

**UNIVERSIDADE ABERTA**



UNIVERSIDADE  
**AbERTA**  
[www.uab.pt](http://www.uab.pt)

**TÉCNICAS DE ANÁLISE DE SOBREVIVÊNCIA PARA SUJEITOS  
INFETADOS COM HIV**

José Portez Sibinde

Mestrado em Bioestatística e Biometria

2022

**UNIVERSIDADE ABERTA**



UNIVERSIDADE  
**AbERTA**  
[www.uab.pt](http://www.uab.pt)

**TÉCNICAS DE ANÁLISE DE SOBREVIVÊNCIA PARA SUJEITOS  
INFETADOS COM HIV**

José Portez Sibinde

Mestrado em Bioestatística e Biometria

Orientador:  
Professor Doutor Pedro Serranho

2022

## Resumo

O presente estudo visou aplicar técnicas de Análise de Sobrevida para o estudo do tempo de vida de sujeitos infectados com o vírus da imunodeficiência humana (HIV) e a influência das covariáveis recolhidas nesse tempo. Os dados são provenientes do centro de saúde de Changanine província de Gaza, Moçambique. Estes foram utilizados mediante autorização ética 115/CNBS/2020 e administrativa 466/GMS/290/021 do Comité Nacional de Bioética para a Saúde e do Ministério de Saúde, respetivamente.

Para obtenção dos resultados foi utilizado o *software* R versão 4.1.1 e o *package survival* no qual foram utilizados os métodos de Kaplan-Meier e Cox. O teste logrank foi aplicado para comparação de curvas de sobrevivência estimadas pelo método de Kaplan-Meier em diferentes grupos das covariáveis categóricas, sendo que foi estimado o tempo de vida mediano em cada um desses grupos. Para obtenção das covariáveis significativas do modelo de Cox foi aplicado o método de eliminação *backward* e após validação do pressuposto de riscos proporcionais pela análise dos resíduos de Schoenfeld e teste de correlação dos resíduos com o tempo, foi estudada a adequação do modelo por análise dos resíduos de cox-snell, resíduos Martingala e resíduos *deviance*.

Verificou-se, pelo modelo de regressão de Cox, que fazer profilaxia INH e CTX bem como integrar os grupos de apoio à adesão comunitária contribui no aumento do tempo de sobrevivência dos pacientes. Por outro lado foi verificado também que o diagnóstico positivo da tuberculose, subir de nível do estadio e da desnutrição, ter diagnóstico positivo de sarcoma de Kaposi, ser do sexo masculino e ter carga viral maior que 10mil cópias, contribui significativamente para redução do tempo de sobrevivência dos sujeitos.

Palavras-chave: Análise de Sobrevida, Virus do HIV, Método de Kaplan-Meier, Modelo de Cox, tempo de sobrevivência.

## Abstract

The present study aimed to apply Survival Analysis techniques to study the life span of subjects infected with the human immunodeficiency virus (HIV) and the influence of the collected covariates during that time period. The data comes from the Changanine health center in Gaza province, Mozambique. These were used under ethical authorization 115/CNBS/2020 and administrative authorization 466/GMS/290/021 from the National Bioethics Committee for Health and the Ministry of Health, respectively.

The results were obtained with the software R version 4.1.1, namely through the survival package, in which the Kaplan-Meier and Cox methods were considered. The logrank test was applied to compare the survival curves estimated by the Kaplan-Meier method in different groups of categorical covariates, and the median life span was estimated in each of these groups. To obtain the significant covariates of the Cox model, the backward elimination method was applied and after validating the proportional hazards assumption by analyzing the Schoenfeld residuals and the residual correlation test with time, the fitness of the model was studied by residual analysis of cox-snell, Martingal residues and deviance residues.

The Cox model showed that performing INH and CTX prophylaxis, as well as integrating support groups for community adherence, contributes to increasing the survival time of patients. On the other hand, it was also verified that the positive diagnosis of tuberculosis, increasing the level of the stage and malnutrition, having a positive diagnosis of Kaposi's sarcoma, being male and having a viral load greater than 10,000 copies, contributes significantly to reducing the time of subjects' survival.

Keywords: Survival Analysis, HIV Virus, Kaplan-Meier Method, Cox Model, survival time.

## Agradecimentos

Vão os meus especiais agradecimentos ao Professor Doutor Pedro Serranho por todo apoio por ele prestado, desde a concepção do título da dissertação, plano de dissertação, protocolo de pesquisa e ao acompanhamento e orientações.

À coordenação do Mestrado em Bioestatística e Biometria pelo fornecimento da carta de cobertura para obtenção dos dados de pesquisa.

Ao Comité Nacional de Bioética para a Saúde e ao Ministério de Saúde de Moçambique pela autorização ética concedida pelo número 115/CNBS/2020 e autorização administrativa dada sob a nota número 466/GMS/290/021 do gabinete do ministro, respetivamente, do uso da base de dados.

À Direção Provincial de Saúde de Gaza pelo fornecimento de todos requisitos necessários para obtenção e uso dos dados.

Ao Centro de Saúde de Changanine pela disponibilidade e apoio na inserção na base de dados de alguns dados a partir dos processos dos pacientes.

## Lista de Abreviaturas e Siglas

AIDS	Síndrome da Imunodeficiência Adquirida
CNBS	Comité Nacional de Bioética para a Saúde
CTX	Cotrimoxazole
Diag_Nutric	Codificação da Variável Diagnóstico da Nutrição
diagn_SK	Codificação da Variável Diagnóstico de Sarcoma de Kaposi
diagn_TB	Codificação da Variável Diagnóstico da Tuberculose
ELISA	<i>Enzyme-Linked Immunosorbent Assay</i>
GAAC	Grupos de Apoio à Adesão Comunitária
grav_lact	Codificação da Variável Grávidas ou Lactantes
HIV	Vírus da Imunodeficiência Humana
IMC	Índice da Massa Corporal
INH	Isoniazida
KM	Kaplan-Meier
OMS	Organização Mundial da Saúde
profil_CTX	Codificação da Variável Profilaxia com Cotrimoxazole
profil_INH	Codificação da Variável Profilaxia com Isoniazida
PVHIV	Pessoas Vivendo com HIV
RR	Risco Relativo
SK	Sarcoma de Kaposi
TARV	Terapia Antirretroviral
TB	Tuberculose
TPT	Tratamento Preventivo da Tuberculose
trat	Codificação da Variável Tratamento
UNAIDS	Programa Conjunto das Nações Unidas sobre HIV/SIDA

## ÍNDICE

Resumo.....	i
Abstract.....	ii
Agradecimentos .....	iii
Lista de Abreviaturas e Siglas.....	iv
Índice de Figuras.....	vii
Índice de Tabelas.....	viii
1. Introdução.....	1
2. Revisão de Literatura .....	4
2.1. A Epidemia do HIV/SIDA no Mundo e em Moçambique .....	4
2.2. Diagnóstico e tratamento .....	5
2.3. Fatores determinantes da sobrevivência de indivíduos HIV positivos .....	7
3. Metodologia de Análise de Sobrevivência.....	11
3.1. Função Taxa de Falha .....	12
3.2. Censura .....	15
3.3. Estimadores não paramétricos para função de sobrevivência.....	16
3.3.1. Estimador Kaplan-Meier para a função de sobrevivência .....	16
3.3.2. Estimador de Nelson- Aalen para a função de sobrevivência .....	17
3.4. Testes de Comparação das funções de sobrevivência.....	18
3.5. Modelos Paramétricos de Análise de Sobrevivência .....	20
3.5.1. Estimação dos Parâmetros do modelo: Método de máxima verosimilhança	21
3.5.2. Modelo Exponencial .....	21
3.5.3. Modelo de Regressão de Weibull.....	21
3.5.4. Modelo de Regressão Log-normal .....	22

3.6.	Modelo de Cox.....	22
3.6.1.	Proporcionalidade de Funções de Risco .....	24
3.6.2.	Validação e Adequação do Modelo .....	25
3.6.3.	Método de Seleção de Variáveis .....	26
3.6.4.	Validação da hipótese de Riscos Proporcionais.....	27
3.6.5.	Resíduos de Cox-Snell.....	27
3.6.6.	Resíduos Martingala.....	28
3.6.7.	Resíduos Deviance .....	28
3.7.	Modelos de Tempo de Vida Acelerados .....	29
4.	Resultados e Discussão .....	31
4.1.	Descrição da Pesquisa .....	31
4.1.1.	Critérios de Inclusão.....	31
4.1.2.	Critérios de Exclusão.....	31
4.1.3.	Covariáveis.....	32
4.2.	Estimativa não Paramétrica de Kaplan-Meier .....	35
4.3.	Ajuste do Modelo de Cox .....	41
4.3.1.	Construção do modelo de Cox .....	49
4.3.2.	Análise da Adequação do Ajuste do Modelo .....	54
5.	Conclusão.....	57
5.1.	Perspetivas Futuras .....	60
6.	Referências Bibliográficas .....	61

## Índice de Figuras

Figura 1: Estimativas de Kaplan-Meier da função de sobrevivência para as covariáveis categóricas.....	37
Figura 2: Resíduos padronizados de Shoenfeld .....	47
Figura 3: Funções de sobrevivência para diferentes valores das covariáveis pelo modelo de Cox.....	53
Figura 4: (a) Resíduos de Cox-Snell ( $ei$ ) versus Taxa de Falha Acumulada $\Lambda(ei)$ , (b) estimativas de Sobrevivência por Kaplan-Meier $SKM(ei)$ , e Exponencial Padrão $SExp(ei)$ versus Resíduos de Cox-Snell ( $ei$ ), (c) Pares ( $SKM(ei)$ , $SExp(ei)$ ) e (d) Resíduos Deviance versus Tempo. ....	55
Figura 5: (a) Resíduos Martingala versus Preditor Linear e (b) Resíduos <i>Deviance</i> versus Preditor Linear .....	56

## Índice de Tabelas

Tabela 1: Descrição das variáveis .....	33
Tabela 2: Análise Descritiva das variáveis Categóricas .....	34
Tabela 3: Análise descritiva das variáveis contínuas .....	35
Tabela 4: Teste Logrank para comparação das funções de sobrevivência estimadas por Kaplan-Meier entre os grupos de cada uma das variáveis categóricas .....	36
Tabela 5: Estimativas de Kaplan-Meier do tempo de vida mediano em cada grupo das variáveis categóricas.....	40
Tabela 6: Modelo de Cox com todas as covariáveis obtido pelo teste de Wald.....	41
Tabela 7: Modelo II obtido a partir do primeiro por eliminação da covariável grav_lact .....	42
Tabela 8: Modelo III obtido a partir do modelo II pela eliminação da covariável (carga_viral) 2 .....	42
Tabela 9: Modelo IV obtido pelo teste de Wald a partir do modelo III pela eliminação da covariável trat.....	43
Tabela 10: Modelo V obtido pelo teste de Wald por eliminação da covariável TPT do modelo IV.....	44
Tabela 11: Modelo VI obtido pelo teste de Wald por eliminação da covariável idade do modelo V.....	44
Tabela 12: Modelo VII obtido pelo teste de Wald por eliminação da covariável (carga_viral) 3 do modelo VI.....	45
Tabela 13: Modelo VIII obtido pelo teste de Wald por eliminação da covariável IMC do modelo VII.....	45
Tabela 14: Teste de correlação dos resíduos com o tempo .....	46
Tabela 15: Estimativas dos coeficientes do modelo de Cox e os intervalos de confiança para o Risco Relativo .....	49

Capítulo 1

Introdução

## 1. Introdução

A análise de sobrevivência é um ramo da estatística munida de técnicas e procedimentos estatísticos capazes de modelar o tempo até a ocorrência de um evento de interesse denominado tempo de falha, como por exemplo tempo até a morte, tempo até o divórcio, tempo até ocorrência do enfarte, etc. A análise de sobrevivência é capaz de lidar com dados censurados, o que a diferencia dos demais métodos da estatística usual visto que nesta, as censuras são tidas como observações omissas. Os dados censurados ou simplesmente censuras correspondem a observações nas quais o evento de interesse não ocorreu no tempo de observação. Casos particulares de censuras são por exemplo os sujeitos terem sobrevivido mais tempo do que o tempo de estudo ou o seu seguimento ter sido interrompido. A aplicação da análise de sobrevivência pode ser encontrada em várias áreas de conhecimento, desde medicina e ciências da saúde, biologia e tempo de vida de componentes elétricos, entre outros.

O objetivo do presente trabalho é aplicar técnicas de Análise de Sobrevivência para o estudo do tempo de vida de sujeitos infetados com o vírus da imunodeficiência humana (HIV), nomeadamente estimar o seu tempo de vida após diagnóstico e a influência dos valores de covariáveis recolhidas. Os sujeitos envolvidos no trabalho são provenientes de uma base de dados secundária do centro de saúde de Changanine, na província de Gaza, Moçambique, sendo o estudo aprovado pelo Comité Nacional de Bioética para a Saúde (CNBS) e registado sob o número 115/CNBS/2020.

A implementação do programa de terapia antirretroviral (TARV) é sem dúvidas um dos factores determinantes da sobrevivência dos pacientes com HIV que tem trazido aos sujeitos uma boa qualidade de vida, devolvendo a esperança de viver e sonhar. Contudo, apesar dos pacientes estarem a cumprir com o TARV verificam-se ainda ocorrências de óbitos durante o seguimento destes pacientes, o que realça de certa forma, a necessidade de investigar sobre quais os factores que mais influem na sobrevivência dos pacientes em TARV. Desta forma, o trabalho busca compreender que covariáveis têm o efeito de aumentar ou de diminuir o tempo de sobrevivência dos pacientes do centro de saúde referido.

Neste trabalho, estimamos a função de sobrevivência pelo método não paramétrico de Kaplan-Meier para diferentes valores de algumas variáveis qualitativas isoladas e o tempo de vida mediano dos sujeitos em cada caso. Aplicamos também o modelo de regressão de Cox para estimar o efeito integrado das várias covariáveis sobre a sobrevivência dos sujeitos, nomeadamente a variação do risco mediante a alteração do valor das covariáveis significativas para o modelo. Todas as análises dos dados são feitas com auxílio do *R*, versão 4.1.1.

Este estudo pode contribuir para uma melhor condução dos programas de TARV, incluindo a conscientização dos gestores de base de dados sobre a necessidade de completar o registo de informação dos pacientes e a implementação de políticas de saúde que visem reduzir as mortalidades dos pacientes em TARV.

A dissertação é organizada em 5 capítulos, sendo o primeiro dos quais este. O segundo capítulo, a revisão da literatura, descreve as abordagens de outros autores sobre as temáticas de análise de sobrevivência, com especial enfoque na aplicação aos dados de sujeitos com HIV. No terceiro capítulo é apresentada a abordagem metodológica da análise de sobrevivência utilizada no trabalho bem como as técnicas de análise de sobrevivência utilizadas para a análise dos dados. No quarto capítulo são apresentados e discutidos os resultados obtidos pela aplicação da metodologia proposta onde são apresentadas as estimativas de Kaplan-Meier para a função de sobrevivência, o modelo de regressão de Cox ajustado e a sua validação.

No quinto capítulo apresentamos as principais conclusões obtidas neste estudo.

## Capítulo 2

### Revisão de Literatura

## 2. Revisão de Literatura

Neste capítulo vamos rever algumas publicações e resultados relevantes sobre Análise de Sobrevivência e HIV, com particular incidência sobre o nosso objetivo de estudo.

### 2.1. A Epidemia do HIV/SIDA no Mundo e em Moçambique

HIV é uma sigla inglesa para *human immunodeficiency virus*. É, portanto, um vírus que ataca o sistema imunológico humano, enfraquecendo-o graças à fusão entre os linfócitos TCD4+ e o vírus, promovendo a disfunção destas células de defesa do organismo [1]. Nessa fusão o HIV insere nos linfócitos TCD4+ o seu RNA que, dada a incompatibilidade deste material genético com o DNA das células de defesa, faz a transformação do seu RNA em DNA pela ação enzimática da transcriptase reversa e duplica o seu material genético provocando a destruição progressiva das células do sistema imunológico [2]. Com profunda imunodepressão o organismo humano fica propenso a várias infeções, cancro ou outras doenças oportunistas e isto caracteriza a síndrome da imunodeficiência adquirida (AIDS) [3].

Segundo a Organização Mundial da Saúde (OMS) e o Programa Conjunto das Nações Unidas sobre HIV/SIDA (UNAIDS) a prevalência global do HIV em adultos na última década foi 0.8%, o correspondente a 33 milhões de pessoas vivendo com HIV (PVHIV) no mundo [4].

A epidemia do HIV/SIDA, que constitui um dos maiores problemas de saúde pública do mundo, tem as consequências mais devastadoras no continente africano, principalmente na África Subsaariana onde mais de 30 milhões de pessoas vivem com HIV, o que corresponde a cerca de dois terços do total mundial [5]. A ação global para o combate ao HIV/SIDA tem um grande impacto nesta região africana onde de todos os casos de HIV em crianças no mundo 90% são africanos e ainda 60% das infeções em adultos ocorrem em mulheres [4]. A África Austral é responsável por cerca de 2/3 dos casos de HIV em África tornando-se assim o epicentro das infeções por HIV neste continente [6]. O número de óbitos atribuídos a HIV/SIDA em 2017 chegou a 336.175, o correspondente ao triplo em relação aos 113.631 óbitos ocorridos em

1990[7].

Moçambique apesar dos esforços concentrados na prevenção e combate ao HIV/SIDA continua a enfrentar a epidemia do HIV. A taxa de prevalência nacional em 2007 entre pessoas das idades compreendidas entre 15 a 49 anos foi de 16 %, variando entre províncias[8]. Nessa altura, a província de Gaza apresentava uma taxa de prevalência de 16.4% [9]. No entanto, este cenário foi lentamente revertido a nível nacional. Em 2019 a prevalência nacional do HIV reduziu para 11.5% na mesma faixa etária dos 15 – 49 anos, sendo as mulheres as mais afetadas, com uma prevalência de 13.1% contra 9.2% dos homens[10]. Na província de Gaza, ao contrário da tendência nacional, a prevalência aumentou para 25.1%[10], sendo a mais alta de todas as províncias do país. A elevada prevalência pode ser explicada não só pelos sistemas de saúde precários mas também pelo estigma, desigualdades e discriminação associada ao género, que constituem barreiras para aceder aos cuidados [11]. Desta forma, a avaliação dos dados de sujeitos portadores de HIV assume especial importância na província de Gaza, sendo por isso relevante o estudo proposto neste trabalho.

