

CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ
ISSN on-line 2317-4536 / ISSN impresso 1413-9022 - v 54, p.1-13, 2023
DOI: 10.12957/cadest.2023.77204

SOBREVIDA DE MULHERES COM CÂNCER NO COLO DO ÚTERO

Thaís Regina Araujo dos Santos
Universidade do Estado do Rio de Janeiro – UERJ
thaisaraujo0290@gmail.com

Ricardo de Melo e Silva Accioly
Universidade do Estado do Rio de Janeiro - UERJ
raccioly@ime.uerj.br

Resumo

O câncer do colo do útero é um problema de saúde pública no Brasil. Neste artigo é analisada a sobrevida de pacientes com câncer do colo do útero, atendidas no Instituto Nacional de Cancer (INCA), Rio de Janeiro, no período entre 1991 e 1995, com o objetivo de identificar os fatores que mais influenciaram na mortalidade das pacientes. O estudo foi realizado em duas etapas, a primeira utilizando uma metodologia não paramétrica através do método de Kaplan-Meier, posteriormente a análise foi ampliada através da utilização de um modelo semi-paramétrico, o modelo de taxas proporcionais de Cox, a partir do qual foi feita uma análise multivariada. Os 362 casos analisados indicaram que o estadiamento do tumor é muito significativo para a sobrevida das pacientes, além desse, dois outros fatores se mostraram marginalmente significativos.

Palavras-chave: Câncer do colo do útero; Modelo de Cox; Análise de Sobrevivência.

1. Introdução

Segundo o INCA (2022), excluídos os tumores de pele não melanoma, o câncer do colo do útero é o terceiro tipo de câncer mais incidente entre as mulheres no Brasil. Nas regiões Norte (20,48/100 mil) e Nordeste (17,59/100 mil) é o segundo mais incidente, na Centro-Oeste (16,66/100 mil) é o terceiro, na região Sul (14,55/100 mil) ocupa a quarta posição e na região Sudeste (12,93/100 mil), a quinta posição.

“Para o ano de 2023 foram estimados 17.010 casos novos, o que representa uma taxa ajustada de incidência de 13,25 casos a cada 100 mil mulheres” (INCA, 2022).

“O câncer do colo do útero é causado por uma infecção persistente pelo papilomavírus humano (HPV), um vírus transmitido sexualmente, muito comum em homens e mulheres” (INCA/Fiocruz, 2018). Estima-se que cerca de 80% das mulheres sexualmente ativas irão adquiri-la ao longo de suas vidas. “A maioria das infecções por HPV em mulheres com menos de 30 anos regride espontaneamente, ao passo que, acima dessa idade, a persistência é mais frequente” (INCA/FIOCRUZ, 2018), se tornando ainda mais comum entre 50 e 60 anos.

Além da infecção pelo HPV e da idade, outros fatores como a imunidade, o tabagismo, a multiplicidade de parceiros sexuais, a multiparidade, o uso de contraceptivos orais, o início precoce da vida sexual, o baixo nível socioeconômico e as deficiências nutricionais também são considerados fatores de risco para o desenvolvimento do câncer do colo do útero. (MS/INCA, 2002).

O principal fator prognóstico do câncer cervical é o estadiamento do tumor. De acordo com a Federação Internacional de Ginecologia e Obstetrícia (FIGO), o estadiamento tumoral está diretamente associado com a sobrevida global e a chance de recorrência local. No entanto, a FIGO não avalia vários outros fatores como a idade, a cor da pele, a condição socioeconômica e o tipo histológico do tumor, que também foram relatados como fatores prognósticos (CARMO & LUIZ, 2011).

Conhecer o estágio do tumor, ajuda a definir o tipo de tratamento e a prever o prognóstico da paciente, uma vez que ele fornece informações sobre a localização, disseminação e se afeta as funções de outros órgãos do corpo.

