

Análise da sobrevida de mulheres diagnosticadas com câncer de mama triplo negativo utilizando modelo de longa duração

Survival analysis of women diagnosed with triple-negative breast cancer using long-term survival model

Eder Angelo Milani * Vinicius F. Calsavara † Juliana Scudilio ‡
Carolina D. Tomazella § Carolina R. Bonini ¶ Caroline C. Antunes †
Danielly G. dos Santos ** Mariana A. Zanotti ††
Vera Tomazella ‡‡

Resumo:

O câncer de mama é uma das principais causas de morte de mulheres no mundo, no Brasil a realidade não é diferente. Neste contexto, o trabalho tem por objetivo estimar a proporção de mulheres que sobrevivem por um longo período de tempo após serem diagnosticadas com câncer de mama, utilizando um modelo de fração de cura. A metodologia proposta foi aplicada a um conjunto de dados de um estudo realizado no A.C. Camargo Cancer Center, São Paulo, Brasil, no período de 2001 a 2013 com mulheres diagnosticadas com câncer de mama triplo negativo, submetidas ao tratamento de quimioterapia neoadjuvante. O modelo proposto apresentou um ajuste adequado aos dados, o qual nos permitiu observar que mulheres com diagnóstico precoce possuem uma maior probabilidade de sobreviver por um longo período de tempo, indicando assim a necessidade de ampliação do rastreamento e diagnóstico precoce do câncer de mama.

Palavras-chave: Análise de Sobrevivência. Câncer de Mama. Modelo de Longa Duração. Modelo de Fração de Cura.

*Instituto de Matemática e Estatística, Universidade Federal de Goiás, edermilani@ufg.br

†A.C. Camargo Cancer Center, Departamento de Epidemiologia e Estatística, viniciusfalsavara@gmail.com

‡Departamento de Estatística, Universidade Federal de São Carlos, juliana-scudilio@uol.com.br

§Faculdade Integrada Padre Albino, catomazella@hotmail.com

¶Faculdade Integrada Padre Albino, carolina.rbonini@gmail.com

†Faculdade Integrada Padre Albino, carolinecantunes@hotmail.com

**Faculdade Integrada Padre Albino, dani_gaspareti@hotmail.com

††Faculdade Integrada Padre Albino, mariana.zanotti_@hotmail.com

‡‡Departamento de Estatística, Universidade Federal de São Carlos, vera@ufscar.br

Abstract: Breast cancer is one of the main causes of death for women worldwide, in Brazil the reality is no different. In this context, the work aims to estimate the proportion of women who survive for a long-term after being diagnosed with breast cancer, using a cure fraction model. The proposed methodology was applied to a data set of a study carried out at A.C. Camargo Cancer Center, São Paulo, Brazil, from 2001 to 2013 with women diagnosed with triple negative breast cancer, who underwent neoadjuvant chemotherapy treatment. The proposed model presented an adequate adjustment to the data, which allowed us to observe that women who had an early diagnosis have a greater probability of surviving for a long period of time, thus indicating the need to expand the screening and early diagnosis of breast cancer.

Keywords: Breast Cancer. Survival Analysis. Long-term Model. Cure Rate Model.

1 Introdução

O corpo humano é constituído de milhões de células vivas que cumprem diferentes funções. Ao longo da vida essas células crescem, sofrem processos de divisão e morrem. O câncer se inicia quando as células do corpo têm um crescimento fora de controle, passando a realizar os processos de crescimento e divisão sem os mecanismos de controle natural de morte celular [16]. De acordo com a Organização Mundial de Saúde (OMS), o câncer é um problema de saúde pública. Em 2013, o câncer foi responsável por mais de oito milhões de mortes em todo o mundo e, em 1990, passou da terceira para a segunda principal causa de morte, atrás apenas das doenças cardiovasculares [9].

De acordo com a Agência Internacional para Pesquisa sobre Câncer [6], em 2018 a estimativa da incidência de câncer de mama ultrapassava os oito milhões de casos no mundo, sendo mais incidente no sexo feminino, ficando atrás apenas do câncer de pulmão quando considerado ambos os sexos. Segundo o Instituto Nacional de Câncer [3], no mesmo ano houve mais de 17 mil óbitos por câncer de mama em mulheres, e a incidência de novos casos em 2020 é de, aproximadamente, 66 mil casos, sendo o câncer mais comum entre mulheres brasileiras.

A importância do diagnóstico precoce do câncer está associada ao acompanhamento clínico adequado que, em números, se traduz em uma elevação da incidência do tumor mamário. Segundo [10], há diversos fatores de risco associados ao câncer de mama, tais como: histórico familiar de câncer de mama, tabagismo, alcoolismo, obesidade e sedentarismo. O aumento da incidência dessa doença está se tornando uma preocupação constante na vida das mulheres brasileiras. Como destacado por [15], a sobrevida (um indivíduo não apresentar o evento de interesse após um determinado tempo) de pacientes diagnosticados com câncer continua a aumentar na maioria dos países. No Brasil, a taxa de sobrevida em cinco anos foi de aproxi-

madamente 75% entre o período de 2010 a 2014, mas esse valor é, possivelmente, subestimado.