## 2.2. Diagnóstico e tratamento

Para o diagnóstico da infecção pelo HIV destacamos os testes laboratoriais imunoenzimáticos do tipo ELISA (*Enzyme-Linked Immunosorbent Assay*). Graças a evolução das metodologias empregues a partir do primeiro imunoensaio em 1985, foram desenvolvidos quatro gerações de testes, denominados ensaios de primeira, segunda, terceira e quarta gerações. Os ensaios de primeira geração baseiam-se na deteção da presença de anticorpos específicos através de um conjugado composto por um anticorpo anti-IgG humana enquanto os de segunda geração se baseiam nos antígenos recombinantes ou peptídeos produzidos pelas proteínas do HIV. Os ensaios de terceira geração permitem detetar simultaneamente anticorpos anti-HIV IgM e IgG enquanto os de quarta geração detetam o antígeno p24 e anticorpos específicos anti-HIV, simultaneamente [12].

Os testes rápidos também são empregues para o diagnóstico do HIV porém, após resultado regente, testes complementares e confirmatórios são requeridos para confirmar o diagnóstico [13].

Atualmente a AIDS é considerada umas das doenças crônicas conhecidas e controláveis. Sua morbidade e mortalidade foram drasticamente reduzidas graças à introdução do programa de TARV[14]. Contudo, taxas elevadas de mortalidade por AIDS continuam sendo registadas devido ao baixo nível de adesão dos pacientes aos programas de tratamento e apontam-se como principais causas da falta da adesão as restrições financeiras, falta de apoio familiar, a depressão, o alcoolismo, o estigma social, o esquecimento da medicação e os efeitos colaterais dos medicamentos [15].

Sendo a adesão um processo colaborativo visando facilitar a aceitação e a integração de um certo regime terapêutico no quotidiano das pessoas em tratamento, esta pressupõe a participação dos intervenientes nas decisões sobre estes regimes terapêuticos [16]. As ações de adesão são norteadas pelos princípios orientadores do sistema único de Saúde, nomeadamente:

- Integridade do Cuidado;
- Universalidade no acesso a insumos, serviços e ações de saúde;
- Equidade e superação de vulnerabilidades específicas;
- Valorização da autonomia e da garantia de participação das pessoas vivendo com HIV/SIDA na construção de estratégias que promovam a adesão ao tratamento;
- Trabalho em redes articuladas e complementares entre os diferentes níveis de atenção à saúde;
- Articulação intra e intersetorial, buscando superar obstáculos, particularmente sociais.

O estadiamento clínico da OMS de infecção por HIV estabelece 4 estágios, nomeadamente o estágio I, II, III e IV. A disposição destes estágios clínicos foi estabelecida mediante a presença e o peso de sintomas no indivíduo diagnosticado

de infecção por HIV, sendo a gravidade destes sintomas atribuída aos estágios mais elevados. O estágio I é definitivamente assintomático enquanto o indivíduo no estágio II apresenta sintomas leves como por exemplo a perda ligeira de peso, infecções recorrentes do trato respiratório e úlceras orais. Os estágios III e IV são os que apresentam sintomas que caracterizam a AIDS. Por exemplo a tuberculose (TB) pulmonar e diarreia crónica estão enquadradas no estágio III, enquanto no estágio IV ocorre, por exemplo, a TB extrapulmonar e o sarcoma de Kaposi (SK) [17].

O programa de tratamento antirretroviral foi projetado para impedir a reprodução do HIV e progressão para a AIDS. Assim, o efeito principal do tratamento é de eliminar a replicação do vírus, garantindo-se a recuperação do sistema imunológico dos infetados e evitando o desenvolvimento do AIDS e conseqüentemente a morte. Os benefícios clínicos do programa de TARV para pacientes diagnosticados com infecção por HIV do ponto de vista da redução de taxas da mortalidade e da melhoria de qualidade de vida é uma realidade bem conhecida em todos os países, mas há variações regionais que são intrínsecas dessas regiões, que levam a que as taxas de mortalidade sejam relativamente diferentes em países em vias de desenvolvimento [18].

### 2.3. Fatores determinantes da sobrevivência de indivíduos HIV positivos

Todas as pesquisas referenciadas neste trabalho no âmbito de análise de sobrevivência de pacientes com HIV, utilizaram o método de Kaplan-Meier para estimar a probabilidade de sobrevivência de pacientes e o modelo de Cox para determinar os fatores determinantes do tempo de sobrevivência. Embora tenham utilizado comumente as variáveis idade, sexo, estadio e CD4, estas não apresentaram o mesmo efeito sobre a sobrevivência de pacientes nas diferentes pesquisas.

Segundo [19] as variáveis idade e sexo não tem efeito significativo no tempo de sobrevivência, com p-valor de 0.12 e 0.23, respetivamente, no teste de Wald. Neste estudo, uma maior sobrevivência está associada ao maior número na contagem de CD4 e ao aumento do peso.

Segundo [20], [21], os pacientes mais jovens apresentam maior probabilidade de sobreviver muito mais tempo do que os mais velhos e no que diz respeito ao sexo foi constatado que as características associadas à menor sobrevivência estão ligadas aos indivíduos de sexo masculino do que do sexo feminino, contrariando o fato destas variáveis não serem significativas reportado em [19].

Outros estudos consideraram outras covariáveis na estimação do tempo de sobrevivência de sujeitos com HIV, para além das descritas acima. O estado civil, religião, nível de escolaridade e o tipo de emprego não apresentam efeito significativo na sobrevivência [22], [23] enquanto as covariáveis TB e o aparecimento de infeções oportunistas, o nível de adesão dos pacientes no programa de tratamento, o consumo de álcool, uso de preservativo e estado funcional estiveram estritamente relacionados com a sobrevivência dos pacientes nestas pesquisas. As infeções oportunistas são apontadas como as causas mais comuns de morte em pacientes nos estadios III e IV e o tratamento antirretroviral veio reduzir significativamente a incidência destas infeções para aqueles pacientes que aderiram ao programa de TARV. Maior adesão, baixo nível do consumo de álcool e uso de preservativo se relacionam com maior sobrevivência.

Os resultados relativamente ao tempo de sobrevivência estimado não estão muito afastadas uns dos outros. Nos estudos efetuados por [20], [23], [24] pacientes submetidos ao TARV tiveram o tempo de vida estimado em 6 anos, embora estas pesquisas tenham utilizado amostras de tamanhos diferentes. Contudo, outras determinaram que o tempo de sobrevivência de pacientes com HIV em tratamento antirretroviral é de cerca de 9 anos [25], 7 anos [19] e 5 anos [21] e os fatores prognósticos da sobrevivência não diferem. Uma peculiaridade apresentada nestas pesquisas é a presença de muitos dados censurados à direita que chegam a 90% do total de observações [24].

Outro fator que influi bastante na baixa sobrevivência, para além do baixo número de linfócitos T no sangue, é o baixo nível de escolaridade [26]. Por outro lado, as taxas de mortalidade verificadas entre toxicod dependentes têm sido muito elevadas e ainda

são por si só significativamente mais elevadas do que as taxas de mortalidade para a toda a população no geral. Embora os usuários de drogas, infectados pelo HIV, sejam menos propensos a aderência ao TARV, sua resposta a este tratamento, em termos de sobrevivência, não difere de outros grupos, se todos estiverem em tratamento.

Capítulo 3  
Metodologia de Análise de Sobrevida

### 3. Metodologia de Análise de Sobrevida

No presente capítulo abordam-se os principais conceitos de análise de sobrevivência bem como os métodos e modelos que serão utilizados neste trabalho. As obras mais consultadas são principalmente [27]–[29]. Os resultados apresentados podem ser consultados em mais detalhe em qualquer destes textos.

A análise de sobrevivência é uma área de estatística que se dedica especificamente ao estudo da sobrevivência dos sujeitos. A variável de interesse denominada tempo de falha, corresponde ao tempo até a ocorrência de um evento de interesse denominado falha. A análise de sobrevivência foi particularmente desenhada para poder tratar de censuras, por exemplo, sujeitos que, por alguma razão, abandonam o estudo antes de ocorrer a falha. Desta forma, os métodos de análise de sobrevivência permitem considerar estes tempos censurados no modelo, aproveitando a informação neles contida de que o tempo de falha é superior ao tempo de censura. Além disso, em geral o tempo de sobrevivência pode ser fortemente influenciado por outras covariáveis, pelo que é importante em análise de sobrevivência considerar modelos que permitam a sua inclusão e determinar de que forma estas podem influenciar o tempo de sobrevivência.

Do ponto de vista formal, consideramos  $T$  uma variável aleatória contínua e com valores não negativos correspondente ao tempo de vida de determinado indivíduo de uma população homogênea. A função  $S(t)$  mostra a sobrevivência em função do tempo  $t$  e é chamada de função de sobrevivência, sendo o seu gráfico conhecido como curva de sobrevivência. Formalmente,  $S(t)$  é a probabilidade de que a variável aleatória  $T$  seja maior do que um tempo  $t$  específico, ou seja,

$$S(t) = P(T \geq t), t \geq 0, \quad (1)$$

ou seja, a probabilidade do tempo de sobrevivência ser superior a  $t$ . Esta é uma função contínua e decrescente com as seguintes propriedades:

- A função  $S$  só está definida para tempo não negativo, isto é,  $t \geq 0$ .
- $S(t)$  é uma função não crescente, ou seja,  $S(t_1) \geq S(t_2)$  para  $t_1 \leq t_2$

- $S(0) = 1$ , ou seja, no instante inicial  $t=0$  a probabilidade de sobreviver é 1.

A função de distribuição de probabilidade da variável aleatória  $T$  por definida por

$$F(t) = P(T \leq t) \quad (1.1)$$

pode ser escrita no contexto de Análise de Sobrevivência como

$$F(t) = 1 - P(T \geq t) \Leftrightarrow F(t) = 1 - S(t) \quad (1.2)$$

uma vez que  $P(T \leq t) = 1 - P(T \geq t)$ . Como consideramos o tempo  $t$  contínuo, a função densidade de probabilidade  $f$  satisfaz

$$F(t) = \int_0^t f(u) du \quad (1.3)$$

pelo que, por diferenciação em  $t$  se obtém

$$F'(t) = -S'(t) = f(t). \quad (1.4)$$

O valor esperado de tempo de vida  $T$  pode ser dado segundo [27], [28], por integração por partes, pela seguinte expressão

$$\mu = E[T] = \int_0^{\infty} tf(t)dt = \int_0^{\infty} S(t)dt \quad (1.5)$$

que equivale à esperança média de vida dos sujeitos. Por outro lado o tempo mediano  $t_{med}$  é o instante de tempo que satisfaz

$$S(t_{med}) = 0.5.$$

### 3.1. Função Taxa de Falha

A Função de Sobrevivência não ilustra de forma clara em que intervalos de tempo a probabilidade de ocorrer a falha é maior. Para esse efeito é importante ter a taxa de falha, pelo que de seguida introduziremos esse conceito e a sua relação com a função de Sobrevivência.

Dados dois instantes  $t_1$  e  $t_2$ , define-se taxa de falha no intervalo de tempo  $[t_1, t_2)$ , como a razão entre a probabilidade de uma falha ocorrer nesse intervalo de tempo condicionada ao facto da falha não ter acontecido antes e a amplitude do intervalo, isto é,

$$\lambda_{[t_1, t_2]} = \frac{P(t_1 \leq T < t_2 | T \geq t_1)}{\Delta t} \quad (2.1)$$

Isto é equivalente a escrever a taxa de falha como (cf. [28])

$$\lambda_{[t_1, t_2]} = \frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)} \quad (2.2)$$

Se nesta expressão o intervalo de tempo  $t_2 - t_1 = \Delta t$  for suficientemente pequeno, isto é, se  $\Delta t \rightarrow 0$ , esta expressão passa a designar-se taxa de falha instantânea e pode ser escrita como

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

o que é equivalente a

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)} \quad (2.4)$$

A função taxa de falha,  $\lambda$ , define a taxa de variação da probabilidade da falha ocorrer em algum instante no intervalo  $[t, t + \Delta t)$  dado que não ocorreu antes do instante  $t$ .

A função taxa de falha  $\lambda(t)$  tem as seguintes propriedades:

- $\lambda(t) \geq 0$  para todo  $t$ ;
- $\lambda(t)$  não tem limite superior;
- $\lambda(t)$  pode assumir qualquer forma.

O significado de  $\lambda(t) = 0$  é de que nenhum evento de interesse acontecerá no infinitesimal que se segue a  $t$ .

A função taxa de falha acumulada descreve o risco acumulado até ao tempo  $t$  e é dado por

$$\Lambda(t) = \int_0^t \lambda(u) du. \quad (2.5)$$

Esta relação também pode ser vista como a quantidade total de risco que foi acumulada até o momento  $t$ . A forma como a expressão acima se apresenta torna a interpretação mais simples e por via disso perdem-se detalhes presentes na taxa de falha. Há uma relação importante entre  $\lambda(t)$ ,  $f(t)$  e  $S(t)$  dada por

$$\lambda(t) = \frac{f(t)}{S(t)}. \quad (2.6)$$

Isto significa que a taxa de falha, a densidade de probabilidade e a sobrevivência não são independentes uns dos outros, a juntar ainda à relação entre  $f$  e  $S$  em (1.4). Para se obter (2.6) partimos da definição da taxa de falha e segundo [28], temos:

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{P(T \geq t) \Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{S(t) \Delta t} \\ \lambda(t) &= \frac{f(t)}{S(t)}. \end{aligned} \quad (2.7)$$

O primeiro passo provém da propriedade de uma probabilidade condicional e o segundo da definição de uma função de sobrevivência. A última etapa mostra a relação desejada.

Além da relação (2.7), há outra relação importante entre  $\lambda(t)$ , (ou  $\Lambda(t)$ ) e  $S(t)$  que demonstraremos de seguida. De (2.7) e (1.4) segue a relação entre  $\lambda(t)$  e  $S(t)$  dada por

$$\lambda(t) = - \left[ \frac{S'(t)}{S(t)} \right]. \quad (2.8)$$

Integrando membro a membro [28] esta equação nos limites de 0 a  $t$  obtém-se:

$$\begin{aligned} \int_0^t \lambda(u) du &= - \int_0^t \frac{S'(u)}{S(u)} du \\ \int_0^t \lambda(u) du &= -[\ln S(t) - \ln S(0)] \\ \int_0^t \lambda(u) du &= -[\ln S(t) - \ln 1] \\ \int_0^t \lambda(u) du &= -\ln S(t) \end{aligned} \quad (2.9)$$

donde se obtém

$$S(t) = e^{-\int_0^t \lambda(u) du}. \quad (2.10)$$

### 3.2. Censura

Para realizar uma análise de sobrevivência, é necessário registrar o tempo  $t_i$  até ao evento de interesse (falha) para os sujeitos  $i \in \{1, \dots, N\}$  de um grupo. No entanto, isso nem sempre é possível e muitas vezes dispomos de apenas informações parciais sobre o tempo até ao evento. Nesse caso, fala-se em censura. Especificamente, um paciente tem um tempo de sobrevivência censurado (à direita) se quando se regista o tempo o evento ainda não ocorreu para esse paciente. Isso pode acontecer quando:

- um paciente desiste do estudo, por exemplo, parar de comparecer à unidade sanitária para exame de rotina;
- o estudo tem uma linha de tempo fixa e o evento ocorre após o tempo limite;
- um paciente alcançou a cura;
- um paciente morre por outra causa que não a em estudo.

Neste estudo iremos nos focalizar na censura à direita. Neste tipo de censura apenas se sabe que o tempo de sobrevivência dos indivíduos em observação excede o tempo  $t$  de observação, uma vez que a observação termina antes da ocorrência do evento de interesse. A análise de sobrevivência difere da estatística usual porque tem esta

peculiaridade de poder lidar com dados censurados, dispondo de ferramentas próprias que permitem analisar este tipo de dados. Os métodos da estatística usual não são aplicáveis a observações censuradas, podendo ser aplicados apenas se as censuras forem eliminadas desses dados, o que significará perder informação valiosa e poder deturpar as conclusões a retirar dos dados.

### 3.3. Estimadores não paramétricos para função de sobrevivência

Apesar de neste trabalho apenas aplicarmos o estimador de Kaplan-Meier de entre os estimadores não paramétricos, iremos nesta secção fazer uma revisão dos dois estimadores mais comumente utilizados em Análise de Sobrevivência.

#### 3.3.1. Estimador Kaplan-Meier para a função de sobrevivência

O estimador de Kaplan-Meier é uma estimativa não paramétrica de máxima verosimilhança que permite estimar a função de sobrevivência. Este estimador é muitas vezes o primeiro passo para realização da análise de sobrevivência porque é uma abordagem mais simples e versátil. Sendo um método não paramétrico, tem as seguintes suposições [30]:

- As censuras são independentes das probabilidades de sobrevivência dos sujeitos;
- Sujeitos recrutados no início e no final do estudo têm as mesmas probabilidades de sobrevivência;
- Os eventos ocorrem em instantes de tempo específicos.

O estimador de Kaplan-Meier (KM) [28] de uma função de sobrevivência  $S_{KM}(t)$  é dado por

$$S_{KM}(t) = \prod_{i:t_i < t} \frac{n_i - d_i}{n_i} = \prod_{i:t_i < t} \left(1 - \frac{d_i}{n_i}\right) \quad (3)$$

para todo  $t > 0$ , em que:

- $n_i$  é o número de sujeitos em risco no instante  $t_i$ ;
- $d_i$  é o número de falhas ocorridas no intervalo  $[t_i, t_{i+1})$ .

Aqui,  $n_i$  corresponde ao número de sujeitos presentes no estudo no instante  $t_i$ . Em contraste, os sujeitos que experimentaram o evento ou a censura em instantes anteriores não estão mais presentes. Como pode ser visto na equação, (3) o argumento do produto considera apenas eventos  $i$  que ocorrem antes do tempo  $t$ , ou seja,  $t_i < t$ . Portanto, a curva de sobrevivência  $S_{KM}(t)$  para o tempo  $t$  considera todos os eventos que aconteceram antes de  $t$ .

