, a distribuição precisa incorporar as características do tipo de estudo, o tipo de variável resposta e a presença de dados censurados.

1.3 Objetivos

Como apresentado na Seção 1.1, a DTG é patologia rara que vem sendo muito estudada para avançarmos em sua compreensão. Assim, o presente estudo tem o objetivo de estudar a letalidade da neoplasia trofoblástica gestacional através do tempo até o óbito.

Com enfoque Bayesiano, será utilizado o modelo paramétrico de análise de sobrevivência para o tempo até o óbito. Para isso, será realizada a análise das características ou fatores associados ao tempo até a ocorrência de um determinado evento, que neste caso é óbito por NTG.

No Capítulo 3, os resultados serão apresentados, utilizando uma abordagem Bayesiana. No Capítulo 4, posteriormente, serão apresentadas as conclusões e as considerações finais.

2 Materiais e Métodos

Este capítulo será dividido em duas partes: a primeira relacionada às características do estudo e dos dados e a segunda sobre a metodologia de análise de dados. Na Seção 2.1, serão evidenciados às características do estudo associado ao desenho e tipo de estudo, a origem e característica dos dados e o software utilizado.

Na última Seção 2.2, será realizada a descrição da metodologia de Análise de Sobrevivência, sendo elucidada as suas definições e funções básicas, a análise exploratória através do estimador produto de Kaplan-Meier, o modelo paramétrico de sobrevivência e o modelo Bayesiano para estimar o tempo até o óbito. Maiores detalhes sobre a metodologia podem ser encontrados em CARVALHO et al. (2011) e COLOSIMO e GIOLO (2014).

2.1 Desenho e tipo do estudo

O desenho do presente estudo foi uma coorte aberta, colaborativa, retrospectiva de base hospitalar, com resposta de natureza longitudinal. Os dados são provenientes de uma consulta de banco de dados, prontuários e protocolos médicos de pacientes com neoplasia trofoblástica gestacional que foram acompanhados em Centros de Referência no Brasil. Para obtenção de mais informações sobre a coleta de dados, é sugerido consultar FREITAS et al. (2020).

Foram coletadas informações, durante 12 meses, de 2181 mulheres portadoras de NTG que atingiram remissão ou foram a óbito. O banco de dados foi construído no programa Excel. As variáveis estudadas se encontram na Tabela 1.

Tabela 1: Variáveis clínicas, histológicas e medicamentosas obtidas das pacientes que participaram do presente estudo

Id	Variável	Tipo de variável
V1	Sobrevida (em dias): tempo até o óbito	Numérica discreta
V2	Óbito: sim, não	Categórica binária
V3	hCG (UI/L) pré-tratamento: dosagem de hCG antes do tratamento começar	Numérica contínua
V4	Dosagem de hCG < 100 mil (U/I): sim, não	Categórica binária
V5	Idade (em anos): idade das pacientes	Numérica discreta
V6	Classificação da idade: ≤ 19 anos, entre 20 e 39 anos, ≥ 40 anos	Categórica ordinal
V7	Histologia: sem histologia, mola invasora, coriocarcinoma	Categórica nominal
V8	Origem: pós-molar, aborto, gravidez ectópica, gestação à termo ou pré-termo	Categórica nominal
V9	Estadiamento: I, II, III, IV	Categórica ordinal
V10	Metástase: sim, não	Categórica binária
V11	Escore prognóstico OMS	Numérica discreta
V12	Classificação Escore prognóstico OMS: ≤ 6 , ≥ 7	Categórica nominal
V13	Tratamento no CR: sim, não	Categórica binária
V14	Tratamento quimioterápico inicial: sim, não	Categórica binária
V15	Classificação do tratamento inicial: único, múltiplo e não quimioterápico	Categórica nominal
V16	Intervalo entre término da gravidez e início do tratamento (em dias)	Numérico contínuo

Todas as análises do presente estudo foram realizadas no software open-source R Studio, versão 4.1.1. Além disso, a fim de avaliar o modelo com o enfoque Bayesiano foi utilizado o software Open-Bugs, versão 3.2.2.

2.2 Análise de sobrevivência

A análise de sobrevivência pode ser caracterizada como um conjunto de técnicas e modelos da estatística quantitativa estocástica associada com a avaliação de características e fatores relacionados ao tempo até a ocorrência do desfecho desejado, além de forçenc

ferramentas para análise de dados censurados, incluindo modelos. Assim, não há perda de informações. Dessa forma, esse método é um tipo de modelo de regressão, em que há presença de variável resposta, covariáveis explicativas, funções de ligação e estrutura de erro (CARVALHO et al., 2011).

A variável resposta pode ser elucidada de três maneiras interconvertíveis, sendo elas a probabilidade de sobrevivência, a taxa de incidência e a taxa de incidência acumulada. A primeira está relacionada com a probabilidade de não ocorrência do desfecho desejado e por definição, no início das observações, essa probabilidade é igual a 1. A taxa de incidência diz respeito à taxa instantânea de ocorrência do evento de interesse no tempo, condicionada com a sua não ocorrência anterior a esse momento. Já a incidência acumulada, refere-se à função de risco acumulado, ou seja, mede o risco de ocorrência do evento até um determinado tempo (CARVALHO et al., 2011).

Nesse sentido, nas próximas seções serão aprofundados os conceitos e funções básicas, análise exploratória e técnicas paramétricas de análise de sobrevivência com enfoque bayesiano.

2.2.1 Definições básicas

Na análise de sobrevivência, o foco do estudo é avaliar o tempo até a ocorrência de um determinado evento, podendo ser o diagnóstico de uma doença, o nascimento, a cura e o óbito. Assim, vale destacar que determinados indivíduos de uma coorte podem não apresentar o desfecho desejado. Nesse sentido, é necessário conceituar a censura (CARVALHO et al., 2011).

A censura é definida como a perda de informação devido à falta de observação da data de ocorrência do desfecho, uma vez que o indivíduo estudado não apresentou o evento de interesse. Segundo CARVALHO et al. (2011), existem 3 tipos de censuras: à direita, à esquerda e intervalar. A censura à direita é a mais comum e é definida pela não ocorrência do desfecho até o final do estudo; à esquerda é quando a observação desejada ocorre antes do tempo de início do estudo, podendo estar relacionada com o não conhecimento do momento da ocorrência; e intervalar quando o desfecho acontece durante o tempo observado, mas não há exatidão quanto ao tempo.

Assim, T é uma variável aleatória que determina o tempo até o desfecho, sendo composto pelo tempo inicial, escala de medida e evento de interesse; L uma variável aleatória que representa o limite inferior de um intervalo de observação; e U uma variável aleatória

associado ao limite superior desse intervalo. Nesse sentido, as censuras supracitadas podem ser representadas da seguinte forma:

- Censura à direita: $L < T$;
- Censura à esquerda: $T \leq U$;
- Censura intervalar: $L \leq T < U$.

Além disso, há a censura informativa e não informativa. A não informativa refere-se à perda de informação devido ao fato dessa ausência não estar relacionada com o desfecho, enquanto, na informativa, há presença de dado censurado, pois o indivíduo estudado decidiu abandonar a pesquisa devido a fatores relacionados ao evento estudado como, por exemplo, abandono por causa da piora do paciente (CARVALHO et al., 2011).

Segundo COLOSIMO; GIOLO (2014), há mais 3 tipos de censura: tipo I, tipo II e tipo III. A censura do tipo I e II estão associadas com a ausência do evento de interesse até o término do estudo, sendo que, no tipo I, esse fim está previamente definido e, no tipo II, a finalização ocorre devido ao alcance de uma quantidade de desfechos pré-estabelecidos. Já a tipo III, é quando o indivíduo é retirado do estudo, mesmo que não apresente o desfecho desejado.

Nesse sentido, a censura e o desfecho podem ser representados a partir da variável indicadora δ_i , como evidenciado a seguir:

$$\delta_i = \begin{cases} 1, & \text{se } T \leq C, \text{ ou seja, é um desfecho,} \\ 0, & \text{se } T > C, \text{ ou seja, é um censura,} \end{cases}$$

sendo C uma variável aleatória independente a T que indica o tempo até a censura e δ_i a ocorrência ou não do desfecho no indivíduo i , ou seja, a variável indicadora de falha ou censura.

Normalmente, os dados de sobrevivência para cada pessoa avaliada são apresentados pelo par (t_i, δ_i) . Caso ocorra a existência de covariáveis, a representação acontece pela trinca $(t_i, \delta_i, \mathbf{x}_i)$.

Nessa perspectiva, vale destacar que é necessário incluir os dados censurados nos estudos e na análise estatística. Conforme COLOSIMO e GIOLO (2014), mesmo que o dado seja censurado, visto informações incompletas, tem-se informações sobre o tempo de sobrevida dos participantes do estudo e, além disso, a falta de conhecimento sobre os dados censurados implica em conclusões viciadas.

Além da censura, observa-se na análise de sobrevivência o truncamento que ocorre quando há exclusão devido a fatores associados com a ocorrência do evento. Nesse sentido, todos os dados do indivíduo são excluídos do estudo, uma vez que não há a apresentação do desfecho desejado dentro de uma janela temporal pré-estabelecida. Assim, pode-se observar o truncamento à direita e à esquerda, que apresentam a ideia semelhante à censura. O truncamento à direita ocorre quando o desfecho não acontece até o fim do estudo, e o truncamento à esquerda quando o desfecho ocorre antes do início do estudo.

Nesse sentido, a censura e o truncamento proporcionam um viés ao estudo, sendo necessário o uso de métodos para mitigá-lo. Dentre essas técnicas, destaca-se a predefinição do perfil do paciente que é aceito para compor o estudo, o tempo de acompanhamento e tempo do estudo, assim como a identificação das causas de censura e de truncamento. Vale elucidar ainda, a utilização de medidas descritivas que permitam a comparação das variáveis coletadas relacionadas aos perfis dos pacientes censurados ou truncados. Os tipos de estudo que são utilizados para esse tipo de análise são: coorte, caso-controle ou de intervenção, os quais permitem acompanhamento dos participantes até o desfecho. No presente estudo, tem-se uma coorte aberta, no qual a data de entrada do indivíduo na pesquisa varia.

Os dados do atual estudo, considerando pacientes com DTG, foram compostos por censura à direita, ou seja, alguns pacientes não foram à óbito até o final do estudo. Além disso, todas as censuras foram do tipo informativas, pois as participantes saíram do estudo devido à remissão, ao óbito ou à falta de informação sobre a sua saída por causa de fatores relacionados ao evento estudado. Também só há censura do tipo I e III, não sendo observada a do tipo II, visto que não houve interesse em consultar uma quantidade pré-definida dos dados.

2.2.2 Funções básicas

2.2.2.1 Função densidade de probabilidade

O tempo até o desfecho, T , é definido como uma variável aleatória contínua positiva, sendo a sua função densidade de probabilidade conhecida como $f(t)$ CARVALHO et al.(2011). Essa função pode ser compreendida como a probabilidade de o indivíduo apresentar o desfecho em um intervalo instantâneo de tempo:

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

onde ϵ é o incremento ou acréscimo de tempo infinitamente pequeno, ou seja, que tende a 0.

Quando não há censura, isto é, quando todos os indivíduos avaliados apresentam o desfecho de interesse no período estudado, a $f(t)$ pode ser estimada a partir do seguinte cálculo:

$$\hat{f}(t) = \frac{N_t(t)}{(N)(\Delta t)}, \quad (2.1)$$

em que t é o índice do intervalo, ou seja, o intervalo de tempo no qual se calcula a função de densidade; N é o número total de ocorrências; $N_t(t)$ é o número de ocorrências no intervalo de tempo t ; e Δt é a amplitude do intervalo de tempo t .

