

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE ESTATÍSTICA

MIRIAN KELLY LEITE SOUZA

**UTILIZAÇÃO DO MODELO DE COX EM DADOS DE PACIENTES  
DIAGNOSTICADOS COM CARCINOMA ESPINOCELULAR NA  
OROFARINGE**

OURO PRETO  
2026

MIRIAN KELY LEITE SOUZA

UTILIZAÇÃO DO MODELO DE COX EM DADOS DE PACIENTES  
DIAGNOSTICADOS COM CARCINOMA ESPINOCELULAR NA OROFARINGE

Monografia apresentada ao Curso de Estatística da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Estatística.

Orientadora: Graziela Dutra Rocha Gouvea

OURO PRETO  
2026



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DE OURO PRETO  
REITORIA  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
COLEGIADO DO CURSO DE ESTATÍSTICA



**FOLHA DE APROVAÇÃO**

Mirian Kely Leite Souza

Utilização do modelo de COX em dados de pacientes diagnosticados com carcinoma espinocelular na orofaringe

Monografia apresentada ao Curso de Estatística da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Estatística

Aprovada em 04 de março de 2026

**Membros da banca**

Dra. Graziela Dutra Rocha Gouvea - Orientadora (Universidade Federal de Ouro Preto)  
Dra. Diana Campos de Oliveira (Universidade Federal de Ouro Preto)  
Dr. Fernando Luiz Pereira de Oliveira (Universidade Federal de Ouro Preto)

Professora Dra. Graziela Dutra Rocha Gouvea, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 04/03/2026



Documento assinado eletronicamente por **Graziela Dutra Rocha Gouvea, PROFESSOR DE MAGISTERIO SUPERIOR**, em 12/03/2026, às 09:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1070374** e o código CRC **D9216870**.

## **AGRADECIMENTOS**

A minha família por todo amor, carinho, apoio e confiança. Em especial, minha irmã, Emily, obrigada pelas incontáveis ligações que venceram as barreiras da distância.

As amigas que a faculdade me trouxe, Maria Luisa e Camila.

As amigas que a vida me deu, Ruth, Nathalia e Jessica. A vida é complicada, mas ao lado de grandes amigos ela se torna bem mais leve e bonita.

A Raquel e Ediney, que sempre me acolheram de braços abertos.

Aos professores do departamento de Estatística da UFOP, por todo ensinamento transmitido.

A minha orientadora, Graziela pela paciência, incentivo e apoio no desenvolvimento deste trabalho.

Obrigada a todos por me acompanharem nesta caminhada!

## **RESUMO**

O objetivo deste estudo foi analisar o tempo de sobrevida de pacientes diagnosticados com carcinoma espinocelular da orofaringe, enfocando especialmente os modelos de Kaplan Meier e de riscos proporcionais de Cox. Foram utilizados dados provenientes de um estudo clínico realizado pelo Radiation Therapy Oncology Group, contendo informações clínicas de 195 pacientes. Realizamos inicialmente a análise descritiva e a comparação das curvas de sobrevivência por meio do estimador de Kaplan-Meier e do teste de log-rank, com posterior dicotomização das covariáveis. Em seguida, ajustou-se o modelo de regressão de riscos proporcionais de Cox para identificar os fatores prognósticos associados ao risco de óbito. Os resultados indicaram que a condição funcional do paciente, o estadiamento tumoral e o comprometimento linfonodal exerceram influência significativa sobre a sobrevida, evidenciando maior risco de morte em pacientes com doença mais avançada e pior estado clínico geral. A avaliação dos resíduos de Schoenfeld foi empregada para verificação da suposição de proporcionalidade dos riscos, indicando adequação global do modelo. Conclui-se que a análise de sobrevivência constitui uma importante ferramenta estatística para compreensão da evolução clínica do carcinoma espinocelular da orofaringe, permitindo identificar fatores relevantes para o prognóstico.

**Palavras-chave:** carcinoma espinocelular, orofaringe, Kaplan-Meier, análise de sobrevivência.

## **ABSTRACT**

The objective of this study was to analyze the survival time of patients diagnosed with oropharyngeal squamous cell carcinoma, with particular emphasis on the Kaplan–Meier and Cox proportional hazards models. Data were obtained from a clinical study conducted by the Radiation Therapy Oncology Group, comprising clinical information from 195 patients. Initially, a descriptive analysis was performed, and survival curves were compared using the Kaplan–Meier estimator and the log-rank test, followed by dichotomization of the covariates. Subsequently, the Cox proportional hazards regression model was fitted to identify prognostic factors associated with the risk of death. The results indicated that patients' functional status, tumor staging, and lymph node involvement had a significant influence on survival, demonstrating a higher risk of death among patients with more advanced disease and poorer overall clinical condition. The evaluation of Schoenfeld residuals was employed to verify the proportional hazards assumption, indicating the overall adequacy of the model. It is concluded that survival analysis constitutes an important statistical tool for understanding the clinical progression of oropharyngeal squamous cell carcinoma, allowing the identification of relevant prognostic factors.

**Keywords:** squamous cell carcinoma, oropharynx, Kaplan–Meier, survival analysis.

## LISTA DE FIGURAS

|                                                                                                                          |    |
|--------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1 – Curvas de sobrevivência (Kaplan Meier) para comparação entre os grupos da covariável GRADE.....               | 9  |
| Figura 2 – Curvas de sobrevivência (Kaplan Meier) para comparação entre os grupos da covariável SITE.....                | 10 |
| Figura 3 – Curvas de sobrevivência (Kaplan Meier) para comparação entre os grupos da covariável T_STAGE.....             | 11 |
| Figura 4 – Curvas de sobrevivência (Kaplan Meier) para comparação entre os grupos da covariável N_STAGE.....             | 12 |
| Figura 5 – Curvas de sobrevivência (Kaplan Meier) para comparação entre os grupos da covariável COND.....                | 14 |
| Figura 6 – Resíduos padronizados de Schoenfeld em função do tempo para as covariáveis GRADE, T_STAGE, N_STAGE, COND..... | 17 |
| Figura 7 – Curvas $[\log(\Lambda(t))]$ versus t para a covariável COND.....                                              | 18 |