É importante perceber que, para avaliar o estimador Kaplan-Meier, apenas os eventos de interesse que ocorrem em  $t_i$  são importantes para a definição dos saltos da estimativa da função de sobrevivência. Isso significa que, entre dois eventos, por exemplo,  $t_i$  e  $t_{i+1}$ , a curva de sobrevivência é constante. No entanto, as censuras que ocorrem em cada intervalo têm influência no valor da estimativa da função de sobrevivência nos intervalos seguintes. Isso permite uma reformulação simples do estimador de Kaplan-Meier para uma formulação recursiva dada por

$$S_{KM}(t_k) = \frac{n_k - d_k}{n_k} S_{KM}(t_{k-1}). \quad (3.1)$$

### 3.3.2. Estimador de Nelson- Aalen para a função de sobrevivência

Outro método não paramétrico usado na literatura é o estimador de Nelson-Aalen para a função de sobrevivência [28]. Em contraste com o estimador de Kaplan-Meier, que é um estimador direto para a função de sobrevivência, o estimador de Nelson-Aalen é um estimador indireto para  $S(t)$ . Para além disso, o estimador Nelson-Aalen faz uma estimativa directa para a função taxa de falha cumulativa dada por

$$\Lambda_{NA}(t) = \sum_{i:t_i \leq t} \frac{d_i}{n_i} = \sum_{i:t_i \leq t} \lambda_{NA,i}(t). \quad (3.2)$$

A partir desta relação, obtemos uma estimativa para a função de sobrevivência por Nelson-Aalen dada por

$$S_{NA}(t) = \exp(-\Lambda_{NA}(t)) = \exp\left(-\sum_{i:t_i \leq t} \frac{d_i}{n_i}\right). \quad (3.3)$$

Em geral pode-se mostrar que

$$S_{KM}(t) \leq S_{NA}(t). \quad (3.4)$$

### 3.4. Testes de Comparação das funções de sobrevivência

Na análise de sobrevivência uma alternativa para comparar os tempos de sobrevivência entre dois ou mais grupos em estudo é a observação gráfica das curvas de funções de sobrevivência estimadas que, como é óbvio, não fornece uma informação objetiva sobre o quão essas diferenças são eventualmente significativas. Desta forma, testes de hipóteses apropriados são empregues com vista a apurar se ocorrem ou não diferenças significativas entre as funções de sobrevivência em grupos distintos. Os testes de comparação das curvas de sobrevivência mais comuns são:

- Teste Logrank [27];
- Teste de Wilcoxon [27] que é um caso especial de um teste Log-rank ponderado.

O teste de Logrank, também conhecido como teste de Mantel-Haenszel, é um teste de hipótese não paramétrico usado para comparar curvas de sobrevivência. Este teste permite testar a hipótese nula de igualdade de funções de sobrevivência entre 2 grupos dada por

$$\begin{cases} S_1(t) = S_2(t), & \text{no intervalo de observação,} \\ S_1(t) \neq S_2(t), & \text{no intervalo de observação} \end{cases}$$

O teste Logrank faz as seguintes suposições:

- Sujeitos censurados e não censurados têm a mesma probabilidade do evento (a censura não é informativa).

- As curvas de Kaplan-Meier dos dois grupos não se devem cruzar (a suposição de riscos proporcionais deve ser mantida).
- Nenhuma distribuição particular para a curva de sobrevivência é assumida (distribuição livre).

Considere-se primeiro a seguinte Tabela que classifica os indivíduos que experimentaram ou não experimentaram o evento de interesse em cada grupo em observação:

Grupo	Falha		Total
	Sim	Não	
1	$d_{1j}$	$r_{1j} - d_{1j}$	$r_{1j}$
2	$d_{2j}$	$r_{2j} - d_{2j}$	$r_{2j}$
Total	$d_j$	$r_j - d_j$	$r_j$

Fixando os valores marginais, então  $d_{1j}$  segue uma distribuição Hipergeométrica sob a hipótese nula de não haver associação entre o evento e o grupo. O valor esperado para o número de falhas no grupo 1 e no  $j$ -ésimo instante, no caso da hipótese nula de igualdade das funções de sobrevivência é dado por

$$E(d_{1j}) = \frac{r_{1j}d_j}{r_j}. \quad (4)$$

O valor de estatística de teste Logrank tem uma distribuição do qui-quadrado com 1 grau de liberdade

$$\chi^2_{\text{logrank}} = \frac{\left[ \sum_{j=1}^k \left( d_{ij} - r_{ij} \frac{d_j}{r_j} \right) \right]^2}{\sum_{j=1}^k \frac{r_{1j}r_{2j}d_j(r_j - d_j)}{r_j^2(r_j - 1)}}. \quad (4.1)$$

O teste logrank assume que as funções taxas de falha para os dois grupos,  $i = 1, 2$ , são proporcionais, o que significa que as razões entre as taxas de falha para os dois grupos são constantes entre todos os instantes. Deste modo, podem ser usadas

curvas de sobrevivência para visualizar se as funções taxas de falha dos dois grupos são ou não proporcionais. Este teste pode também ser generalizado para a comparação de 3 ou mais funções de sobrevivência.

Outra possibilidade é o Teste de Gehan, que é um teste de duas amostras sem distribuição teórica e é uma generalização do teste de Wilcoxon que mostra observações censuradas à direita. O valor de estatística de teste é dado por

$$\chi^2_{\text{Gehan}} = \frac{\left[ \sum_j^k n_j \left( d_{ij} - r_{ij} \frac{d_j}{r_j} \right) \right]^2}{\sum_j^k n_j^2 \frac{r_{1j} r_{2j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}} \quad (4.2)$$

Assim, o teste de Gehan pondera cada parcela em função do número  $n_j$  de sujeito em risco, sendo esta a principal diferença em relação ao teste Logrank.

### 3.5. Modelos Paramétricos de Análise de Sobrevivência

Em alternativa a um modelo não paramétrico, fazer uma suposição sobre a distribuição de probabilidade da taxa de falha permite modelar parametricamente uma função de sobrevivência com mais detalhe e precisão. Há, no entanto, o perigo de fazer tais suposições quando elas não são justificadas pelos dados. Apesar não aplicarmos métodos paramétricos neste trabalho, mas para garantir uma revisão alargada das possibilidades de estimação em Análise de Sobrevivência, apresentamos nesta secção quatro modelos paramétricos comumente utilizados em análise de sobrevivência.

De notar que, uma vez que a sobrevivência de sujeitos é influenciada por fatores inerentes a estes, os modelos de regressão paramétricos permitem incorporar covariáveis de modo que sejam levadas em consideração no tempo de sobrevivência. Neste caso, a variável independente será o vetor das covariáveis  $x = (x_1, x_2, \dots, x_p)'$  e a variável dependente será o tempo. O vector dos parâmetros  $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$  é determinado em função do modelo e das covariáveis consideradas.

### 3.5.1. Estimação dos Parâmetros do modelo: Método de máxima verosimilhança

Para estimar os parâmetros é comum utilizar-se o método da máxima verosimilhança, definindo a função de verosimilhança dada por

$$\mathcal{L}(t_i, \theta) = \prod_{i=1}^n [f(t_i, \theta)]^{\delta_i} \cdot [S(t_i, \theta)]^{1-\delta_i} \quad (5)$$

onde  $t_i, i = 1, 2, \dots, n$  são os tempos de falha recolhidos e  $\theta = (\theta_1, \theta_2, \dots, \theta_m)$  são parâmetros a estimar e  $\delta_i$  assume os valores 1 ou 0 consoante o instante  $t_i$  é uma falha ou uma censura, respetivamente. Assim, assumindo-se a distribuição da taxa de falha, faz-se a substituição da função densidade de probabilidade e função de sobrevivência na função de verosimilhança e determina-se os parâmetros da distribuição que minimizam este problema. Em algumas distribuições a determinação destes parâmetros resultam em equações sem soluções explícitas para todos os parâmetros. Nestes casos são utilizados métodos iterativos para estimação desses parâmetros recorrendo-se a softwares estatísticos.

### 3.5.2. Modelo Exponencial

No modelo de regressão exponencial a taxa de falha e a função de sobrevivência são dados, respetivamente por

$$\lambda(t|x) = e^{-\beta(x)} \quad (5.1)$$

$$S(t|x) = \exp\left(-\frac{t}{e^{\beta(x)}}\right), \quad t \geq 0 \quad (5.2)$$

onde  $\beta(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ .

### 3.5.3. Modelo de Regressão de Weibull

Este modelo de regressão considera que a taxa de falha é dada por

$$\lambda(t|x) = \frac{\gamma}{e^{\gamma\beta(x)}} t^{\gamma-1}, \quad t \geq 0 \quad (5.3)$$

e a função de sobrevivência por

$$S(t|x) = \exp\left(-\left(\frac{t}{e^{\beta(x)}}\right)^\gamma\right), t \geq 0. \quad (5.4)$$

Para  $\gamma = 1$ , o modelo de Weibull reduz-se ao exponencial, pelo que o modelo de Weibull é uma generalização do anterior e tem, assim, melhor capacidade de se adaptar aos dados.

#### 3.5.4. Modelo de Regressão Log-normal

Para a regressão log-normal, consideramos a regressão linear da média  $\mu$  da variável  $Y = \ln(T)$  em função das covariáveis, isto é,

$$\mu = \beta(x) + \varepsilon \quad (5.5)$$

em que  $\varepsilon$  tem uma distribuição normal com média nula, o que se traduz no modelo de regressão log-normal para a função de sobrevivência dado por

$$S(t|x) = \Phi\left(\frac{\beta(x) - \ln t}{\sigma}\right). \quad (5.6)$$

Assim, assumimos que as covariáveis apenas afetam a média, mas não o desvio padrão.

#### 3.6. Modelo de Cox

Os modelos de regressão paramétricos descritos na sessão anterior supõem que uma distribuição estatística do tempo de sobrevivência seja conhecida a priori. A escolha correta do tipo de distribuição é um fator determinante para obtenção de estimativas confiáveis nestes modelos de regressão. Um modelo de regressão mais flexível, que não pressupõe um conhecimento sobre o tipo de distribuição, foi proposto por Cox em 1972 e é conhecido como modelo de Cox ou modelo de riscos proporcionais ou ainda modelo de taxas de falha proporcionais. Este é um modelo semi-paramétrico, sendo a influência das covariáveis estimada por um fator paramétrico que afeta uma função de risco basal estimada de forma não paramétrica.

Um dos interesses em análise de sobrevivência, para além de estimação do tempo de vida e do efeito das covariáveis na sobrevivência dos sujeitos, é o risco de ocorrência de falha e o modelo de Cox consegue modelar diretamente a função de risco.

O Modelo de Cox é um modelo de regressão semi-paramétrico que assume:

- riscos proporcionais;
- independência temporal das covariáveis;
- linearidade nas covariáveis.

Considere-se um modelo com apenas uma covariável e sejam  $\lambda_1(t)$  e  $\lambda_0(t)$  as funções de risco no decorrer do tempo para indivíduos no novo tratamento e no tratamento tradicional. De acordo com o pressuposto de riscos proporcionais tem-se

$$\frac{\lambda_1(t)}{\lambda_0(t)} = k \quad (6)$$

onde  $k$  é o risco relativo ou o factor com que um sujeito na situação 1 aumenta o risco de falha em relação a um sujeito na situação basal 0. Esta constante é positiva pelo que se assume da forma  $k = \exp(\beta x)$ , sendo  $x$  uma covariável e  $\beta$  o parâmetro a estimar que define a variação do risco (proporcional) da situação basal 0 para a situação 1.

Portanto, este pressuposto de riscos proporcionais sugere a mesma razão entre os riscos definidos pela taxa de falha em cada situação em todos os instantes e que um cruzamento ou outras variações observadas ao longo das suas curvas podem sugerir a não existência de proporcionalidade.

Considerando  $p$  covariáveis,  $x_i, i = 1, 2, \dots, p$  o modelo de taxas de riscos proporcionais é dado por:

$$\lambda(t|X) = \lambda_0(t) \exp\left(\sum_{i=1}^p \beta_i x_i\right) \quad (6.1)$$

e a respectiva função de sobrevivência é

$$S(t|X) = S_0(t) \exp(\sum_{i=1}^p \beta_i x_i). \quad (6.2)$$

O modelo de Cox é denominado semi-paramétrico porque tem duas componentes, uma não-paramétrica e outra paramétrica. A componente não paramétrica,  $\lambda_0(t)$ , é a função de risco basal, não-negativa, e assumida num modelo com covariáveis de valor nulo, ou seja, a função de risco assumida quando  $X = 0$ :

$$\lambda(t, X) = \lambda_0(t). \quad (6.3)$$

### 3.6.1. Proporcionalidade de Funções de Risco

O fundamento do modelo de Cox é o de proporcionalidade das funções de risco entre valores distintos de uma mesma covariável. Considere-se dois indivíduos distintos  $i$  e  $j$  cuja sobrevivência é influenciada por um vector das  $p$  covariáveis. O modelo assume que a razão das taxas de falha entre estes indivíduos é independente do tempo, e é dada por

$$\frac{\lambda(t|x^{(i)})}{\lambda(t|x^{(j)})} = \frac{\lambda_0(t) e^{\beta_1 x_1^{(i)} + \beta_2 x_2^{(i)} + \dots + \beta_p x_p^{(i)}}}{\lambda_0(t) e^{\beta_1 x_1^{(j)} + \beta_2 x_2^{(j)} + \dots + \beta_p x_p^{(j)}}}$$

$$\frac{\lambda(t|x^{(i)})}{\lambda(t|x^{(j)})} = e^{\beta_1(x_1^{(i)} - x_1^{(j)}) + \beta_2(x_2^{(i)} - x_2^{(j)}) + \dots + \beta_p(x_p^{(i)} - x_p^{(j)})}. \quad (6.4)$$

Deste modo se, no início de seguimento, o risco de indivíduo  $i$  experimentar a falha for o dobro do risco do indivíduo  $j$ , essa razão irá se manter até ao fim do seguimento destes indivíduos. Os coeficientes  $\beta$  são muito importantes na componente paramétrica do modelo pois medem o verdadeiro efeito das covariáveis sobre a função de risco.

A fim de realizar inferências sobre o modelo, um novo método de estimação foi introduzido por Cox, nomeadamente o método de verosimilhança parcial que é mais adequado neste modelo devido à presença da componente não-paramétrica. Esta função é dada pela expressão seguinte:

$$\mathcal{L}(\beta) = \prod_{i=1}^n \left( \frac{e^{\beta_1 x_1^{(i)} + \beta_2 x_2^{(i)} + \dots + \beta_p x_p^{(i)}}}{\sum_{j \in R(t_i)} e^{\beta_1 x_1^{(j)} + \beta_2 x_2^{(j)} + \dots + \beta_p x_p^{(j)}}} \right)^{\delta_i} \quad (6.5)$$

Nesta expressão o índice do somatório  $j \in R(t_i)$  indica indivíduos em risco no  $i$ -ésimo instante e  $\delta_i$  indica falhas ou censuras, como descrito anteriormente. As estimativas de máxima verosimilhança parcial são os valores dos coeficientes  $\beta$  que maximizam a função  $\mathcal{L}(\beta)$ .

### 3.6.2. Validação e Adequação do Modelo

A seleção de variáveis no modelo permite a redução do número de parâmetros do modelo de modo a potenciar a eficácia e a interpretabilidade do modelo. Por isso torna-se necessário, em primeiro, lugar obter os coeficientes do modelo utilizando-se testes apropriados que permitam decidir se dada covariável é potencialmente útil ou não no modelo a considerar. Os testes de Wald, de razão de verosimilhanças e o de escores são usados para avaliar se dada covariável deve ou não ser incluída no modelo, por meio da análise da significância do seu coeficiente. Estes visam testar a hipótese nula de que o coeficiente  $\beta$  da covariável  $X$  tem um efeito nulo no modelo. Seja  $\hat{\beta}_j$ , a estimativa do efeito da  $j$ -ésima covariável de forma que esta maximiza a função de verosimilhança  $\mathcal{L}(\beta)$ . Esta estimativa é comparada com um valor hipotético  $\beta_{0j}$  de forma que o quociente entre o quadrado da distância ponderada  $\hat{\beta}_j - \beta_{0j}$  e o erro padrão do coeficiente irrestrito,  $\hat{\beta}_j$  tem distribuição assintótica  $\chi^2$  sobre a hipótese nula. Assim, a estatística de Wald, no caso de um só parâmetro, é dada por

$$W = \frac{(\hat{\beta}_j - \beta_{0j})^2}{Var(\hat{\beta}_j)} \sim \chi_1^2. \quad (6.6)$$

No caso de teste em vários parâmetros, a estatística de Wald é dada por

$$W = (\hat{\beta}_j - \beta_{0j})^T I(\hat{\beta})(\hat{\beta}_j - \beta_{0j}) \sim \chi_n^2 \quad (6.7)$$

onde  $\hat{\beta}$  é a estimativa da matriz de variância/covariância de  $\hat{\beta}_j$  e  $I(\hat{\beta})$  é denominada matriz de informação observada [27], dada por

$$I(\hat{\beta}) = \left( \frac{\partial^2 \log\{\mathcal{L}(\beta_{0j})\}}{\partial \beta_{0j}^2} \right)^{-1}.$$

Uma alternativa ao teste de Wald é o teste de razão de verosimilhanças. Este teste faz a comparação entre logaritmo da verosimilhança do modelo com covariáveis e logaritmo da verosimilhança do modelo sem covariáveis, como também é utilizado para comparar modelos distintos. A estatística de Teste de Razão de Verosimilhanças (TRV) é

$$TRV = -2 \log \left[ \frac{\mathcal{L}(\hat{\beta}_j)}{\mathcal{L}(\beta_{0j})} \right] = 2 \log[\mathcal{L}(\beta_{0j}) - \mathcal{L}(\hat{\beta}_j)]. \quad (6.8)$$

A hipótese nula é rejeitada para valores de  $TRV > \chi_{n^2, 1-\alpha}$

O teste Escore provém da função escore  $U(\beta)$  e da respectiva matriz de variâncias/covariâncias. Sendo

$$U(\beta) = \frac{\partial \log \mathcal{L}(\beta)}{\partial \beta} \quad (6.9)$$

tem-se a estatística de teste dada por

$$S = U'(\beta_0)[I(\beta_0)]^{-1}U(\beta_0). \quad (6.10)$$

E a hipótese nula será rejeitada a  $100\alpha\%$  de significância, em grandes amostras, se  $S > \chi_{n^2, 1-\alpha}$

### 3.6.3. Método de Seleção de Variáveis

O método de eliminação *backward* [31] começa com um modelo com todas as covariáveis e a partir deste, começa-se a eliminação, em cada passo, de cada uma das covariáveis menos significativas. Este processo só termina quando todas as covariáveis restantes contribuem significativamente no modelo. O processo inverso

do método *backward* é o método de seleção *forward* que não abordamos no presente estudo. Este é também construído passo a passo, incluindo em cada fase a variável que mais reduz a estatística TRV a partir de um modelo simples sem covariáveis.