2.2.2.2 Função de sobrevivência e função de risco

Geralmente, T é elucidado pela função de sobrevivência, $S(t)$, ou pela função de risco, $\lambda(t)$. A função de sobrevivência é uma das funções mais importantes da Análise de Sobrevivência, uma vez que ela representa a probabilidade de uma observação sobreviver pelo menos o tempo t , definida como:

$$S(t) = P(T \geq t).$$

Da mesma maneira que na função densidade, caso não haja ocorrência de censura, a função de sobrevivência para o intervalo de tempo t pode ser estimada a partir do seguinte cálculo:

$$\hat{S}(t) = \frac{R_t(t)}{N}, \quad (2.2)$$

onde R_t é o número de indivíduos sob risco no início do intervalo de tempo t .

Pode-se notar, que pela sua natureza, essa função é monótona não crescente, sendo decorrente da função de distribuição acumulada, $F(t)$. Essa função é definida pela probabilidade de uma observação falhar até o tempo t , sendo que a sua expressão pode ser representada da seguinte maneira:

$$\begin{aligned} S(t) &= P(T \geq t) \\ &= 1 - P(T < t) \\ &= 1 - F(t). \end{aligned}$$

Como supracitado, a função de sobrevivência está associada com a probabilidade de sobreviver pelo menos até o momento t . Assim, também pode-se encontrar a probabilidade

do desfecho ocorrer em um intervalo de tempo $[t_1, t_2)$, como determinado a seguir:

$$\begin{aligned} P(t_1 \leq T < t_2) &= P(T < t_2) - P(T < t_1) \\ &= [1 - P(T \leq t_1)] - [1 - P(T \leq t_2)] \\ &= S(t_1) - S(t_2). \end{aligned}$$

Nesse sentido, a função de sobrevivência pode ser utilizada na comparação do tempo de vida de dois ou mais grupos de indivíduos diferentes, tal como a relação entre o tempo e a sobrevivência deles. Dessa forma, é possível estabelecer o tempo para os percentis ou a porcentagem de pessoas que sobreviveram em um determinado período.

Além da função de sobrevivência, tem-se a função de risco (hazard), que pode ser definida conforme estabelecido por COLOSIMO E GIOLO (2014) e CARVALHO et al. (2011). Para COLOSIMO e GIOLO (2014), essa função é caracterizada como a probabilidade da observação alcançar o desfecho em um intervalo de tempo $[t_1, t_2)$, desde que o desfecho não ocorra antes de t_1 , sendo necessário considerar a medida do intervalo. No entanto, conforme CARVALHO et al. (2011), $\lambda(t)$ não é uma probabilidade e sim uma taxa, visto que pode assumir valores maiores do que 1. Além disso, a partir do resultado de $S(t_1) - S(t_2)$, tem-se a taxa de incidência, que é elucidada da seguinte forma:

$$\frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)}$$

Devido a questões matemáticas, o intervalo pode ser redefinido de $[t_1, t_2)$ para $[t, t + \epsilon)$. Dessa forma, a expressão evidenciada acima, pode ser apresentada da seguinte maneira:

$$\frac{S(t) - S(t + \epsilon)}{\epsilon S(t)}$$

Assumindo que ϵ é infinitamente pequeno, ou seja, próximo de 0, tem-se que a taxa de risco instantânea no tempo t , dado que o desfecho ainda não ocorreu até esse tempo, é definida:

$$\lambda(T) = \lim_{\epsilon \rightarrow 0} \frac{P(t \leq T \leq t + \epsilon | T \geq t)}{\epsilon}. \quad (2.3)$$

Caso não haja nenhum dado censurado, estima-se $\lambda(t)$ da seguinte maneira:

$$\hat{\lambda}(t) = \frac{N_t(t)}{R_t(t)(\Delta t)}, \quad (2.4)$$

Vale destacar ainda, que a função de risco possui uma relação com a função de sobrevivência, de densidade de t , $f(t)$ e/ou de probabilidade acumulada de t , $F(t)$. Assim,

a partir da definição 2.3, tem-se:

$$\begin{aligned}
 \lambda(x) &= \lim_{\epsilon \rightarrow 0} \frac{P(t \leq T \leq t + \epsilon | T \geq t)}{\epsilon} \\
 &= \lim_{\epsilon \rightarrow 0} \frac{P(t \leq T \leq t + \epsilon, T \geq t)}{\epsilon P(T \geq t)} \\
 &= \lim_{\epsilon \rightarrow 0} \frac{P(t \leq T \leq t + \epsilon)}{\epsilon P(T \geq t)} \\
 &= \frac{1}{P(T \geq t)} \lim_{\epsilon \rightarrow 0} \frac{P(t \leq T \leq t + \epsilon)}{\epsilon} \\
 &= \frac{f(t)}{S(t)}.
 \end{aligned}$$

A partir disso, tem-se:

(a)

$$\begin{aligned}
 \lambda(t) &= \frac{f(t)}{S(t)} \\
 &= \frac{1}{S(t)} f(t) \\
 &= -\frac{1}{S(t)} \frac{d}{dt} (1 - F(t)) \\
 &= \frac{1}{S(t)} \frac{d}{dt} S(t) \\
 &= -\frac{d}{dt} \ln(S(t))
 \end{aligned}$$

(b)

$$\begin{aligned}
 \lambda(t) &= \frac{f(t)}{S(t)} \\
 &= \frac{f(t)}{1 - F(t)}
 \end{aligned}$$

A função de risco $\lambda(t)$ é utilizada para explicar a distribuição de sobrevivência dos pacientes, visto que ela elucidada as mudanças da taxa de falha ao longo do tempo. Vale destacar que essa função pode apresentar diferentes formatos, podendo ser crescente, constante, decrescente e até mesmo mudar sua classificação conforme o tempo.

Assim, a partir da taxa de risco instantânea, tem-se a taxa de risco acumulada, representada por $\Lambda(t)$ e definida como:

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

No caso de ausência de dados censurados, assim como ocorreu com $\lambda(t)$, a função $\Lambda(t)$ pode ser estimada a partir do cálculo da seguinte expressão:

$$\hat{\Lambda}(t) = \sum_{x-1}^{t-1} \hat{\lambda}_t(t) \Delta t, \quad (2.5)$$

em que Δt é a amplitude do intervalo de tempo t . Assim, as funções 2.1, 2.2, 2.4 e 2.5 são estimadores para o intervalo de tempo t , não sendo instantâneas como as suas definições.

A função de risco acumulado, tal como a função de risco instantâneo, possui relação com a função de sobrevivência a partir da seguinte definição:

$$\begin{aligned} \Lambda(t) &= \int_0^t \lambda(u) du \\ &= \int_0^t \frac{-d}{du} \ln(S(u)) du \\ &= -\ln(S(t)). \end{aligned}$$

Pode-se destacar que apesar de pouca utilização na prática, a taxa de falha acumulada possui um grau de importância. Isso acontece devido ao fato de a estimação não paramétrica de $\Lambda(t)$ apresentar um estimador com excelentes propriedades, sendo útil para estimar a taxa de falha, $\lambda(t)$ quando a sua estimação é difícil de ser encontrada. Além das funções supracitadas, outras informações importantes de se estimar são: tempo médio de vida, tm e a vida média residual; $vmr(t)$ definido com o restante médio de vida esperado dado que o indivíduo sobreviveu até o momento t . No entanto, ambos podem ser encontrados através de técnicas básicas de integração.

2.2.3 Análise exploratória

A análise exploratória é uma etapa importante para iniciar o estudo dos dados. Assim, é possível resumir e extrair o máximo de informações necessárias, norteando o entendimento do estudo. Na atual pesquisa, há a presença de dados censurados e, nesse sentido, há inviabilização de algumas análises que normalmente são utilizadas como o cálculo da média e determinadas técnicas gráficas. Nessa perspectiva, é necessário realizar adaptações como, por exemplo, a utilização de gráficos de boxplot que apesar da interpretação possuir um certo grau de dificuldade devido aos dados censurados, ele tem um poder de análise descritiva válida para as técnicas de Análise de Sobrevivência.

Como supracitado, a Análise de Sobrevivência possui uma função de sobrevivência a qual representa o principal elemento para a análise descritiva. Assim, é possível es-

timá-la e, com isso, obter estatísticas de interesse como os quantis. Na abordagem não paramétrica não é realizado nenhuma suposição sobre a distribuição probabilística do tempo de sobrevivência, assim, não são estimados parâmetros estatísticos (CARVALHO et al., 2011).

Nesse contexto, vale destacar que o método não-paramétrico é uma técnica simples para estimação do tempo de sobrevivência e pode ser utilizado para auxiliar na comparação gráfica de um ou mais grupos. No entanto, apresenta algumas limitações, principalmente em relação à falta de controle direto sobre as variáveis estudadas e de utilidade quando uma variável é dependente do tempo (RENGANATHAN, 2016).

2.2.3.1 Estimador produto de Kaplan-Meier

O estimador não paramétrico de Kaplan-Meier, também conhecido como estimador produto-limite, é o indicado para estimar a função de sobrevivência $S(t)$, sendo considerado o modo mais simples para essa estimação, uma vez que é possível extrair informações sem que os dados censurados sejam desconsiderados. O estimador de Kaplan-Meier (KM), nesse sentido,

Utiliza os conceitos de independência de eventos e de probabilidade condicional para desdobrar a condição de sobreviver até o tempo t em uma sequência de elementos independentes que caracterizam a sobrevivência em cada intervalo de tempo anterior a t , cuja probabilidade é condicional aos que estão em risco em cada período (CARVALHO et al., 2011).

Assim, a expressão do estimador produto está associada com a ocorrência dos eventos independentes e, dessa maneira, a função de sobrevivência é estimada a partir da utilização do produto das probabilidades de sobrevivência até o tempo t (CARVALHO et al., 2011).

Nessa perspectiva, a curva de sobrevivência de Kaplan-Meier é conceituada como a probabilidade de o indivíduo sobreviver em um determinado momento t , dado que o tempo é dividido em pequenos intervalos. Isso acontece, visto que nesse estimador a definição de tempo está relacionado com o tempo até a ocorrência do desfecho (CARVALHO et al., 2011). No entanto, para esse tipo de análise, três critérios precisam ser contemplados (GOEL; KHANNA; KISHORE, 2010):

- 1) Os indivíduos censurados devem ter a mesma expectativa de sobrevivência que aqueles que continuam no estudo;
- 2) A probabilidade de sobrevivência é um fator atemporal para aqueles que ingressam

no estudo em períodos diferentes;

- 3) Os eventos ocorrem no momento específico.