Nas últimas décadas, ocorreram importantes avanços no tratamento do câncer de mama. A quimioterapia surgiu em meados dos anos 1950. No final dos anos 1970, a poliquimioterapia foi aplicada com uso de várias drogas em conjunto para diminuir o risco de recidiva. Para o avanço no desenvolvimento de pesquisas sobre o câncer, em 1973, foi criado o Instituto Nacional de Câncer [3]. Na década de 1990 surgiram drogas mais efetivas, e a radioterapia e cirurgias conservadoras ganharam espaço.

Em 2003, a era da genômica se iniciou e, após 2010, o desenvolvimento de terapias específicas e a personalização do tratamento tornou-se realidade, trazendo resultados superiores (para mais detalhes ver [7]).

Quando o diagnóstico é precoce, o tratamento tem uma maior capacidade de cura. As modalidades de tratamento para câncer de mama podem ser divididas em duas: sendo a primeira um tratamento local, em que um procedimento mais simples com cirurgia e radioterapia é realizado, e a segunda um tratamento sistêmico, que é um procedimento mais árduo para o paciente e é realizado quando a doença já se espalhou pelo corpo (metástases), fazendo o uso de quimioterapia, hormonioterapia e terapia biológica. O reconhecimento da progressão precoce da doença pelo sangue levou ao desenvolvimento de um tratamento prévio à cirurgia conhecido como quimioterapia neoadjuvante. Este tipo de tratamento tem como objetivo o controle precoce da doença micrometastática, a diminuição do tamanho do tumor (*downstaging*) e a testagem da quimiossensibilidade do tumor *in vivo*. A resposta à quimioterapia neoadjuvante é um preditor de resposta a longo prazo e fornece informações prognósticas após um curto período de acompanhamento, ao contrário dos estudos adjuvantes que fornecem resultados em 5-10 anos de acompanhamento [11, 14].

O câncer de mama triplo-negativo é definido como um tumor sem expressão de receptor de estrogênio (ER), receptor de progesterona (PR) e HER2. Estes incluem os chamados tumores *claudin-low*, que são relatados como enriquecidos com células com propriedades semelhantes às células-tronco e com características de transição epitelial para mesenquimal; o subgrupo rico em interferon, que engloba tumores com um prognóstico consideravelmente melhor do que aquele associado a outros cânceres de mama triplo-negativos; e o subgrupo semelhante a mama normal, que pode ser um artefato. O câncer de mama triplo-negativo tem menos opções de tratamento não invasivo do que outros tipos de câncer de mama. A cirurgia é uma opção de tratamento, caso o tumor não tenha se disseminado para outros órgãos. A quimioterapia neoadjuvante pode ser administrada antes da cirurgia para diminuir

o tamanho do tumor [1].

Com os avanços na área médica, um número maior de pacientes passou a ser considerado “curado” ou imune à doença em estudo. Um indivíduo é considerado imune ou “curado” quando não apresenta o evento de interesse no tempo de observação definido, dessa maneira um indivíduo imune sempre tem sua observação censurada (não observado o evento de interesse durante o estudo). Dizer que um paciente foi “curado” de um câncer significa que não houve recidiva da doença em um determinado período de tempo, por exemplo, cinco anos após o diagnóstico da doença. No entanto, dependendo do tipo da doença, bem como diversos outros fatores de risco, a doença pode reincidir. Portanto, dizer no sentido literal que o paciente foi curado e que não terá mais recidiva é uma afirmação difícil de se fazer na prática. Entretanto, para os nossos estudos, dizemos que existe uma fração de cura quando há observações no banco de dados que representam indivíduos que são imunes ao evento de interesse. A partir dos modelos tradicionais de análise de sobrevivência, não é possível estimar essa fração de indivíduos. Nesse contexto, uma metodologia utilizada em análise de sobrevivência que leva em conta a proporção de curados são os modelos de longa duração, também chamados de modelos com fração de cura. Esse tipo de modelagem têm sido amplamente empregado na literatura atraindo a atenção dos pesquisadores e constituindo um campo de pesquisa importante dentro da área médica, para mais detalhes ver [2].

Na literatura é possível encontrar um extensa discussão sobre modelos de longa duração ou fração de cura, sendo que boa parte dos autores discutem a criação de modelos envolvendo mistura de distribuições. Dentre esses modelos, o trabalho precursor foi proposto por [4], que utilizou o método de máxima verossimilhança para estimar a proporção de sobreviventes em uma população de 121 mulheres com câncer de mama em um estudo com duração de 14 anos. A partir do trabalho de [4], Berkson e Gage [2] propuseram um modelo de mistura com o objetivo de estimar a proporção de curados em uma amostra submetida a um tratamento de câncer de estômago. Recentemente, diferentes modelos foram considerados em estudos clínicos em que há uma proporção de pacientes imunes ao evento de interesse. O modelo de mistura padrão proposto por [2] e suas extensões no contexto de fragilidade [17, 5] podem explicar o tempo de vida dos pacientes, enquanto modelos mais complexos de longa duração como de [20, 8, 12], entre outros, vêm sendo explorados com o objetivo de explicar melhor os mecanismos biológicos envolvidos.