## LISTA DE TABELAS

|                                                                                                                                            |    |
|--------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tabela 1 - Agrupamento das covariáveis pela semelhança ou distinção entre elas.....                                                        | 8  |
| Tabela 2 - Resultados da comparação 2 a 2 pelo teste de log rank para a covariável Grade.....                                              | 10 |
| Tabela 3 - Resultados da comparação 2 a 2 pelo teste de log rank para a covariável SITE.....                                               | 11 |
| Tabela 4 - Resultados da comparação 2 a 2 pelo teste de log rank para a covariável T-STAGE.....                                            | 12 |
| Tabela 5 - Resultados da comparação 2 a 2 pelo teste de log rank para a covariável N-STAGE.....                                            | 13 |
| Tabela 6 - Resultados da comparação 2 a 2 pelo teste de log rank para a covariável COND.....                                               | 14 |
| Tabela 7 - Ajuste do modelo final com as estimativas dos coeficientes, suas razões de risco e os respectivos erros-padrão e valores p..... | 15 |
| Tabela 8 - Avaliação da Proporcionalidade das taxas de falha no modelo.....                                                                | 16 |

## SUMÁRIO

|                                                 |    |
|-------------------------------------------------|----|
| 1 INTRODUÇÃO .....                              | 1  |
| 2 OBJETIVOS .....                               | 2  |
| 3 REVISÃO DE LITERATURA .....                   | 2  |
| 3.1 Carcinoma espinocelular da orofaringe ..... | 2  |
| 3.2 Análise de sobrevivência .....              | 3  |
| 3.3 Curvas de Kaplan-Meier .....                | 4  |
| 3.4 Teste de Log-Rank .....                     | 5  |
| 3.5 Modelo de Cox .....                         | 5  |
| 4 MATERIAL E MÉTODOS .....                      | 6  |
| 5 RESULTADOS E DISCUSSÕES .....                 | 8  |
| 5.1 Bondade do ajuste .....                     | 15 |
| 6 CONCLUSÕES .....                              | 18 |
| REFERÊNCIAS .....                               | 20 |

## 1. INTRODUÇÃO

Em estudos clínicos na área da saúde, é comum o interesse em analisar o tempo decorrido até a ocorrência de um evento de interesse, como, por exemplo, o óbito ou a recidiva de uma doença. Entretanto, nem todas as observações são acompanhadas durante todo o período do estudo, resultando na presença de dados censurados. Este fato inviabiliza a aplicação de métodos estatísticos tradicionais, exigindo o uso de técnicas mais específicas.

A Análise de Sobrevivência constitui um conjunto de métodos estatísticos desenvolvidos para lidar com dados dessa natureza, em que ocorrem observações censuradas, permitindo, então, a estimação do tempo até a ocorrência do evento de interesse, a comparação entre grupos e a avaliação do efeito de covariáveis sobre o risco associado ao desfecho. Dentre essas técnicas, destacam-se as curvas de Kaplan–Meier, o teste de log-rank e o modelo de riscos proporcionais de Cox.

Muitos estudos na área de saúde apresentam dados censurados e, neste contexto, a Análise de Sobrevivência frequentemente é utilizada para modelar dados relacionados ao tempo de sobrevida de pacientes que possuem as mais diversas doenças, como hipertensão, diabetes, problemas no coração, câncer, entre outros.

Uma doença relevante clinicamente é o carcinoma espinocelular (CEC), definida como uma neoplasia maligna da pele que surge através da reprodução anormal de células atípicas da camada espinhosa da epiderme. O CEC, além de apresentar um comportamento agressivo, tem alto risco de invasão local e metástase.

Nesse contexto, por meio do presente estudo, foi possível utilizar os métodos de análise de sobrevivência para analisar o tempo de sobrevivência de pacientes diagnosticados com carcinoma espinocelular na região da orofaringe, identificando os fatores que influenciam no tempo de sobrevida dos pacientes com tal doença.

## **2. OBJETIVOS**

Aplicar as técnicas de análise de sobrevivência para explicar as covariáveis e o tempo de sobrevida dos pacientes diagnosticados com carcinoma espinocelular na orofaringe, pelo modelo de regressão de riscos proporcionais de Cox (Cox, 1972). Para isto, definimos os seguintes objetivos específicos:

- Descrever o conjunto de dados do estudo, caracterizando o perfil clínico dos pacientes e quantificar a presença de censuras;
- Dicotomizar as covariáveis por meio da comparação das curvas de sobrevivência utilizando o método de Kaplan Meier (Kaplan e Meier, 1958) e o teste de log-rank (Mantel, 1966);
- Ajustar o modelo de regressão de Cox aos dados e identificar fatores associados a sobrevivência;
- Avaliar a adequação do modelo ajustado por meio da análise de resíduos;

## **3. REVISÃO DE LITERATURA**

### **3.1 Carcinoma Espinocelular da Orofaringe**

O carcinoma espinocelular pertence ao grupo não melanoma, tipo de tumor que se desenvolve em células que não produzem melanina. Dos tipos de tumor cutâneo, o não melanoma é o mais comum e frequente no Brasil apresentando uma mortalidade baixa, mas pode resultar em mutilações graves se não receber o tratamento adequado (INCA,2022).

O CEC origina-se na epiderme, camada mais superficial da pele e resulta no desenvolvimento de uma lesão robusta, avermelhada e descamativa que apresenta hemorragia e não cicatriza. Podendo ser semelhante a uma verruga. Afeta, em geral, indivíduos acima dos 50 anos, sendo mais frequente em homens fumantes e com alta exposição solar, especialmente aqueles de pele, olhos e cabelos claros (PETRI, 2009).

No âmbito específico da cavidade oral, a orofaringe compreende estruturas da base da língua, o palato mole, as amígdalas e a parede faríngea. O CEC nessa região apresenta particular importância clínica devido à proximidade com estruturas vitais e à elevada frequência de metástases linfonodais. Fatores como a extensão do tumor primário, o comprometimento linfonodal e o grau histológico são determinantes para o desfecho clínico (ENDO; TANIOKA; MIYACHI, 2011). Adicionalmente, a condição funcional do paciente pode levar a um aumento de comorbidades e da mortalidade (WEDDING et al., 2007).