#### 3.6.4. Validação da hipótese de Riscos Proporcionais

Todas as técnicas utilizadas para a avaliação da adequação do modelo só serão úteis se o pressuposto de riscos proporcionais do modelo for válido. Uma alternativa para a avaliação da hipótese de proporcionalidade dos riscos é a análise gráfica dos resíduos de Schoenfeld porém, dada a subjetividade na interpretação por esta técnica, testes de hipóteses são necessários para maior segurança. O teste de hipóteses sobre as correlações dos resíduos com o tempo é usado e visa testar a hipótese nula da não existência de correlações entre estes.

Tendo-se verificado a validade da hipótese dos riscos proporcionais os resíduos de Cox-Snell são utilizados para a avaliação da adequação global do modelo, resíduos Martingala para avaliação da forma funcional adequada das covariáveis e resíduos *deviance* para facilitar a identificação de existência de *outliers* que podem influenciar o modelo.

#### 3.6.5. Resíduos de Cox-Snell

Tendo definido o modelo final com as covariáveis singificativas e portanto com impacto no tempo de sobrevivência e tendo verificado que estas satisfazem o pressuposto de riscos proporcionais para o modelo de Cox através da análise dos resíduos de Shoenfeld, há necessidade de se verificar a qualidade do ajuste global do modelo aos dados através da análise dos resíduos de Cox-Snell. Este resíduos são definidos por

$$\hat{\epsilon}_i = \hat{\Lambda}_0(t_i) \exp(X_i^T \hat{\beta}). \quad (6.11)$$

Uma forma de verificar a validade do modelo é através da comparação das estimativas da taxa de falha acumulada,  $\hat{\Lambda}_0(t_i)$ , com a distribuição exponencial padrão. O procedimento consiste na comparação das curvas de Kaplan-Meier e do modelo exponencial padrão, obtidas em cada  $\hat{\epsilon}_i$ . Visto que a taxa de falha para dados

exponencialmente distribuídos é constante, então se o gráfico de  $\hat{\Lambda}(\hat{e}_i)$  for aproximadamente linear considera-se um bom ajuste.

Os resíduos de Cox-Snell verificam a adequação global do modelo aos dados, mas existem formas de verificar aspetos particulares do ajuste, conforme descreveremos de seguida.

### 3.6.6. Resíduos Martingala

Os resíduos Martingala são usados para determinar se a representação está correta em função das covariáveis, isto é se esta dependência tem forma linear, quadrática ou outra. No caso da forma ser linear nenhuma transformação é necessária, uma vez que este é o pressuposto no modelo de Cox. Caso contrário, será necessária uma transformação adequada da variável antes desta ser inserida no modelo de Cox.

No modelo de Cox os resíduos Martingala são dados por

$$\hat{m}_i = \delta_i - \hat{\Lambda}_i(t_i). \quad (6.12)$$

Distribuídos no intervalo  $(-\infty, 1]$ , valores positivos dos resíduos Martingala indicam todas as falhas que ocorreram durante o tempo de seguimento enquanto valores negativos indicam os sujeitos restantes, ou seja, os que sobreviveram.

### 3.6.7. Resíduos Deviance

Uma abordagem com vista a detectar os possíveis *outliers* que podem influenciar bastante nos resultados da estimação é o estudo dos resíduos *deviance*. Visto que os resíduos Martingala são geralmente assimétricos e extremamente enviesados, a transformação destes no sentido de reduzir o enviesamento e assimetrias resulta nos resíduos *deviance*, dados por

$$\hat{d}_i = \text{sgn}(\hat{m}_i) \sqrt{-2(\hat{m}_i + \delta_i \ln(\delta_i - \hat{m}_i))}. \quad (6.13)$$

Os resíduos deviance tendem a centralizar todas as observações em torno do zero. Um bom ajuste do modelo é obtido quando as observações ficam distribuídas no intervalo de  $(-3,3)$ .

### 3.7. Modelos de Tempo de Vida Acelerados

Apesar de não serem utilizados neste trabalho, vamos também rever os conceitos básicos dos modelos de tempo de vida acelerado. Estes são modelos de análise de sobrevivência mais recentes que relacionam a distribuição do tempo de vida com as variáveis explicativas. Nestes modelos as covariáveis têm um efeito multiplicativo sobre o tempo de vida, acelerando ou desacelerando-o, isto é, existe um fator  $k$  tal que  $T = kT_0$ . A magnitude do fator  $k$ , denominado factor de aceleração, está relacionada com o vetor das covariáveis  $x$  através da relação

$$k = e^{-\beta x}. \quad (7)$$

Se  $k > 1 \Leftrightarrow \beta x < 0$ , o efeito das covariáveis acelera o tempo de sobrevivência.

Se  $k < 1 \Leftrightarrow \beta x > 0$ , o efeito das covariáveis desacelera o tempo de sobrevivência.

A probabilidade de sobrevivência de indivíduos com covariáveis  $x$ , com tempo de vida acelerado, é modelada pela expressão

$$S(t; x) = S_0(e^{\beta x} t). \quad (7.1)$$

Esta expressão traduz o facto de que a probabilidade de um sujeito com covariáveis  $x$  estar vivo no tempo  $t$  iguala a probabilidade de que um sujeito de referência esteja vivo no tempo  $te^{\beta x}$ , o que significa um tempo a passar tão depressa ou lentamente em um fator  $e^{\beta x}$ , conforme esse fator cresce ou decresce.

Como a taxa de falha é dada por

$$\lambda(t) = \lambda_0(te^{x\beta})e^{x\beta}. \quad (7.2)$$

Significa que o sujeito em causa está exposto a qualquer idade, por exemplo, ao dobro do risco de um sujeito de referência duas vezes mais velho.

Capítulo 4  
Resultados e Discussão

## 4. Resultados e Discussão

### 4.1. Descrição da Pesquisa

Este estudo é observacional, utiliza uma base de dados secundária de pacientes com HIV em seguimento num programa de TARV do centro de saúde de Changanine, em Moçambique, província de Gaza no distrito de Chibuto. O estudo foi aprovado pelo Comité Nacional de Bioética para a Saúde (CNBS) de Moçambique e foi registado sob o número 115/CNBS/2020 e tem por objetivo estabelecer um modelo de análise de sobrevivência de pacientes com HIV em função das covariáveis disponíveis na base de dados.

A base de dados considerada contém 1301 pacientes seguidos no período de 1 de Janeiro de 2000 a 20 de Março de 2020. Foram excluídos da pesquisa todos os sujeitos seguidos durante o período anterior a 2013 e os seguidos após 2019, por apresentarem informação incompleta nas covariáveis de interesse. Procedeu-se igualmente para os restantes sujeitos da base de dados obtida, conforme os critérios de inclusão/exclusão do protocolo de pesquisa, que destacamos a seguir, tendo-se obtido uma base de dados final com 1122 sujeitos. Destes, 816 sobreviveram em TARV até ao fim do estudo. Por outro lado, perdeu-se o seguimento de 135 sujeitos durante o estudo, sendo os 171 correspondentes a óbitos.

#### 4.1.1. Critérios de Inclusão

Foram incluídos na base de dados

- Pacientes do centro de saúde de Changanine seguidos entre 2013 a 2019.
- Pacientes HIV positivos com tempo de seguimento não inferior a 6 meses;
- Pacientes com idades compreendidas entre 16 e 65 anos no período de diagnóstico;
- Pacientes com dados completos para as covariáveis de interesse;

#### 4.1.2. Critérios de Exclusão

Foram excluídos da base de dados

- Transferidos para a unidade sanitária num período inferior a 6 meses até à data do estudo;
- Pacientes com ausências superiores a 400 faltas durante todo o período de estudo (2013 – 2019).

O evento de interesse é a morte, que se supõe que esteja relacionada com a infeção por HIV, embora este aspeto não possa ser garantido uma vez que a base de dados não foi criada especificamente para estes estudo. Assim, procura-se determinar e quantificar o efeito de diversas covariáveis sobre a sobrevivência dos pacientes neste programa de tratamento, utilizando-se o método de Kaplan-Meier (com covariáveis qualitativas ou por classes) e o de Cox (para covariáveis preditivas quantitativas e qualitativas, avaliando simultaneamente o efeito de vários fatores de risco no tempo de sobrevivência). A descrição das covariáveis em análise é apresentada nas Tabelas 1, 2 e 3.

O tempo de sobrevivência foi contado a partir da data de admissão do paciente ao programa de TARV, tendo esta ocorrido no período de estudo que decorreu entre 1 de Janeiro de 2013 e 31 de Dezembro de 2019. Este tempo é o tempo de vida dos sujeitos, sendo portanto terminado no óbito ou censura do sujeito. As censuras correspondem às situações de abandonos, de não ter ocorrido a falha durante o período do estudo, a transferências para outros centros de saúde ou outras saídas do estudo diferentes de óbitos por HIV.

Dos 1122 sujeitos seguidos no período em referência, 171 foram óbitos e dos restantes 951 censurados, 135 sujeitos foram perdidos no seguimento.

#### 4.1.3. Covariáveis

Na base de dados disponibilizada interessou analisar 15 covariáveis das quais apenas 3 foram tratadas como contínuas e as restantes 12 foram tratadas como categóricas. Fez-se a análise descritiva de todas covariáveis para se obter uma visão sobre a distribuição das frequências das categorias que compõem estas covariáveis categóricas e também verificar a dispersão das covariáveis contínuas.

A carga viral foi considerada covariável categórica devido à existência da possibilidade da carga viral ser detetável ou indetetável por teste laboratorial padrão, tendo sido dividida em 4 categorias.

Tabela 1: Descrição das variáveis

<b>Código da Variável</b>	<b>Codificação</b>	<b>Tipo</b>	<b>Descrição</b>
sexo	F – Feminino M – Masculino	Dicotómica	
idade		Quantitativa Contínua	Idade, e anos
IMC		Quantitativa Contínua	Índice de massa corporal
Diag_Nutric	1 – Normal 2 – Ligeira 3 – Moderada 4 – Grave	Ordinal	Diagnóstico da Nutrição
estadio	1 – Primeiro estadio 2 – Segundo estadio 3 – Terceiro estadio 4 – Quarto estadio	Ordinal	Estádios da OMS
Carga_viral	1 – abaixo de 1000 2 – entre 1000 a 5000 3 – entre 5000 a 10000 4 – acima de 10000	Ordinal	Carga Viral (número de cópias do vírus)
CD4		Quantitativa Contínua	Número de células CD4
trat	0 – Tratamento Padrão 1 – Mudança de linha	Dicotómica	Tratamento com anti-retroviral
TPT	0 – Não fez TPT 1 – Fez TPT	Dicotómica	Tratamento preventivo da tuberculose
GAAC	0 – Não pertence a GAAC 1 – Pertence a um GAAC	Dicotómica	Grupos de Apoio à Adesão Comunitária
diagn_SK	0 – Diagnóstico Negativo 1 – Diagnóstico Positivo	Dicotómica	Diagnóstico do Sarcoma de Kaposi
grav_lact	0 – Não grávida nem lactante 1 – Grávida ou lactante	Dicotómica	Grávida ou Lactante

diagn_TB	0 – Sem diagnóstico de TB 1 – com diagnóstico de TB	Dicotômica	Diagnóstico de Tuberculose
profil_CTX	0 – Não fez tratamento CTX 1 – Fez tratamento CTX	Dicotômica	Tratamento profilático com Cotrimoxazol
profil_INH	0 – Não fez tratamento INH 1 – Fez tratamento INH	Dicotômica	Tratamento profilático com Isoniazida

Tabela 2: Análise Descritiva das variáveis Categóricas

Variável	Níveis		Censuras	Falhas	Frequencia	
Diag_Nutri	1	Normal	793	82	875	78%
	2	Ligeira	21	2	23	2.1%
	3	Moderada	100	63	163	14.5%
	4	Grave	37	24	61	5.4%
estadio	1	Primeiro estadio	557	35	592	52.8%
	2	Segundo estadio	229	33	262	23.4%
	3	Terceiro estadio	154	77	231	20.6%
	4	Quarto estadio	11	26	37	3.3%
sexo	F	Feminino	749	77	826	73.6%
	M	Masculino	202	94	296	26.4%
trat	0	Tratamento Padrão (1ª linha)	942	170	1112	99.1%
	1	Mudança de linha	9	1	10	0.9%
TPT	0	Não fez TPT	927	170	1097	97.8%
	1	Fez TPT	24	1	25	2.2%
GAAC	0	Não pertence a GAAC	729	159	888	79.1%
	1	Pertence a um GAAC	222	12	234	20.9%
Diagn_SK	0	Diagnóstico Negativo	649	73	722	64.3%
	1	Diagnóstico Positivo	302	98	400	35.7%
Grav_lact	0	Não grávida nem lactante	742	161	903	80.5%
	1	Grávida ou lactante	209	10	219	19.5%
Diagn_TB	0	Sem diagnóstico de TB	942	121	1063	94.7%
	1	Com diagnóstico de TB	9	50	59	5.3%
profil_CTX	0	Não fez tratamento CTX	476	166	642	57.2%
	1	Fez tratamento CTX	475	5	480	42.8%
profil_INH	0	Não fez tratamento INH	561	168	729	65%
	1	Fez tratamento INH	390	3	393	35%

Carga_viral	1	Carga viral abaixo de 1000	880	60	940	83.8%
	2	Cargaviral entre 1000 a 5000	28	2	30	2.7%
	3	Carga Viral entre 5000 a 10000	11	5	16	1.4%
	4	Carga Viral acima de 10000	32	104	136	12.1%

Tabela 3: Análise descritiva das variáveis contínuas

Variável	Média	Mínimo	Máximo	Desvio Padrão
Idade (anos)	40.18	16	90	12.8
IMC	20.87	13	58	3.4
CD4	437.48	2	2540	304.55

#### 4.2. Estimativa não Paramétrica de Kaplan-Meier

Foram feitas estimativas da função de sobrevivência pelo método de Kaplan-Meier para cada variável qualitativa distinta, para comparar e descrever o comportamento da função de sobrevivência considerando as categorias da variável qualitativas como grupos distintos, para os quais se estimou a função de sobrevivência. A função de sobrevivência representa a estimativa de percentagem, a cada instante, de sujeitos vivos em cada grupo, não considerando o efeito das restantes covariáveis.

Para efeitos de comparação das curvas de sobrevivência por Kaplan-Meier foi feito o teste Logrank para cada uma das variáveis categóricas. Este teste visa testar a hipótese nula de igualdade das funções de sobrevivência para os diferentes grupos, tendo esta não sido rejeitada para o caso da covariável TPT e a da covariável tratamento, com valores-p de 0.4 e 0.3, respetivamente, o que nos sugere a inexistência de diferenças entre os grupos que compõem cada uma destas covariáveis.

Importa referir que na variável TPT, 1097 sujeitos não fizeram o tratamento preventivo da tuberculose enquanto apenas 25 o fizeram, o que pode ter contribuído bastante para o resultado obtido pelo teste Logrank visto que há muito poucos casos no primeiro grupo em relação ao segundo. O mesmo ocorre em relação à variável tratamento

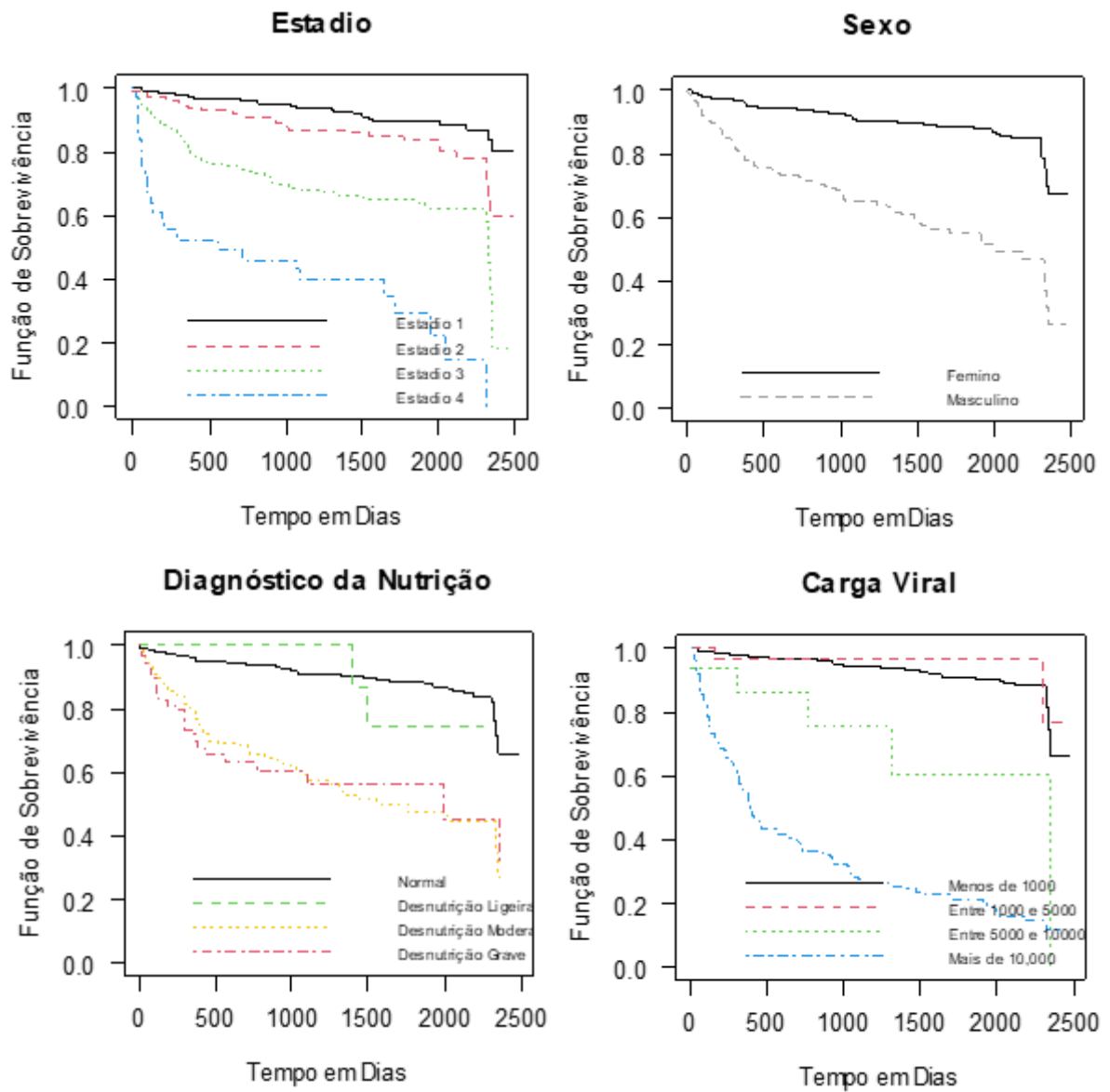
composta pelos 1112 sujeitos na primeira linha de tratamento e 10 sujeitos que mudaram de linha.