Face ao que foi exposto acima, o principal objetivo deste trabalho é propor um modelo de sobrevivência que consiga capturar a proporção das pacientes que não vieram a óbito durante a realização do estudo, levando em consideração a segmentação da amostra por fatores de riscos associado ao câncer de mama. A saber, o

modelo de mistura padrão, proposto por [2], é utilizado. Ilustramos a relevância prática do modelo em um conjunto de dados reais de mulheres diagnosticadas com câncer de mama triplo negativo e que foram submetidas ao tratamento de quimioterapia neoadjuvante no período de 2001 a 2013 no A.C.Camargo Cancer Center, São Paulo, Brasil. Além disso, o objetivo secundário é divulgar os resultados obtidos de forma a contribuir para o aumento da confiança das pacientes em tratamento contra o câncer.

O presente trabalho está organizado da seguinte forma: na Seção 2, apresentamos uma breve descrição da metodologia de análise de sobrevivência e um modelo estatístico que permite estimar a proporção de pacientes com longa duração na população. Na Seção 3, descrevemos o conjunto de dados oriundo de um estudo de pacientes com câncer de mama diagnosticadas no A.C.Camargo Cancer Center, bem como os resultados do modelo ajustado aos dados. Na Seção 4, apresentamos as conclusões do trabalho.

2 Métodos

2.1 Análise de Sobrevivência

O objetivo principal em análise de sobrevivência é o tempo até a ocorrência de um evento de interesse, comumente conhecido como tempo de falha. Os exemplos de estudos que envolvem dados com esta característica são muitos, podemos destacar, o tempo do diagnóstico até o óbito de pacientes com uma doença específica, o tempo até a recidiva de um câncer, entre outros.

Uma peculiaridade dos dados de análise de sobrevivência é a presença de observações incompletas, geralmente chamadas de censuras. Este fato ocorre devido a vários motivos, dentre eles, o abandono do tratamento, a saída do estudo por motivo diferente ao estudado, ou ainda a não ocorrência do evento de interesse antes do término do experimento. A utilização de informações parciais são de extrema importância, pois mesmo de maneira incompleta, estas trazem informações sobre o tempo até a ocorrência do evento de interesse.

Seja $T \geq 0$ uma variável aleatória não negativa, contínua, representando o tempo até a ocorrência do evento de interesse. A função de sobrevivência é definida como sendo a probabilidade de um paciente sobreviver pelo menos até um instante pré-estabelecido t , isto é,

$$S(t) = P(T \geq t) = \int_t^{\infty} f(u)du = 1 - F(t),$$

em que $f(t)$ é a função de densidade de probabilidade e $F(t)$ é a função de distribuição acumulada. A função $S(t)$ é dita própria se satisfaz as seguintes condições:

- $\lim_{t \rightarrow \infty} S(t) = 0$,
- $\lim_{t \rightarrow 0} S(t) = 1$.

No entanto, se $\lim_{t \rightarrow \infty} S(t) > 0$, a função de sobrevivência é dita ser imprópria.

Outra importante função em análise de sobrevivência é a taxa instantânea de falha, geralmente chamada de função de risco, sendo definida por

$$h(t) = -\frac{d}{dt} \log S(t) = \frac{f(t)}{S(t)}.$$

Em análise de sobrevivência, o conjunto de dados observados pode ser representado por $(\mathbf{t}, \boldsymbol{\delta}, \mathbf{X})$, em que \mathbf{t} representa o tempo observado, \mathbf{X} corresponde ao vetor de covariáveis observadas, e $\boldsymbol{\delta}$ é a variável indicadora de falha, ou seja, assume o valor 1, se o tempo é de falha e 0, se é um tempo censurado.

2.2 Modelo de Mistura Padrão

Tradicionalmente, o modelo de mistura padrão proposto por [2] é ajustado na análise de dados com longa duração devido a sua simplicidade e facilidade na interpretação dos parâmetros. A função de sobrevivência populacional é construída a partir de uma mistura de duas funções de sobrevivência, uma imprópria, representando os pacientes imunes ao evento de interesse, e outra própria, associada aos indivíduos em risco (não imunes).

O modelo de mistura padrão é construído considerando uma variável de Bernoulli não observável M_i ao indivíduo, em que

$$M_i = \begin{cases} 0, & \text{se o indivíduo } i \text{ não está em risco,} \\ 1, & \text{se o indivíduo } i \text{ está em risco,} \end{cases} \quad i = 1, 2, \dots, n,$$

com $P(M_i = 0) = p_0$ e $P(M_i = 1) = 1 - p_0$.

Como existem duas subpopulações (curados e não curados) a função de sobrevivência populacional é dada por

$$S_{pop}(t) = p_0 + (1 - p_0)S(t), \quad t \geq 0, \quad (1)$$

em que $S(t)$ representa a função de sobrevivência própria associada aos indivíduos em risco, $S_{pop}(t)$ a função de sobrevivência imprópria e $p_0 \in (0, 1)$ a proporção de imunes ou curados.