Devido a presença de dados censurados, os métodos estatísticos tradicionais se tornam inadequados para este tipo de investigação, sendo a análise de sobrevivência a mais indicada. Nesse contexto, o tempo até a ocorrência de desfechos clínicos, como óbito ou recidiva da doença, torna-se uma informação central para a compreensão da evolução do carcinoma espinocelular.

### **3.2 Análise de Sobrevivência**

Na análise de sobrevivência, a variável resposta é, frequentemente, definida como o tempo transcorrido até a ocorrência de um evento de interesse. Esse intervalo, conhecido como tempo até a falha, pode representar, por exemplo, o tempo até o óbito de um paciente submetido a determinado tratamento, o tempo até a cura ou recidiva de uma doença (PEREIRA, 2002).

Entretanto, nem sempre é possível observar o tempo exato de ocorrência do evento para todos os indivíduos acompanhados. Em muitos estudos, alguns participantes não apresentam o evento durante o período de observação ou têm seu acompanhamento interrompido antes do término da pesquisa. Nesses casos, surgem os chamados dados censurados, nos quais o tempo de ocorrência do evento não é completamente observado, sendo conhecido apenas que ele excede ou antecede determinado momento.

Entre os principais mecanismos, destaca-se a censura à direita, mais comum em estudos de sobrevivência, na qual o tempo observado é inferior ao

tempo real do evento. Já na censura à esquerda, o evento de interesse ocorre antes do início da observação, tornando o seu instante exato desconhecido. Adicionalmente, a censura intervalar caracteriza-se quando a ocorrência é detectada apenas dentro de um intervalo de tempo entre duas avaliações consecutivas (COLOSIMO; GIOLO, 2024).

Segundo os autores, as censuras também podem ser classificadas conforme o critério de interrupção do estudo. Na Tipo I, o acompanhamento é encerrado em um tempo previamente fixado, podendo resultar em observações censuradas para indivíduos que não apresentaram o evento até o término do estudo. Já na Tipo II, o monitoramento é finalizado quando é atingido um número previamente estabelecido de ocorrências do evento de interesse.

### 3.3 Curvas de Kaplan Meier

O método de Kaplan Meier, conforme estabelecido pelos seus proponentes, Kaplan e Meier (1958), é uma técnica não paramétrica que permite calcular a probabilidade de sobrevivência de uma amostra ao longo do tempo, lidando de forma eficaz com as censuras.

Representando graficamente a taxa de sobrevivência em função do tempo através de uma curva. Quanto maior a inclinação da curva, maior a taxa de mortalidade e pior o prognóstico. Já curvas menos acentuadas representam taxas de eventos reduzidas e maior chance de sobrevida. Áreas planas sinalizam fases de estabilidade.

A equação do estimador de Kaplan-Meier pode ser definida como:

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

Estabelecendo  $t_1 < t_2 < \dots < t_k$  como os instantes sucessivos de falha, onde  $d_j$  indica o total de eventos no período e  $n_j$  o montante de indivíduos suscetíveis ao risco no tempo  $t_{j-1}$ .

### 3.4 Teste de Log-Rank

O Teste de log-rank (Mantel, 1966) é comumente utilizado para fazer comparação entre grupos, considerando o tempo do início do estudo até o evento, e descobrir se possuem alguma diferença significativa entre eles, para tal testamos as seguintes hipóteses:

Hipótese nula ( $H_0$ ): Os grupos têm curvas de distribuição idênticas.

Hipótese alternativa ( $H_1$ ): Os grupos apresentam curvas de distribuição diferentes.

Para cada tempo  $t_i$  em que ocorre um evento, monta-se uma tabela de contingência  $2 \times 2$ . A estatística do teste é então derivada da soma das diferenças entre eventos observados e esperados ao longo de todo o período, seguindo uma distribuição Qui-quadrado.

Para que o teste seja validado, devem ser respeitadas as premissas de riscos proporcionais, onde assume que a razão de risco entre os grupos permanece constante ao longo do tempo. Ademais, o mecanismo de censura deve ser independente do tempo até a ocorrência do evento, de modo que não introduza viés na comparação das curvas de sobrevivência (KLEINBAUM; KLEIN, 2012).

### 3.5 Modelo de Cox

O modelo de regressão proporcional de Cox (Cox, 1972) viabiliza a análise de estudos cujo desfecho é o tempo até a ocorrência de um evento de interesse, incorporando o efeito de covariáveis no modelo.

Na regressão de Cox, o vetor de coeficientes  $\beta$  expressa a influência das covariáveis na taxa de falha. A determinação do modelo depende da estimação desses parâmetros com base em observações amostrais, o que exige a adoção de um procedimento de estimação (COLOSIMO; GIOLO, 2024).

O modelo geral de Cox é expresso por:

$$\lambda(x) = \lambda_0(t) g(\beta'x)$$

onde:

- $\lambda(x)$ : taxa de falha no tempo  $t$ , dado o vetor de covariáveis
- $\lambda_0(t)$ : função de risco basal
- $\beta$ : vetor de coeficientes
- $x$ : vetor de covariáveis
- $g(\beta'x)$ : componente paramétrico não negativo, que atua da seguinte forma multiplicativa:  $g(\beta'x) = \exp \exp(\beta'x) = \exp \exp(\beta_1 x_1 + \dots + \beta_p x_p)$ .

#### 4. MATERIAL E MÉTODOS

Foi utilizada uma base de dados obtida de a partir de estudo clínico realizado pelo grupo “Radiation Therapy Oncology Group”. O estudo considerou pacientes de 16 instituições diagnosticados com carcinoma espinocelular da boca e garganta, contudo, na base apenas dados de três partes da orofaringe foram localizados. A amostra conteve um total de 195 observações. Para o estudo os pacientes foram selecionados aleatoriamente e submetidos a dois tipos de radioterapia, isolada e a associada à quimioterapia, tendo como objetivo principal avaliar o tempo de sobrevivência.

O tempo de falha é definido pela variável “Sobrevivência”, correspondente ao número de dias entre o diagnóstico inicial e o óbito do paciente. A variável “Status” identifica a censura do tipo à direita, codificada como 0, enquanto o valor 1 indica a ocorrência do evento de interesse.

