

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



Identificação de fatores que influenciam a persistência na terapêutica de uma doença crónica

ANA RITA DOS SANTOS GODINHO

Mestrado em Bioestatística

Dissertação orientada por:
Professora Doutora Cristina Maria Tristão Simões Rocha
Zilda Maria Severino Mendes

2016

Resumo

Com o crescente envelhecimento da população tem-se verificado um aumento da incidência de doenças crónicas, geralmente associadas a condições debilitantes ou fisicamente dolorosas, levando por isso a uma redução da qualidade de vida dos doentes. Vários estudos mostram que a adesão e a persistência à terapêutica de doenças crónicas são fundamentais na melhoria dos resultados em saúde, sendo essencial compreender os fatores que as influenciam.

No presente estudo pretende-se identificar os fatores que influenciam o tempo até à não persistência à terapêutica de uma determinada doença crónica. Diferentes famílias de modelos de sobrevivência são também exploradas, por forma a averiguar qual a que melhor descreve o tempo até ao acontecimento em estudo.

Foram analisados dados de um estudo observacional prospetivo, no qual uma coorte de 360 indivíduos com uma dada doença crónica foi seguida por um período de 18 meses. Foram considerados na análise os fatores sociodemográficos e de saúde recolhidos aquando do recrutamento dos indivíduos.

Numa fase preliminar da análise de sobrevivência foram utilizados métodos não paramétricos, através dos quais se verificou que apenas as variáveis Idade, Vive sozinho e Terapêutica se mostraram relevantes. A influência destes fatores no tempo até à não persistência foi estudada através do ajustamento e comparação dos resultados obtidos através do modelo de Cox, dos modelos paramétricos de Weibull, log-normal e log-logístico e dos modelos paramétricos flexíveis propostos por Royston e Parmar. Com a modelação paramétrica espera-se obter uma melhor compreensão do perfil de risco dos doentes ao longo do tempo.

Com o modelo de Cox foram selecionadas apenas as variáveis Idade, Vive sozinho e Terapêutica como as que influenciam de forma significativa o tempo até à não persistência à terapêutica. A comparação dos gráficos das estimativas das funções de sobrevivência, obtidas pelos modelos paramétricos com uma covariável e pelos modelos flexíveis, mostra como a introdução de um *spline* cúbico com m nós internos aumenta bastante a flexibilidade dos modelos paramétricos. O aumento do número de nós internos incluídos no modelo (até um máximo de 3 nós) contribui para a obtenção de estimativas mais precisas, fazendo também diminuir o valor de AIC associado a cada modelo. Segundo este critério e optando pelo modelo mais parcimonioso, é o modelo flexível de riscos proporcionais com 2 nós internos que se revela o mais adequado, entre os modelos ajustados.

Toda a análise estatística foi feita no *software* R versão 3.0.1.

Palavras-chave: Persistência à terapêutica para doença crónica, Análise de Sobrevivência, Modelos Flexíveis de Royston e Parmar, Modelos Paramétricos, Modelo de Cox.

Abstract

Due to population ageing, there has been an increase in chronic diseases usually associated with debilitating or physically painful conditions, leading to a reduction in the patients' quality of life. Several studies show that medication adherence and persistence play a crucial role in improving health results, therefore being essential to understand the factors that influence them.

The present study aims to identify the factors that influence the time to non-persistence in a particular chronic disease treatment. Different families of survival analysis models are also explored in order to determine which one best describes the event in study.

We analyzed data from a prospective observational study, on which a cohort of 360 individuals with a specific chronic disease was followed for 18 months. The social demographic and health factors collected during patient recruitment, were considered in the analysis.

In a preliminary phase of the survival analysis non-parametric methods were used, through which only Age, Living alone and Treatment were identified as relevant variables. The influence of this factors on time to non-persistence was studied through the adjustment and the results comparison of the Cox model, Weibull, log-normal and log-logistic parametric models and the flexible parametric models proposed by Royston and Parmar. With the parametric modulation we expect to get a better understanding of patients' risk profile over time.