O modelo de mistura padrão (1) possui as seguintes propriedades:

- $S_{pop}(0) = 1$;
- $S_{pop}(t)$ é não crescente;
- se $p_0 = 0$, então $S_{pop}(t) = S(t)$;
- $\lim_{t \rightarrow \infty} S_{pop}(t) = p_0$.

A última propriedade retrata o fato da função de sobrevivência populacional ser imprópria, pois a curva de sobrevivência estabiliza em p_0 .

A função densidade de probabilidade populacional é dada por

$$f_{pop}(t) = p_0 f(t),$$

em que $f(t)$ representa a função de densidade própria relativa ao grupo dos indivíduos em risco.

A função de risco populacional é dada por

$$h_{pop}(t) = \frac{f_{pop}(t)}{S_{pop}(t)} = \frac{p_0 f(t)}{1 - p_0 + p_0 S(t)}.$$

Note que,

$$\lim_{t \rightarrow \infty} h_{pop}(t) = \lim_{t \rightarrow \infty} \frac{p_0 f(t)}{S_{pop}(t)} = \left(\frac{p_0}{1 - p_0} \right) \lim_{t \rightarrow \infty} f(t) = 0. \quad (2)$$

O resultado em (2) indica que a função de risco populacional converge para o valor zero quando o tempo é grande, retratando o fato da curva de sobrevivência populacional estabilizar em um determinado valor, chamado de fração de cura. Este resultado indica que uma parcela dos indivíduos em estudo não falharam e, possivelmente, estão “curados” ou imunes ao evento de interesse.

2.3 Modelo de Mistura Padrão Weibull (MMW)

Nesta seção formulamos o modelo de mistura padrão com distribuição Weibull para a função de sobrevivência dos indivíduos em risco. A distribuição de Weibull [21] é amplamente conhecida na literatura para modelar tempo de vida e possui dois parâmetros, um de escala (λ) e o outro de forma (γ). O parâmetro de escala refere-se à média da distribuição, enquanto o parâmetro de forma determina o formato da função densidade.

Uma variável aleatória T tem distribuição Weibull com parâmetros $\lambda > 0$ e $\gamma > 0$, se sua função densidade de probabilidade é dada por

$$f(t) = \gamma \lambda^\gamma t^{\gamma-1} \exp[-(\lambda t)^\gamma], \quad t \geq 0.$$

A respectiva função de sobrevivência é expressa por

$$\begin{aligned} S(t) &= \int_t^\infty \gamma \lambda^\gamma x^{\gamma-1} \exp[-(\lambda x)^\gamma] dx \\ &= \exp[-(\lambda t)^\gamma], \end{aligned} \quad (3)$$

e sua função de risco é

$$h(t) = \gamma \lambda^\gamma t^{\gamma-1}.$$

O modelo de sobrevivência Weibull é uma extensão do modelo exponencial, em que a inclusão de um segundo parâmetro permite uma maior flexibilidade da função de risco. A função de risco associada ao modelo Weibull é crescente se $\gamma > 1$, decrescente quando $\gamma < 1$ e constante se $\gamma = 1$. De acordo com [22], a conveniência do modelo Weibull se deve pela flexibilidade e também pela simplicidade das funções de risco e sobrevivência.

O modelo de mistura padrão Weibull (MMW) é obtido substituindo (3) em (1). A função de sobrevivência populacional é dada por

$$S_{pop}(t) = p_0 + (1 - p_0) \exp[-(\lambda t)^\gamma], \quad (4)$$

e a respectiva função densidade é

$$f_{pop}(t) = p_0 \gamma \lambda^\gamma t^{\gamma-1} \exp[-(\lambda t)^\gamma].$$

É muito comum em estudos clínicos a coleta de informações referentes as características dos pacientes como, sexo, idade, estadiamento da doença, tipo de tratamento, entre outras informações relevantes para explicar a ocorrência do evento de interesse. Neste sentido, modelos de sobrevivência que incluem covariáveis são fundamentais para entender como elas influenciam no tempo de vida.

Considerando que $p_0 \in (0, 1)$ pode ser explicada pelas covariáveis observadas, a função de ligação *logit* será considerada neste trabalho, mas outras funções de ligação como *probit* e *complementar log-log* podem também ser consideradas [19].

A fração de cura na presença de covariáveis considerando a função de ligação

logit é expressa por

$$p_0(\mathbf{x}) = \frac{\exp(\mathbf{x}^\top \boldsymbol{\beta})}{1 + \exp(\mathbf{x}^\top \boldsymbol{\beta})},$$

em que $\mathbf{x}^\top = (1, x_1, \dots, x_q)$ e $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_q)^\top$ são o conjunto de covariáveis e seus respectivos coeficientes da regressão.

A função de sobrevivência populacional na presença de covariáveis pode ser reescrita por

$$S_{pop}(t|\mathbf{x}) = p_0(\mathbf{x}) + [1 - p_0(\mathbf{x})]S(t),$$

em que $S(t)$ é a função de sobrevivência dos pacientes não imunes, como dado em (3).