A base de dados original constava das seguintes informações sobre as covariáveis, tempo de falha e censura:

- Sexo: 1 = masculino, 2 = feminino.
- Tratamento: 1 = padrão, 2 = teste.
- Grau: 1 = bem diferenciado, 2 = moderadamente diferenciado, 3 = pouco diferenciado.
- Idade: Em anos no momento do diagnóstico.

- Condição: 1 = sem incapacidade, 2 = trabalho restrito, 3 = necessita de assistência para cuidados pessoais, 4 = acamado.
- Local: 1 = arco facial, 2 = fossa tonsilar, 3 = pilar posterior, 4 = língua faríngea, 5 = parede posterior.
- Estadiamento T: 1 = tumor primário medindo 2 cm ou menos no maior diâmetro; 2 = tumor primário medindo de 2 a 4 cm no maior diâmetro, infiltração mínima em profundidade; 3 = tumor primário medindo mais de 4 cm; 4 = tumor invasivo maciço.
- Estadiamento N: 0 = nenhuma evidência clínica de metástase em linfonodos; 1 = linfonodo positivo único com 3 cm ou menos no maior diâmetro, não fixo; 2 = linfonodo positivo único com mais de 3 cm de diâmetro, não fixo; 3 = múltiplos linfonodos positivos ou linfonodos fixos.
- Data de entrada: Dia do ano e ano.
- Status: 0 = censurado, 1 = óbito.
- Sobrevivência: Em dias a partir do dia do diagnóstico.

Como algumas das covariáveis apresentaram mais de uma categoria, tornou-se necessário dicotomizá-las empregando o Teste de log-rank, para comparação das curvas de sobrevivência. Os resultados do teste foram interpretados com base no nível de significância de 0,05. Variáveis com  $p \geq 0,05$  resultaram na aceitação da hipótese nula, ao passo que valores de  $p < 0,05$  indicaram a rejeição da mesma, confirmando a existência de diferenças significativas entre os grupos comparados.

A Tabela 1 contém as covariáveis dicotomizadas.

Tabela 1 - Agrupamento das covariáveis pela semelhança ou distinção entre elas.

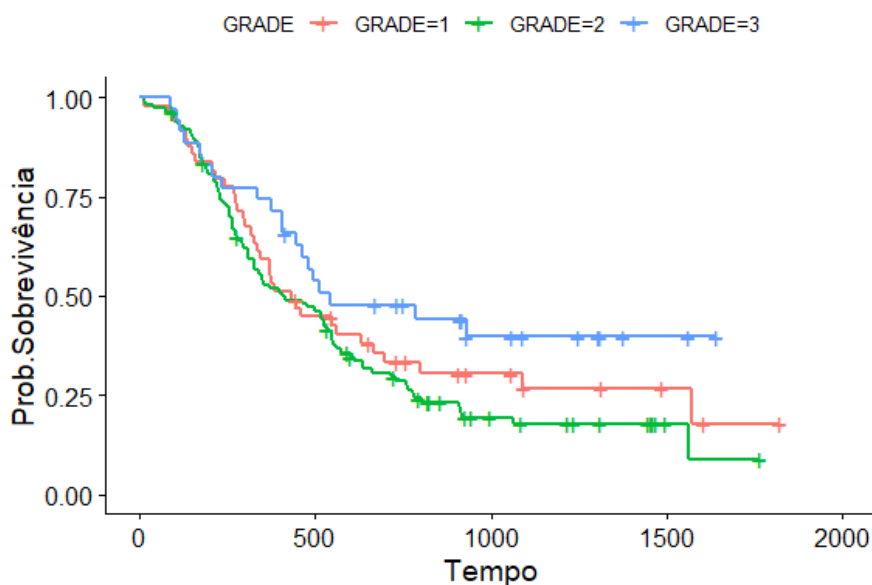
| <b>Covariável</b>           | <b>Valor 0</b>                                                                                                                                                                                 | <b>Valor 1</b>                                     | <b>Legenda</b> |
|-----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|----------------|
| <b>Grau da doença</b>       | Moderadamente diferenciado; bem diferenciado                                                                                                                                                   | Pouco diferenciado                                 | GRADE          |
| <b>Local da doença</b>      | Arco facial                                                                                                                                                                                    | Fossa tonsilar; língua faríngea                    | SITE           |
| <b>Estadiamento T</b>       | Tumor primário medindo 2 cm ou menos no maior diâmetro; tumor primário de 2 a 4 cm no maior diâmetro, infiltração mínima em profundidade; e tumor primário mais de 4 cm                        | Tumor invasivo maciço                              | T_STAGE        |
| <b>Estadiamento N</b>       | Nenhuma evidência clínica de metástase em linfonodos; linfonodo positivo único com 3 cm ou menos no maior diâmetro, não fixo; linfonodo positivo único com mais de 3 cm de diâmetro, não fixo; | Múltiplos linfonodos positivos ou linfonodos fixos | N_STAGE        |
| <b>Condição do paciente</b> | Trabalho restrito; precisa de assistência para cuidados pessoais; acamado                                                                                                                      | Sem incapacidade                                   | COND           |
| <b>Sexo</b>                 | Feminino                                                                                                                                                                                       | Masculino                                          | SEX            |
| <b>Tratamento</b>           | Teste                                                                                                                                                                                          | Padrão                                             | TRT            |

Fonte: Elaborado pela autora.

## 5. RESULTADOS E DISCUSSÕES

De acordo com Bland e Altman (1998), a curva de Kaplan-Meier é caracterizada graficamente como uma função em degraus, em que cada queda vertical representa a ocorrência do evento de interesse, enquanto as marcas horizontais (se presentes) indicam os dados censurados, permitindo uma visualização direta da evolução da sobrevida livre de eventos.