O objetivo deste artigo é analisar a sobrevida de pacientes com câncer do colo do útero, atendidas em hospitais do INCA, Rio de Janeiro, no período entre 1991 e 1995,

buscando identificar os fatores de prognóstico que mais influenciaram na mortalidade das pacientes.

Nas análises estatísticas foi utilizado o pacote estatístico R v. 4.2.3 e o pacote survival (THERNEAU, 2023)

2. Metodologia

A Análise de Sobrevivência é um ramo da Estatística que está relacionado com o estudo de variáveis aleatórias que medem o tempo até um evento. Na área médica, o evento pode ser o tempo até a morte, cura ou reincidência de uma doença.

Uma característica determinante que distingue a Análise de Sobrevivência de outros ramos da Estatística é a presença de dados censurados. A censura se refere a situações nas quais não foi possível se observar o tempo até o evento, seja pelo término do estudo, ou o acompanhamento do paciente foi interrompido como, por exemplo, um paciente que mudou de cidade e por este motivo não poderá mais participar de um estudo.

2.1 Análise não paramétrica

O estimador de Kaplan-Meier, muito usado em estudos clínicos, é uma técnica não-paramétrica que permite estimar a função de sobrevivência na presença de censura à direita. A função de sobrevivência estimada através dele é dada por(1):

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

onde d_j é o número de mortes ocorridas em t_j , m_j são os tempos censurados no intervalo $[t_j, t_{j+1})$ e $n_j = (m_j + d_j) + \dots + (m_k + d_k)$ representa o número de indivíduos sob risco exatamente antes do tempo $t_j (j = 1, \dots, k)$.

A estimativa de Kaplan-Meier (1), utilizada em conjunto com o teste do log-rank, nos permite avaliar se existe diferença entre curvas de sobrevivências, ou seja, se a taxa de mortalidade é diferente entre duas ou mais curvas de sobrevivência.

Esta análise inicial nos permite ter uma visão preliminar das possíveis variáveis, ou fatores de prognóstico, que possam estar afetando a mortalidade das pacientes. Esta análise deve ser complementada por uma análise conjunta, com todas as variáveis disponíveis, para que possamos ter uma percepção mais adequada da significância dos fatores de prognóstico.

2.1 Análise semiparamétrica

No modelo de Cox as variáveis explicativas, x_1, \dots, x_p , agem de forma multiplicativa na função taxa de falhas base. Os efeitos das variáveis explicativas, ou fatores de prognóstico, não muda ao longo do tempo. Por exemplo, elas podem duplicar ou reduzir à metade a taxa de falhas base, a depender dos fatores de riscos dos pacientes. O modelo de Cox é representado por (2):

$$h(t|X) = h_0(t)e^{(x_1\beta_1 + \dots + x_p\beta_p)} \quad (2)$$

A função $h_0(t)$, que representa a função taxa de falhas base, é uma função não especificada e arbitrária. Por este motivo o modelo de Cox (2) é definido como sendo um modelo semiparamétrico.

Um dos principais resultados obtidos a partir do modelo de Cox é a razão de risco ou “*hazard ratio*” que permite avaliarmos os efeitos de um fator de prognóstico em seus diferentes níveis. A razão de risco (RR ou HR) para um indivíduo com variáveis explicativas X^* comparado com outro indivíduo com variáveis explicativas X é (3):

$$RR(X^*: X) = \frac{e^{(X^*\beta)}}{e^{(X\beta)}} = e^{\{(X^*-X)\beta\}} \quad (3)$$

A razão de risco (3), RR (ou HR) mostra o quanto o risco aumenta ($RR > 1$) ou diminui ($RR < 1$) quando aumentamos em 1 unidade o valor de uma variável binária ou contínua.

3. Análise Exploratória

Os dados utilizados são provenientes dos Registros Hospitalares de Câncer (RHC) do INCA e referem-se às pacientes com o diagnóstico de câncer do colo do útero atendidas nos hospitais que compõem a rede do INCA, no Rio de Janeiro, no período compreendido entre 1º de janeiro de 1991 e 31 de dezembro de 1995 (SANTOS & MOTTA, 2005; SANTOS, 2023).