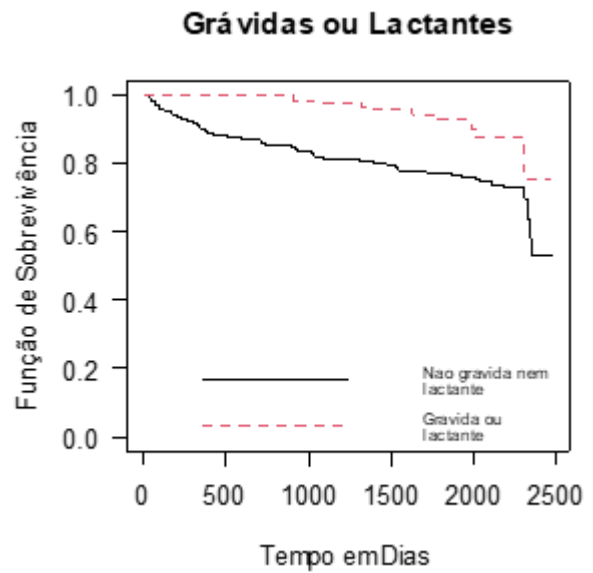
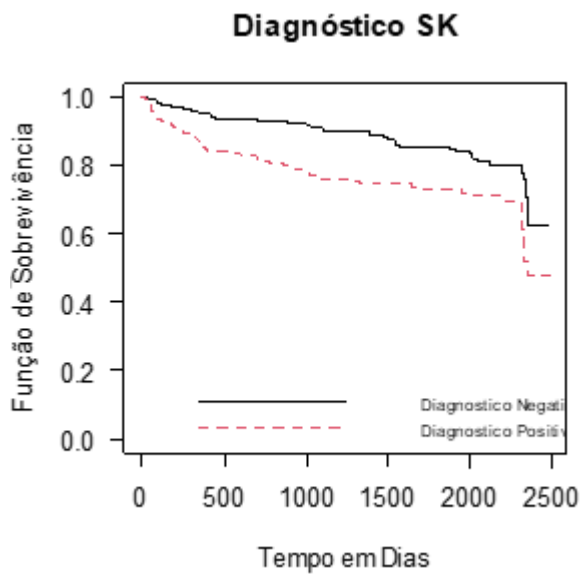
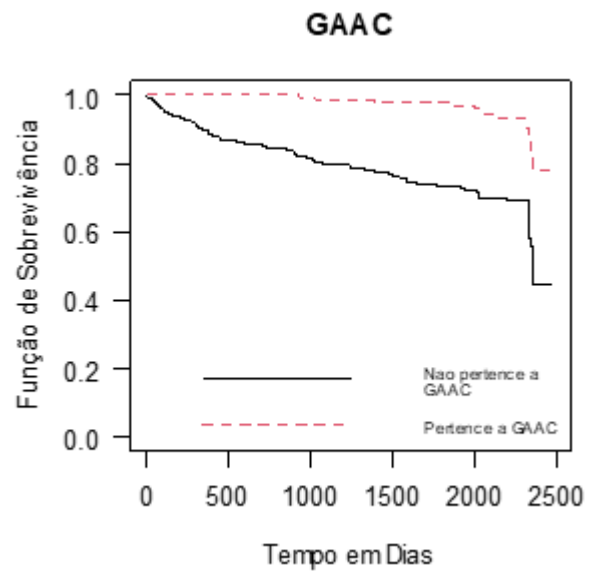
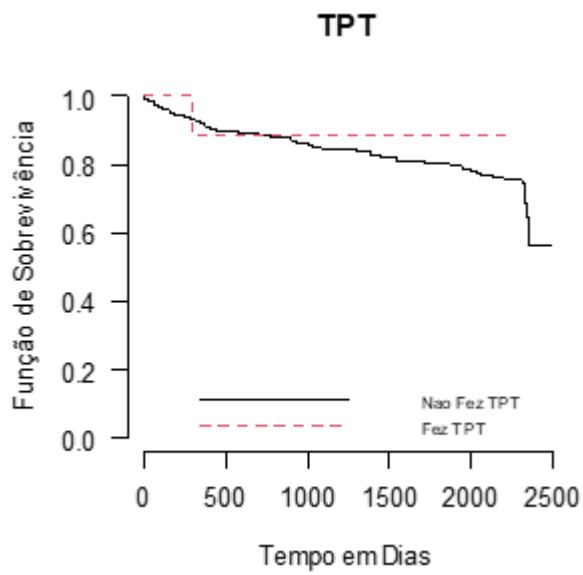
O teste Logrank foi significativo para as restantes covariáveis, sendo de p valor de  $10^{-6}$  para a covariável grav\_lact,  $10^{-7}$  para diagn\_SK,  $2 \times 10^{-12}$  para GAAC e de  $2 \times 10^{-16}$  para as restantes, conforme a Tabela 4.

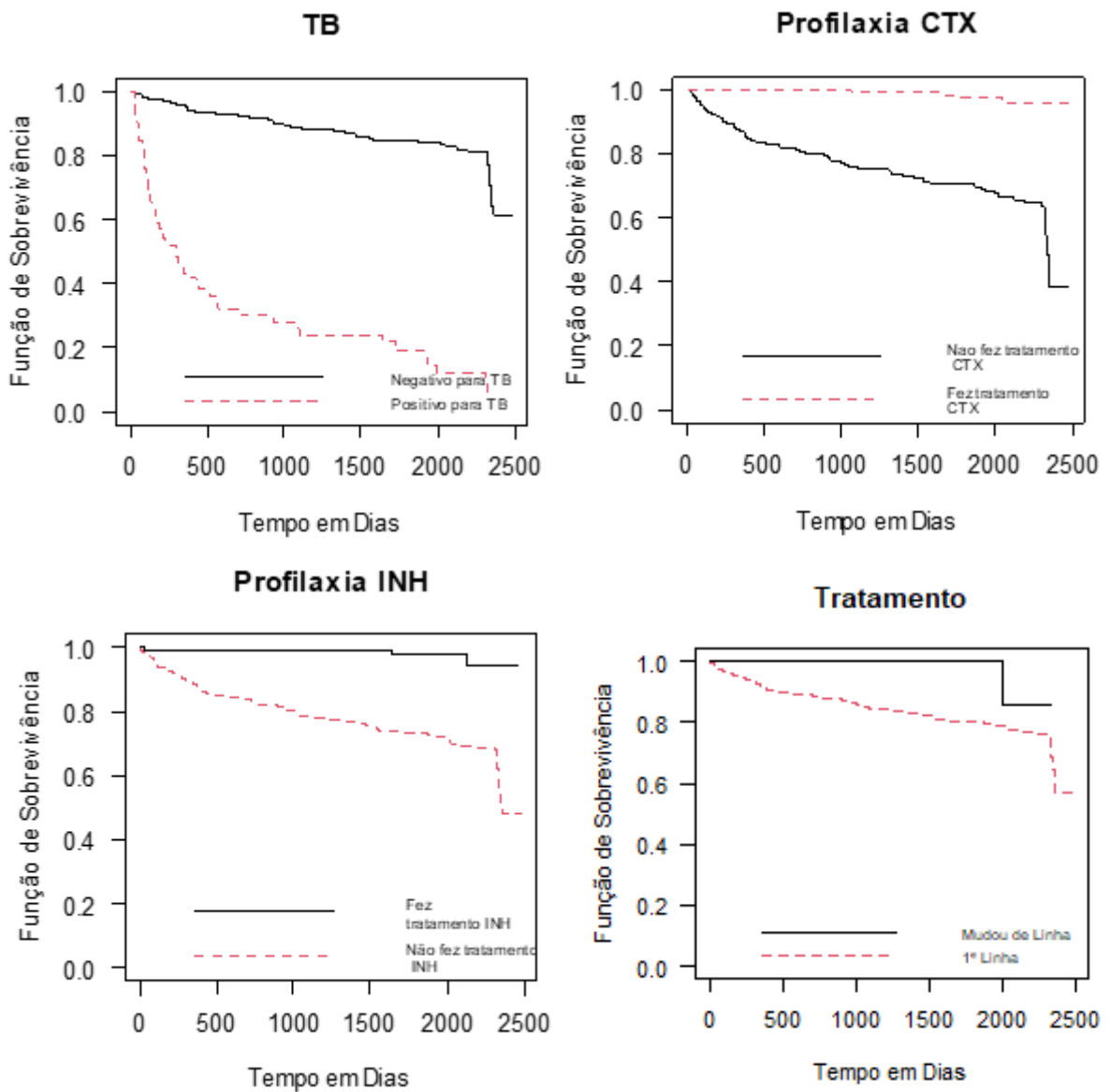
Tabela 4: Teste Logrank para comparação das funções de sobrevivência estimadas por Kaplan-Meier entre os grupos de cada uma das variáveis categóricas

<b>Covariável</b>	$\chi^2$	<b>Graus de Liberdade</b>	<b>p- value</b>
TPT	0.6	1	0.4
trat	0.9	1	0.3
carga_viral	569	3	$2 \times 10^{-16}$
estadio	203	3	$2 \times 10^{-16}$
Diag_Nutric	150	3	$2 \times 10^{-16}$
GAAC	49.5	1	$2 \times 10^{-12}$
diagn_SK	27.9	1	$10^{-7}$
grav_lact	23.4	1	$10^{-6}$
Diagn_TB	324	1	$2 \times 10^{-16}$
profil_CTX	108	1	$2 \times 10^{-16}$
profil_INH	68.3	1	$2 \times 10^{-16}$

Figura 1: Estimativas de Kaplan-Meier da função de sobrevivência para as covariáveis categóricas







Baseando-se na Figura 1, parece razoável afirmar que as covariáveis em causa podem ser usadas como fatores prognósticos da sobrevivência dos pacientes com HIV, embora pareça violado o pressuposto de proporcionalidade de riscos para o teste de Logrank nos casos em que ocorrem cruzamentos das curvas.

Tabela 5: Estimativas de Kaplan-Meier do tempo de vida mediano em cada grupo das variáveis categóricas

Variável	Níveis		Tempo mediano (dias)
Diag_Nutri	1	Normal	>2500
	2	Ligeira	>2500
	3	Moderada	1543
	4	Grave	1994
estadio	1	Primeiro estadio	>2500
	2	Segundo estadio	>2500
	3	Terceiro estadio	2326
	4	Quarto estadio	563
sexo	F	Feminino	>2500
	M	Masculino	2016
trat	0	Tratamento Padrão (1ª linha)	>2500
	1	Mudança de linha	>2500
TPT	0	Não fez TPT	>2500
	1	Fez TPT	>2500
GAAC	0	Não pertence a GAAC	2353
	1	Pertence a um GAAC	>2500
Diagn_SK	0	Diagnóstico Negativo	>2500
	1	Diagnóstico Positivo	2350
Grav_lact	0	Não grávida nem lactante	>2500
	1	Grávida ou lactante	>2500
Diagn_TB	0	Sem diagnóstico de TB	>2500
	1	Com diagnóstico de TB	300
profil_CTX	0	Não fez tratamento CTX	2339
	1	Fez tratamento CTX	>2500
profil_INH	0	Não fez tratamento INH	2356
	1	Fez tratamento INH	>2500
Carga_viral	1	Carga viral abaixo de 1000	>2500
	2	Cargaviral entre 1000 a 5000	>2500
	3	Carga Viral entre 5000 a 10000	2353
	4	Carga Viral acima de 10000	386

Na tabela 5, apresentam-se as estimativas por KM do tempo mediano de vida. Através da mesma pode se notar que em todos os casos das covariáveis trat, TPT e grav\_lact o tempo de vida mediano não se dá no período de estudo pelo que estas podem não ser significativas no modelo de Cox. Por outro lado, as covariáveis trat e TPT não foram significativas no teste logrank, com valor p de 0.3 e 0.4, respetivamente.

#### 4.3. Ajuste do Modelo de Cox

Foi considerado um modelo de Cox com todas as covariáveis em primeiro lugar dada à relevância clínica destas e em particular para o programa do HIV. O modelo final foi obtido a partir do primeiro pela retirada sucessiva das variáveis menos significativas uma a uma, até se obter um modelo de Cox com todas as covariáveis significativas, utilizando-se o método de eliminação *backward* descrito na sessão 3.6.3.

O primeiro modelo, o modelo com todas as covariáveis foi obtido e, pelo método de eliminação *backward* foram sucessivamente obtidos outros modelos por eliminação da covariável menos significativa. De acordo com o modelo 1 (Tabela 6), o segundo modelo foi obtido por eliminação da covariável grav\_lact visto que esta é a menos significativa com p-valor de 0.770902, o maior observado.

Tabela 6: Modelo de Cox com todas as covariáveis obtido pelo teste de Wald

Variável	$\hat{\beta}_i$	p-value	Variável	$\hat{\beta}_i$	p-value
idade	-0.005985	0.369533	(Diag_Nutric)4	0.586226	0.056956
(profil_INH)1	-1.163385	0.073773	(estadio)2	0.492996	0.047443
(profil_CTX)1	-1.956422	0.000128	(estadio)3	0.489836	0.036224
(diagn_TB)1	0.790927	0.001112	(estadio)4	0.521702	0.144243
(grav_lact)1	0.108086	0.770902	(diagn_SK)1	0.306825	0.070313
(GAAC)1	-1.418385	1.50e-05	(carga_viral)2	-0.287674	0.690496
(TPT)1	0.419712	0.679499	(carga_viral)3	0.578353	0.231962
(trat)1	-0.464725	0.665575	(carga_viral)4	1.655829	4.87e-16
(Diag_Nutric)2	1.081710	0.142755	IMC	-0.034988	0.258053
(Diag_Nutric)3	0.751307	0.000515	(sexo)M	0.882926	1.18e-07

Os valores de referência considerados para a *baseline* das covariáveis categóricas que compõem o modelo com todas as covariáveis são profil\_INH 0, profil\_CTX 0, diagn\_TB 0, grav\_lact 0, GAAC 0, TPT 0, trat 0, Diag\_Nutric 1, estadio 1, diagn\_SK 0, carga\_viral 1 e sexo F.

De seguida com a retirada da covariáveis grav\_lact obtemos o segundo modelo (Tabela 7) pelo teste de Wald.

Tabela 7: Modelo II obtido a partir do primeiro por eliminação da covariável grav\_lact

Variável	$\hat{\beta}_i$	p-value	Variável	$\hat{\beta}_i$	p-value
idade	-0.006404	0.325668	(estadio)2	0.488809	0.048964
(profil_INH)1	-1.153616	0.074999	(estadio)3	0.484345	0.037872
(profil_CTX)1	-1.950457	0.000126	(estadio)4	0.516655	0.147793
(diagn_TB)1	0.788780	0.001134	(diagn_SK)1	0.306867	0.070363
(GAAC)1	-1.417242	1.50e-05	(carga_viral)2	-0.278069	0.700016
(TPT)1	0.429370	0.672366	(carga_viral)3	0.572495	0.236226
(trat)1	-0.412070	0.696508	(carga_viral)4	1.651877	5.15e-16
(Diag_Nutric)2	1.108140	0.130141	IMC	-0.033602	0.270585
(Diag_Nutric)3	0.752771	0.000501	factor(sexo)M	0.873222	8.29e-08
(Diag_Nutric)4	0.590412	0.054643			

O modelo III será obtido por eliminação da covariável (carga\_viral) 2 visto que é a menos significativa no modelo II (Tabela 7). Pelo mesmo processo de eliminação *backward* obtemos o novo modelo (Tabela 8).