Como supracitado, o KM está associado com a função de sobrevivência da Equação 2.2 e utiliza definições de independência e probabilidade condicional. Assim, ele fica definido da seguinte forma:

$$\begin{aligned}\hat{S}_{km}(t) &= \frac{R(t_1) - \Delta N(t_1)}{R(t_1)} X \frac{R(t_2) - \Delta N(t_2)}{R(t_2)} X \dots X \frac{R(t_m) - \Delta N(t_m)}{R(t_m)} \\ &= \prod_{j:t_j} \frac{R(t_j) - \Delta N(t_j)}{R(t_j)} \\ &= \prod_{j:t_j} 1 - \frac{\Delta N(t_j)}{R(t_j)}.\end{aligned}$$

Além disso, o KM pode ser apresentado na sua forma recursiva, como descrito abaixo:

$$\hat{S}_{km}(t_j) = \hat{S}_{km}(t_j - 1) X \frac{R(t_j) - \Delta N(t_j)}{R(t_j)},$$

em que $R(t_j)$ é o número de pessoas no grupo de risco no tempo t_j , isto é, o número de indivíduos que sobrevivem até pelo menos o tempo t_j e $\Delta N(t_j)$ o número total de eventos ocorridos em t_j . Essas definições podem ser entendidas a partir de um exemplo. No tempo $t = 0$, $R(0)$ será o número de pacientes no início do estudo, em $t = t_1$. Assim, $R(1)$ será o número de indivíduos que iniciou o estudo menos os óbitos e censurados do momento anterior.

De forma gráfica, a curva de sobrevivência do KM apresenta um formato de escada, que permanece constante até a ocorrência de um evento. Dessa forma, a “altura de cada degrau” possui dependência com a quantidade de indivíduos que apresentaram o desfecho e da quantidade de pessoas em risco.

Nesse sentido, é importante destacar algumas propriedades do KM, sendo elas (COLSIMO; GIOLO, 2014):

- 1) É um estimador não viciado para grandes amostras;
- 2) É fracamente consistente;
- 3) Converge assintoticamente para um processo gaussiano;
- 4) É um estimador de máxima verossimilhança da função de sobrevivência, $S(t)$.

As funções básicas de Análise de Sobrevivência, como já foi destacado, possuem relação entre si. Dessa forma, a estimação dessas funções pode ser realizada a partir da utilização da função de sobrevivência, que é estimada pelo método de Kaplan-Meier. Assim, para o risco acumulado, utilizando-se da expressão $\Lambda(t) = -\ln(S(t))$, tem-se:

$$\hat{\Lambda}_{km}(t) = -\ln(\hat{S}_{km}(t))$$

Com o método de estimação de Kaplan-Meier não é possível estimar o efeito das variáveis ou covariáveis influenciadoras, uma vez que se utiliza essa técnica para comparar as curvas de sobrevivência. Nesse sentido, um modelo de regressão pode ser utilizado com o objetivo de estimar os efeitos das covariáveis sobre uma variável resposta, a partir da aplicação de métodos paramétricos (CARVALHO et al., 2011).

Teste Log-Rank

Para obter a comparação das curvas de sobrevivência de maneira formal, deve-se utilizar o teste de hipótese. Nesse sentido, pode-se utilizar o teste de Mantel-Haenzel, ou log-rank, no qual compara os valores observados e esperados de cada estrato a partir da hipótese nula de que o risco é o mesmo em todos os grupos. Dessa forma, quando as curvas de sobrevivência são equivalentes entre si, tem-se que a incidência dos eventos é semelhante, logo, existe evidência de que a curva de sobrevivência é a mesma. Assim, a hipótese nula é descrita da seguinte forma:

$$H_0 : \lambda_1(t) = \lambda_2(t) = \dots = \lambda_k(t),$$

sendo k o número de estratos. Observa-se que quando ocorre a rejeição da hipótese nula, tem-se que pelo menos uma curva é, significativamente, diferente das demais em algum momento do tempo.

O teste log-rank se propõe a comparar a distribuição da ocorrência dos eventos observados em cada estrato com a distribuição que seria esperada caso a incidência fosse igual em todos os estratos. Assim, caso a distribuição observada for similar à distribuição esperada, têm-se que a curva de sobrevivência dos pacientes pertencentes aquele estrato é similar à dos pacientes em geral. Dessa maneira, essa variável em questão não afeta a sobrevivência.

No cálculo da distribuição esperada de eventos, o total de eventos no tempo t , $\Delta N(t)$, precisa ser redistribuído pelos k estratos, de forma proporcional ao tamanho amostral.

Dessa maneira, para o estrato k , têm-se:

$$E_k(T) = \Delta N(t) \frac{R_k(t)}{R(t)} \quad (2.6)$$

sendo $\Delta N(t)$ o número total de eventos observados em t , $R_k(t)$ o número de pessoas em risco no estrato k no tempo t e $R(t)$ o número total de indivíduos em risco no estudo no tempo t .

Para calcular o log-rank, realiza-se a diferença entre o número total de eventos observados e número total de eventos esperados:

$$\text{Log-rank} = \frac{(O_1 - E_1)^2}{\text{Var}(O_1 - E_1)},$$

sendo E_1 o total de eventos esperados no estrato 1 e O_1 o total de eventos observados no estrato 1.

Vale destacar que o resultado do teste para um estrato se estende para outro de forma simétrica. Além disso, a estatística de teste segue uma distribuição χ^2 com um grau de liberdade.

2.2.4 Modelo paramétrico de sobrevivência

A técnica paramétrica possui benefícios, uma vez que, quando a função de risco é estimada corretamente, o método se torna eficiente, permitindo a predição e análise (RENGANATHAN, 2016).

Como variável resposta, utiliza-se tempo de sobrevivência, uma variável contínua não negativa que pode apresentar assimetria. Assim, é necessário encontrar distribuições que contemplem essas informações. Nesse sentido, pode-se utilizar a distribuição Weibull, uma vez que a sua função de risco pode ser crescente ou decrescente, permitindo mais flexibilidade para a modelagem. Vale destacar ainda, que pode-se utilizar outras distribuições como a exponencial, Log-normal, Gama, Gama Generalizada, no entanto, essas outras distribuições não serão aprofundadas na atual pesquisa.

2.2.4.1 Distribuição Weibull

A distribuição de Weibull pode ser utilizada para analisar dados censurados, uma vez que a sua função de risco não é necessariamente constante, podendo ser monótona crescente ou decrescente. Assim, devido a isso, é uma distribuição muito utilizada em

epidemiologia. Possui relação com a função exponencial, uma vez que essa função é um caso particular da Weibull com parâmetros α e $\gamma = 1$.

As funções de densidade de probabilidade, considerando os parâmetros α e γ , de sobrevivência, de risco e de risco acumulado da distribuição Weibull são apresentadas, respectivamente, a seguir:

$$\begin{aligned} f(t) &= \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}, \\ S(t) &= \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}, \\ \lambda(t) &= \frac{\gamma}{\alpha^\gamma} t^{\gamma-1}, \\ \Lambda(t) &= -\ln(S(t)) \\ &= - \left(\frac{t}{\alpha} \right)^\gamma, \end{aligned}$$

sendo, $t \geq 0$, $\alpha > 0$ e $\gamma > 0$.

A função de risco pode ter sua forma definida pelo parâmetro γ , como demonstrado abaixo:

$\gamma < 1$ a função de risco decresce (estritamente decrescente),

$\gamma > 1$ a função de risco cresce (estritamente crescente),

$\gamma = 1$ a função de risco é constante.

Além disso, vale destacar que a esperança e variância da distribuição Weibull são calculadas da seguinte forma:

$$\begin{aligned} E(T) &= \alpha \Gamma \left[1 + \frac{1}{\gamma} \right], \\ Var(T) &= \alpha^2 \left[\Gamma \left[1 + \left(\frac{2}{\gamma} \right) \right] - \Gamma \left[1 + \left(\frac{1}{\gamma} \right) \right]^2 \right], \end{aligned}$$

sendo, $\Gamma(r) = (r-1)!$ para r inteiro

Nesse sentido, a função de verossimilhança para a distribuição Weibull de forma genérica, pode ser observada abaixo:

$$L(\gamma, \alpha) = \prod_{i=1}^n \frac{\gamma}{\alpha^\gamma} t^{\gamma-1} \exp \left\{ - \left(\frac{t}{\alpha} \right)^\gamma \right\}$$

Para uma amostra de dados de tempo de vida, associado a dados censurados, utili-

zando uma distribuição Weibull a função de verossimilhança é obtida através das seguintes expressões:

$$\begin{aligned} L(\gamma, \alpha) &= \prod_{i=1}^n \left[\frac{\gamma}{\alpha^\gamma} t_i^{\gamma-1} \exp \left\{ - \left(\frac{t_i}{\alpha} \right)^\gamma \right\} \right]^{\delta_i} \left[\exp \left\{ - \left(\frac{t_i}{\alpha} \right)^\gamma \right\} \right]^{1-\delta_i} \\ &= \prod_{i=1}^n \left[\frac{\gamma}{\alpha^\gamma} t_i^{\gamma-1} \right]^{\delta_i} \exp \left\{ - \left(\frac{t_i}{\alpha} \right)^\gamma \right\} \end{aligned}$$

sendo δ_i a variável indicadora de falha ou censura.

Dessa maneira, o logaritmo dessa função é dado pela seguinte fórmula:

$$\begin{aligned} \log(L(\gamma, \alpha)) &= \log \left[\prod_{i=1}^n \left[\frac{\gamma}{\alpha^\gamma} t_i^{\gamma-1} \right]^{\delta_i} \exp \left\{ - \left(\frac{t_i}{\alpha} \right)^\gamma \right\} \right] \\ &= \sum_{i=1}^n \delta_i \log(\gamma) - \sum_{i=1}^n \delta_i \gamma \log(\alpha) + (\gamma - 1) \sum_{i=1}^n \delta_i \log(t_i) - \alpha^{-\gamma} \sum_{i=1}^n t_i^\gamma \end{aligned}$$

2.2.5 Enfoque Bayesiano

2.2.5.1 Aplicação

A inferência Bayesiana é utilizada para lidar com situações mais complexas, apresentando uma abordagem simples para estimar parâmetros desconhecidos. Diferente do que é considerado na inferência clássica, que os parâmetros desconhecidos são quantidades fixas, no enfoque Bayesiano as distribuições de probabilidade são utilizadas para sintetizar a informação existente sobre a quantidade de interesse desconhecida e, geralmente, não observável θ .

2.2.5.2 Teorema de Bayes

Toda a informação disponível sobre a quantidade de interesse θ desconhecida e, geralmente, não observável pode ser resumida a partir de uma distribuição a priori $p(\theta)$ que resume de forma probabilística as informações disponíveis sobre θ . Utilizando a quantidade aleatória X , esse entendimento pode ser ampliado, uma vez que $p(x|\theta)$ define a relação entre X e θ . A distribuição amostral $p(x|\theta)$ define essa relação. Nesse sentido, ao observar $X = x$, as informações sobre o parâmetro θ aumentam e, para isso, é utilizado o Teorema de Bayes como regra de atualização para quantificar esse aumento da informação

sobre θ . O Teorema de Bayes, nesse sentido, é aplicado da seguinte forma:

$$p(\theta|x) = \frac{p(x, \theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(\theta|x)d\theta}$$

Observe que $\frac{1}{p(x)}$ não depende de θ , dessa forma, atua como uma constante, tendo como objetivo normalizar $p(\theta|x)$. Nessa perspectiva, a função de verossimilhança $L(\theta; x) = p(x|\theta)$, que aplicada a cada um dos valores de θ , indica quais são os valores mais plausíveis para serem utilizados. Além disso, os valores de $p(\theta)$ são combinadas, gerando a distribuição a posterior de θ , caracterizada como $p(\theta|x)$. Dessa maneira, a forma usual de representar a equação supracitada nesse item, é:

$$p(\theta|x) \propto L(\theta; x)p(\theta)$$

Essa representação significa que a distribuição a *posteriori* é proporcional a função de verossimilhança multiplicado pela distribuição a *priori*.