O banco de dados representa uma coorte com 362 registros de pacientes atendidas. No estudo foram utilizadas oito das variáveis existentes, elas foram selecionadas buscando ampliar a indicação da FIGO. Na Tabela 1 abaixo estão descritas as variáveis utilizadas.

Tabela 1 - Descrição das variáveis do banco de dados.

Variável	Descrição	Categorias originais	Categorias usadas no estudo
conjugal	Estado conjugal	Casada Separada Solteira Viúva	Casada Separada/Viúva Solteira
cor	Cor da pele	Amarela Branca Indígena Parda Preta	Branca Não branca
instruca	Grau de instrução	Nenhum Fundamental incompleto Fundamental completo Médio incompleto Nível Superior	Nenhum Fundamental incompleto/completo Médio incompleto/ Nível superior
estadiamod	Estadiamento do tumor	Estágios 1 a 4	Estágios 1 a 4
dfumante	Tabagismo	Ex fumante Não Sim	Ex fumante Não Sim
papanico	Papanicolau realizado	Não Sim	Não Sim
npartos	Número de partos	0 a 22 partos	0 1 a 4 5 ou mais
idade	Idade	24 a 87 anos	Menos de 65 65 ou mais

Fonte: Elaborado pelos autores

Foi feita uma análise exploratória para organizar e examinar os dados resumindo as suas características principais com o objetivo de identificar o perfil das pacientes da coorte. Os fatores demográficos e clínicos das pacientes podem ser vistos na Tabela 2. A partir desta análise inicial, foi eliminada a variável Papanicolau por conter 117 dados faltantes nos 362 registros.

Tabela 2 - Fatores demográficos e clínicos das pacientes.

	Categorias	N	%
Total		362	100
Estado conjugal			
	Casada	155	42,8
	Outros	207	57,8
Cor da pele			
	Branca	225	62,2
	Não branca	137	37,8
Grau de instrução			
	Nenhum	67	18,5
	Fundamental incompleto/completo	248	68,5
	Médio incompleto/Nível superior	47	13,0
Estadiamento do tumor			
	1	112	30,9
	2	102	28,2
	3	138	38,1
	4	10	2,8
Tabagismo			
	Ex fumante	14	3,9
	Não	249	68,8
	Sim	99	27,3
Número de partos			
	0	35	9,7
	1 a 4	224	61,9
	5 ou mais	103	28,4
Idade			
	Menos de 65	290	80,1
	65 ou mais	72	19,9