Tabela 8: Modelo III obtido a partir do modelo II pela eliminação da covariável (carga\_viral) 2

Variável	$\hat{\beta}_i$	p-value	Variável	$\hat{\beta}_i$	p-value
idade	-0.006389	0.326217	(estadio)2	0.488606	0.049065
(profil_INH)1	-1.153346	0.075339	(estadio)3	0.487002	0.036803
(profil_CTX)1	-1.949875	0.000129	(estadio)4	0.522388	0.143343

(diagn_TB)1	0.784586	0.001193	(diagn_SK)1	0.306665	0.070611
(GAAC)1	-1.416770	1.51e-05	(carga_viral)3	0.580874	0.229196
(TPT)1	0.439465	0.665065	(carga_viral)4	1.661795	2.31e-16
(trat)1	-0.407367	0.699810	IMC	-0.033496	0.272526
(Diag_Nutric)2	1.118707	0.126338	factor(sexo)M	0.875165	7.75e-08
(Diag_Nutric)3	0.755118	0.000485			
(Diag_Nutric)4	0.595361	0.052592			

Eliminamos agora a terceira covariável menos significativa que, a partir da tabela 8 pode ser visto que a covariável é trat com valor p de 0.69981 e obtém-se o modelo IV apresentado na Tabela 9 seguinte:

Tabela 9: Modelo IV obtido pelo teste de Wald a partir do modelo III pela eliminação da covariável trat

Variável	$\hat{\beta}_i$	p-value	Variável	$\hat{\beta}_i$	p-value
idade	-0.006212	0.337874	(estadio)3	0.488399	0.036132
(profil_INH)1	-1.174786	0.067717	(estadio)4	0.529556	0.137296
(profil_CTX)1	-1.957275	0.000110	(diagn_SK)1	0.313842	0.062887
(diagn_TB)1	0.772894	0.001296	(carga_viral)3	0.572960	0.235240
(GAAC)1	-1.445891	6.39e-06	(carga_viral)4	1.656790	2.41e-16
(TPT)1	0.441176	0.663856	IMC	-0.033940	0.266889
(Diag_Nutric)2	1.121761	0.125316	factor(sexo)M	0.877836	7.09e-08
(Diag_Nutric)3	0.752049	0.000504			
(Diag_Nutric)4	0.583582	0.056598			
(estadio)2	0.478919	0.053064			

A covariável menos significativa é a agora TPT, a ser eliminada do modelo IV. Assim obtém-se pelo mesmo teste de Wald o modelo V (Tabela 10).

Tabela 10: Modelo V obtido pelo teste de Wald por eliminação da covariável TPT do modelo IV

Variável	$\hat{\beta}_i$	<i>p-value</i>	Variável	$\hat{\beta}_i$	<i>p-value</i>
idade	-0.006305	0.330494	(estadio)4	0.534880	0.133682
(profil_INH)1	-1.174518	0.067591	(diagn_SK)1	0.315571	0.061458
(profil_CTX)1	-1.958437	0.000108	(carga_viral)3	0.566348	0.240454
(diagn_TB)1	0.768107	0.001386	(carga_viral)4	1.649772	2.44e-16
(GAAC)1	-1.443643	6.64e-06	IMC	-0.033401	0.273908
(Diag_Nutric)2	1.116536	0.127	factor(sexo)M	0.878252	7.01e-08
(Diag_Nutric)3	0.752203	0.000501			
(Diag_Nutric)4	0.589770	0.053949			
(estadio)2	0.477323	0.053827			
(estadio)3	0.489707	0.035645			

Pela eliminação da covariável idade do modelo anterior, sendo que esta é a menos significativa, obtemos o modelo VI (Tabela 11).

Tabela 11: Modelo VI obtido pelo teste de Wald por eliminação da covariável idade do modelo V

Variável	$\hat{\beta}_i$	<i>p-value</i>	Variável	$\hat{\beta}_i$	<i>p-value</i>
(profil_INH)1	-1.15321	0.071929	(diagn_SK)1	0.30648	0.068893
(profil_CTX)1	-1.95967	0.000103	(carga_viral)3	0.58139	0.227226
(diagn_TB)1	0.77542	0.001229	(carga_viral)4	1.65921	< 2e-16
(GAAC)1	-1.47369	3.88e-06	IMC	-0.04061	0.173349
(Diag_Nutric)2	1.09084	0.135736	factor(sexo)M	0.85296	1.15e-07
(Diag_Nutric)3	0.68574	0.000833			
(Diag_Nutric)4	0.52803	0.078808			
(estadio)2	0.48431	0.050218			
(estadio)3	0.49630	0.033307			
(estadio)4	0.53116	0.135455			

A covariável menos significativa é agora (carga\_viral) 3 pelo que será eliminada para se obter o modelo VII. Pelo mesmo teste de Wald obtém-se o modelo da Tabela 12 seguinte:

Tabela 12: Modelo VII obtido pelo teste de Wald por eliminação da covariável (carga\_viral) 3 do modelo VI

Variável	$\hat{\beta}_i$	<i>p-value</i>	Variável	$\hat{\beta}_i$	<i>p-value</i>
(profil_INH)1	-1.18095	0.065507	(diagn_SK)1	0.31062	0.064828
(profil_CTX)1	-1.93677	0.000125	(carga_viral)4	1.61398	< 2e-16
(diagn_TB)1	0.77350	0.001268	IMC	-0.03932	0.186469
(GAAC)1	-1.50157	2.36e-06	factor(sexo)M	0.87513	4.13e-08
(Diag_Nutric)2	1.07397	0.141694			
(Diag_Nutric)3	0.69281	0.000727			
(Diag_Nutric)4	0.56570	0.057048			
(estadio)2	0.49246	0.046365			
(estadio)3	0.51491	0.027379			
(estadio)4	0.54154	0.128328			

Visto que continuamos tendo covariáveis não significativas o processo de eliminação continua. A covariável menos significativa no modelo VII é IMC. Por eliminação desta covariável obtemos o modelo VIII (Tabela 13)

Tabela 13: Modelo VIII obtido pelo teste de Wald por eliminação da covariável IMC do modelo VII

Variável	$\hat{\beta}_i$	<i>p-value</i>	Variável	$\hat{\beta}_i$	<i>p-value</i>
(profil_INH)1	-1.1823	0.06486	(diagn_SK)1	0.3231	0.05470
(profil_CTX)1	-1.9593	0.0001	(carga_viral)4	1.6113	< 2e-16
(diagn_TB)1	0.7126	0.00241	factor(sexo)M	0.8603	7.36e-08
(GAAC)1	-1.5021	2.33e-06			
(Diag_Nutric)2	1.1945	0.09965			
(Diag_Nutric)3	0.8297	4.46e-06			

(Diag_Nutric)4	0.7933	0.00121
(estadio)2	0.4815	0.05166
(estadio)3	0.5292	0.02342
(estadio)4	0.6452	0.06273

Tendo-se o modelo com covariáveis significativas (Tabela 13) prossegue-se de seguida a sua validação, começando-se pela avaliação da hipótese de proporcionalidade dos riscos.

Para averiguar se o pressuposto de riscos proporcionais é cumprido, respetivamente, foi feita a análise gráfica dos resíduos de Schoenfeld (Figura 2), assim como foi adaptada uma reta de regressão linear aos mesmos e estimado o respetivo coeficiente de correlação, por forma a averiguar quantitativamente se a tendência é horizontal (implicando correlação nula e riscos proporcionais).

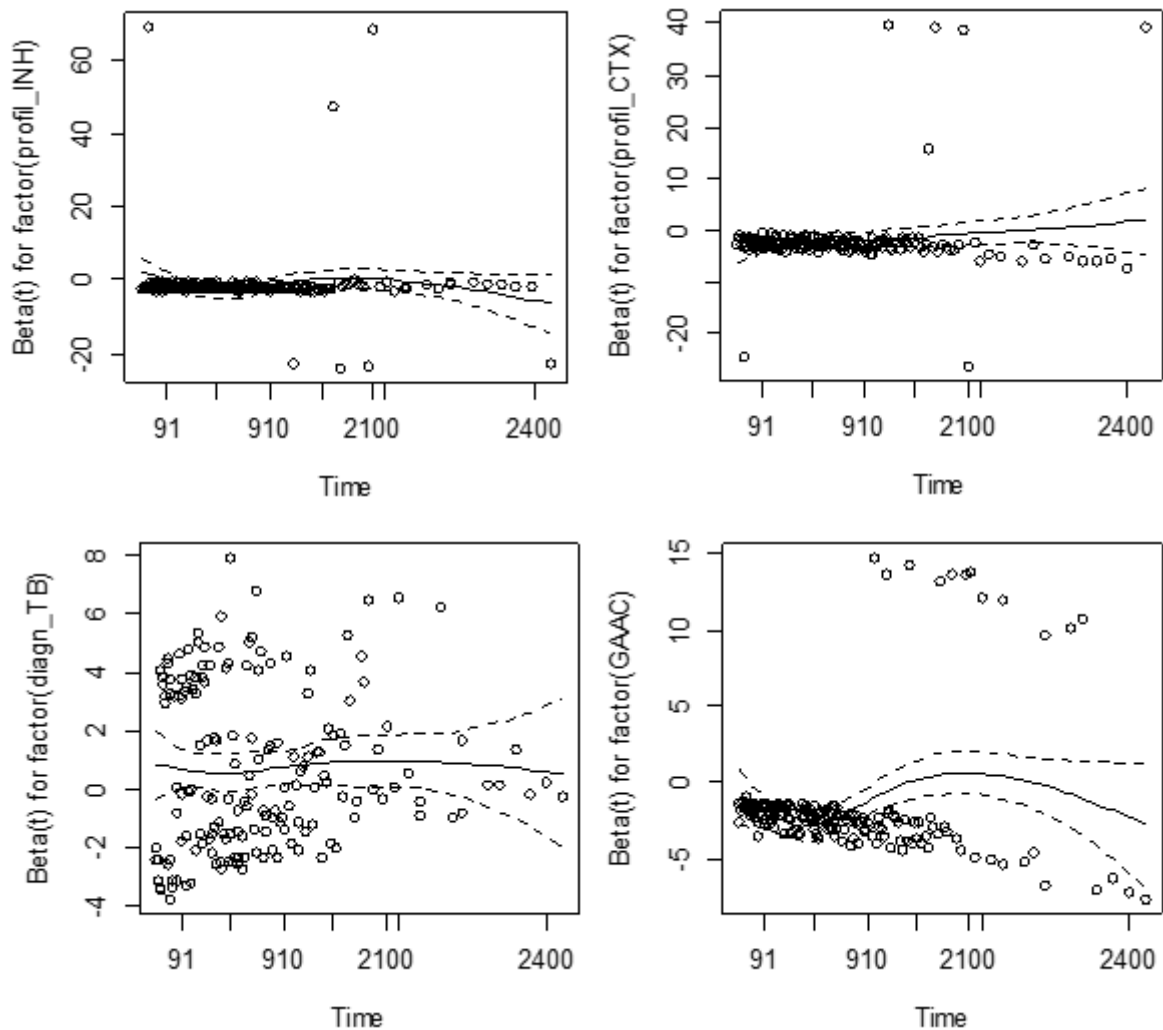
Tabela 14: Teste de correlação dos resíduos com o tempo

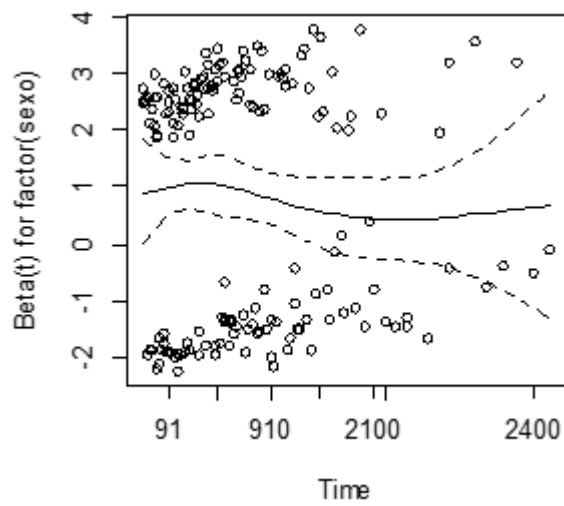
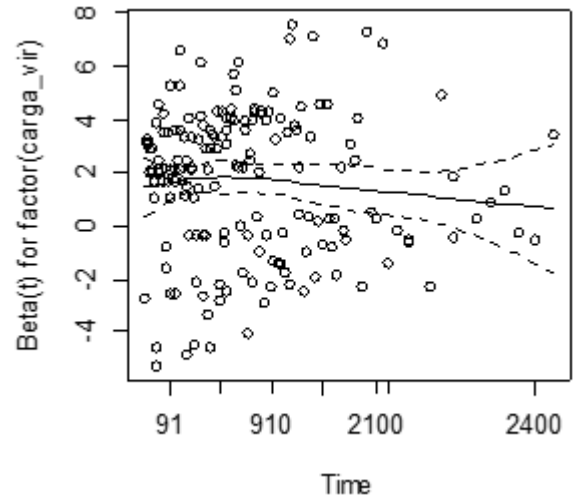
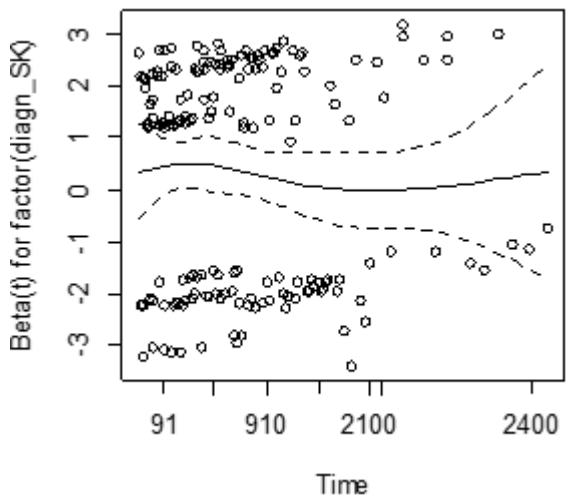
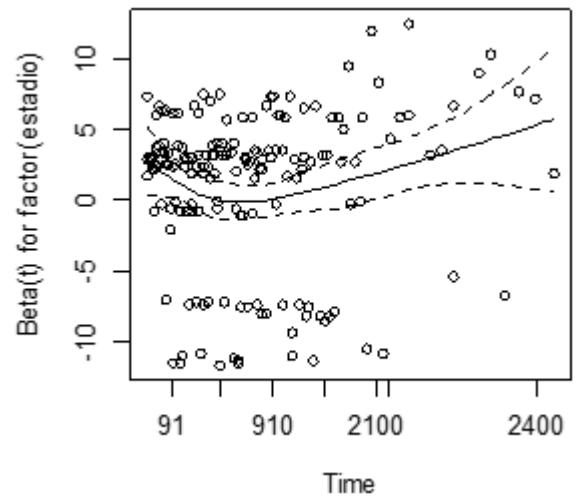
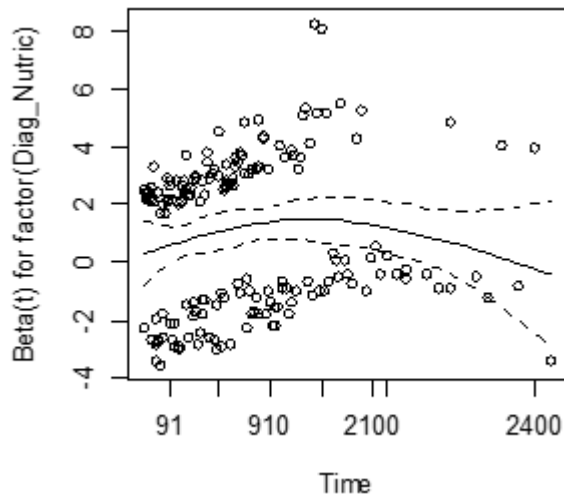
<b>Covariável</b>	$\chi^2$	<b>Graus de liberdade</b>	<b>de p valor</b>
Profil_INH	0.0528	1	0.818
Profil_CTX	2.8219	1	0.093
Diagn_TB	0.0664	1	0.797
GAAC	2.5718	1	0.109
Diagn_Nutric	4.6617	3	0.198
estadio	1.8129	3	0.612
Diagn_SK	0.5114	1	0.475
Carga_viral	1.3112	1	0.252
sexo	2.2434	1	0.134
<b>Global</b>	<b>17.4766</b>	<b>12</b>	<b>0.178</b>

Pela Tabela 14 pode notar-se que o teste de correlações não é significativo, portanto, através deste teste não existe nenhuma tendência entre os resíduos padronizados e

o tempo de sobrevivência transformado pois em todos os casos das covariáveis no modelo tem-se o valor p acima de significância de 0.05 utilizada no teste. Deste modo, a hipótese de proporcionalidade das funções de risco não é violada. O mesmo é visto pela análise gráfica dos resíduos de Schoenfeld (Figura 2), em todas as covariáveis os resíduos de Schoenfeld distribuem-se em torno de uma linha horizontal mostrando inexistência de qualquer tendência.

Figura 2: Resíduos padronizados de Schoenfeld





### 4.3.1. Construção do modelo de Cox

Uma vez comprovada a hipótese de proporcionalidade foi determinado o risco associado a cada covariável bem como os intervalos de confiança para os mesmos.

Tabela 15: Estimativas dos coeficientes do modelo de Cox e os intervalos de confiança para o Risco Relativo

Variável		$\hat{\beta}_i$	<i>p-value</i>	Risco Relativo	IC (95%)	
(profil_INH)1	$x_1$	-1.1823	0.06486	0.3066	0.08738	1.0756
(profil_CTX)1	$x_2$	-1.9593	0.0001	0.141	0.05252	0.3783
(diagn_TB)1	$x_3$	0.7126	0.00241	2.0392	1.28695	3.2313
(GAAC)1	$x_4$	-1.5021	0.00000233	0.2227	0.11938	0.4153
(Diag_Nutric)2	$x_5$	1.1945	0.0995	3.3018	0.79663	13.6851
(Diag_Nutric)3	$x_6$	0.8297	0.00000446	2.2927	1.60852	3.2678
(Diag_Nutric)4	$x_7$	0.7933	0.00121	2.2106	1.36733	3.5741
(Estadio)2	$x_8$	0.4815	0.05166	1.6185	0.99655	2.6287
(Estadio)3	$x_9$	0.5292	0.02342	1.6976	1.07421	2.6828
(Estadio)4	$x_{10}$	0.6452	0.06273	1.9064	0.96630	3.7611
(diagn_SK)1	$x_{11}$	0.3231	0.05470	1.3814	0.99351	1.9209
(carga_vir)4	$x_{12}$	1.6113	<2e-16	5.0094	3.41007	7.3587
(sexo)M	$x_{13}$	0.8603	7.36e-08	2.3638	1.72803	3.2334

A partir da Tabela 15 obtém-se a função de risco dada por

$$\lambda(t|X) = \lambda_0(t)\exp(-1.1823x_1 - 1.9593x_2 + 0.7126x_3 - 1.5021x_4 + 1.1945x_5 + 0.8297x_6 + 0.7933x_7 + 0.4815x_8 + 0.5292x_9 + 0.6452x_{10} + 0.3231x_{11} + 1.6113x_{12} + 0.8603x_{13})$$

O teste de Wald obtido para as covariáveis do modelo apresenta para cada covariável o valor p altamente significativo (Tabela 15), o que indica que o modelo proposto é significativo. Este teste avalia a hipótese nula de que todos os coeficientes  $\beta_i$  são nulos. Assim, no caso do presente estudo esta hipótese é rejeitada visto que o teste apresenta valores significativos (com significância 5%), ou seja, as covariáveis têm efeito significativo sobre a função de risco.