2.2.5.3 Distribuição a *priori* e a *posteriori*

A distribuição a *priori* $p(\theta)$ precisa caracterizar todo o conhecimento existente sobre o parâmetro θ . Se há conhecimento prévio devido à realização de experimentos anteriores ou a informações de algum especialista acerca do parâmetro desconhecido, a definição da distribuição a *priori* pode ser uma tarefa mais simples. No entanto, na maioria das vezes, pouco se sabe a respeito dos parâmetros. Dessa maneira, uma forma de elucidar a distribuição a *priori*, subjetivamente, é atribuindo a ela uma variância grande o suficiente, tornando-a vaga e pouco informativa.

Para a obtenção do modelo Bayesiano de regressão utilizou-se o software Open-Bugs (OPENBUGS, 2014), assim, utilizou-se a função de ligação de $\frac{1}{\alpha} = \exp^{-\beta t_i}$. Dessa forma, após essa aplicação, a função densidade da Weibull se apresenta da seguinte maneira:

$$f(t_i, z_i) = r e^{\beta' z_i} t_i^{r-1} \exp(-e^{\beta' z_i} t_i^r)$$

em que t_i é o tempo até o óbito com vetor de covariáveis z_i e β um vetor de coeficientes de regressão desconhecidos, os quais é preciso definir *prioris*.

Com isso, tem-se que a função de risco de linha de base é apresentada da seguinte maneira:

$$\lambda_0(t_i) = r t_i^{r-1}$$

Parametrização a partir da aplicação de $\mu_i = e^{\beta z_i}$, tem-se que:

$$t_i \sim Weibull(\gamma, \mu_i)$$

Dessa maneira, a função de verossimilhança para dados não censurados é dada pela seguinte expressão:

$$L(t_i, z_i) = \prod_{i=1}^n r \exp^{\beta' z_i} t_i^{r-1} \exp(-e^{\beta' z_i} t_i^r)$$

Dessa maneira, o logaritmo dessa função é descrita da seguinte maneira:

$$\begin{aligned} \log(L(t_i, z_i)) &= \log \left[\prod_{i=1}^n r \exp^{\beta' z_i} t_i^{r-1} \exp(-e^{\beta' z_i} t_i^r) \right] \\ &= \sum_{i=1}^n \log(n) + \sum_{i=1}^n \beta' z_i + \sum_{i=1}^n (r-1) \log(t_i) - \sum_{i=1}^n e^{\beta' z_i} t_i^r \\ &= \log(n) + \beta' \sum_{i=1}^n z_i + (r-1) \sum_{i=1}^n \log(t_i) - \sum_{i=1}^n e^{\beta' z_i} t_i^r \end{aligned}$$

Dessa forma, para definir a *priori* dos parâmetros da Weibull de acordo com a parametrização do OpenBugs, no presente estudo, no que diz respeito à distribuição dos parâmetros β , foi utilizada uma distribuição normal com média 0 e precisão 0,001, caracterizando uma *priori* não informativa, ou seja, que possui pouca influência no modelo.

Já o parâmetro de forma r para a distribuição de sobrevivência recebe uma distribuição Gama(1, 0,0001) a *priori*, que está diminuindo lentamente na linha real positiva (OPENBUGS, 2014). No entanto, utilizando esses parâmetros apresentados no Openbugs, não foi possível estabelecer um modelo. Assim, com o objetivo de encontrar os parâmetros da distribuição do r foi selecionada uma amostra aleatória de tamanho 50, aplicando-se a esperança e a variância da distribuição Weibull. Assim, a ideia foi encontrar valores para que essa distribuição apresentasse convergência. Após essa aplicação, observou-se que o parâmetro r segue uma gamma com parâmetro α igual a 1 e o parâmetro λ variando de 1 até 2. Assim, como a média ($1/\alpha$) depende dos β , têm-se *prioris* não informativas.

Para a obtenção da distribuição a *posteriori*, partiu-se do pré-suposto que se tem todas as informações sobre o experimento. Como supracitado, tem-se que a distribuição a *posteriori* é proporcional a função de verossimilhança multiplicado pela distribuição a *priori*. Assumindo que os parâmetros são independentes a *priori*, que apresentam distri-

buição normal, temos que a *posteriori* se apresenta da seguinte forma:

$$\log(p(\theta|x)) = \log \left(\frac{1}{\sigma_{\beta_0} \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\beta_0 - \mu}{\sigma_{\beta_0}} \right)^2} \times \dots \times \frac{1}{\sigma_{\beta_{13}} \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\beta_{13} - \mu}{\sigma_{\beta_{13}}} \right)^2} \right) + \\ \log(n) + \beta \sum_{i=1}^n + (r-1) \sum_{i=1}^n \log(t_i) - \sum_{i=1}^n e^{\beta z_i t_i^r} + k$$

sendo k uma constante normalizadora na escala logarítmica.

Como é possível observar, a distribuição a *posteriori* é analiticamente intratável. Nesse sentido, é necessário utilizar o método Monte Carlo via Cadeias de Markov.

Nesses casos supracitados, têm-se a situação de ausência de dados censurados. No entanto, para observações censuradas, na análise de sobrevivência, é utilizada uma distribuição Weibull, com limite inferior correspondente ao tempo de censura.

2.2.5.4 Cadeia de Markov

Em situações complexas, realizar a elaboração de contas analíticas para encontrar a distribuição a *posteriori*, pode ser inviável. Nessa perspectiva, é necessário utilizar os métodos de Monte Carlo via Cadeias de Markov (MCMC) em substituição a métodos não iterativos. A partir disso, pode-se obter amostras da distribuição a *posterior* de θ com o objetivo de calcular as estimativas amostrais da distribuição de interesse. Assim, há possibilidade de realizar inferência sobre o vetor de parâmetros desconhecidos. Para isso, é suficiente que o núcleo da expressão da função densidade ou de probabilidade dessas distribuições seja conhecido (GAMERMAN; LOPES, 2006).

Monte Carlo via Cadeias de Markov

Uma Cadeia de Markov é um exemplo de processo estocástico, de forma que as distribuições condicionais de estados futuros, X_{n+j} para $j \leq 1$ dado X_1, \dots, X_n , dependa somente de X_n e não de estados passado X_1, \dots, X_{n-1} , ou seja:

$$P(X_{n+1}|X_0, X_1, \dots, X_n) = P(X_{n+1}|X_n)$$

Conforme GAMERMAN(2006), para que o método de MCMC possa ser utilizado, as cadeias devem ser: homogêneas, irredutíveis ou aperiódicas. As cadeias homogêneas estão associadas as probabilidades de transição estacionárias, sendo constantes ao longo do tempo; as irredutíveis cada estado pode ser atingido a partir de qualquer outro estado; e as aperiódicas não possuem estados absorventes.

Na prática, valores iniciais são gerados e, à medida que o número de iterações aumenta, os valores iniciais são esquecidos, caracterizando o período de aquecimento, até que haja a convergência para a distribuição de equilíbrio da cadeia de Markov.

Amostrador de Gibbs

O Amostrador de Gibbs é uma Cadeia de Markov que não tem um mecanismo de aceitação-rejeição, ou seja, a cadeira sempre irá se mover para um novo valor. Assim, esse algoritmo irá fornecer uma amostra da distribuição conjunta de interesse através das distribuições condicionais completas.

Assim, considerando que $\pi(\theta)$ seja a distribuição que se deseja amostrar, em que θ é um vetor ou matriz que está dividido em d componentes $\theta = (\theta_1, \dots, \theta_d)$. As transições de um estado para outro ocorrem a partir das distribuições condicionais completas a *posteriori* $\pi(\theta_i|\theta_{(-i)})$, para $i = 1, \dots, d$ e $\theta_{(-i)} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$. O Amostrador de Gibbs opera da seguinte maneira:

Primeiro passo: Iniciar o contador de iterações de cadeia em $t = 0$;

Segundo passo: Definir valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})$;

Terceiro passo: A partir de $\theta_{(t-1)}$, deve-se obter um novo valor de $\theta(t)$, através da geração sucessiva de valores.

Quarto passo: Incrementar em uma unidade o contador t e voltar para o passo 2 até obterá convergência.

A convergência da Cadeia de Markov acontece após a fase de aquecimento e é identificada, facilmente, analisando o gráfico da Cadeia. Os valores resultantes formam a amostra $\pi(\theta)$.

Algoritmo de Metropolis-Hastings

O Algoritmo de Metropolis-Hasting utiliza a ideia dos métodos de rejeição em que um valor é gerado e aceito com certa probabilidade, a partir de uma distribuição auxiliar. Esse método assegura a convergência da cadeia para uma determinada distribuição de equilíbrio, que, sob a abordagem Bayesiana, pode ser a distribuição a *posteriori*. Considerando uma cadeia no estado θ e um novo valor θ' fornecido por uma distribuição proposta $q(\cdot|\theta)$. Note que essa distribuição pode depender do atual estado da cadeia. Assim, a probabilidade de aceitação do valor gerado θ' é:

$$\alpha(\theta, \theta') = \min\left(1, \frac{\pi(\theta')q(\theta, \theta')}{\pi(\theta)q(\theta', \theta)}\right)$$

em que π é a distribuição de interesse.

Um fator importante para as aplicações com a abordagem Bayesiana, em que não há conhecimento amplo sobre a distribuição a *posteriori*, é que é preciso conhecer π de forma parcial, a não ser que seja uma constante, uma vez que a probabilidade de $\alpha(\theta, \theta')$ é inalterável. Nessa perspectiva, o Algoritmo de Metropolis-Hastings, na prática, procede da seguinte maneira:

Primeiro passo: Iniciar o contador de iterações em $t=0$ e definir o valor inicial $\theta(0)$;

Segundo passo: Gerar um novo valor de θ' da proposta $q(\cdot|\theta)$;

Terceiro passo: Calcular a probabilidade de aceitação de $\alpha(\theta, \theta')$ e gerar $\mu \sim U(0, 1)$;

Quarto passo: Aceitar o novo valor de $\mu \leq \alpha$ e fazer $\theta(t+1) = \theta'$. Caso contrário, rejeitar e fazer $\theta(t+1) = \theta$;

Quinto passo: Acrescentar ao contador t uma unidade e voltar para o segundo passo, até que a convergência seja atingida.

A escolha da distribuição $q(\cdot|.)$, pode ser feita de forma arbitrária, entretanto, é fundamental para o funcionamento do algoritmo. Na abordagem Bayesiana, normalmente, as *prioris* são utilizadas como distribuição proposta e a *posteriori* é a distribuição de interesse, ou seja, $\pi = p(\theta|x)$ e a probabilidade de aceitação assume a seguinte forma:

$$\alpha(\theta, \theta') = \min\left(1, \frac{p(x|\theta')p(\theta')q(\theta|\theta')}{p(x|\theta)p(\theta)q(\theta'|\theta)}\right)$$

Assim, no presente estudo, para a obtenção de amostras das *posteriori* dos parâmetros, foi utilizado o software Open Bugs, que permite obter essas amostras através do algoritmo de Amostrador de Gibbs de forma automática. No software, foram simuladas 10000 rodadas de aquecimento, conhecido como *burn-in* e, posteriormente, mais 10000 amostras foram geradas para a obtenção das distribuições a *posteriori* dos parâmetros de interesse. O script do software OpenBugs se encontra no Apêndice 6.