Fonte: Elaborado pelos autores

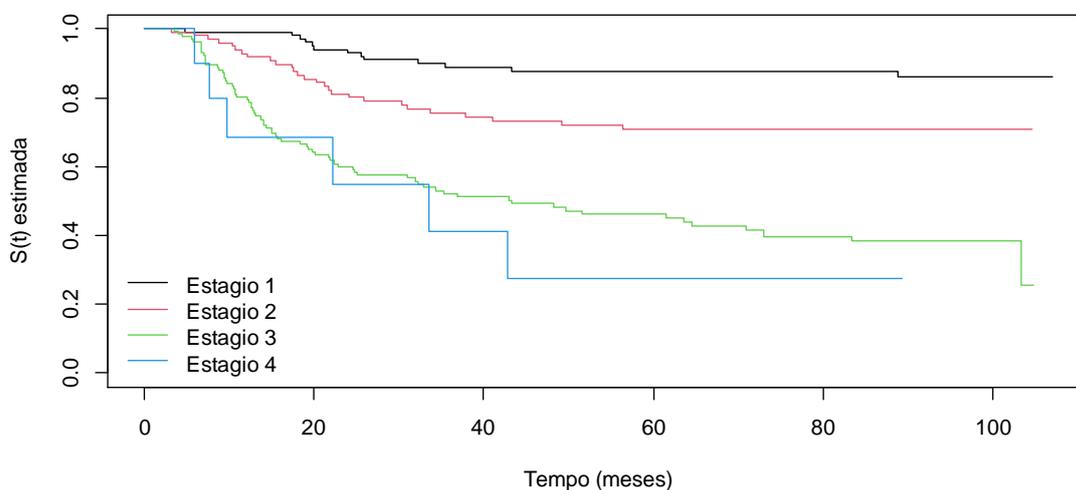
4. Análise não paramétrica

Todas as variáveis da Tabela 2 tiveram suas curvas de sobrevivência estimadas através do método de Kaplan-Meier (1). Em todos os casos foi realizado o teste do log-rank para testar a hipótese de igualdade das curvas de sobrevivência em cada uma das categorias existentes. O nível de significância adotado foi de 0,05 (5%).

Vamos apresentar os resultados das que se mostraram com diferenças significativas no teste do log-rank.

Na Figura 1 abaixo vemos o resultado da variável estadiamento do tumor que se mostrou relevante e bastante significativa no teste do log-rank.

Figura 1 - Gráfico de sobrevivência da variável estadiamento do tumor.



Fonte: Elaborado pelos autores

Na Tabela 3 abaixo vemos o resultado do teste do log-rank para a variável estadiamento.

Tabela 3 - Teste do log-rank da variável estadiamento do tumor.

Estadiamento	N	Observado	Esperado	(O-E) ² /E	(O-E) ² /V
1	112	13	43,8	21,63	34,26
2	102	27	36,5	2,47	3,55
3	138	74	37,5	35,39	51,90
4	10	6	2,2	6,59	6,73

Chisq= 67.1 em 3 graus de liberdade, p= 2e-14

Fonte: Elaborado pelos autores

O teste do log-rank aplicado nas demais variáveis indicou que o número de partos, o grau de instrução e o estado conjugal eram marginalmente significativas. As figuras 2, 3 e 4, além das Tabelas 4,5 e 6 apresentam os resultados obtidos. As demais variáveis não apresentaram significância estatística no nível de significância de 0,05.

Figura 2 - Gráfico de sobrevivência da variável estado conjugal.

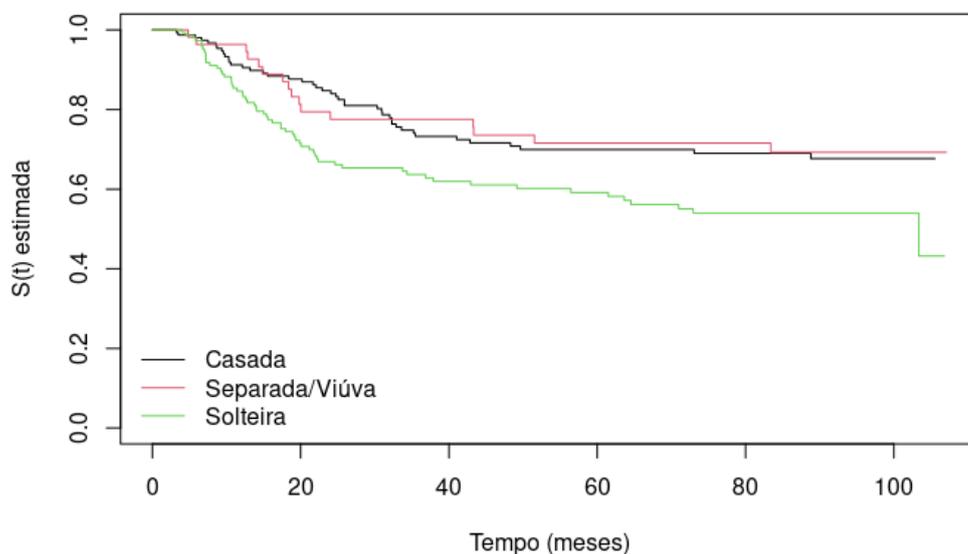


Tabela 4 - Teste do log-rank da variável estado conjugal.