With the Cox Model, the only variables significantly associated to time to non-persistence were Age, Living alone and Treatment. The comparison of estimated survival functions' graphs, obtained with the univariate parametric models and with the flexible models, shows that the introduction of a cubic spline with m internal knots greatly increases the parametric models' flexibility. The increase in the number of internal knots included in the model (to a maximum of 3 knots) leads to more precise estimates and decreasing as well the AIC value associated with each model. According to these criteria and opting for the most parsimonious model, the one that seems to be the most appropriate between the adjusted models is the flexible proportional hazards model with 2 internal knots.

All statistical analysis was performed using R Statistical Software v3.0.1.

Keywords: Chronical Disease Treatment Persistence, Survival Analysis, Royston and Parmar Flexible Models, Parametric Models, Cox Model.

“It always seems impossible until it is done.”
Nelson Mandela

Agradecimentos

Escrever uma tese de mestrado não foi tarefa fácil e não o poderia ter feito sem a ajuda e o apoio que várias pessoas me deram no decorrer deste longo percurso.

Gostaria de agradecer à Professora Doutora Cristina Simões Rocha, orientadora desta tese de Mestrado, pelas ideias, pela partilha de conhecimento, pela disponibilidade e paciência durante todo este tempo. Pelo rigor, precisão e atenção investida na leitura desta tese, que me fez aumentar o nível de exigência com o meu próprio trabalho e me permitiu crescer tanto a nível pessoal como técnico. Muito, muito obrigada!

Gostaria de agradecer à Doutora Zilda Mendes, que além de coorientadora desta tese, coordena a equipa de estatística do Centro de Estudos e Avaliação em Saúde (CEFAR), da qual tenho muito gosto em fazer parte. Muito obrigada pela motivação diária, pela confiança no meu trabalho, pela empatia, pela partilha de conhecimento, pela amizade e especialmente pelas oportunidades que me tem dado nos últimos anos, pelas quais tenho a maior gratidão.

Gostaria de agradecer também ao CEFAR, sem o qual este trabalho não seria possível. Aos meus colegas e amigos pelo trabalho fantástico que desenvolvemos juntos todos os dias e pelos momentos de descontração e bom convívio.

Gostaria de agradecer à minha mãe que desde sempre acreditou em mim e sempre me encorajou nesta aventura que é fazer um mestrado e uma tese.

Por último, gostaria de agradecer ao meu namorado Vladyslav pois sem ele esta tese definitivamente não estaria concluída. Pela paciência infinita que tem para mim, pela amizade, companheirismo e por todas as vezes que me motivou, encorajou e pressionou a escrever e concluir esta tese.

Conteúdo

INTRODUÇÃO	1
1.1 ENQUADRAMENTO	1
1.2 DEFINIÇÃO DO PROBLEMA E OBJETIVOS	2
METODOLOGIA	3
2.1 RECOLHA DE DADOS	3
2.2 VARIÁVEIS EM ESTUDO	3
2.2.1 VARIÁVEIS SOCIODEMOGRÁFICAS	3
2.2.2 VARIÁVEIS DE SAÚDE	4
2.2.3 VARIÁVEIS DE TRATAMENTO	4
2.3 CÁLCULO DO TEMPO ATÉ À NÃO PERSISTÊNCIA	5
2.3.1 TERAPÊUTICA MENSAL	5
2.3.2 TERAPÊUTICA SEMANAL	6
2.4 ANÁLISE ESTATÍSTICA	7
2.4.1 FATORES ASSOCIADOS À NÃO PERSISTÊNCIA	7
2.4.2 FATORES ASSOCIADOS AO TEMPO ATÉ À NÃO PERSISTÊNCIA	7
FUNDAMENTOS TEÓRICOS	9
3.1 REGRESSÃO LOGÍSTICA	9
3.1.1 AJUSTAMENTO DO MODELO	9
3.1.2 DIAGNÓSTICO DO MODELO	10
3.2 ANÁLISE DE SOBREVIVÊNCIA	11
3.2.1 ESTIMAÇÃO NÃO PARAMÉTRICA	12
3.3 MODELO DE REGRESSÃO DE COX	14
3.3.1 AJUSTAMENTO DO MODELO	14
3.3.2 DIAGNÓSTICO DO MODELO	15
3.4 MODELOS DE SOBREVIVÊNCIA PARAMÉTRICOS	16
3.4.1 MODELOS DE TEMPO DE VIDA ACELERADO	16
3.4.2 MODELOS DE POSSIBILIDADES PROPORCIONAIS	19
3.5 MODELOS FLEXÍVEIS DE ROYSTON E PARMAR	21
3.5.1 MODELO FLEXÍVEL DE RISCOS PROPORCIONAIS	22
3.5.2 MODELO FLEXÍVEL DE POSSIBILIDADES PROPORCIONAIS	23
3.5.3 AJUSTAMENTO DO MODELO	23
3.5.4 DIAGNÓSTICO DO MODELO	24
RESULTADOS	25
4.1 MODELO DE REGRESSÃO LOGÍSTICA	25
4.1.1 AJUSTAMENTO DO MODELO LOGÍSTICO	27
4.1.2 INTERPRETAÇÃO DOS COEFICIENTES DO MODELO FINAL	29
4.1.3 DIAGNÓSTICO DO MODELO	30
4.2 ANÁLISE DE SOBREVIVÊNCIA	31
4.2.1 ESTIMAÇÃO NÃO PARAMÉTRICA	31
4.3 MODELO DE RISCOS PROPORCIONAIS DE COX	35
4.3.1 AJUSTAMENTO DO MODELO	35
4.3.2 INTERPRETAÇÃO DOS COEFICIENTES DO MODELO FINAL	37
4.3.3 DIAGNÓSTICO DO MODELO	38

