

Pedro F. Dutra¹, Núbia K. O. Almeida²

¹Universidade Federal Fluminense

²Departamento de Estatística

pefrazao@id.uff.br

Introdução

A técnica de Análise de Sobrevida é usualmente utilizada em estudos da área da saúde que desejam investigar fatores referentes a características pessoais dos enfermos, da doença e do tratamento que influenciam a sobrevida de pessoas com certo tipo de doença. Dessa forma, surge naturalmente o interesse de se investigar doenças que acometem brasileiros, dentre elas, o câncer de mama. A identificação de fatores que possam contribuir para sobrevida de mulheres em tratamento de câncer de mama, dada a evolução dos tratamentos ao longo dos tempos, é tema relevante que justifica estudo contínuo.

Objetivos

Avaliar se os atributos estadiamento clínico, ocorrência de múltiplos tumores e lateralidade do tumor geram efeito sobre o tempo de sobrevida de mulheres no seu 1º tratamento contra o câncer de mama, a partir das suas funções de sobrevida.

Materiais e métodos

Dados

Os dados utilizados são referentes a 36.823 mulheres brasileiras que foram diagnosticadas e iniciaram seu tratamento entre 2010 e 2019, e cujas informações estão disponíveis publicamente e de forma anônima pelo Instituto Nacional do Câncer (Integrador RHC). Desta forma, o estudo é uma coorte aberta, pois o início de acompanhamento das pacientes varia.

Definições básicas

Denote por T a variável aleatória definida pelo tempo até a ocorrência de um evento, que será denominado falha. A função de sobrevida de T é dada por:

$$S(t) = P(T > t), -\infty < t < +\infty$$

Por razões técnicas, define-se

$$S(a^-) = \lim_{t \rightarrow a^-} S(t), a \in \mathbb{R}$$

Censuras à direita

Frequentemente, em estudos de populações de tempos de sobrevida, alguns indivíduos são censurados, isto é, têm os seus tempos exatos de falha desconhecidos. A censura à direita [1] ocorre quando este tempo não é observado, seja por saída precoce do estudo, ou então pela não ocorrência da falha até o término do período de acompanhamento. Neste cenário, definem-se as quantidades abaixo:

- O número de indivíduos presentes no estudo (n)
- Os tempos distintos de falhas observados
 $t_1 < t_2 < \dots < t_k$
- A quantidade de falhas ocorridas (d_j) no tempo t_j e, c_j , a quantidade de censuras que ocorreram no intervalo $[t_j, t_{j+1})$ nos tempos $t_{j1}, t_{j2}, \dots, t_{jcj}$ para $j = 1, \dots, k$.
- A quantidade de indivíduos que estão sob risco ($n_j = (c_j + d_j) + \dots + (c_k + d_k)$), num tempo imediatamente anterior a t_j .

Para correta utilização das técnicas que se seguem, assume-se que os tempos de censura são aleatórios e independentes, ou seja, que não estão associados à falha de nenhum indivíduo.

Função de verossimilhança

Constrói-se a função de verossimilhança dos dados, considerando que informação cada observação da amostra aleatória (incompleta) dos tempos de falha, T_1, T_2, \dots, T_n , traz. Se durante o intervalo $[t_j, t_{j+1})$, o i -ésimo indivíduo falhar, então, $T_i = t_j$. Agora, se ele for censurado à direita, então, obtém-se a informação de que $T_i > t_{jl}$, onde t_{jl} é o tempo até a sua censura. Deste modo, a verossimilhança, em termos dos tempos de falha, será:

$$L = \prod_{j=1}^k \left\{ [S(t_j^-) - S(t_j)]^{d_j} \prod_{l=1}^{c_j} S(t_{jl}) \right\}$$

Pode-se mostrar [2] que o estimador de máxima verossimilhança (não paramétrico) de $S(t)$ é

$$\hat{S}(t) = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j}$$

Teste logrank

O teste logrank [3] serve para identificar se todos os estratos de um fator possuem a mesma função de sobrevida. Deste modo, suas hipóteses são:

$$\begin{aligned} H_0: \quad & S_1(t) = \dots = S_r(t) \\ H_1: \quad & \exists u, v \text{ tais que } S_u(t) \neq S_v(t) \end{aligned}$$

Sendo k a quantidade total de tempos de falha, o teste baseia-se em k tabelas de contingência, uma para cada um deles. Assim, considerando o tempo de falha t_j , os dados são organizados da seguinte forma:

Estrato 1	...	Estrato m	...	Estrato r	Total
Falhas	d_{1j}	\dots	d_{mj}	\dots	d_{rj}
Sobreviventes	$n_{1j} - d_{1j}$	\dots	$n_{mj} - d_{mj}$	\dots	$n_{rj} - d_{rj}$
Sob risco	n_{1j}	\dots	n_{mj}	\dots	n_{rj}