Com base nas estimativas dos coeficientes  $\beta_i$  o Risco Relativo (RR), ou seja, a razão entre a probabilidade da falha ocorrer num grupo exposto e a probabilidade da falha ocorrer num grupo não exposto, dado por  $e^{\beta x}$ , apresenta relações diferentes para diferentes covariáveis. Coeficientes negativos, nas covariáveis *profil\_INH*, *profil\_CTX* e *GAAC* associam valores de RR inferiores a 1 enquanto os coeficientes positivos das restantes covariáveis associam o RR superior à unidade. Assim, covariáveis associadas aos coeficientes  $\beta_i$  negativos têm o efeito retardante para ocorrência de falha enquanto os positivos a aceleram. Especificamente, mantidas constantes as restantes covariáveis, um sujeito que fez tratamento profilático com Isoniazida ( $x_1$ ) reduziria o risco de morte por um fator multiplicativo de 0.3066, ou seja, uma redução da taxa de risco em 69.34% comparativamente ao sujeito que não fez, ou seja, aumentaria suas chances de sobrevivência do que um sujeito que não fez este tratamento, porém o intervalo de 95% de confiança para o risco relativo,  $RR = 0.3066$ , contém 1 pelo que, não há evidências para assumir a diferença dos riscos entre sujeitos tratados com Isoniazida e os não tratados.

Analogamente, ter feito o tratamento profilático com cotrimoxazole ( $x_2$ ), cujo coeficiente no modelo é  $\beta_2 = -1.9593$  está associado à redução de mortalidade por um fator de 0.141, o equivalente a uma redução em cerca de 86% em comparação com os sujeitos que não fizeram o tratamento. Para esta covariável, o intervalo de confiança não inclui 1, o que nos permite assumirmos a existência de diferenças entre o grupo de sujeitos que tiveram o tratamento profilático com CTX e o grupo dos que não tiveram. Para a covariável *GAAC* ( $x_4$ ), que se refere a Grupos de Apoio à Adesão

Comunitária, com  $\beta_4 = -1.5021$  e  $RR = 0.2227$  tem-se que pertencer a este grupo reduz o risco de morte em 77.73% em comparação com sujeitos que não pertencem ao grupo.

A análise do risco relativo foi feita também nas covariáveis cujos coeficientes  $\beta_i$  são positivos e obteve-se o seguinte:

- Ser diagnosticado com tuberculose está associado ao aumento do risco de morte dos sujeitos com HIV em 2.0392 vezes que um sujeito sem este diagnóstico. Assim, visto que o intervalo de confiança de 95% não inclui 1, temos evidência estatística para que a diferença entre o grupo de sujeitos com TB e os sem TB seja assumida.
- O risco de morte, quando o sujeito tem desnutrição ligeira é 3.3018 vezes maior do que sujeitos HIV positivos sem nenhuma desnutrição. O intervalo de 95% de confiança para  $RR = 3.3018$  inclui 1 pelo que não é de assumir a diferença significativa entre sujeitos com desnutrição ligeira e os sem desnutrição.
- Um sujeito com desnutrição moderada incorre um risco de morte de 2.2927 vezes que um sujeito sem desnutrição. O intervalo de confiança de 95% para este risco nesta covariável não inclui 1, pelo que a diferença entre indivíduos desta categoria e os indivíduos normais é assumida.
- Um sujeito com o diagnóstico de desnutrição grave tem o risco de morte 2.2106 vezes maior do que sujeitos normais, sem desnutrição. Esta diferença é significativa visto que com grau de confiança de 95% o intervalo de confiança para  $RR = 2.2106$  não inclui 1.
- Pacientes dos estadios de ordem inferior incorrem menor risco de ocorrência de falha do que os dos estadios de ordem superior. Um paciente que se encontra no estadio 2 teria um risco de óbito de 1.6185 vezes superior que o paciente do estadio 1 porém, o intervalo de confiança para o RR contém 1 pelo

que não há evidência para se concluir que o risco de um paciente se encontrar no estadio 2 difira do risco dele estar no estadio 1. Um paciente do estadio 3 tem um risco de falha de 1.6976 vezes que do estadio 1 e dado que o intervalo de confiança para este risco de falha não inclui 1 pode-se assumir a diferença entre grupo de pacientes do estadio 3 com os do primeiro estadio. Os pacientes do estadio 4 incorreriam o risco de morte de 1.9064 vezes que os do estadio 1 mas tem-se que o intervalo de confiança para o RR inclui 1 pelo que não existe evidência estatística de diferença entre os estadios 1 e 4.

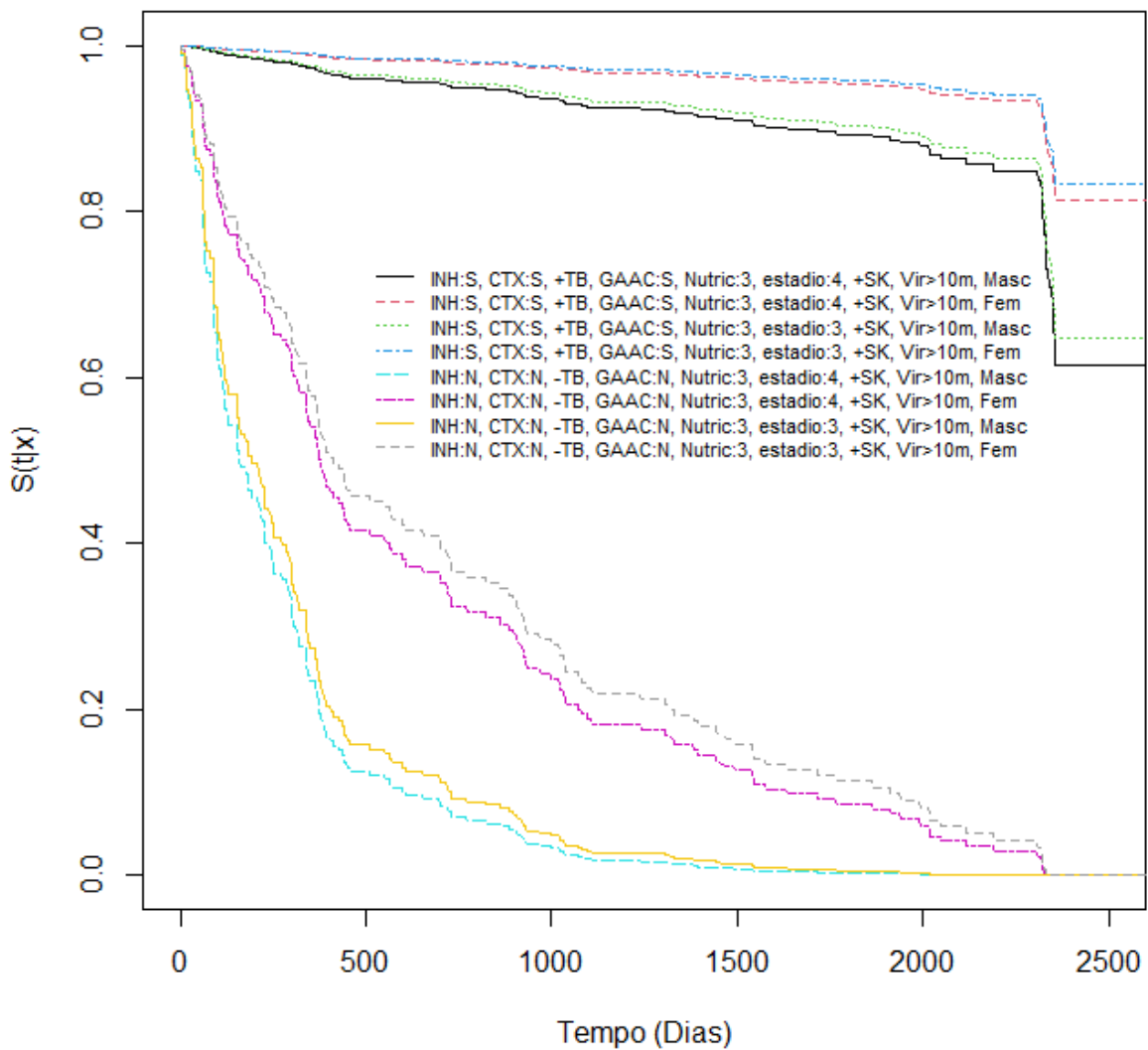
- Um sujeito com HIV diagnosticado com Sarcoma de Kaposi tem risco de evoluir para óbito 1.3814 vezes que o sujeito sem este diagnóstico porém, dado que o intervalo de 95% de confiança para o RR nesta covariável inclui 1 não é de aceitar a diferença entre grupos de sujeitos com diagnóstico positivo de SK e os com diagnóstico negativo.
- Sujeitos com carga viral acima de 10 mil cópias tem 5.0094 vezes mais risco de alcançarem a falha que os que tiverem a carga viral abaixo de 10 mil cópias. De notar que para este caso, visto que o intervalo de confiança de 95% não inclui 1, assume-se a diferença entre sujeitos com carga viral de 10 mil cópias e os sujeitos com carga viral abaixo de 10 mil cópias.
- Sujeitos de sexo masculino têm risco de morte 2.3638 vezes superior que sujeitos de sexo feminino. A diferença entre o grupo de sujeitos de sexo masculino e o grupo de sujeitos de sexo feminino é evidente já que o  $RR = 2.3638$  se encontra num intervalo de confiança de 95% que não inclui 1.

A função de sobrevivência é muito influenciada pelas covariáveis descritas acima. O sexo é um fator que deve ser levado em consideração na comparação de sobrevivência de sujeitos pois através da Figura 3 pode notar-se as discrepâncias nas funções de sobrevivência entre sujeitos de sexo masculino e feminino. Pela mesma figura pode ser visto que sujeitos nos estadios 3 e 4, mantidas constantes as restantes

covariáveis, não apresentam grandes diferenças quanto à sua sobrevivência. Contudo, uma grande diferença está entre sujeitos que passaram pelas profilaxias INH e CTX, com diagnóstico positivo para TB, pertencentes a GAAC em relação aos sujeitos que não tiveram nenhuma profilaxia, sendo seu diagnóstico de TB negativo e que não aderiram a GAAC.

Figura 3: Funções de sobrevivência para diferentes valores das covariáveis pelo modelo de Cox

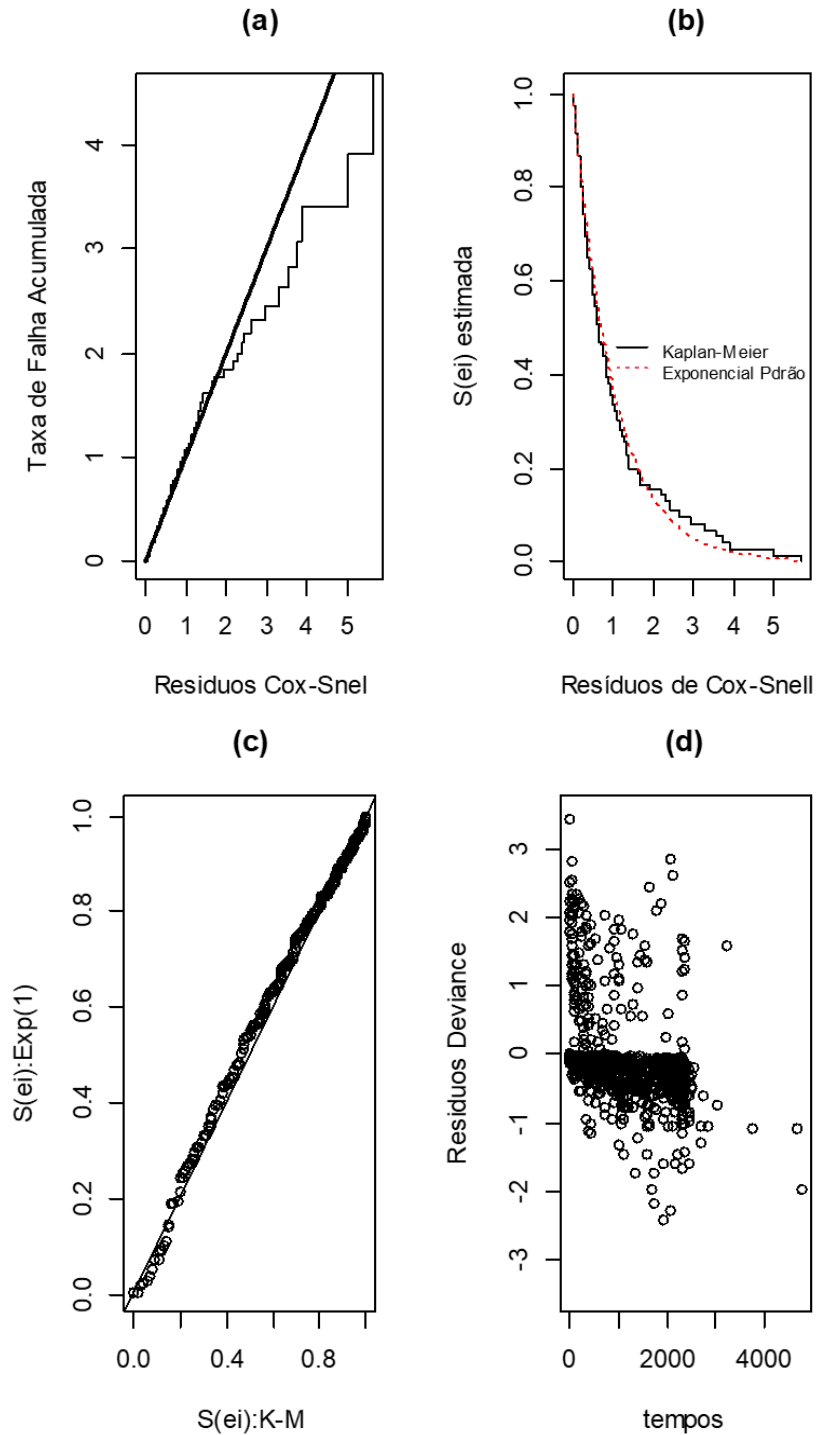
### Comparação das funções de sobrevivência para o modelo de Cox



#### 4.3.2. Análise da Adequação do Ajuste do Modelo

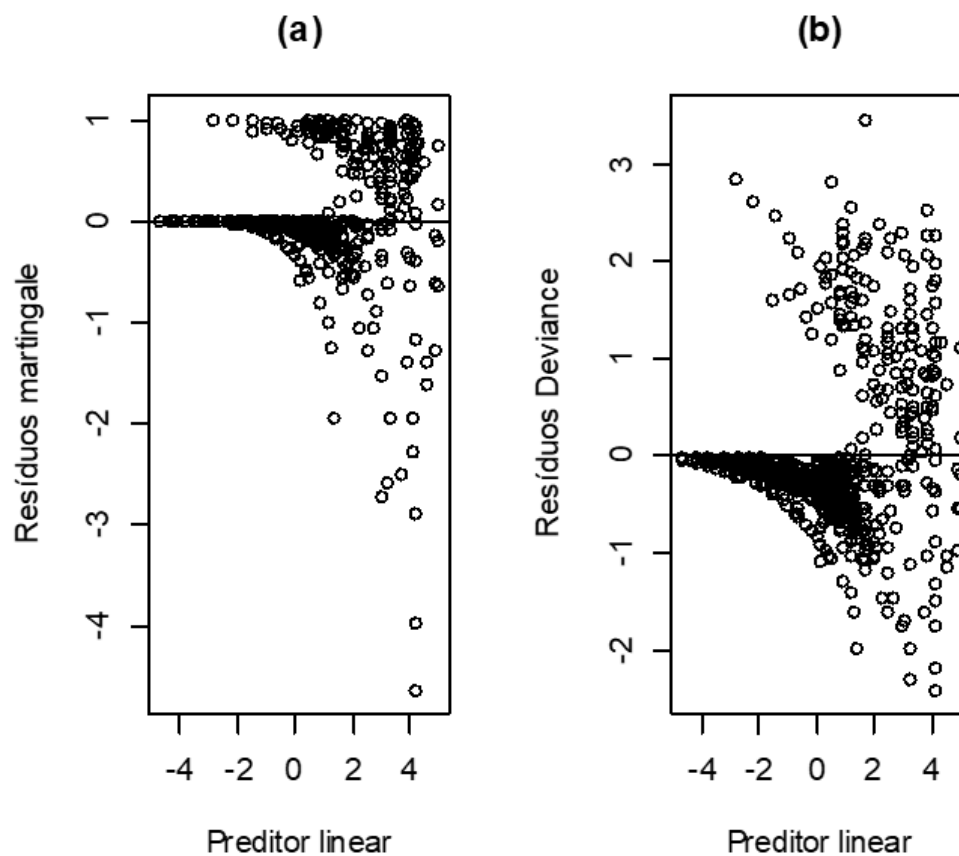
A qualidade do ajuste do modelo foi verificada mediante a análise gráfica dos resíduos de Cox-Snell, resíduos *deviance* e resíduos Martingala. Por análise gráfica dos resíduos de Cox-Snell notou-se que é razoável admitir que o modelo obtido tem um bom ajuste visto que os resíduos de Cox-Snell têm uma distribuição aproximada à distribuição exponencial padrão em quase todos os seus pontos, como mostra a Figura 4(b). Os resíduos de Cox-Snell avaliam a qualidade do ajuste global da adequação do modelo aos dados dos sujeitos com HIV.