Estadiamento	N	Observado	Esperado	(O-E)^2/E	(O-E)^2/V
Casada	155	43	53.9	2.19	3.97
Separada/Viúva	55	16	20.7	1.08	1.3
Solteira	152	61	45.4	5.34	8.62

Chisq= 8.6 on 2 degrees of freedom, p= 0.01

Fonte: Elaborado pelos autores

Figura 3 - Gráfico de sobrevivência da variável grau de instrução.

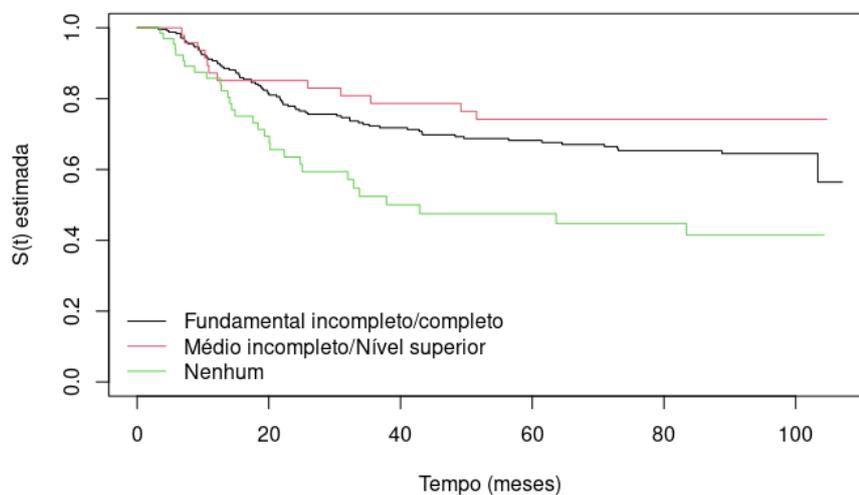


Tabela 5 - Teste do log-rank da variável grau de instrução.

Estadiamento	N	Observado	Esperado	$(O-E)^2/E$	$(O-E)^2/V$
Fundamental incompleto/completo	248	78	84.5	0.50	1.69
Médio incompleto/Nível superior	47	12	18.6	2.35	2.79
Nenhum	67	30	16.9	10.18	11.92

Chisq= 13.1 on 2 degrees of freedom, p= 0.001

Fonte: Elaborado pelos autores

Figura 4 - Gráfico de sobrevivência da variável número de partos.

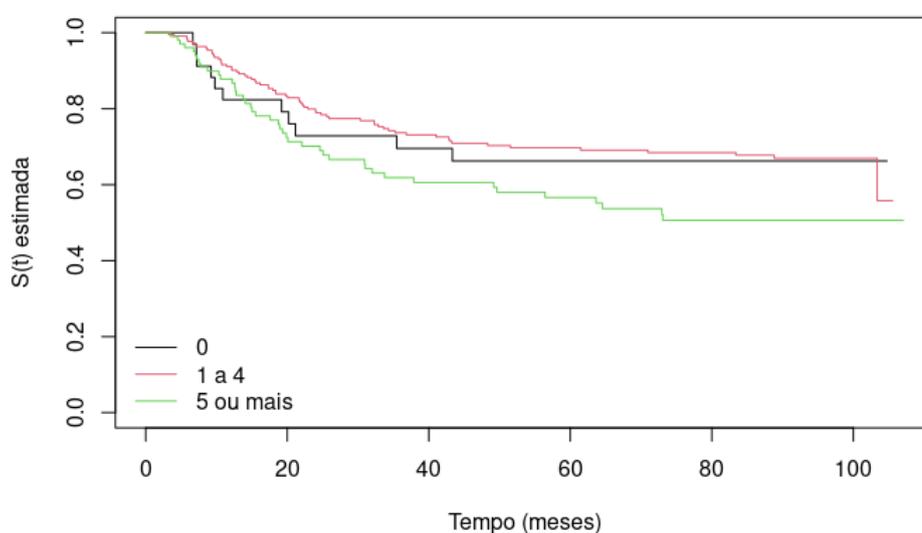


Tabela 6 - Teste do log-rank da variável número de partos.