2.2.5.5 Seleção do modelo

Para analisar a qualidade do ajuste de um modelo de regressão foi utilizada a soma dos quadrados dos resíduos. Note que quanto menor o SQResp, melhor é o ajuste do modelo.

$$SQResp = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}$$

3 Resultados

3.1 Análise descritiva

Foram acompanhadas 2181 pacientes, desde o primeiro registro em um CR até o óbito ou remissão, sendo que dessas 89 foram a óbito devido à NTG, o que representa 4,08% de alcance do desfecho. Em contrapartida, foi observado 2092 dados censurados, o que representa 95,92%. Além disso, vale destacar que as pacientes que faleceram devido à causa diferente da NTG foram classificadas como truncamento e os seus dados não foram considerados no estudo.

A Tabela 2 elucida as características clínicas dos dados categóricos das pacientes estudadas. A partir disso, é possível observar que a maioria apresentou valores de hCG menores que 100.000 UI/L, idade entre 20 e 40 anos, ausência de dados histológicos (o que significa diagnóstico de NTG por dosagens de hCG em ascensão ou platô), sendo a origem relacionada à mola (pós-molar), com estadiamento de I e escore prognóstico OMS menor ou igual a 6, sem metástase e com tratamento quimioterápico inicial com agente único. Além disso, a maior parte da amostra realizou o tratamento no centro de referência (CR).

Tabela 2: Características clínicas das pacientes estudadas

Característica	Óbitos (%)	Não óbitos (%)	Total (%)
hCG < 100.000 UI/L			
Sim	47 (52,81)	1.735 (82,93)	1.782 (81,71)
Não	42 (47,19)	357 (17,07)	399 (18,29)
Idade			
≤ 19 anos	12 (13,48)	379 (18,12)	391 (17,93)
20 a 39 anos	64 (71,91)	1.470 (70,27)	1.534 (70,33)
≥40 anos	13 (14,61)	243 (11,62)	256 (11,74)
Histologia			
Sem histologia	44 (49,44)	1.930 (92,26)	1.974 (90,51)
Mola invasora	3 (3,37)	20 (0,96)	23 (1,05)
Coriocarcinoma	42 (47,19)	142 (6,79)	184 (8,44)
Origem			
Pós-molar	57 (64,04)	1.927 (92,11)	1.984 (90,97)
Aborto	9 (10,11)	96 (4,59)	105 (4,81)
Gestação ectópica	2 (2,25)	6 (0,29)	8 (0,37)
Gestação a termo ou pré-termo	21 (23,60)	63 (3,01)	84 (3,85)
Estadiamento			
I	9 (10,11)	1.753 (83,80)	1.763 (80,79)
II	9 (10,11)	56 (2,68)	65 (2,98)
III	44 (49,44)	257 (12,28)	301 (13,80)
IV	27 (30,34)	26 (1,24)	53 (2,43)
Metástase			
Sim	82 (92,13)	334 (15,97)	416 (19,07)
Não	7 (7,87)	1.758 (84,03)	1.765 (80,93)
Escore prognóstico OMS			
≤ 6	26 (29,21)	1.877 (89,72)	1.903 (87,25)
≥ 7	63 (70,79)	215 (10,28)	278 (12,75)
Tratamento no Centro de Referência			
Sim	60 (67,42)	1.969 (94,12)	2.029 (93,03)
Não	29 (32,58)	123 (5,88)	152 (6,97)
Tratamento inicial			
Único	25 (28,09)	1.832 (87,57)	1.857 (85,14)
Múltiplo	46 (51,69)	251 (12,00)	297 (13,62)
Não quimioterápico	18 (20,22)	9 (0,43)	27 (1,24)

A distribuição das variáveis numéricas estudadas entre óbitos e não óbitos, encontra-se na Figura 1. Apesar de dados associados ao tempo até o óbitos não serem muito bem representados por gráfico boxplot, em relação à sobrevida observa-se que há indícios de diferença, pois as pacientes que foram a óbito costumam falecer em tempos menores, permanecendo menos tempo no estudo do que as que não foram a óbito. Sobre a idade, graficamente, parece não haver diferença. Em relação à distribuição do Escore prognóstico OMS entre óbitos e não óbitos, pode-se observar que, graficamente, o escore prognóstico

das que faleceram é maior que das que não faleceram, apesar dos maiores valores serem observados por não óbitos (outliers).

Sobre a distribuição do intervalo entre o término da gravidez e o início do tratamento, parece que as mulheres que foram a óbito possuem um intervalo de tempo maior, o que pode estar relacionado com demora no diagnóstico. Vale destacar que esse fator é um indicativo para óbito.

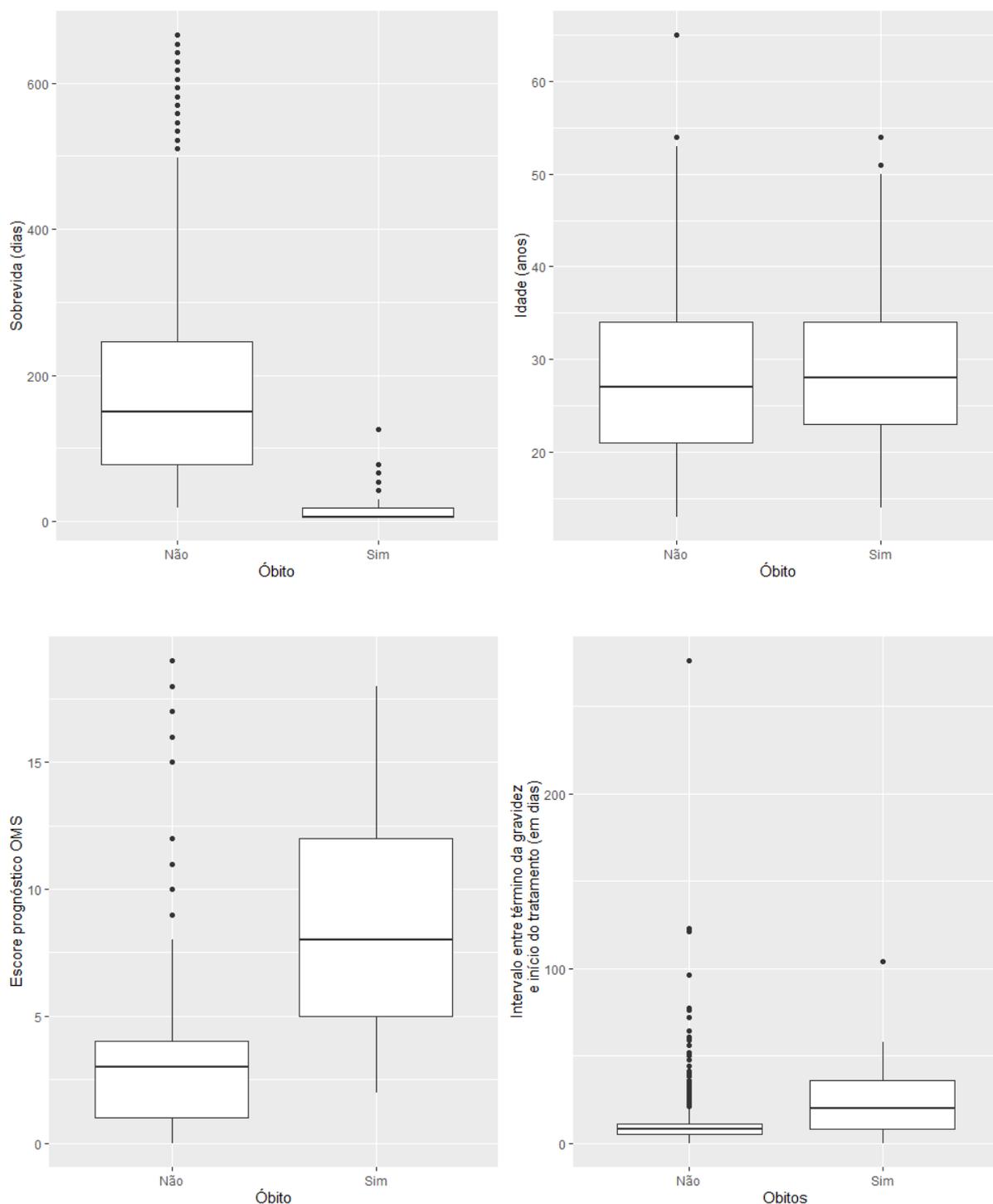


Figura 1: Distribuição dos dados entre os óbitos e não óbitos

O gráfico da função de sobrevivência estimada pelo método de Kaplan-Meier pode ser observado na Figura 2. Com isso, podemos notar, por exemplo, que no $t=18$, a probabilidade de sobreviver 18 dias, sabendo que a paciente não veio a óbito em pelo

menos 6 dias, é de 0,966.

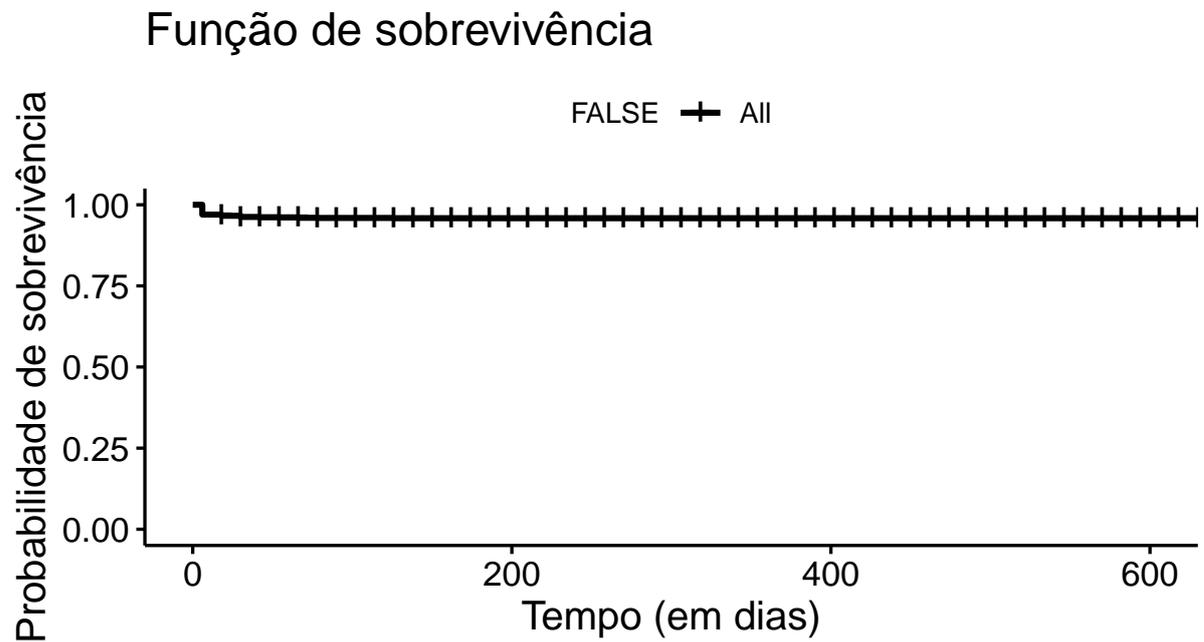


Figura 2: Função de sobrevivência estimada pelo método de Kaplan-Meier

Os gráficos que elucidam as funções de sobrevivência estimada pelo método de Kaplan-Meier das variáveis avaliadas podem ser observado na Figura 3.