2.4 Inferência

A estimação dos parâmetros do modelo MMW (4) é realizada por meio do procedimento clássico de máxima verossimilhança. Seja $\mathbf{D} = (\mathbf{t}, \boldsymbol{\delta}, \mathbf{X})$ os dados observados de uma amostra de tamanho n , em que $\mathbf{t} = (t_1, \dots, t_n)^\top$ representa o vetor dos tempos de vida observados, $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^\top$ a variável indicadora de falha, sendo $\delta_i = 1$ se ocorreu o evento de interesse e $\delta_i = 0$ se ocorreu censura e \mathbf{X} a matriz de covariáveis observadas.

Seja $\boldsymbol{\theta} = (\lambda, \gamma, p_0)$ o vetor de parâmetros do modelo MMW (4). A função de máxima verossimilhança de $\boldsymbol{\theta}$ considerando que a censura é não informativa, é dada por

$$\begin{aligned} L(\boldsymbol{\theta}; \mathbf{D}) &\propto \prod_{i=1}^n f_{pop}(t_i, \boldsymbol{\theta})^{\delta_i} S_{pop}(t_i, \boldsymbol{\theta})^{1-\delta_i} \\ &= \prod_{i=1}^n [(1 - p_0)\gamma\lambda^\gamma t^{\gamma-1} \exp[-(\lambda t)^\gamma]]^{\delta_i} [p_0 + (1 - p_0)\exp(-(\lambda t)^\gamma)]^{1-\delta_i}. \end{aligned}$$

Devido à complexidade da função de verossimilhança, não é possível obter os estimadores de máxima verossimilhança de forma analítica, desta forma a estimativa dos parâmetros são obtidas por meio de procedimentos numéricos, como o algoritmo BFGS implementado no *software* R Core Team [18].

3 Aplicação a dados reais

O conjunto de dados refere-se a um estudo de sobrevida de pacientes com diagnóstico de câncer de mama e tratadas no A.C.Camargo Cancer Center, São Paulo, Brasil, o qual é referência no diagnóstico, tratamento, pesquisa e ensino do câncer. Neste banco de dados contém informações de mulheres diagnosticadas com câncer de mama triplo negativo e que foram submetidas à quimioterapia neoadjuvante no período de 2001 a 2013. O evento de interesse foi definido como sendo a morte causada por câncer de mama, segundo a declaração de óbito. O tempo de vida foi registrado a partir da data do diagnóstico da doença até a data do óbito. As pacientes que morreram devido a outras causas ou que não apresentaram o evento de interesse até o final do estudo contribuíram com o seu tempo de acompanhamento, tendo recebido o indicativo de censura.

A amostra contém informações sobre o tempo de vida de 78 pacientes e três covariáveis observadas: infiltrado inflamatório tumoral, características do tumor primário e as características dos linfonodos das cadeias de drenagem linfática. Um total de 25 (32,1%) pacientes foram a óbito devido ao câncer de mama, enquanto 53 (67,9%) não foram a óbito. O menor tempo observado foi de nove meses, enquanto o maior tempo observado foi de 162,87 meses. O tempo médio de acompanhamento foi de 52,43 meses.

A inflamação exerce impacto em cada etapa da tumorigênese, desde a iniciação tumoral até a instalação de doença metastática. A presença dos linfócitos infiltrados no tumor (TILs) tem sido correlacionada com o tamanho, estágio do tumor e a sobrevida do paciente em uma variedade de neoplasias. Diversos sistemas de estadiamento poderiam ser concebidos neste estudo, mas levou-se em conta apenas as características do tumor primário (T) e as características dos linfonodos das cadeias de drenagem linfática do órgão em que o tumor se localiza (N). Estes parâmetros recebem graduações, geralmente de T0 a T4 e de N0 a N3, sendo que T0 significa que não existe evidência de tumor primário (não pode ser encontrado), enquanto que as siglas T1, T2, T3 e T4 descrevem o tamanho do tumor e/ou a disseminação da doença nas proximidades, posto que quanto maior o número, maior o tumor e/ou maior a disseminação pelos tecidos próximos. O mesmo ocorre para a posição do tumor N, sendo que N0 significa que os linfonodos vizinhos não contêm câncer, enquanto que N1, N2 ou N3 podem descrever a localização e/ou o número dos linfonodos com a doença. Quanto maior o número, mais o câncer se espalhou para os linfonodos. Na Tabela 1 apresentamos as variáveis observadas nesse estudo.

Na Tabela 2 apresentamos a segmentação das variáveis TIL, N e T em relação ao *status*. Observamos que 63,64% das mulheres diagnosticadas com câncer de mama

Tabela 1: Descrição das variáveis observadas.

Variável	Categoria	Codificação
Infiltrado inflamatório tumoral (TIL)	$\leq 10\%$	0
	$> 10\%$	1
Características do tumor primário (T)	T1 ou T2	0
	T3 ou T4	1
Localização do tumor (N)	N0	0
	N1 ou N2 ou N3	1
<i>Status</i> (indicador de falha)	Não Óbito	0
	Óbito	1

Tabela 2: Tabela cruzada entre as variáveis TIL, N e T segundo a ocorrência ou não do evento de interesse.