Estadiamento	N	Observado	Esperado	$(O-E)^2/E$	$(O-E)^2/V$
0	35	11	11.9	0.0705	0.0786
1 a 4	224	66	77.3	1.6652	4.6932
5 ou mais	103	43	30.7	4.8951	6.5954

Chisq= 6.6 on 2 degrees of freedom, p= 0.04

Fonte: Elaborado pelos autores

O próximo passo será avaliar se esta análise univariada é confirmada numa análise conjunta através do modelo de Cox.

4. Modelo de Cox

Na avaliação das variáveis explicativas através do modelo de Cox (2), foi usado um nível de significância de 0,05. Neste contexto, as variáveis grau de instrução, tabagismo, idade e cor não se mostraram estatisticamente significativas, no conjunto de dados analisado. As variáveis número de partos e estado conjugal eram marginalmente significativas ($p < 0,05$). O estadiamento do tumor se mostrou muito significativo ($p < 0,001$) na sobrevida das pacientes. A complexidade dos dados indica o uso de um modelo que possibilite sua análise conjunta (COLOSIMO & GIOLO, 2006).

Após a análise inicial, foram mantidas apenas as variáveis significativas. Na Tabela 7 temos os resultados do modelo de Cox com as variáveis estadiamento do tumor, número de partos e estado conjugal.

Tabela 7 – Resultados do Modelo de Cox.

	coef	Erro padrão(coef)	z	p-valor
Estadiamento 2	0,917	0,338	2,714	0,0067
Estadiamento 3	1,900	0,302	6,296	$3,06 \times 10^{-10}$
Estadiamento 4	2,468	0,501	4,928	$8,31 \times 10^{-7}$
Número de partos	0,063	0,024	2,667	0,008
Estado conjugal=Outros	0,461	0,194	2,384	0,017

Fonte: Elaborado pelos autores

A Associação Americana de Câncer (AMERICAN CANCER SOCIETY, 2023) e o MS/INCA (2002) apontam diversos riscos para o desenvolvimento de câncer na região cervical, entre eles estão listados, o histórico sexual e o número de gestações completas. O número de partos tem uma correlação imediata com o número de gestações completas, mas o histórico sexual não tem uma relação direta com o estado conjugal. A variável estado conjugal pode estar representando um efeito que não é seu.

Para avaliar se estava ocorrendo alguma interação entre a variável estado conjugal e número de partos, foi testada esta interação através do modelo de Cox. Os resultados desta interação não se mostram significativos. Em função dos resultados obtidos em Carmo e Luiz (2011) e os fatores de riscos apontados pela Associação Americana de Câncer e o MS/INCA, foi definida a exclusão desta variável do modelo.

Na Tabela 8 abaixo estão listados a razão de risco (3) dos níveis das variáveis estadiamento e número de partos.

Tabela 8 – Razão de Risco (Hazard Ratio) do Modelo de Cox.

	HR
Estadiamento 1	1,00
Estadiamento 2	2,51
Estadiamento 3	6,65
Estadiamento 4	10,22
Número de partos	1,058

Fonte: Elaborado pelos autores

Os resultados da Tabela 8 indicam um efeito multiplicativo na taxa de mortalidade base de 2,51 no caso de uma paciente no estágio 2, 6,65 no estágio 3 e de 10,22 no estágio 4. O efeito do número de partos foi bem menor e representou um efeito 1,058 a cada nova gestação completa.

5. Conclusões

É importante destacar que este estudo apresenta algumas limitações. A mais significativa está no fato da base de dados abranger o período de 1991 a 1995, com somente 362 pacientes, apresentando diversas variáveis com dados ausentes. Os dados foram coletados seguindo a metodologia do Instituto Nacional do Câncer (INCA), seguindo as definições de controle de qualidade correspondentes ao período da coleta. A presença de dados ausentes é fato comum em diversas bases de dados, especialmente os dados demográficos e clínicos. O uso de uma base de dados como a utilizada em Carmo e Luiz (2011) permitiria uma análise mais ampla, principalmente considerando o tamanho da coorte utilizada.