4.4 MODELOS PARAMÉTRICOS	40
4.4.1 MODELOS DE TEMPO DE VIDA ACELERADO	40
4.4.2 MODELOS DE POSSIBILIDADES PROPORCIONAIS	46
4.4.3 COMPARAÇÃO DOS MODELOS PARAMÉTRICOS	48
4.5 MODELOS FLEXÍVEIS DE ROYSTON E PARMAR	49
4.5.1 MODELO FLEXÍVEL DE RISCOS PROPORCIONAIS	49
4.5.2 MODELO FLEXÍVEL DE POSSIBILIDADES PROPORCIONAIS	53
<u>DISCUSSÃO E CONCLUSÕES</u>	<u>59</u>
<u>BIBLIOGRAFIA</u>	<u>61</u>
<u>ANEXO A</u>	<u>63</u>
<u>ANEXO B</u>	<u>66</u>

Lista de Figuras

FIGURA 2. 1: EXEMPLO DO CÁLCULO DO TEMPO ATÉ À NÃO PERSISTÊNCIA, DE UM DOENTE COM TERAPÊUTICA MENSAL.....	5
FIGURA 2. 2: EXEMPLO DO CÁLCULO DO TEMPO ATÉ À NÃO PERSISTÊNCIA, DE UM DOENTE COM TERAPÊUTICA SEMANAL.....	6
FIGURA 4. 1: DISTRIBUIÇÃO DOS INDIVÍDUOS, POR ESTADO DE PERSISTÊNCIA NO FINAL DO ESTUDO SEGUNDO A CLASSE DE IDADE (A) E SEGUNDO VIVE OU NÃO SOZINHO (B).....	26
FIGURA 4. 2: CURVA ROC.....	30
FIGURA 4. 3: CURVA DE SOBREVIVÊNCIA (A) E ESTIMATIVA DA FUNÇÃO DE RISCO CUMULATIVA (B).....	31
FIGURA 4. 4: GRÁFICOS DAS ESTIMATIVAS $\hat{S}(t)$ (À ESQUERDA) E $\ln[-\ln \hat{S}(t)]$ (À DIREITA), PARA AS VARIÁVEIS IDADE (A), VIVE SOZINHO (B) E TERAPÊUTICA (C).....	34
FIGURA 4. 5: GRÁFICOS DOS RESÍDUOS DE SCHOENFELD PADRONIZADOS <i>VERSUS</i> O TEMPO (À ESQUERDA) E <i>VERSUS</i> O LOGARITMO DO TEMPO (À DIREITA), PARA AS VARIÁVEIS IDADE – “60 A 70 ANOS” (A), IDADE – “MAIS DE 70 ANOS” (B), VIVE SOZINHO (C) E TERAPÊUTICA (D).....	39
FIGURA 4. 6: CURVAS DE SOBREVIVÊNCIA: COMPARAÇÃO ENTRE AS ESTIMATIVAS DE KAPLAN-MEIER E AS CURVAS ESTIMADAS PELO MODELO DE WEIBULL PARA AS VARIÁVEIS IDADE (A), VIVE SOZINHO (B) E TERAPÊUTICA (C).....	41
FIGURA 4. 7: CURVAS DE SOBREVIVÊNCIA: COMPARAÇÃO ENTRE AS ESTIMATIVAS DE KAPLAN-MEIER E AS CURVAS ESTIMADAS PELO MODELO LOG-NORMAL PARA AS VARIÁVEIS IDADE (A), VIVE SOZINHO (B) E TERAPÊUTICA (C).....	45