Variável	Categoria	<i>Status</i>		Total
		Censura	Óbito	
TIL	$\leq 10\%$	20(57,14%)	15(42,86%)	35(63,64%)
	$>10\%$	19(82,61%)	4(17,39%)	23(45,46%)
N	N0	18(85,71%)	3(14,29%)	21(26,82%)
	N1, N2 ou N3	35(61,40%)	22(39,60%)	57(73,08%)
T	T1 ou T2	24(82,76%)	5(17,24%)	29(33,33%)
	T3 ou T4	28(58,33%)	20(41,67%)	48(67,67%)

têm infiltrado inflamatório ($TIL \leq 10\%$), destas 42,86% foram a óbito e 57,14% foram censuradas. As mulheres diagnosticadas com câncer de mama e com os linfonodos vizinhos não contendo câncer (N0) correspondem a 26,82%, destas 14,29% foram a óbito e 85,71% não experimentaram o evento de interesse. Analisando as características do tumor primário, tem-se que 67,67% das pacientes estão em T3 ou T4, destas 41,67% foram a óbito. Vale destacar a presença de dados faltantes nas covariáveis TIL e N, sendo que na covariável TIL a ausência ocorreu em 20 pacientes, enquanto que na covariável N não houve informação disponível para apenas uma paciente.

3.1 Resultados

O estimador de Kaplan-Meier (KM) [13], é uma técnica não-paramétrica utilizada para estimar a função de sobrevivência na presença ou não de observações censuradas. Neste trabalho as curvas estimadas por meio do modelo MMW serão comparadas

com as curvas estimadas utilizando o estimador de KM.

Na Tabela 3 apresentamos a estimativa de máxima verossimilhança (EMV), o erro padrão (EP) e o intervalo de confiança assintóticos de 90% de confiança (IC(90%)) dos parâmetros do modelo MMW, sem considerar covariáveis. De acordo com os resultados, todos os parâmetros são significativos e a proporção de pacientes não susceptíveis ao evento de interesse foi de 0,622.

Na Figura 1 apresentamos as curvas de sobrevivência estimadas por meio do modelo MMW e por KM. As duas curvas estimadas estão próximas, indicando que o modelo MMW se ajusta razoavelmente bem aos dados. Observamos ainda que a estimativa intervalar do parâmetro de forma γ não inclui o valor 1, o qual indica evidências contra a distribuição exponencial.

Tabela 3: Estimativas dos parâmetros do modelo MMW na ausência de covariáveis.

Parâmetros	EMV	EP	IC(90%)
λ	0,031	0,003	(0,025; 0,036)
γ	2,208	0,359	(1,616; 2,800)
p_0	0,622	0,063	(0,518; 0,727)

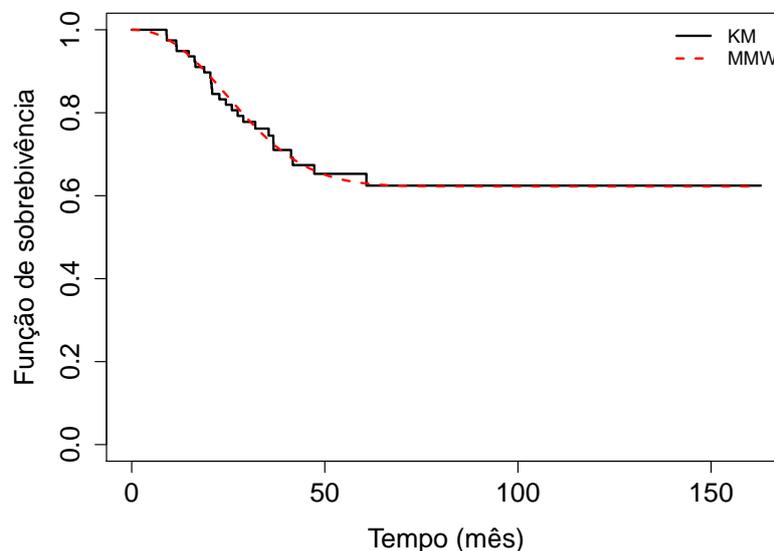


Figura 1: Função de sobrevivência estimada via KM e por meio do modelo MMW.

Para avaliar o efeito das covariáveis na sobrevivência das pacientes, o modelo MMW foi ajustado considerando cada uma das covariáveis. Na Tabela 4 apresentamos as estimativas dos parâmetros para os três modelos, sendo considerado as covariáveis TIL, N e T.

De acordo com os resultados todos os parâmetros foram significativos, exceto o intercepto (β_0) no ajuste com a covariável TIL. Destacamos ainda que todas as frações de cura também foram significativas.