Sob a hipótese nula, espera-se que a proporção de falhas em cada estrato seja igual à proporção de falhas em toda a amostra, para $j = 1, \dots, k$. Assim, denotando por $e_{ij} = n_{ij}(\frac{d_j}{n_j})$, a quantidade de falhas esperadas no grupo i no tempo j , defini-se o vetor de desvios ocorridos no tempo j , $w_j' = (d_{1j} - e_{1j}, \dots, d_{rj} - e_{rj})$, e, daí, o vetor que contém todos os desvios do estudo, $w = \sum_{j=1}^k w_j$. Sendo W sua matriz de covariância-variancia, demonstra-se que a estatística de teste

$$w' W^{-1} w \xrightarrow{n \rightarrow \infty} \chi_{r-1}^2$$

Resultados

Foram realizados testes de log-rank para comparar as funções de sobrevida segundo as variáveis estadiamento clínico, ocorrência de múltiplos tumores e lateralidade do tumor. Em todos, a hipótese nula foi rejeitada ao nível de 5% de significância. Na Figura 1, estão as funções de sobrevida estimadas da variável estadiamento clínico. Elas sugerem que, quanto mais avançado o estádio, menor a probabilidade de sobreviver.

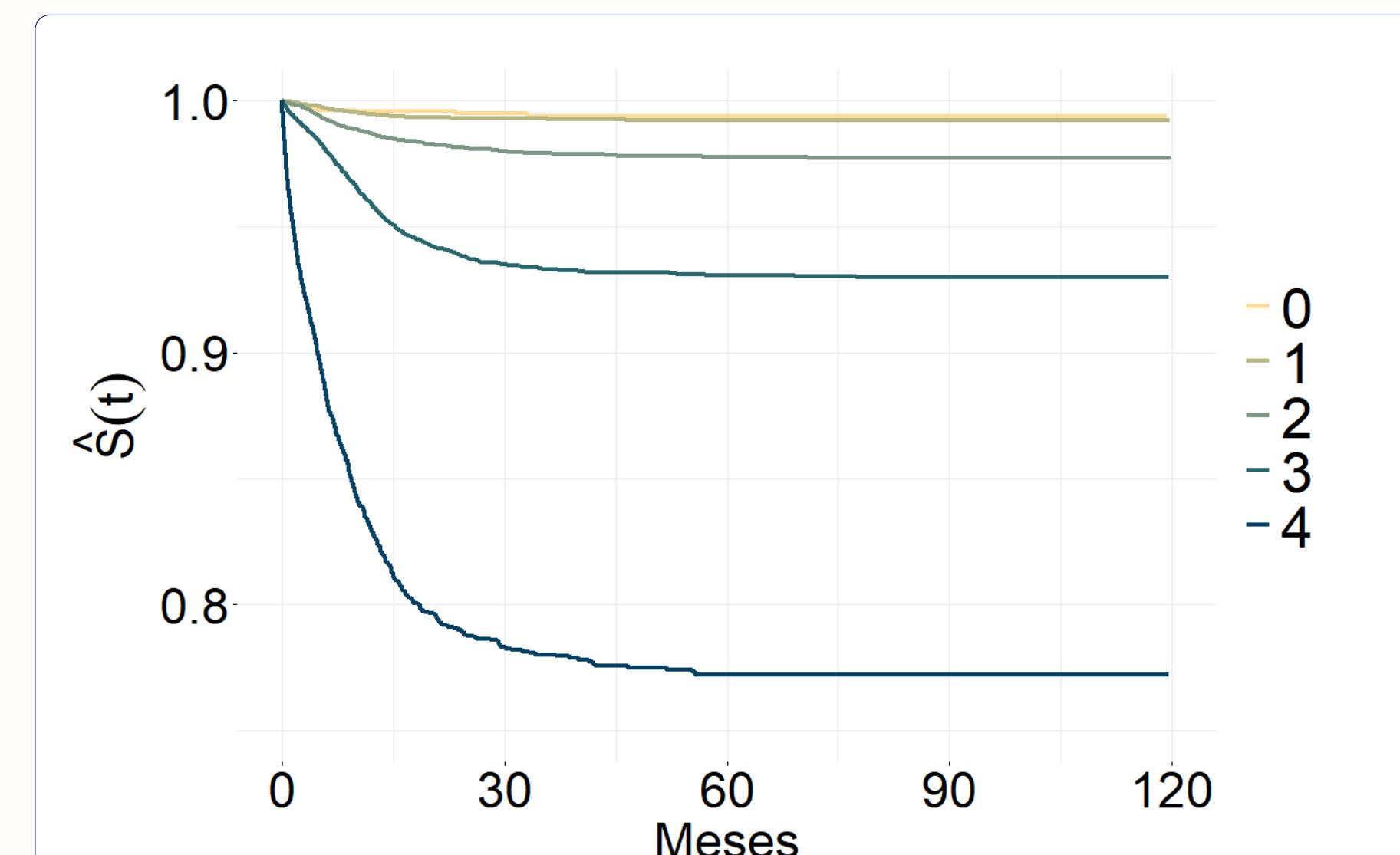


Figura 1: Sobrevida segundo o estadiamento clínico

Na Figura 2 é exposto que pacientes com múltiplos tumores apresentaram maior risco de óbito.