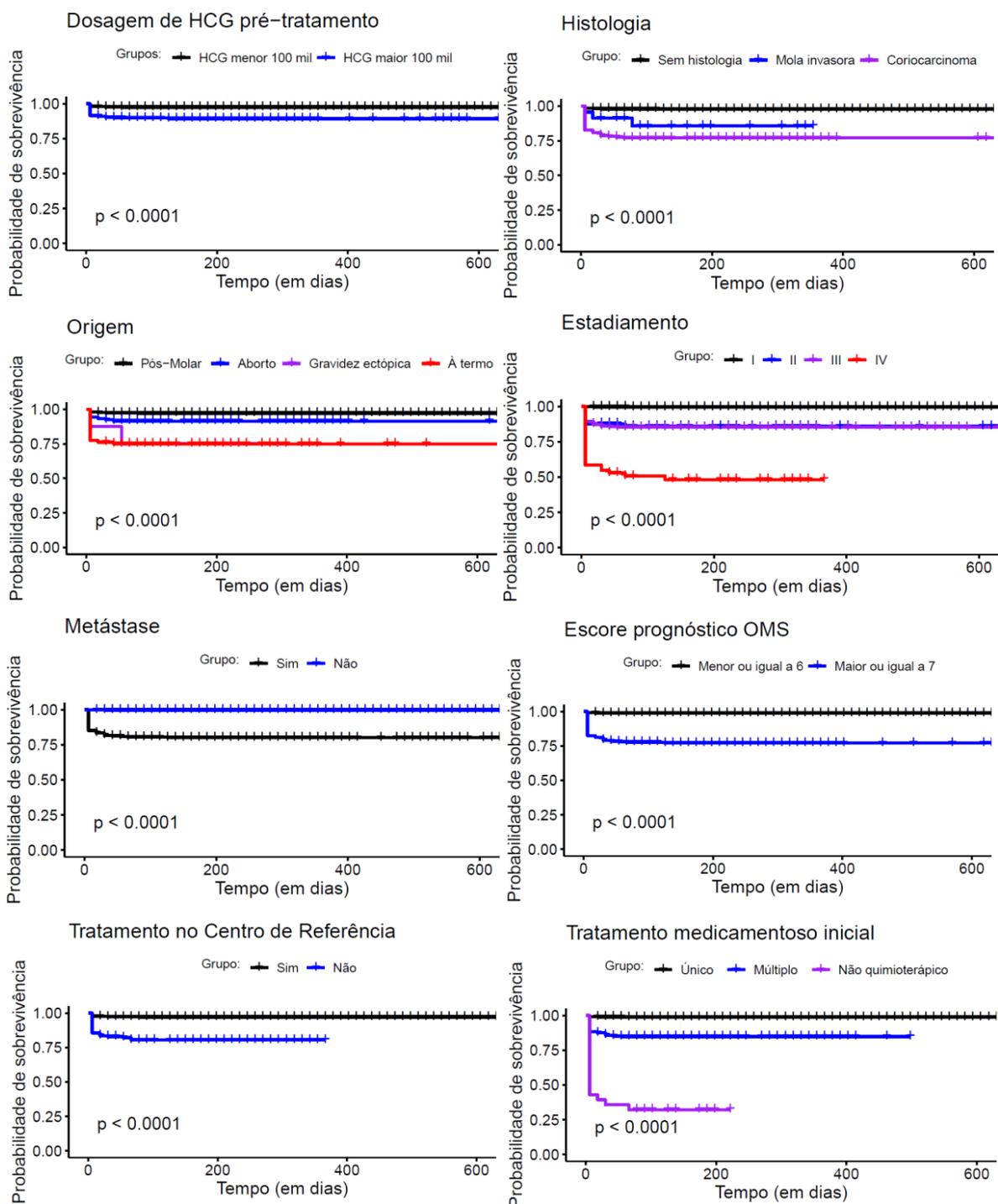


Figura 3: Função de sobrevivência estimada pelo método de Kaplan-Meier das variáveis avaliadas

Observa-se que o tempo de sobrevida é menor nas pacientes com hGC maior que 100 mil (UI/L) ($p < 0,0001$). Com isso, pode-se notar, por exemplo, que nesse grupo $t=18$, a probabilidade de sobreviver 18 dias, sabendo que a paciente não veio a óbito em pelo

menos 6 dias, é de 0,910. Por outro lado, as pacientes com valores de hCG menores que 100 mil (UI/L), no mesmo período de tempo avaliado, tem probabilidade de 0,979. De fato, esse achado já era esperado, visto que a neoplasia trofoblástica gestacional possui como marcador biológico a concentração de hCG.

Em relação à histologia, observa-se que houve rejeição da hipótese nula, ou seja, pelo menos em um momento t uma curva é diferente das demais ($p < 0,0001$). Com isso, pode-se notar, por exemplo, que nesse grupo $t=18$, a probabilidade de sobreviver 18 dias, sabendo que a paciente não veio a óbito em pelo menos 6 dias, é de 0,826 nos casos com histologia de coriocarcinoma. Por outro lado, as pacientes com mola invasiva, no mesmo período de tempo avaliado, tem probabilidade de 0,957 e as sem histologia 0,983. De fato, esse achado já era esperado, visto que a mola invasora é considerada um tumor de menor gravidade do que o coriocarcinoma, que é um tumor maligno que apresenta relação com óbito, inclusive em função da maior associação com metástases.

Sobre a origem da NTG, observa-se que houve rejeição da hipótese nula, ou seja, pelo menos em um momento t uma curva é diferente das demais ($p < 0,0001$). Com isso, pode-se notar, por exemplo, que nesse grupo em $t=18$, a probabilidade de sobreviver 18 dias, sabendo que a paciente não veio a óbito em pelo menos 6 dias, é de 0,762 nos casos cuja gestação antecedente foi a termo. Por outro lado, no pós-molar, no mesmo período de tempo avaliado, tem probabilidade de 0,980, as com aborto 0,933. Nos casos de gestação ectópica, a probabilidade de sobreviver 54 dias, sabendo que a paciente não veio a óbito em pelo menos 6 dias, é de 0,750.

No que diz respeito ao estadiamento da doença, observa-se que houve rejeição da hipótese nula, ou seja, pelo menos em um momento t uma curva é diferente das demais ($p < 0,0001$). Com isso, pode-se notar, por exemplo, que nesse grupo $t=18$, a probabilidade de sobreviver 18 dias, sabendo que a paciente não veio a óbito em pelo menos 6 dias, é de 0,826. Esse achado era esperado, visto que quanto o maior estadiamento maior é o grau de severidade da doença.

Sobre a presença ou não de metástase, observa-se que o tempo de sobrevida é menor nas pacientes com metástase ($p < 0,0001$). Com isso, pode-se notar, por exemplo, que nesse grupo com $t=18$, a probabilidade de sobreviver 18 dias, sabendo que a paciente não veio a óbito em pelo menos 6 dias, é de 0,851. Por outro lado, as pacientes sem metástase, no mesmo período de tempo avaliado, tem probabilidade de 0,998. Esse achado também foi esperado, visto que a metástase contribui para um pior prognóstico, estando diretamente associada ao óbito.

Avaliando a função de sobrevivência estimada pelo método de Kaplan-Meier conforme o escore prognóstico da OMS, observa-se que o tempo de sobrevida é menor nas pacientes com escore maior ou igual a 7 ($p < 0,0001$). Com isso, pode-se notar, por exemplo, que nesse grupo com $t=18$, a probabilidade de sobreviver 18 dias, sabendo que a paciente não veio a óbito em pelo menos 6 dias, é de 0,851. Esse fator também condiz com o que está relatado na literatura, de fato, o escore prognóstico preve o potencial de desenvolvimento de resistência ao tratamento quimioterápico com agente único, que quando maior que 7 demonstra que a paciente possui alto risco e necessita de quimioterapia combinado, o que prejudica o prognóstico e contribui para os casos de óbito.

De acordo com o local de tratamento (dentro ou fora do centro de referência), Observa-se que o tempo de sobrevida é menor nas pacientes tratadas fora dos centros de referência ($p < 0,0001$). Com isso, pode-se notar, por exemplo, que nesse grupo com $t=18$, a probabilidade de sobreviver 18 dias, sabendo que a paciente não veio a óbito em pelo menos 6 dias, é de 0,855. Por outro lado, as pacientes tratadas no centro de referência, no mesmo período de tempo avaliado, tem probabilidade de 0,978. Esse achado está de acordo com a literatura, pois no tratamento em centros de referência as pacientes são atendidas de forma mais específica, podendo diminuir a mortalidade e promovendo maior taxa de sobrevivência, como observado no atual estudo.

Sobre o tratamento inicial utilizado, observa-se que houve rejeição da hipótese nula, ou seja, pelo menos em um momento t uma curva é diferente das demais ($p < 0,0001$). Com isso, pode-se notar, por exemplo, que nesse grupo com $t=18$, a probabilidade de sobreviver 18 dias, sabendo que a paciente não veio a óbito em pelo menos 6 dias, é de 0,429. Clinicamente, este achado é esperado, pois o tratamento adequado da NTG é com quimioterapia, e, desta forma, as pacientes que não receberam quimioterápicos, em alguns casos por não chegarem a um CR a tempo ou por estarem muito debilitadas, não puderam ser adequadamente tratadas

As tabelas com a função de sobrevivência das variáveis analisadas podem ser observadas no Apêndice 5.

3.2 Modelo de regressão sob enfoque Bayesiano

Para a geração do modelo, foram consideradas as variáveis: classificação quanto à dosagem de hCG < 100 mil (UI/L), histologia, origem, escore prognóstico OMS, metástase, classificação do escore prognóstico, tratamento no CR, tratamento inicial e intervalo entre

término da gravidez e início do tratamento. Sendo a sobrevida (em dias) a variável de interesse.

Vale destacar que a variável concentração de hCG (UI/L) pré-tratamento não foi utilizada, visto que quando incluída no modelo, não foi possível estimar os parâmetros. Dessa forma, utilizou-se apenas a classificação quanto à dosagem de hCG, que também apresenta relevância clínica.

Para observar quais variáveis explicativas seriam significativa, em um primeiro momento, avaliou-se a associação de cada dado supracitado com a variável de interesse. Depois, outras variáveis significativas foram incorporadas, até a obtenção do modelo final. Para avaliar se a variável apresentava associação significativa, foi observado se o intervalo de credibilidade (IC) possuía ou não o 1. Se o 1 não estiver contido nesse intervalo, há influência significativa.

Avaliando o modelo Bayesiano, utilizando cada variável explicativa de forma isolada com a variável resposta, foram formados 10 modelos individuais. Pode-se observar que houve associação significativa nas variáveis: sem histologia, histologia de mola invasora e de coriocarcinoma, origem pós-molar, aborto, origem associada à gestação ectópica, à termo ou pré-termo, escore prognóstico OMS, escore OMS menor ou igual a 6, tratamento no CR e tratamento inicial único, múltiplo e não quimioterápico (Tabela 3). Assim, podemos observar que a histologia, origem, escore prognóstico OMS e tratamento inicial influenciam diminuindo o tempo de sobrevivência, enquanto escore OMS menor ou igual a 6 e tratamento no CR as outras variáveis significativas, aumentam. Note que o efeito é a exponencial dos coeficientes β da regressão, o que foi apresentado na Seção 2.2.5.3, logo, o que está apresentado não é o valor dos β e sim a exponencial desse β , sendo representado como o efeito que esse parâmetro tem na sobrevida das pacientes. Nesse sentido, vale destacar que quando o efeito é igual a 1 implica em não significância, uma vez que para uma variável binária, esse efeito igual a 1 significa que ambos os fatores dessa variável tem o mesmo efeito no tempo até o óbito. Dessa forma, o exponencial de β é igual a 1.