O gráfico de Kaplan-Meier referente ao Grau da Doença (GRADE), evidencia diferenças no comportamento das curvas de sobrevivência entre os grupos analisados. O que é confirmado através do teste de log-rank na Tabela 2, que indica diferença estatisticamente significativa,  $p = 0,03$ , entre os grupos 2 e 3. Porém entre os grupos 1 e 2 e entre os grupos 1 e 3 não houve diferenças. Logo, pode-se concluir que pacientes classificados com grau pouco diferenciado apresentam menor probabilidade de sobrevivência ao longo do tempo quando comparados aos demais grupos.



Fonte: Elaborado pela autora.

Figura 1 – Curvas de sobrevivência (Kaplan Meier) para comparação entre os grupos da covariável GRADE.

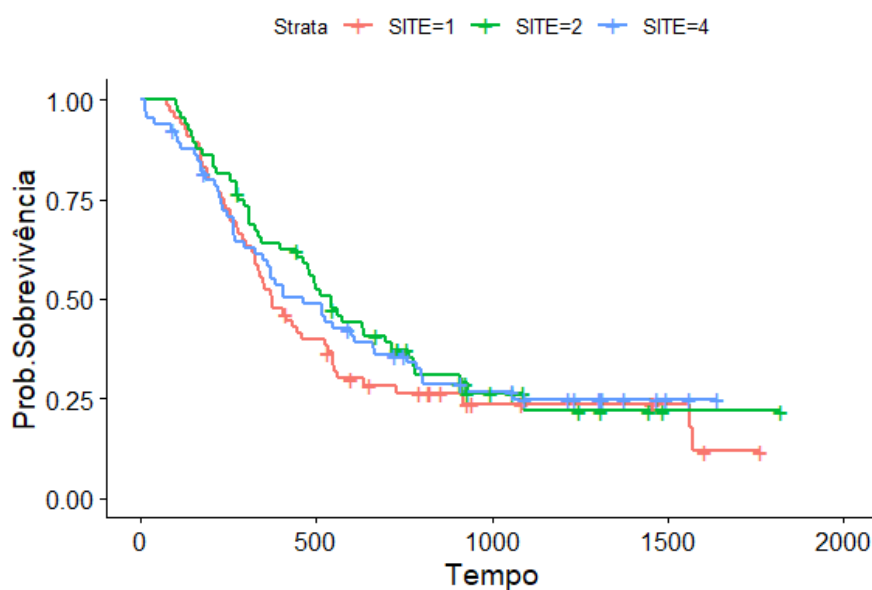
Tabela 2 – Resultados da comparação 2 a 2 pelo teste de log rank para a covariável Grade.

| GRADE        | Estatística do teste | p-valor |
|--------------|----------------------|---------|
| <b>1 X 2</b> | 1                    | 0,313   |
| <b>1 X 3</b> | 1,6                  | 0,2     |
| <b>2 X 3</b> | 4,9                  | 0,03    |

Fonte: Elaborado pela autora.

No que se refere ao Local da Doença (SITE), as curvas de sobrevivência são semelhantes não demonstrando associação significativa com o tempo de sobrevida nesta amostra. Os resultados do teste de log-rank, na Tabela 3, também não indicaram diferenças estatisticamente significativas.

Dessa forma, para a dicotomização fez-se a menor separação possível, agrupando as covariáveis correspondentes aos locais 2 e 4, fossa tonsilar e língua faríngea respectivamente.



Fonte: Elaborado pela autora.

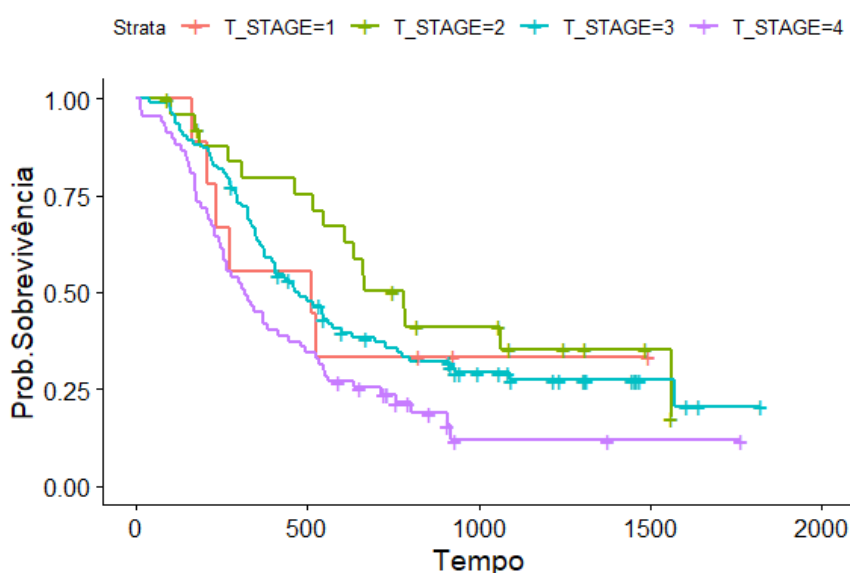
Figura 2 – Curvas de sobrevivência (Kaplan Meier) para comparação entre os grupos da covariável SITE.

Tabela 3 - Resultados da comparação 2 a 2 pelo teste de log rank para a covariável SITE.

| SITE         | Estatística do teste | p-valor |
|--------------|----------------------|---------|
| <b>1 X 2</b> | 1,3                  | 0,2     |
| <b>1 X 4</b> | 0,4                  | 0,5     |
| <b>2 X 4</b> | 0,1                  | 0,7     |

Fonte: Elaborado pela autora.

Para o estadiamento T, (T-STAGE), nota-se uma distinção pronunciada entre pacientes com tumores invasivo maciço (T4) e aqueles classificados nos estágios iniciais (T1, T2 e T3). As trajetórias demonstradas pelas curvas significam que os indivíduos categorizados como T4 apresentam diminuição na probabilidade de sobrevivência com o passar do tempo. O teste log-rank (Tabela 4) ressaltou as disparidades estatisticamente significativas entre os grupos T2 × T4 ( $p = 0,004$ ) e T3 × T4 ( $p = 0,01$ ), reforçando assim a lógica de manter os estágios iniciais em um grupo único, e o estágio avançado como uma classificação separada.



Fonte: Elaborado pela autora.