Figura 4: (a) Resíduos de Cox-Snell ( $\hat{e}_i$ ) versus Taxa de Falha Acumulada  $\hat{\Lambda}(\hat{e}_i)$ , (b) estimativas de Sobrevivência por Kaplan-Meier  $\hat{S}_{KM}(\hat{e}_i)$ , e Exponencial Padrão  $\hat{S}_{Exp}(\hat{e}_i)$  versus Resíduos de Cox-Snell ( $\hat{e}_i$ ), (c) Pares ( $\hat{S}_{KM}(\hat{e}_i)$ ,  $\hat{S}_{Exp}(\hat{e}_i)$ ) e (d) Resíduos Deviance versus Tempo.



Por análise dos resíduos *deviance*, Figura 4(d) e Figura 5(b), pode notar-se que não há evidência de observações atípicas dado que estes resíduos se encontram distribuídos aleatoriamente em torno do valor zero e dentro do intervalo  $(-3, 3)$ , com exceção de apenas uma observação que se encontra fora deste intervalo.

Figura 5: (a) Resíduos Martingala versus Preditor Linear e (b) Resíduos *Deviance* versus Preditor Linear



## 5. Conclusão

Neste estudo analisaram-se dados provenientes do Centro de Saúde de Changanine, referentes aos pacientes infetados com o HIV, seguidos no programa de TARV durante o período de 01 de Janeiro de 2013 a 31 de Dezembro de 2019 que, após a observância dos critérios de inclusão, foi possível obter uma base de dados com 1122 sujeitos. O estudo visou analisar a sobrevivência dos pacientes diagnosticados de infeção por HIV, o que implicou estudar o tempo desde a sua admissão no programa de TARV até à ocorrência de óbito por causas relacionadas com a doença.

Com vista a aplicar técnicas de análise de sobrevivência para o estudo do tempo até óbito e a influência dos valores de covariáveis recolhidas de sujeitos infetados com HIV foi necessário utilizar o método não paramétrico de KM através do qual se determinou, em cada covariável categórica, os grupos com maior ou menor sobrevivência por avaliação do tempo de vida mediano obtido. Assim, tem-se que:

- O tempo de vida mediano dos sujeitos com desnutrição ligeira bem como o dos sem desnutrição é superior a 2500 dias (cerca de 7 anos) enquanto sujeitos com desnutrição moderada tiveram tempo de vida mediano de 1543 dias e os sujeitos com desnutrição grave tiveram 1994 dias do tempo de vida mediano.
- Nos dois primeiros estádios o tempo de vida mediano foi superior a 2500 dias, sendo de 2326 dias no estadio III e 563 no estadio IV.
- Indivíduos do sexo feminino tiveram tempo de vida mediano superior a 2500 dias enquanto os de sexo masculino tiveram 2016 dias.
- Pacientes que aderiram aos GAAC tiveram o tempo de vida mediano superior a 2500 dias enquanto os que não fizeram parte destes grupos seu tempo de vida mediano caiu para 2353 dias.
- Sujeitos com diagnóstico positivo de Sarcoma de Kaposi tiveram uma sobrevivência menor em relação aos de diagnóstico negativo visto que o seu tempo de vida mediano foi de 2350 e 2500 dias, respetivamente.

- Sujeitos com TB tiveram tempo de vida mediano de 300 dias enquanto os que não contraíram TB tiveram tempo mediano de vida superior a 2500 dias.
- Sujeitos que realizaram tratamento profilático de CTX tiveram tempo de sobrevivência mediano superior a 2500 dias contra 2339 dias dos que não tiveram o tratamento.
- Sujeitos que tiveram tratamento profilático de INH tiveram tempo de vida mediano superior a 2500 dias enquanto o grupo de sujeitos que não teve este tratamento seu tempo de vida mediano foi de 2356.
- Sujeitos com carga viral abaixo de 5000 cópias tiveram tempo de vida mediano superior a 2500 dias enquanto o tempo de vida mediano de sujeitos com carga viral situada entre 5000 e 10000 foi de 2353 dias e de 386 dias para sujeitos com mais de 10000 cópias.

Ainda sobre a estimativa de KM foi efetuado o teste logrank para a comparação de grupos de sujeitos em cada uma das variáveis categóricas, tendo-se concluído que cada uma das variáveis Diag\_Nutric, GAAC, diagn\_SK, diagn\_TB, profil\_CTX, profil\_INH, estadio e sexo, diferem significativamente.

A regressão de Cox foi usada para medir e quantificar o efeito preditivo das covariáveis, isto é, se aumentam ou diminuem o risco de falha e por que fator. Deste modo as covariáveis significativas no modelo de Cox, foram profilaxia com INH, profilaxia com CTX, diagnóstico de TB, GAAC, diagnóstico de nutrição, estadio, diagnóstico de SK, carga viral maior que 10000 copias e sexo. Não foram significativas, o que não era de esperar, as covariáveis, (carga\_viral) 2, (carga\_viral) 3, TPT e IMC visto que o aumento no número de carga viral tenderia a agravar o estado clínico dos sujeitos, enquanto o tratamento preventivo da tuberculose (TPT) aumentaria as chances de sobreviver mais tempo uma vez que reduziria o risco de contrair a TB nos estádios 3 e 4. Outras covariáveis não significativas que não parecem terem grande impacto na sobrevivência dos sujeitos com HIV são: grav\_lact que refere simplesmente se a mulher é gestante ou lactante, trat que refere se o sujeito está na primeira linha ou mudou de linha de tratamento e a idade que não define a faixa etária de maior

exposição embora se espere que sujeitos de idade superior a 60 anos fossem de menor exposição que o outro grupo. Para todas as covariáveis significativas era esperado que o fossem com exceção da covariável sexo cuja diferença quanto aos indivíduos do sexo masculino e os femininos não parece óbvia. Talvez se explique pelas relações extra conjugais que podem ser mais praticadas pelos sujeitos de sexo masculino do que os do sexo feminino.

A partir da análise descritiva dos dados foi possível verificar que a sobrevivência dos homens é baixa em comparação com a das mulheres visto que apenas 9.3% (correspondente a 77 mulheres) das 826 mulheres envolvidas na pesquisa, evoluíram para óbito contra 31.76% dos óbitos de indivíduos de sexo masculino. A covariável sexo foi altamente significativa no modelo de Cox com valor p inferior a 0.001 pelo que esta tem grande influência no tempo de vida dos sujeitos com HIV do Centro de Saúde de Changanine e, através deste modelo, ser paciente do sexo masculino aumentaria o risco de morte 2.32 vezes do que ser paciente do sexo feminino. Por um lado, este resultado contrasta com os resultados obtidos por [19] para os quais sexo não foi significativo, mas por outro mesmo é semelhante aos obtidos por [20] e [21] quanto à covariável sexo e com discordâncias quanto à covariável idade.

A idade média dos pacientes envolvidos na nossa pesquisa é de 40.18 anos, sendo 1025 pacientes com idade inferior a 60 anos e 97 pacientes com idades a partir dos 60 anos, o que nos sugere que maior parte dos pacientes é jovem. Contudo, o nosso estudo não encontrou relação entre o fator idade dos pacientes e sua sobrevivência uma vez que esta variável não foi significativa no modelo de Cox.

Tendo em vista o objetivo do presente estudo pode se dizer que foi alcançado na íntegra dado que a aplicação de técnicas de análise de sobrevivência para o estudo do tempo de vida dos sujeitos infetados com HIV foi feita mediante a aplicação do método de Kaplan-Meier, tendo-se determinado o tempo de vida mediano em cada covariável categórica. Foi aplicado o modelo de Cox para determinar a influência dos valores das covariáveis tendo sido concluído que a covariáveis que contribuem para o

aumento do tempo de sobrevivência são fazer profilaxias CTX e INH e aderir aos grupos de apoio à adesão comunitária enquanto as covariáveis que constituem fator de risco e que conseqüentemente têm efeito de reduzir o tempo de vida dos sujeitos são: ter diagnóstico da tuberculose, subir de nível na classificação do estado de desnutrição e de estadios clínicos da OMS, ter valor de carga viral acima de 10 mil cópias, ter diagnóstico de sarcoma de Kaposi e ser indivíduo de sexo masculino.

### 5.1. Perspetivas Futuras

Constituíram limitações deste estudo a falta de informação em algumas covariáveis como por exemplo: crenças religiosas, estado civil, alcoolismo, uso de drogas, uso de preservativo, nível de escolaridade, estado funcional e a profissão. Desta forma, nas abordagens futuras serão incluídas estas covariáveis, para além das consideradas neste estudo.

Os modelos paramétricos pressupõem que seja conhecida a priori, a distribuição teórica de probabilidade para o tempo. Este fato torna estes modelos menos flexíveis caso os dados não sejam coerentes com a distribuição considerada e não devem ter resultados tão adequados quanto o modelo de regressão de Cox, pelo que o trabalho futuro não incluirá estes modelos, no entanto, uma hipótese de trabalho futuro poderá ser estudar a adequação destes modelos aos dados, o que não foi possível neste trabalho por limitação de tempo. Poderá ser objeto de estudo futuro o modelo de Cox com variáveis dependentes de tempo uma vez que as variáveis como carga viral, IMC e a nutrição podem variar ao longo do tempo.

## 6. Referências Bibliográficas

- [1] U. Nations e T. Families, «AIDS and HIV Infection Information for United Nations Employees and Their Families», *UNAIDS*, 2000, [Em linha]. Disponível em: [https://data.unaids.org/publications/irc-pub01/jc306-un-staff-rev1\\_en.pdf](https://data.unaids.org/publications/irc-pub01/jc306-un-staff-rev1_en.pdf).
- [2] M. Bellini e P. C. Frasson, «Ciências e seu ensino: o que dizem os cientistas e os livros didáticos sobre o HIV/AIDS?», *Ciência Educ.*, vol. 12, n. 3, pp. 261–274, 2006, doi: 10.1590/s1516-73132006000300002.
- [3] F. Zanetti e M. K. Pettenon, «SÍNDROME DA IMUNODEFICIÊNCIA ADQUIRIDA ASSOCIADA A PNEUMONIA ADQUIRIDA NA COMUNIDADE : APRIMORANDO», pp. 1–5, 2020.
- [4] M. Vitoria *et al.*, «The global fight against HIV/AIDS, tuberculosis, and malaria: Current status and future perspectives», *Am. J. Clin. Pathol.*, vol. 131, n. 6, pp. 844–848, 2009, doi: 10.1309/AJCP5XHDB1PNAEYT.
- [5] M. S. Cohen, N. Hellmann, J. A. Levy, K. DeCock, e J. Lange, «The spread, treatment, and prevention of HIV-1: evolution of a global pandemic», *J. Clin. Invest.*, vol. 118, pp. 1244–1254, 2008.
- [6] M. . J. Van Rensburg, «Measuring the quality of life of residents in SADC communities affected by HIV», *Taylor Fr.*, vol. 21, n. 9, pp. 1132–1140, 2009, [Em linha]. Disponível em: <https://doi.org/10.1080/09540120802705875>.
- [7] P. N. Gona *et al.*, «Burden and changes in HIV/AIDS morbidity and mortality in Southern Africa Development Community Countries, 1990-2017», *BMC Public Health*, vol. 20, n. 1, pp. 1–14, 2020, doi: 10.1186/s12889-020-08988-9.
- [8] C. M. Audet, J. Burlison, T. D. Moon, M. Sidat, A. E. Vergara, e S. H. Vermund, «Sociocultural and epidemiological aspects of HIV/AIDS in Mozambique», *BMC Int. Health Hum. Rights*, vol. 10, n. 1, 2010, doi: 10.1186/1472-698X-10-15.
- [9] Conselho Nacional de Combate ao HIV/SIDA, «Plano Estratégico Nacional De

Combate Ao Hiv/Sida», 2004.

- [10] P. J. Feldblum *et al.*, «HIV prevalence and incidence in a cohort of women at higher risk for HIV acquisition in Chókwè, Southern Mozambique», *PLoS One*, vol. 9, n. 5, pp. 1–6, 2014, doi: 10.1371/journal.pone.0097547.
- [11] A. Hardon *et al.*, «From Access to Adherence: The Challenges of Antiretroviral Treatment - Studies from Botswana, Tanzania and Uganda», *Who*, pp. 1–320, 2006, [Em linha]. Disponível em:  
<http://apps.who.int/medicinedocs/pdf/s13400e/s13400e.pdf>.
- [12] M. I. Pardini *et al.*, «Diagnóstico da Infecção pelo HIV em Adultos e Crianças», *Ministério da Saúde, Secr. Vigilância em Saúde, Dep. Vigilância, Prevenção e Control. das Infecções Sex. Transm. do HIV/Aids e das Hepatites Virais*, 2018, [Em linha]. Disponível em:  
[http://www.aids.gov.br/system/tdf/pub/2016/57787/manual\\_tecnico\\_hiv\\_27\\_11\\_2018](http://www.aids.gov.br/system/tdf/pub/2016/57787/manual_tecnico_hiv_27_11_2018).
- [13] Wers, E. S. S. On, L. R. Mota, M. L. Bazzo, M. Franchini, e O. da C. F. Junior, «MANUAL TÉCNICO PARA O DIAGNÓSTICO DA INFECÇÃO PELO HIV», 2013, [Em linha]. Disponível em:  
[https://bvsmis.saude.gov.br/bvs/publicacoes/manual\\_tecnico\\_diagnostico\\_infeccao\\_hiv.pdf](https://bvsmis.saude.gov.br/bvs/publicacoes/manual_tecnico_diagnostico_infeccao_hiv.pdf).
- [14] K. Danforth, R. Granich, e D. Wiedeman, «Global Mortality and Morbidity of HIV/AIDS», 2017, [Em linha]. Disponível em:  
<https://www.ncbi.nlm.nih.gov/books/NBK525184/> doi: 10.1596/978-1-4648-0524-0\_ch2.
- [15] B. Achappa, D. Madi, U. Bhaskaran, J. T. Ramapuram, S. Rao, e S. Mahalingam, «Adherence to antiretroviral therapy among people living with HIV», *N. Am. J. Med. Sci.*, vol. 5, n. 3, pp. 220–223, 2013, doi: 10.4103/1947-2714.109196.

- [16] MINISTERIO DE SAUDE, *Manual de adesão ao tratamento para pessoas vivendo com HIV e AIDS*. Brasília, 2008.
- [17] WHO, «HIV / AIDS Programme WHO CASE DEFINITIONS OF HIV FOR SURVEILLANCE AND REVISED CLINICAL STAGING AND IMMUNOLOGICAL CLASSIFICATION OF HIV-RELATED DISEASE IN ADULTS AND CHILDREN», 2007, [Em linha]. Disponível em: <https://www.who.int/hiv/pub/guidelines/HIVstaging150307.pdf>.
- [18] A. C. Medici e W. Bank, «Globalização , Doenças Transmissíveis e Desigualdade Mundial», n. January 2003, 2015.
- [19] P. Amanzi, C. Michelo, C. Simoonga, R. Dambe, e G. Chongwe, «Survival of people on antiretroviral treatment in Zambia: A retrospective cohort analysis of HIV clients on ART», *Pan Afr. Med. J.*, vol. 24, pp. 1–11, 2016, doi: 10.11604/pamj.2016.24.144.6004.
- [20] L. Santos Wolmer de Melo, H. R. Lacerda, E. Campelo, E. Moraes, e R. A. de Alencar Ximenes, «Survival of AIDS patients and characteristics of those who died over eight years of highly active antiretroviral therapy, at a referral center in Northeast Brazil», *Brazilian J. Infect. Dis.*, vol. 12, n. 4, pp. 269–277, 2008, doi: 10.1590/s1413-86702008000400003.
- [21] M. B. Jackson, «SURVIVAL ANALYSIS OF ADULT HIV/AIDS PATIENTS AND STOCHASTIC MODELLING OF AIDS DISEASE PROGRESSION: A CASE STUDY OF JIMMA UNIVERSITY SPECIALIZED HOSPITAL, ETHIOPIA M.», *J. Control. Release*, vol. 156, pp. 315–322, 2011.
- [22] A. W. Alemu e M. S. Sebastián, «Determinants of survival in adult HIV patients on antiretroviral therapy in Oromiyaa, Ethiopia», *Glob. Health Action*, vol. 3, n. 1, p. 5398, 2010, doi: 10.3402/gha.v3i0.5398.
- [23] K. Kebebew e E. Wencheko, «Survival analysis of HIV-infected patients under antiretroviral treatment at the Armed Forces General Teaching Hospital, Addis

- Ababa, Ethiopia», *Ethiop. J. Heal. Dev.*, vol. 26, n. 3, pp. 186–192, 2012, doi: 10.4314/ejhd.v26i3.
- [24] T. Assefa e E. Wencheke, «Survival analysis of patients under chronic HIV-care and antiretroviral treatment at Tikur Anbessa Specialized Hospital, Addis Ababa, Ethiopia», n. 3.
- [25] M. C. de Melo, M. R. Donalisio, e R. C. Cordeiro, «Sobrevida de pacientes com AIDS e coinfeção pelo bacilo da tuberculose nas regiões Sul e Sudeste do Brasil», *Cienc. e Saude Coletiva*, vol. 22, n. 11, pp. 3781–3792, 2017, doi: 10.1590/1413-812320172211.26352015.
- [26] E. T. de S. Helena, M. L. Mafra, e M. Simes, «Fatores associados à sobrevida de pessoas vivendo com aids no Município de Blumenau, Estado de Santa Catarina, Brasil, 1997-2004», *Epidemiol. e Serviços Saúde*, vol. 18, n. 1, pp. 45–53, 2009, doi: 10.5123/s1679-49742009000100005.
- [27] E. A. Colosimo e S. R. Giolo, *Análise de Sobrevivência Aplicada*. 2006.
- [28] P. Serranho, «Notas de Análise de Sobrevivência», 2015.
- [29] D. G. Kleinbaum e M. Klein, *Survival analysis, A Self-Learning Text*, vol. 33, n. 4. 2012.
- [30] D. Koletsi e N. Pandis, «Survival analysis, part 2: Kaplan-Meier method and the log-rank test», *Am. J. Orthod. Dentofac. Orthop.*, vol. 152, n. 4, pp. 569–571, 2017, doi: 10.1016/j.ajodo.2017.07.008.
- [31] E. Christensen, «Multivariate Survival Analysis Using Cox's Regression Model», vol. 7, n. 6, pp. 1346–1358, 1987, [Em linha]. Disponível em: <https://aasldpubs.onlinelibrary.wiley.com/doi/pdf/10.1002/hep.1840070628>.