Tabela 3: Efeito das variáveis no tempo até o óbito, considerando o modelo simples

Modelo	Variável	Efeito	Intervalo de Credibilidade
Modelo 1	hCG < 100 mil (UI/L)	0,921	[0,843; 1,008]
Modelo 2*	Sem histologia	$0,135 \times 10^{-6}$	$[0,132 \times 10^{-6}; 0,155 \times 10^{-6}]$
	Mola invasora	$0,111 \times 10^{-6}$	$[0,086 \times 10^{-6}; 0,163 \times 10^{-6}]$
	Coriocarcinoma	$0,113 \times 10^{-6}$	$[0,094 \times 10^{-6}; 0,136 \times 10^{-6}]$
Modelo 3*	Pós-molar	0,040	[0,038; 0,041]
	Aborto	0,040	[0,035; 0,047]
	Gestação ectópica	0,035	[0,021; 0,059]
	Gestação à termo ou pré-termo	0,031	[0,028; 0,038]
Modelo 4	Estadiamento	0,998	[0,981; 1,024]
Modelo 5	Metástase	1,001	[0,942; 1,061]
Modelo 6	Escore prognóstico OMS	0,965	[0,956; 0,965]
Modelo 7	Escore OMS ≤ 6	1,692	[1,560; 1,815]
Modelo 8	Tratamento no CR	1,727	[1,605; 1,821]
Modelo 9	Único	$0,621 \times 10^{-13}$	$[0,573 \times 10^{-13}; 0,673 \times 10^{-13}]$
	Múltiplo	$0,446 \times 10^{-13}$	$[0,408 \times 10^{-13}; 0,514 \times 10^{-13}]$
	Não quimioterápico	$0,209 \times 10^{-13}$	$[0,156 \times 10^{-13}; 0,260 \times 10^{-13}]$
Modelo 10	Intervalo entre término da gravidez e início do tratamento	0,999	[0,995; 1,001]

*Os modelos 2, 3 e 9 são compostos pelas variáveis categóricas associadas à histologia (sem histologia, mola invasora e coriocarcinoma), à origem (pós-molar, aborto, gestação ectópica e à gestação à termo ou pré-termo) e ao tratamento inicial (único, múltiplo e não quimioterápico). Assim, foram testadas juntas como variáveis dummy.

O intervalo de credibilidade do hCG menor que 100 mil (UI/L), estadiamento, metástase e intervalo entre término da gestação e início do tratamento contém o valor 1. Dessa forma, essas variáveis não possuem associação significativa, podendo ser retiradas do modelo completo. Como foi utilizada uma abordagem Bayesiana para estimar o tempo até o óbito, pode-se extrair mais informações. Com isso, avaliando o modelo simples, apesar de não ter significância, o hCG menor que 100 mil (UI/L) diminui em 7,9% o tempo até o óbito. Em relação as outras variáveis não significativas, a probabilidade do efeito foi bem pequena, por volta de 0,001 ou 0,1

Alguns parâmetros, que quando testados separadamente apresentaram associação significativa, quando incluído no modelo múltiplo deixou de possuir tal significância, como

os relacionados à origem (pós-molar, aborto, gestação ectópica e gestação à termo ou pré-termo).

Além disso, foi necessário realizar alguns ajustes e observações a fim de selecionar o modelo mais robusto, sendo sempre avaliada a convergência dos parâmetros. Nesse sentido, o modelo final conteve os parâmetros: sem histologia, escore OMS menor ou igual a 6 e tratamento inicial único. Assim, antes de verificar o ajuste referente à estimação dos parâmetros do modelo com base no método MCMC, avaliou-se a convergência nas cadeias geradas, como realizado em todos os modelos estudados. Dessa forma, esse item foi cumprido e pode ser observado na Figura 4.

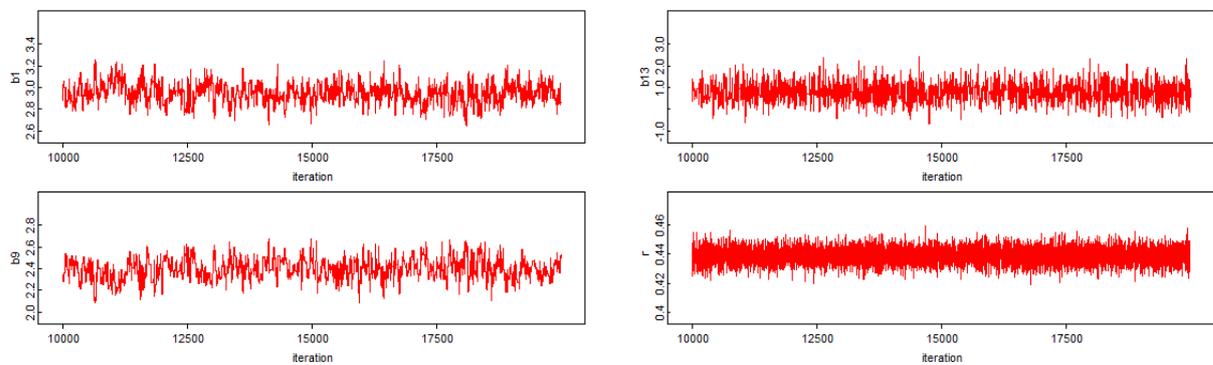


Figura 4: Monitoramento da Convergência dos Parâmetros do Modelo Completo

A Figura 5 demonstra a densidade *a posteriori* dos parâmetros avaliados.

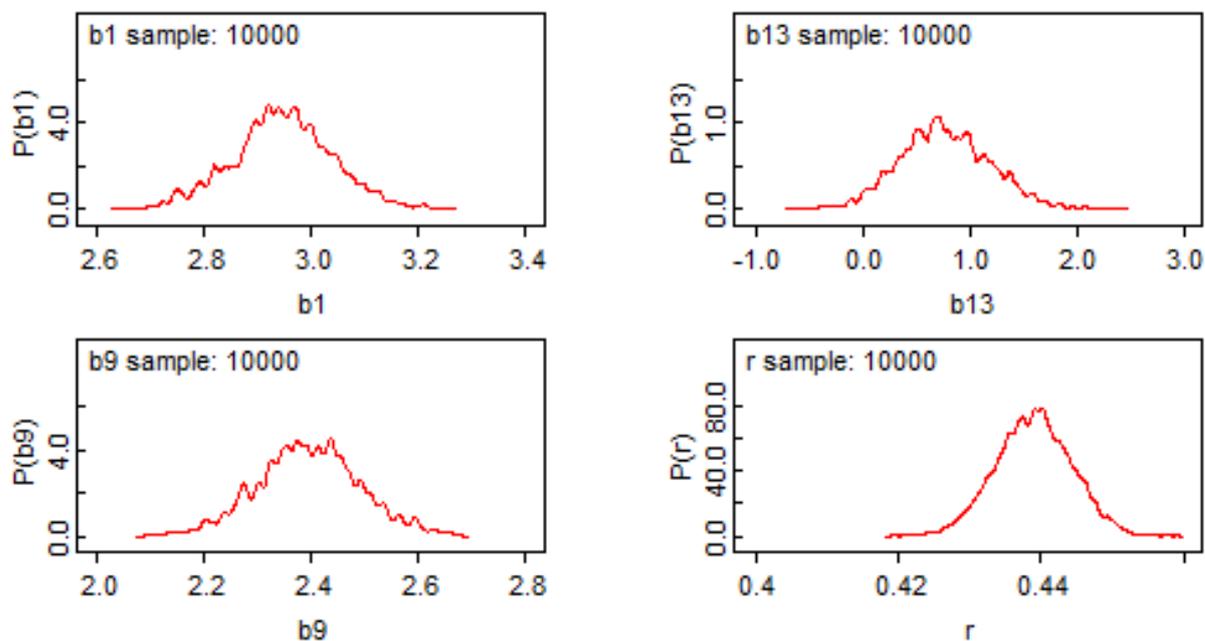


Figura 5: Densidade a *posteriori* dos parâmetros do Modelo Completo

A partir da observação da Figura 4, pode-se notar, a partir de uma inspeção visual, indícios de que os parâmetros atingiram convergência, uma vez que as cadeias estão contidas dentro de um intervalo. Além disso, a Figura 5 elucidada que as *posteriori* possuem formato similar aos das *priori*.

Assim, após avaliar as suposições, o modelo múltiplo foi gerado e pode ser observado na tabela 4 e o script se encontra no Apêndice 7.

Tabela 4: Estimativas do parâmetros do Modelo Completo

Variável	Efeito	Intervalo de credibilidade
Sem histologia	19,317	[15,975; 22,966]
Escore OMS ≤ 6	10,881	[9,272; 13,223]
Tratamento único	2,216	[1,045; 5,033]

Assim, observou-se que sem histologia contribui para um aumento de 19,317 vezes do tempo até o óbito em relação a quem tem histologia; um escore OMS menor ou igual a 6 promove aumento de 10,881 vezes desse mesmo tempo em relação a quem tem escore OMS maior ou igual a 7; e o tratamento inicial único aumenta em 2,216 vezes desse mesmo tempo em relação a quem não fez tratamento único.

Esses achados corroboram a literatura, de fato uma paciente cujo diagnóstico de NTG

foi feito com base nas dosagens, ascendentes ou em platô, do hCG (sem histologia, portanto) possivelmente apresenta caso mais leve da doença, o que contribui para o aumento do tempo de sobrevida. Além disso, o escore OMS preve o potencial de desenvolvimento de resistência ao tratamento quimioterápico com agente único, que é considerado menos severo. Assim, um valor menor ou igual a 6 elucida uma paciente com baixo risco que necessita de um tratamento menos complicado, o que contribui para o aumento do tempo até o óbito.

Com a finalidade de avaliar a qualidade do ajuste dos modelos, foi calculado o SQResp (soma dos quadrados dos resíduos) para cada um dos modelos com as variáveis incluídas no modelo final e para o modelo final. Esse resultado pode ser observado na Tabela 5. Notou-se que o modelo final apresentou menor valor de SQResp, elucidando ser o melhor ajuste.

Tabela 5: Avaliação da qualidade do modelo completo e dos modelos com apenas uma variável explicativa

Modelo	Variáveis	SQResp
Modelo final	Sem histologia, escore OMS ≤ 6 e tratamento único	50270
Modelo 11	Sem histologia	50360
Modelo 12	Escore OMS ≤ 6	50380
Modelo 13	Tratamento único	52010

4 Conclusão

O presente estudo apresentou como objetivo a utilização do enfoque Bayesiano como um modelo paramétrico de análise de sobrevivência para o tempo até o óbito devido à NTG. Assim, as análises foram realizadas para observar os fatores associados ao tempo até a ocorrência de um determinado evento, que neste caso é óbito por NTG.

Como a amostra é considerada grande foi utilizado o estimador de Kaplan-Meier para estimar a curva de sobrevivência. A partir do resultado destas curvas foram realizados testes de log rank, cuja hipótese nula é de igualdade entre as curvas de sobrevivência. Observou-se, que todas as variáveis estudadas apresentaram valor de p menor que 0,05, ou seja, houve rejeição da hipótese nula. Assim, pelo menos um momento t uma curva de risco diferente das demais.