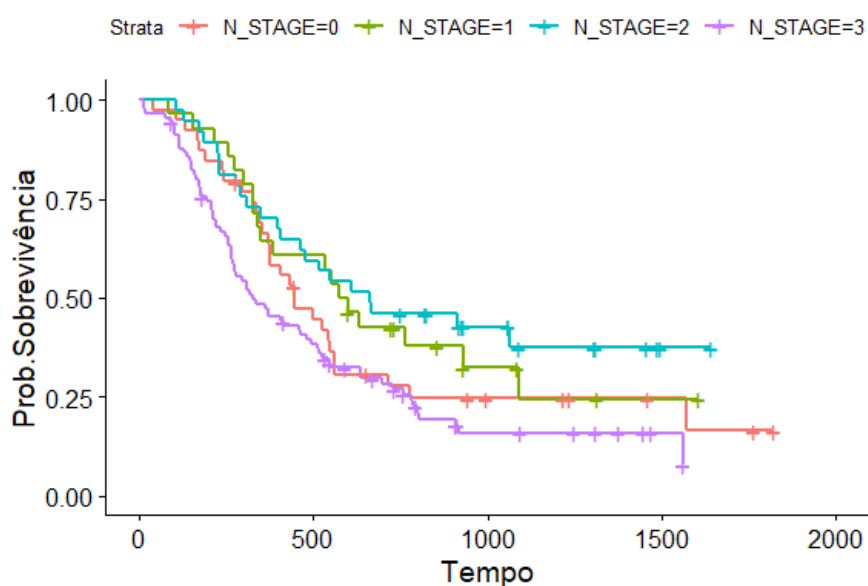
Figura 3 – Curvas de sobrevivência (Kaplan Meier) para comparação entre os grupos da covariável T\_STAGE.

Tabela 4 - Resultados da comparação 2 a 2 pelo teste de log rank para a covariável T-STAGE.

| <b>T STAGE</b> | <b>Estatística do teste</b> | <b>p-valor</b> |
|----------------|-----------------------------|----------------|
| <b>1 X 2</b>   | 0,7                         | 0,4            |
| <b>1 X 3</b>   | 0                           | 0,9            |
| <b>1 X 4</b>   | 1,1                         | 0,3            |
| <b>2 X 3</b>   | 1,3                         | 0,3            |
| <b>2 X 4</b>   | 8,1                         | 0,004          |
| <b>3 X 4</b>   | 6,7                         | 0,01           |

Fonte: Elaborado pela autora.

No gráfico referente ao estadiamento linfonodal (N-STAGE), verifica-se que as curvas mostram uma separação progressiva ao longo do tempo, o que é confirmado pelos resultados do teste de log-rank, (Tabela 5) que apontaram diferenças significativas entre os grupos N1 e N3 ( $p = 0,04$ ) e os grupos N2 e N3 ( $p = 0,004$ ). Demonstrando que pacientes com múltiplos linfonodos comprometidos ou linfonodos fixos apresentam pior prognóstico quando comparados àqueles sem evidência de metástase linfonodal ou com comprometimento limitado.



Fonte: Elaborado pela autora.

Figura 4 – Curvas de sobrevivência (Kaplan Meier) para comparação entre os grupos da covariável N\_STAGE.

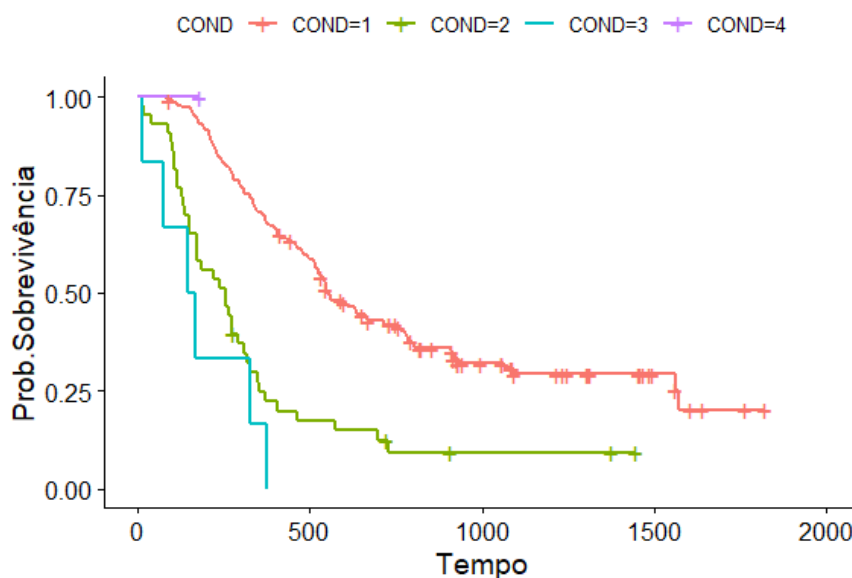
Tabela 5 - Resultados da comparação 2 a 2 pelo teste de log rank para a covariável N-STAGE.

| <b>N STAGE</b> | <b>Estatística do teste</b> | <b>p-valor</b> |
|----------------|-----------------------------|----------------|
| <b>0 X 1</b>   | 0,7                         | 0,4            |
| <b>0 X 2</b>   | 2,2                         | 0,1            |
| <b>0 X 3</b>   | 1,9                         | 0,2            |
| <b>1 X 2</b>   | 0,3                         | 0,6            |
| <b>1 X 3</b>   | 4,1                         | 0,04           |
| <b>2 X 3</b>   | 8,2                         | 0,004          |

Fonte: Elaborado pela autora.

O gráfico de Kaplan–Meier para a covariável Condição (COND) denota diferenças expressivas entre os grupos.

Os resultados do teste de log-rank (Tabela 6) indicaram diferenças bem significativas entre os grupos 1 e 2 e os grupos 1 e 3 ( $p < 0,001$ ), confirmando a forte associação entre condição e risco de óbito. E acentuando que pacientes sem incapacidade apresentam maior probabilidade de sobrevivência ao longo do tempo do que pacientes com limitações funcionais mais severas. Esse achado destaca a condição clínica geral do paciente como um dos principais determinantes do prognóstico.



Fonte: Elaborado pela autora.

Figura 5 – Curvas de sobrevivência (Kaplan Meier) para comparação entre os grupos da covariável COND.