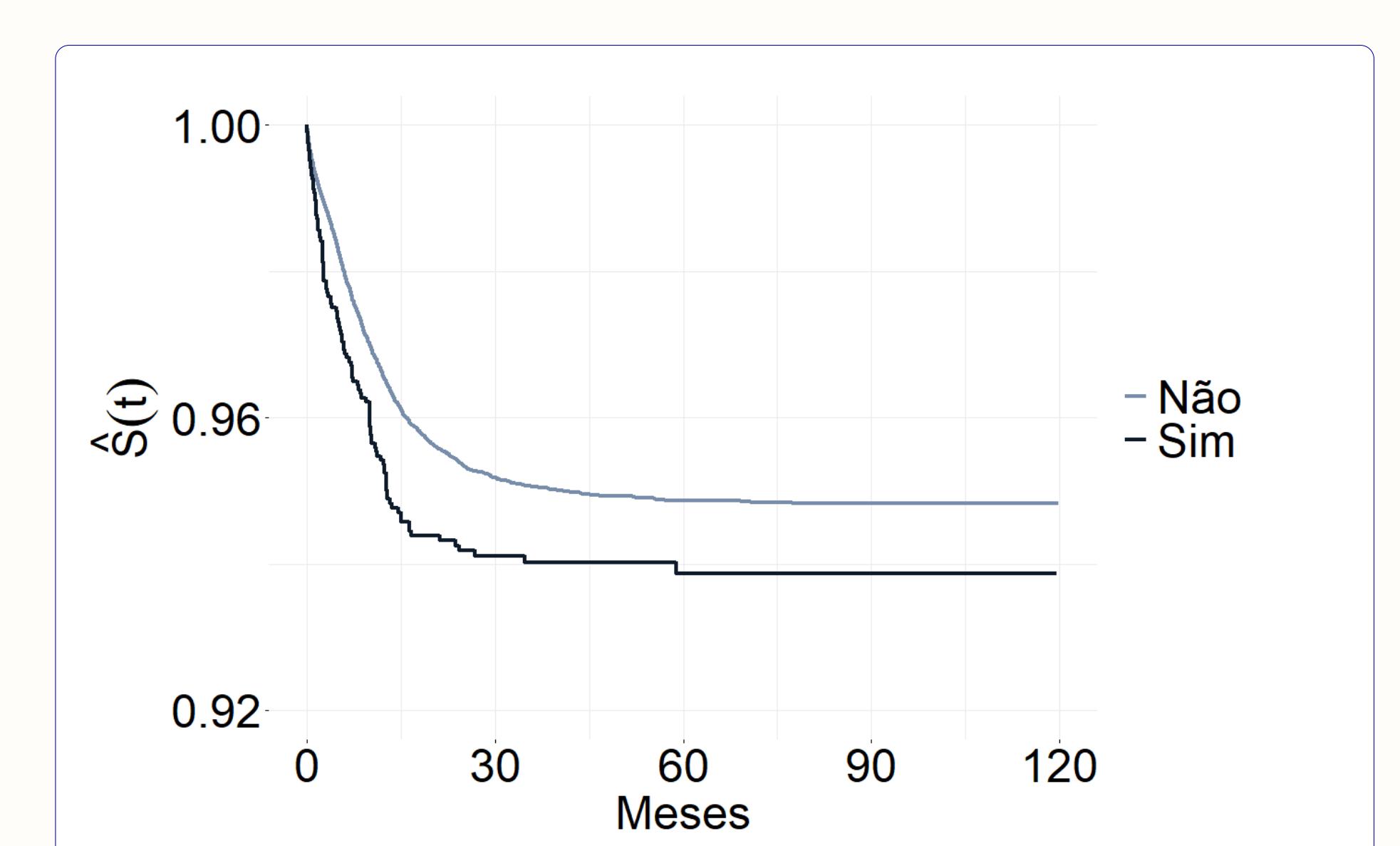


Figura 2: Sobrevida segundo a ocorrência de múltiplos tumores

A Figura 3 mostra que o tumor bilateral é o mais prejudicial à sobrevida.