Conforme os resultados encontrados e elucidados no presente estudo, conclui-se que o modelo Bayesiano pode ser considerado consistente para modelar o tempo até o óbito de mulheres com neoplasia trofoblástica gestacional. A partir do método de MCMC, observou-se cadeias bem comportadas e as amostras geradas são oriundas das distribuições *a posteriori*. Além disso, o modelo Bayesiano ocasionou uma interpretação fácil para um tipo de modelagem altamente complexa.

Em um primeiro momento, quando as variáveis foram testadas de forma individual, observou-se associação significativa nos fatores: sem histologia, histologia de mola invasora e de coriocarcinoma, origem pós-molar, aborto, origem associada à gestação ectópica, à termo ou pré-termo, escore prognóstico OMS, escore $\text{OMS} \leq 6$, tratamento no CR e tratamento inicial único, múltiplo e não quimioterápico, sendo elucidado que a histologia, origem, escore prognóstico OMS e tratamento inicial influenciam diminuindo o tempo de sobrevivência, enquanto escore ≤ 6 e tratamento no CR as outras variáveis significativas, aumentam. Após realização de ajustes, o modelo final conteve as variável sem histologia, escore $\text{OMS} \leq 6$ e tratamento inicial único, todas contribuindo para aumentar o tempo de sobrevida.

Nesse sentido, é importante obter dados sobre as pacientes com NTG, uma vez que o conhecimento sobre esta doença promove maior divulgação científica sobre o assunto e, com isso, contribui para um fácil e rápido diagnóstico o que corrobora o aumento do tempo de sobrevida. No entanto, vale destacar que é importante considerar a histologia, o escore OMS e o tratamento inicial, visto que esses fatores contribuem para o aumento do tempo de sobrevida.

A partir dessas considerações, pode-se destacar a importância de investir em modelos para avaliar o tempo até o óbitos de pacientes, sendo um fator relevante na área da saúde e em outras áreas. Assim, é possível refletir de forma eficiente uma situação altamente complexa a partir de uma interpretação mais simples.

Referências

- BERKOWITZ, R. S.; GOLDSTEIN, D. P. Current management of gestational trophoblastic diseases. *Gynecologic Oncology*, v. 112, p. 654–662, 2009.
- BRAGA, A. et al. Doença trofoblástica gestacional. protocolo febrasgo. *Obstetrícia*, 2018.
- CARDOSO, F. Letalidade da neoplasia trofoblástica gestacional em mulheres brasileiras. 2017.
- CARVALHO, L. V. P. de. Modelos de análise de sobrevivência aplicados a dados de neoplasia trofoblástica gestacional. 2019.
- CARVALHO, M. S. et al. *Análise de Sobrevivência - Teoria e aplicações em saúde*. 2. ed. [S.l.]: Editora Fiocruz, 2011.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de Sobrevivência Aplicada*. 2. ed. [S.l.]: Editora Edgard Blücher Ltda, 2014.
- EBELING, C. E. *An introduction to reliability and maintainability engineering*. [S.l.]: McGraw-Hill, 1997.
- FEBRASGO. Doença trofoblástica gestacional: Uma história para ser contada. 2017. Disponível em: <https://www.febrasgo.org.br/pt/noticias/item/204-doenca-trofoblastica-gestacional-uma-historia-para-ser-contada>.
- FREITAS, F. et al. Gestational trophoblastic neoplasia lethality among brazilian women: A retrospective national cohort study. *Gynecologic Oncology*, v. 1, 2020.
- GAMERMAN, D.; LOPES, H. *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. [S.l.]: Chapman & Hall/CRC, 2006.
- GOEL, M. K.; KHANNA, P.; KISHORE, J. Understanding survival analysis: Kaplan-meier estimate. *International Journal of Ayurveda Research*, v. 1, 2010.
- KLEINBAUM, D. G.; KLEIN, M. *Survival Analysis: A Self-Learning Text*. [S.l.]: Springer-Verlag, 1996.
- OPENBUGS. Openbugs examples. 2014. Disponível em: https://www.mrc-bsu.cam.ac.uk/wp-content/uploads/WinBUGS_Vol1.pdf.
- PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. *Estatística Bayesiana*. Av. de Berna, Lisboa: Fundação Calouste Gulbenkian, 2003.
- RENGANATHAN, V. Overview of frequentist and bayesian approach to survival analysis. *Applied Medical Informatics*, p. 25–38, 2016.

5 Apêndice A - Tabelas com a função de sobrevivência das variáveis analisadas

Tabela 6: Função de sobrevivência estimada pelo método de Kaplan-Meier

Tempo, em dias	Nº de pessoas sob risco	Nº de eventos	Estimador KM	IC (95%)
6	2181	66	0,970	0,963-0,977
18	2115	8	0,966	0,959-0,974
30	2101	8	0,962	0,954-0,970
42	2024	2	0,961	0,953-0,970
54	1894	1	0,961	0,953-0,969
66	1754	2	0,960	0,952-0,968
78	1661	1	0,959	0,951-0,968
126	1297	1	0,959	0,950-0,967.

Tabela 7: Função de sobrevivência estimada pelo método de Kaplan-Meier conforme hCG
Pacientes com hGC menor que 100 mil (U/I)

Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	1782	32	0,982
18	1750	6	0,979
30	1739	5	0,976
42	1672	1	0,975
54	1559	1	0,975
66	1448	1	0,974
78	1370	1	0,973

Pacientes com hGC maior que 100 mil (U/I)			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	399	34	0,915
18	365	2	0,910
30	362	3	0,902
42	352	1	0,900
66	306	1	0,897
126	220	1	0,893

Tabela 8: Função de sobrevivência estimada pelo método de Kaplan-Meier conforme histologia

Pacientes sem histologia			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	1974	33	0,983
18	1941	4	0,981
30	1931	4	0,979
42	1864	1	0,979
66	1614	1	0,978
126	1202	1	0,977
Pacientes com mola invasiva			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	23	1	0,957
18	22	1	0,913
78	16	1	0,856
Pacientes com coriocarcinoma			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	184	32	0,826
18	152	3	0,810
30	149	4	0,788
42	140	1	0,782
54	134	1	0,777
66	122	1	0,770

Tabela 9: Função de sobrevivência estimada pelo método de Kaplan-Meier conforme origem da NTG

Paciente pós-molar			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	1984	40	0,980
18	1944	6	0,977
30	1932	7	0,973
66	1613	2	0,972
78	1525	1	0,971
126	1191	1	0,971
Pacientes com aborto			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	105	6	0,943
18	99	1	0,933
30	98	1	0,924
42	96	1	0,914
Pacientes com gestação ectópica			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	8	1	0,875
54	7	1	0,750
Pacientes com gravidez à termo			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	84	19	0,774
18	65	1	0,762
42	63	1	0,750

Tabela 10: Função de sobrevivência estimada pelo método de Kaplan-Meier conforme estadiamento

Pacientes com estadiamento I			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	1765	4	0,998
18	1758	2	0,997
30	1753	2	0,995
78	1379	1	0,995
Pacientes com estadiamento II			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	65	8	0,877
66	51	1	0,860
Pacientes com estadiamento III			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	301	32	0,894
18	269	6	0,874
30	260	4	0,860
42	248	1	0,857
54	238	1	0,853
Pacientes com estadiamento IV			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	53	22	0,585
30	31	2	0,547
42	29	1	0,528
66	24	1	0,506
126	21	1	0,482

Tabela 11: Função de sobrevivência estimada pelo método de Kaplan-Meier conforme presença ou não de metástase

Pacientes sem metástase			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	1765	4	0,998
18	1761	1	0,997
30	1757	1	0,997
78	1381	1	0,996
Pacientes com metástase			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	416	62	0,851
18	354	7	0,834
30	344	7	0,817
42	329	2	0,812
54	315	1	0,810
66	296	2	0,804
126	210	1	0,800

Tabela 12: Função de sobrevivência estimada pelo método de Kaplan-Meier conforme escore prognóstico OMS

Pacientes com escore prognóstico ≤ 6			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	1903	17	0,991
18	1886	5	0,988
30	1875	2	0,987
66	1568	1	0,987
78	1485	1	0,986
Pacientes com escore prognóstico ≥ 7			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	278	49	0,824
18	229	3	0,813
30	226	6	0,791
42	215	2	0,784
54	202	1	0,780
66	186	1	0,776
126	128	1	0,770

Tabela 13: Função de sobrevivência estimada pelo método de Kaplan-Meier conforme local de tratamento

Pacientes tratadas no centro de referência			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	2029	44	0,978
18	1985	5	0,976
30	1975	7	0,972
42	1907	2	0,971
78	1569	1	0,971
126	1221	1	0,970
Pacientes tratadas fora do centro de referência			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	152	22	0,855
18	130	3	0,836
30	126	1	0,829
54	109	1	0,821
66	98	2	0,805

Tabela 14: Função de sobrevivência estimada pelo método de Kaplan-Meier conforme tratamento inicial utilizado

Pacientes tratadas com agente simples			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	1857	15	0,992
18	1842	5	0,989
30	1832	2	0,988
66	1523	1	0,988
78	1441	1	0,987
126	1161	1	0,987
Pacientes tratadas com agente múltiplo			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	297	36	0,882
18	261	2	0,875
30	258	2	0,858
42	246	2	0,851
54	236	1	0,847
Pacientes não tratadas com quimioterápicos			
Tempo, em dias	Número de pessoas sob risco	Número de eventos	Estimador KM
6	27	15	0,429
18	12	1	0,393
30	11	1	0,357
66	10	1	0,321

6 Apêndice B - *Script* do modelo implementado no *OpenBugs*

```
model
{
b0 ~ dnorm(0,0.001)
b1 ~ dnorm(0,0.001)
b2 ~ dnorm(0,0.001)
b3 ~ dnorm(0,0.001)
b4 ~ dnorm(0,0.001)
b5 ~ dnorm(0,0.001)
b6 ~ dnorm(0,0.001)
b7 ~ dnorm(0,0.001)
b8 ~ dnorm(0,0.001)
b9 ~ dnorm(0,0.001)
b10 ~ dnorm(0,0.001)
b11 ~ dnorm(0,0.001)
b12 ~ dnorm(0,0.001)
b13 ~ dnorm(0,0.001)
for(i in 1:2181) {
sobrevida[i] ~ dweib(r,lambda[i])C(t.cen[i],)
lambda[i] <- exp(-mu[i]*r)
```

```
mu[i]<- b0 + b1 * semhisto[i]+b2 * mola[i] + b3 * corio[i]+b4 * posmolar[i]+ b5 *
aborto[i]+ b6 * ectopica[i] + b7 * atermo[i] + b8 * escoreOMS[i] + b9 * escoremenor6[i]+
b10 * tratamentocr[i]+b11 * ntquimio[i]+ b12 * multiplo[i] + b13 * unico[i]

dif [i] <- pow(sobrevida[i] - mu[i],2)

}

r~dgamma(1,1.8)

sqresp <- sum(dif [])/2181

}
```

7 Apêndice C - Modelos Final

```
model
{
b1 ~ dnorm(0,0.001)
b9 ~ dnorm(0,0.001)
b13 ~ dnorm(0,0.001)
for(i in 1:2181) {
sobrevida[i] ~ dweib(r,lambda[i])C(t.cen[i],)
lambda[i] <- exp(-mu[i]*r)
mu[i]<- b1 * semhisto[i] + b9 * escoremenor6[i]+ b13 * unico[i]
dif [i] <- pow(sobrevida[i] - mu[i],2)
}
r~dgamma(1,1.8)
sqresp <- sum(dif [])/2181
}
```