Tabela 4: Estimativas dos parâmetros do modelo MMW considerando as covariáveis TIL, N e T, respectivamente

Parâmetros	EMV	EP	IC(90%)
λ	0,032	0,005	(0,024; 0,039)
γ	2,004	0,366	(1,399; 2,609)
β_0	-0,019	0,399	(-0,678; 0,639)
β_1 (TIL > 10%)	1,385	0,686	(0,252; 2,518)
p_{00} (TIL \leq 10%)	0,495	0,100	(0,331; 0,660)
p_{01} (TIL > 10%)	0,797	0,093	(0,644; 0,950)
λ	0,031	0,003	(0,025; 0,037)
γ	2,195	0,360	(1,602; 2,789)
β_0	1,561	0,646	(0,495; 2,627)
β_1 (N1, N2 ou N3)	-1,351	0,708	(-2,519; -0,182)
p_{00} (N0)	0,827	0,093	(0,674; 0,979)
p_{01} (N1, N2 ou N3)	0,552	0,075	(0,428; 0,677)
λ	0,031	0,003	(0,026; 0,037)
γ	2,208	0,356	(1,620; 2,796)
β_0	1,303	0,517	(0,450; 2,156)
β_1 (T3 ou T4)	-1,201	0,604	(-2,197; -0,205)
p_{00} (T1 ou T2)	0,786	0,087	(0,616; 0,957)
p_{01} (T3 ou T4)	0,526	0,081	(0,367; 0,684)

Na Figura 2 são apresentadas as curvas da função de sobrevivência estimadas via KM e MMW, para cada uma das três covariáveis em estudo. De acordo com os resultados, mulheres com TIL > 10% apresentam uma sobrevida maior em relação aquelas com TIL \leq 10%. Observamos ainda uma melhor sobrevida para as pacientes em que os linfonodos vizinhos não contêm câncer (N0), em comparação com as mulheres que tiveram os linfonodos classificados como N1, N2 ou N3. Além disso, maior sobrevida foi observada para as pacientes em T1 ou T2, conforme esperado. Em todos os gráficos, a curva de sobrevivência estimada pelo modelo MMW está relativamente próxima da curva estimada via KM, o que sugere um bom ajuste do modelo aos dados.

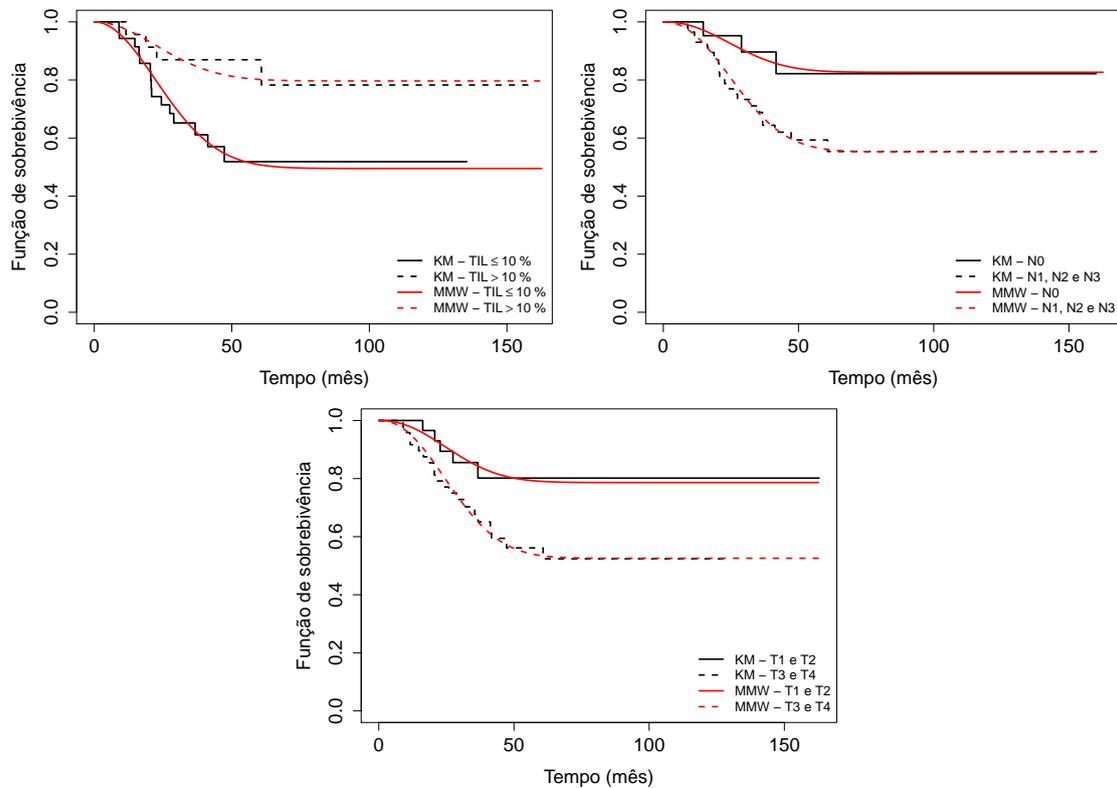


Figura 2: Curvas de sobrevivência estimada via KM e por meio do modelo MMW de acordo com as covariáveis TIL, N e T, respectivamente.

4 Conclusão

A utilização do modelo de mistura padrão com distribuição Weibull possibilitou uma melhor caracterização do perfil e da sobrevivência das pacientes diagnosticadas com câncer de mama triplo negativo e que foram submetidas à quimioterapia neoadjuvante. A realização desta pesquisa reforça a importância de se trabalhar com as informações disponíveis nos serviços de saúde responsáveis pelo atendimento ao paciente com câncer no Brasil. Os resultados mostraram que mulheres diagnosticadas com câncer de tamanho T1 ou T2, com os linfonodos vizinhos sem câncer e linfócitos infiltrados $>10\%$, apresentaram as melhores sobrevida. Notamos que em cenários de melhores prognósticos, a taxa de mulheres que sobreviveram por um longo período de tempo após serem diagnosticadas com câncer é de aproximadamente 80%, enquanto que em um cenário de pior prognóstico essa taxa é próxima de 50%. Assim, torna-se evidente a necessidade de ampliação do rastreamento do câncer de mama, bem como a garantia de tratamento oportuno para os casos diagnosticados, o que poderá influenciar positivamente o prognóstico da doença e influenciar positivamente na longevidade das pacientes.