Tabela 6 - Resultados da comparação 2 a 2 pelo teste de log rank para a covariável COND.

| COND         | Estatística do teste | p-valor     |
|--------------|----------------------|-------------|
| <b>1 X 2</b> | 33,5                 | 0,000000007 |
| <b>1 X 3</b> | 29,6                 | 0,00000005  |
| <b>1 X 4</b> | 0,1                  | 0,8         |
| <b>2 X 3</b> | 1,4                  | 0,2         |
| <b>2 X 4</b> | 0,5                  | 0,5         |
| <b>3 X 4</b> | 0,9                  | 0,3         |

Fonte: Elaborado pela autora.

De modo geral, os gráficos de Kaplan–Meier, aliados aos resultados do teste de log-rank, permitiram identificar quais covariáveis apresentaram associação significativa com o tempo de sobrevida. A partir dessas análises, procedeu-se à dicotomização das variáveis relevantes, as quais foram posteriormente consideradas no ajuste do modelo de regressão de riscos proporcionais de Cox.

A seleção das variáveis para o modelo multivariado fundamentou-se em um ajuste inicial de regressão de Cox com todos os potenciais preditores.

Adotou-se como critério de permanência no modelo final a significância estatística ao nível de 0,05. Assim, o modelo final ajustado manteve as covariáveis GRADE, T-STAGE, N-STAGE e COND, conforme apresentado na Tabela 7.

Tabela 7 – Ajuste do modelo final com as estimativas dos coeficientes, suas razões de risco e os respectivos erros-padrão e valores p.

| <b>Covariável</b> | <b>Estimativa<br/>(<math>\beta</math>)</b> | <b><math>exp(\beta)</math></b> | <b><math>exp(-\beta)</math></b> | <b>Erro<br/>padrão</b> | <b>Valor-p</b> |
|-------------------|--------------------------------------------|--------------------------------|---------------------------------|------------------------|----------------|
| <b>GRADE</b>      | -0,4639                                    | 0,6288                         | 1,5903                          | 0,2515                 | 0,06           |
| <b>T-STAGE</b>    | 0,4096                                     | 1,5062                         | 0,6639                          | 0,1794                 | 0,02           |
| <b>N-STAGE</b>    | 0,3384                                     | 1,4027                         | 0,7129                          | 0,1804                 | 0,06           |
| <b>COND</b>       | -1,0343                                    | 0,3555                         | 2,8132                          | 0,1909                 | <0,001         |

Fonte: Elaborado pela autora.

Por meio dos resultados obtidos, pode-se observar que, pacientes que possuem grau de doença (GRADE) moderadamente ou muito diferenciado, tem risco de morte aproximadamente 1,6 vezes maior do que aqueles que possuem grau de doença pouco diferenciado.

Quanto ao estágio do tumor (T-STAGE), observa-se que o risco de morte de pacientes com tumor maciço invasivo é 1,5 vezes maior do que aqueles com estágio de tumor classificados em 1,2 e 3.

Em relação à presença de linfonodos (N-STAGE), nota-se que, aqueles que possuem múltiplos linfonodos possuem risco de morte 40% maior do que aqueles que possuem nenhuma metástase ou linfonodo positivo com 3 cm ou mais.

Por fim, no que se refere à condição do paciente, aqueles que se encontram acamados, possuem risco de morte 2,8 vezes maior quando comparados com aqueles que não possuem incapacidade, ou estão em trabalho restrito.

## 5.1 Bondade do ajuste

O modelo de Cox é considerado um elemento bastante flexível, contudo, é preciso avaliar sua adequação pois o mesmo pode não se ajustar bem a algumas situações clínicas (COLOSIMO; GIOLO, 2024).

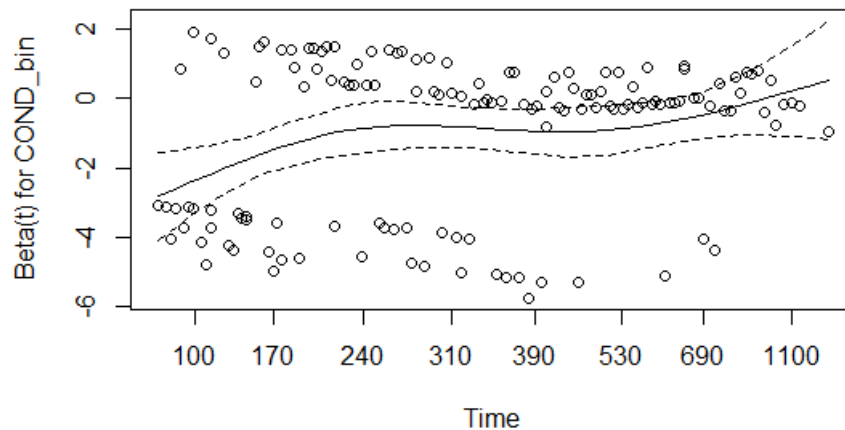
Visando compreender se o modelo concorda com o conjunto de dados, foi verificada a suposição de riscos proporcionais por meio dos resíduos de Schoenfeld. Com base nos resultados apresentados na Tabela 8, observa-se que a covariável COND e o teste global apresentaram violação estatisticamente significativa, todavia as covariáveis GRADE, T-STAGE e N-TAGE não apresentaram evidências de violação da suposição ( $p$  valor  $> 0,05$ ).

Tabela 8 – Avaliação da Proporcionalidade das taxas de falha no modelo.

|                | <b>chisq</b> | <b>df</b> | <b>p</b> |
|----------------|--------------|-----------|----------|
| <b>GRADE</b>   | 1.55         | 1         | 0.2131   |
| <b>T STAGE</b> | 2.89         | 1         | 0.0890   |
| <b>N STAGE</b> | 2.97         | 1         | 0.0850   |
| <b>COND</b>    | 10.34        | 1         | 0.0013   |
| <b>GLOBAL</b>  | 14.08        | 1         | 0.0070   |

Fonte: Elaborado pela autora.