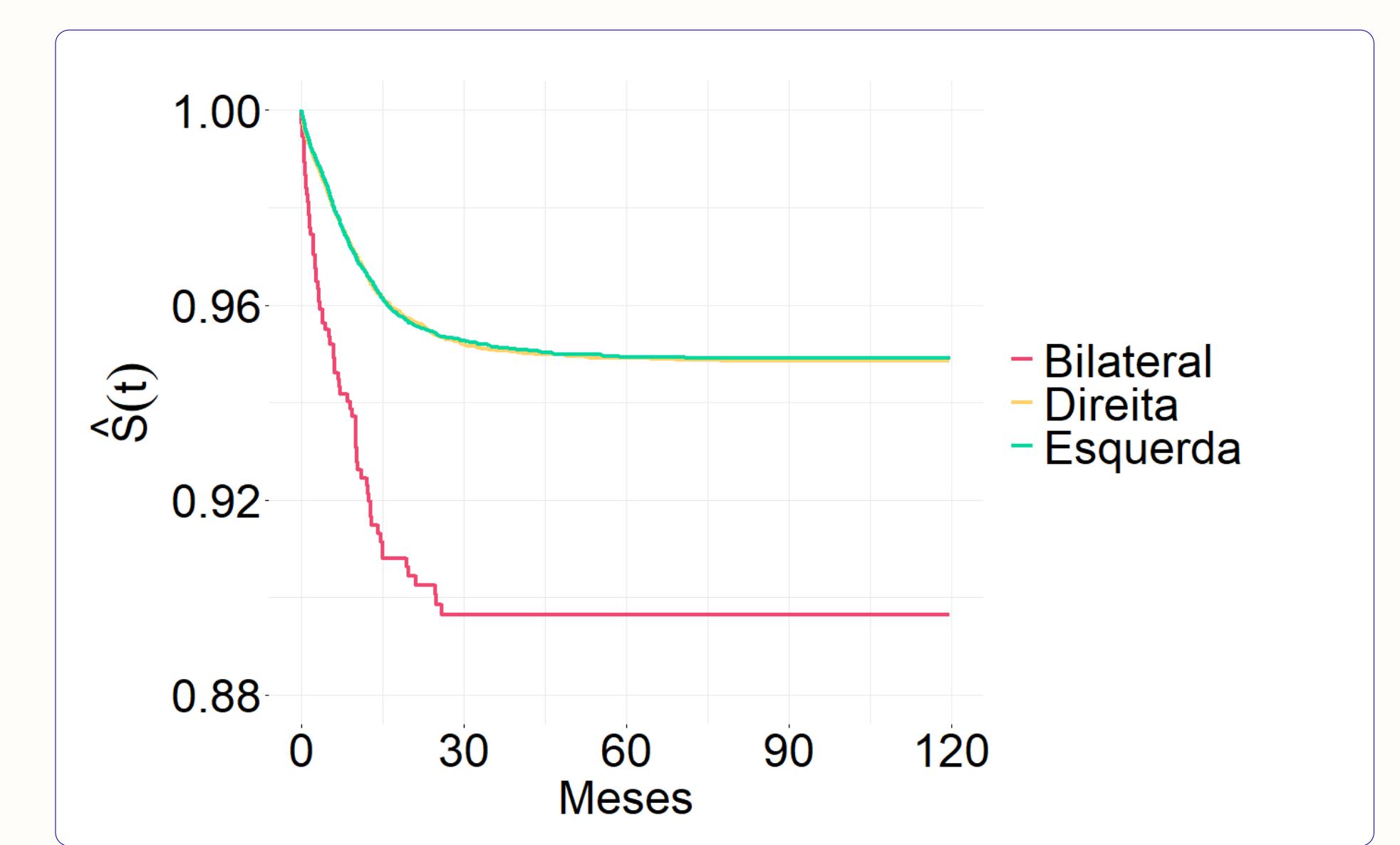


Figura 3: Sobrevida segundo a lateralidade do tumor

Conclusão

Todos os fatores avaliados influenciam o tempo de sobrevida das pacientes.

Agradecimentos

Este trabalho contou com o apoio da Fundação de Amor à Pesquisa do Estado do Rio de Janeiro (FAPERJ), por meio de bolsa de iniciação científica.

Referências

- [1] CARVALHO, Marilia Sá; ANDREOZZI; [et al.]. *Análise de sobrevida: teoria e aplicações em saúde*. 2. ed. Rio de Janeiro: Editora Fiocruz, 2011.
[2] KAPLAN, E. L.; MEIER, P. *Nonparametric estimation from incomplete observations*. Journal of the American Statistical Association, v. 53, n. 282, p. 457-481, 1958.
[3] KALBFLEISCH, J. D.; PRENTICE, R. L. *The statistical Analysis of Failure Time Data*. New Jersey: John Wiley & Sons, 2002.