Agradecimentos:

A autora Juliana Scudilio agradece à CAPES pelo suporte financeiro.

Referências

- [1] ANDERS, C. K.; CAREY, L.A. **ER/PR negative, HER2-negative (triple-negative) breast cancer**. Disponível em <https://www.uptodate.com/contents/er-pr-negative-her2-negative-triple-negative-breast-cancer>. Acesso em 05 jan. 2021.
- [2] BERKSON, J.; GAGE, R. P. Survival curve for cancer patients following treatment, **Journal of the American Statistical Association**, v. 47, n. 259, p. 501-515, 1952.
- [3] BRASIL. Ministério da Saúde. **Instituto Nacional do Câncer José Alencar Gomes da Silva**, Disponível em: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>. Acesso em: 10 jun. 2020.
- [4] BOAG, J. W. Maximum likelihood estimates of the proportion of patients cured by cancer therapy, **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 11, n. 1, p. 15-53, 1949.
- [5] CALSAVARA, V. F.; TOMAZELLA, V. L. D.; FOGO, J. C. The effect of frailty term in the standard mixture model, **Chilean Journal of Statistics**, v. 4, p. 95-109, 2013.
- [6] CANCER TODAY **International Agency for Research on Cancer**. Disponível em: <https://gco.iarc.fr/today/home>. Acesso em: 19 jun. 2020.
- [7] CAVALCANTE, F. P.; MILLEN, E. C.; ZERWES, F. P.; NOVITA, G. G. Evolução do tratamento local do câncer de mama: Revisão narrativa. **Rev Ginecol Obstet**, v.42, n. 6, p. 356-364, 2020.
- [8] CHEN, M.; IBRAHIM, J. G.; SINHA, D. A new Bayesian model for survival data with a surviving fraction, **Journal of the American Statistical Association**, v. 94, n. 447, p. 909-919, 1999.
- [9] FITZMAURICE, C. *et al.* The global burden of cancer 2013 **JAMA Oncology**, v. 1, n. 4, p. 505-527, 2015.

-
- [10] GONÇALVES, A. T. C. *et al.* Câncer de mama: mortalidade crescente na Região Sul do Brasil entre 1980 e 2002, **Cadernos de Saúde Pública**, v. 23, n. 8, p. 1785-1790, 2007.
- [11] GRALOW J. R. *et al.* Preoperative therapy in invasive breast cancer: pathologic assessment and systemic therapy issues in operable disease. **Journal of Clinical Oncology: official journal of the American Society of Clinical Oncology**, vol. 26, n. 5, p. 814–819, 2008.
- [12] IBRAHIM, J. G.; CHEN, M.; SINHA, D. Bayesian semiparametric models for survival data with a cure fraction, **Biometrics**, v. 57, n. 2, p. 383-388, 2011.
- [13] KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations, **Journal of the American Statistical Association**, v.53, p. 457–481, 1958.
- [14] KILLELEA B. K. *et al.* Neoadjuvant chemotherapy for breast cancer increases the rate of breast conservation: results from the National Cancer Database. **Journal of the American College of Surgeons**, v. 220, n. 6, p. 1063–1069, 2015.
- [15] MAKDISSI, F. B. *et al.* Breast cancer survival in a Brazilian cancer center: a cohort study of 5095 patients, **Mastology**, v. 29, n. 1, p. 37–46, 2019.
- [16] PRECOCE, D. Controle do Câncer de Mama-Documento de Consenso, **Revista Brasileira de Cancerologia**, v. 50, n. 2, p. 77-90, 2004.
- [17] PRICE, D. L.; MANATUNGA, A. K. Modelling survival data with a cured fraction using frailty models, **Statistics in Medicine**, v. 20, p. 1515-1527, 2001.
- [18] R CORE TEAM **R: A Language and Environment for Statistical Computing**, R Foundation for Statistical Computing, Vienna, Austria, 2020. <https://www.R-project.org/>
- [19] TACONELI, J. P. **Modelo de mistura paramétrico com fragilidade na presença de covariáveis**, Dissertação (Mestrado em Estatística) DEs, UFS-Car, São Carlos - SP, 2013.
- [20] YAKOVLEV, A. Y.; TSODIKOV, A. D. **Stochastic models of tumor latency and their biostatistical applications**, World Scientific, 1996.



- [21] WEIBULL, W. A statistical theory of the strength of materials, Roy, **Swed. Inst. Eng. Res**, v. 151, n. 545, 1939.
- [22] WIENKE, A. **Frailty models in survival analysis**, CRC press, 2010.

Submetido em 30 jun. 2020.

Aceito em 13 maio 2021.