Apesar destas limitações, os resultados obtidos estão coerentes com os resultados dos artigos de Carmo e Luiz (2011) e Alves *et. al.* (2017), em que o estadiamento do tumor se mostrou como o fator de prognóstico mais significativo. As razões de risco (ou *hazard ratios* – HR), para este fator de prognóstico, indicaram um efeito significativo no aumento da mortalidade para os níveis mais elevados do estadiamento.

Os resultados obtidos ressaltam a necessidade de detecção precoce deste tipo de câncer. Quando diagnosticado de forma precoce, o câncer do colo do útero é uma das neoplasias com maior chance de cura.

Referências

ALVES, R. J. V.; WATTE, G.; GARCEZ, A. S.; ARMANDO, A.; MOTTA, N. W.; ZELMANOWIC, A. M. Sobrevida de mulheres com câncer de colo uterino em um centro de referência do sul do Brasil. **Brazilian Journal of Oncology**, Porto Alegre, v. 13, n. 46, p. 1-7, 2017.

AMERICAN CANCER SOCIETY. Disponível em: <https://www.cancer.org/cancer/types/cervical-cancer/causes-risks-prevention/risk-factors.html>. Acesso em: maio de 2023.

CARMO, C. C; LUIZ, R. R. Sobrevida de mulheres com câncer de colo uterino diagnosticadas em um centro brasileiro. **Revista Saúde Pública**, Rio de Janeiro, v. 45, n. 4, p. 661-667, 2011.

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

INSTITUTO NACIONAL DE CÂNCER. Dados e números sobre câncer do colo do útero: **relatório anual 2022**. Rio de Janeiro: INCA, 2022.

INSTITUTO NACIONAL DE CÂNCER. **Sistema de Registro Hospitalar de Câncer (SisRHC)**. Disponível em: <https://irhc.inca.gov.br/RHCNet/>. Acesso em: janeiro de 2023.

MINISTÉRIO DA SAÚDE. INSTITUTO NACIONAL DE CÂNCER. FUNDAÇÃO OSWALDO CRUZ. **A mulher e o câncer do colo do útero**. Rio de Janeiro: INCA/FIOCRUZ, 2018.

MINISTÉRIO DA SAÚDE. SECRETARIA DE ASSISTÊNCIA À SAÚDE. INSTITUTO NACIONAL DE CÂNCER. **Falando sobre câncer do colo do útero**. Rio de Janeiro: MS/INCA, 2002.

SANTOS, N.A; MOTTA, R. L. F. Aplicação do modelo de taxas proporcionais de Cox no estudo de caso de câncer do colo do útero. **Monografia** (Bacharelado em Estatística) – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2005.

SANTOS, T.R.A. Sobrevida de mulheres com câncer cervical atendidas em hospitais do INCA entre 1991 e 1995 e fatores prognósticos associados. **Monografia** (Bacharelado em Estatística) – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2023.

THERNEAU, T. **A Package for Survival Analysis in R**. R package version 3.5-0, <https://CRAN.R-project.org/package=survival>., 2023.

SURVIVAL ANALYSIS OF WOMEN WITH CERVICAL CANCER

Abstract

Cervical cancer is a public problem in Brazil. This study analyzes the survival of patients with cervical cancer treated at Instituto Nacional de Cancer (INCA), Rio de Janeiro, between 1991 and 1995, with the aim of identifying the factors that most influenced the patients' mortality. The study was done in two stages, the first using a non-parametric methodology through the Kaplan-Meier method, later the analysis was expanded using a semi-parametric model, the Cox proportional hazards model, from which a multivariate analysis was performed. The 362 analyzed cases indicated that tumor staging is very significant for patient survival, two other factors that were marginally significant.

Keywords: *Cervical cancer; Cox Model; Survival Analysis.*