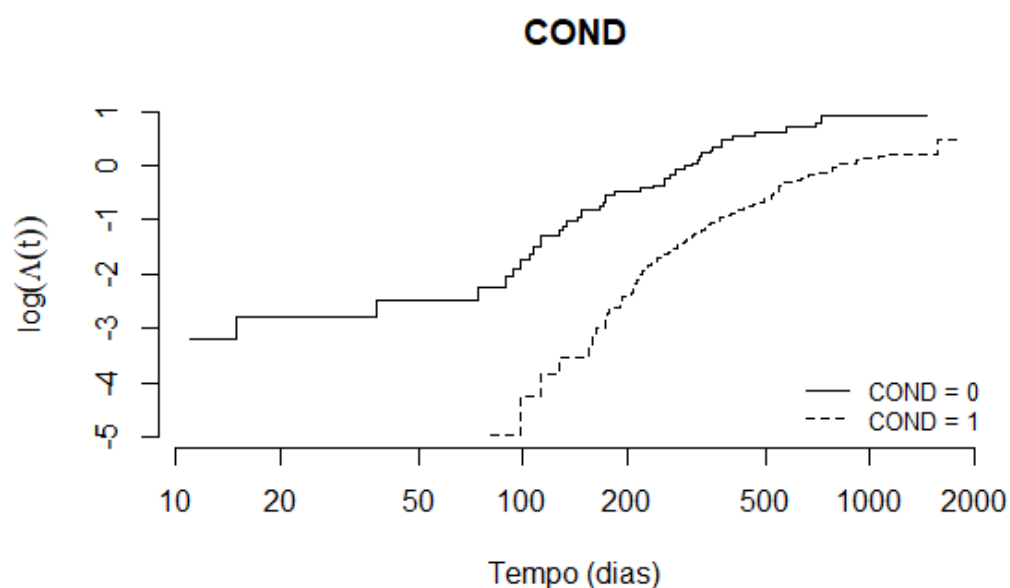
A análise gráfica do logaritmo da função de risco acumulada (Figura 6) sustenta a suposição de riscos proporcionais para as covariáveis T\_STAGE, N\_STAGE e COND, dado o paralelismo observado entre as curvas. Embora a variável GRADE apresente cruzamentos iniciais, a ausência de significância estatística no teste de resíduos de Schoenfeld ( $p = 0,2131$ ) confirma que não há violação, tornando a suposição aceitável. Ao analisar a Figura 6, não se observa inclinação importante, indicando que não há violação da suposição, uma vez que, inclinação nula, indica evidências a favor da proporcionalidade.



Fonte: Elaborado pela autora.

Figura 6 – Resíduos padronizados de Schoenfeld em função do tempo para as covariáveis GRADE, T\_STAGE, N\_STAGE, COND.

Como a covariável COND não apresentou proporcionalidade de riscos, como observado na Tabela 8, optou-se por complementar a inspeção gráfica para tal covariável, construindo o gráfico do Logaritmo do Risco Acumulado versus o tempo, para verificar se as curvas se cruzam. Como pode ser visto na Figura 7, não há indícios de violação máxima da suposição de riscos proporcionais, uma vez que as curvas não se cruzam.



Fonte: Elaborado pela autora.

Figura 7 – Curvas  $[log \log (\Lambda(t))]$  versus  $t$  para a covariável COND.

Assim, apesar da não proporcionalidade observada para a covariável COND, por meio do teste de proporcionalidade apresentado na Tabela 8, e, analisando a Figura 7, optou-se por admitir o modelo de Cox para os dados deste estudo. Os resultados apresentados permitem a comparabilidade com estudos prévios e a avaliação dos efeitos médios das covariáveis sobre o desfecho, em consonância com a natureza exploratória da análise.

## 6. CONCLUSÕES

No presente estudo, foram aplicados métodos de análise de sobrevivência para avaliar fatores associados ao tempo de sobrevida de pacientes diagnosticados com carcinoma espinocelular da orofaringe. Os resultados apontam que a condição funcional do paciente, bem como os estadiamentos tumoral e linfonodal, são os principais fatores que exercem influência significativa sobre o risco de óbito.

Observou-se também que apesar das limitações inerentes ao uso de dados secundários e à necessidade de dicotomização de algumas covariáveis, os achados obtidos estão em concordância com a literatura e reforçam a importância da análise de sobrevivência como ferramenta estatística na investigação de desfechos clínicos.

## REFERÊNCIAS BIBLIOGRÁFICAS

BLAND, J. M.; ALTMAN, D. G. Survival probabilities (the Kaplan-Meier method). *BMJ*, v. 317, n. 7172, p. 1572–1580, 1998.

COLOSIMO, Enrico A. *Análise de sobrevivência aplicada*. 2. ed. São Paulo: Blucher, 2024. E-book. p. 151. ISBN 9788521221975. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788521221975/>. Acesso em: 03 fev. 2026.

COX, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 34, n. 2, p. 187–220, 1972.

ENDÓ, Y.; TANIOKA, M.; MIYACHI, Y. Prognostic factors in cutaneous squamous cell carcinoma: is patient delay in hospital visit a predictor of survival? *ISRN Dermatology*, v. 2011, art. 285289, 2011. DOI: 10.5402/2011/285289.

INSTITUTO NACIONAL DE CÂNCER (INCA). Câncer de pele não melanoma. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/pele-nao-melanoma>. Acesso em: 03 fev. 2026.

KALBFLEISCH, J. D.; PRENTICE, R. L. *The statistical analysis of failure time data*. New York: John Wiley & Sons, 2003.

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

KLEINBAUM, David G.; KLEIN, Mitchel. *Survival analysis: a self-learning text*. 3. ed. New York: Springer, 2012.

MANTEL, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, v. 50, n. 3, p. 163–170, 1966.

PEREIRA, Janaina. *Métodos de comparação de curvas de sobrevivência*. Piracicaba: Escola Superior de Agricultura Luiz de Queiroz, 2002.

PETRI, Valeria. *Dermatologia prática*. Rio de Janeiro: Guanabara Koogan, 2009.

WEDDING, Ulrich et al. Age, severe comorbidity and functional impairment independently contribute to poor survival in cancer patients. *Journal of Cancer Research and Clinical Oncology*, v. 133, n. 12, p. 945–950, 2007.