FIGURA 4. 8: CURVAS DE SOBREVIVÊNCIA: COMPARAÇÃO ENTRE AS ESTIMATIVAS DE KAPLAN-MEIER E AS CURVAS ESTIMADAS COM O MODELO LOG-LOGÍSTICO PARA AS VARIÁVEIS IDADE (A), VIVE SOZINHO (B) E TERAPÊUTICA (C).....	47
FIGURA 4. 9: CURVAS DE SOBREVIVÊNCIA: COMPARAÇÃO ENTRE AS ESTIMATIVAS DE KAPLAN-MEIER E AS CURVAS ESTIMADAS PELOS MODELOS FLEXÍVEIS DE RISCOS PROPORCIONAIS COM M NÓS PARA AS VARIÁVEIS IDADE (A), VIVE SOZINHO (B) E TERAPÊUTICA (C).....	51
FIGURA 4. 10: CURVAS DE SOBREVIVÊNCIA: COMPARAÇÃO ENTRE AS ESTIMATIVAS DE KAPLAN-MEIER E AS CURVAS ESTIMADAS PELOS MODELOS FLEXÍVEIS DE POSSIBILIDADES PROPORCIONAIS COM M NÓS PARA AS VARIÁVEIS IDADE (A), VIVE SOZINHO (B) E TERAPÊUTICA (C).....	56
FIGURA A. 1: ANÁLISE BIVARIADA: DISTRIBUIÇÃO DOS INDIVÍDUOS, POR ESTADO DE PERSISTÊNCIA NO FINAL DO ESTUDO SEGUNDO A CLASSE DE CADA VARIÁVEL INDEPENDENTE EM ANÁLISE.	64
FIGURA B. 1: GRÁFICOS DAS ESTIMATIVAS S_t PARA CADA VARIÁVEL INDEPENDENTE EM ANÁLISE.	66

Lista de Tabelas

TABELA 3. 1: POSIÇÃO DOS NÓS INTERNOS EM MODELOS <i>SPLINE</i> , PARA DIFERENTES GRAUS DE LIBERDADE.....	22
TABELA 4. 1: TESTES DO QUI-QUADRADO DE INDEPENDÊNCIA	26
TABELA 4. 2: ANÁLISE UNIVARIADA: <i>ODDS RATIO</i> E VALOR P DOS TESTES DE WALD	27
TABELA 4. 3: ANÁLISE UNIVARIADA: TESTES DA RAZÃO DE VEROSIMILHANÇAS.....	28
TABELA 4. 4: VALOR DE AIC E TESTE DA RAZÃO DE VEROSIMILHANÇAS, EM CADA PASSO DA SELEÇÃO DE VARIÁVEIS.....	29
TABELA 4. 5: MODELO MÚLTIPLO FINAL: ESTIMATIVA DE CADA β_j , RESPECTIVO TESTE DE WALD E <i>ODDS RATIO</i>	29
TABELA 4. 6: TESTES PARA AVALIAÇÃO DA QUALIDADE DO AJUSTAMENTO.	30
TABELA 4. 7: ESTIMAÇÃO DOS QUARTIS DO TEMPO DE PERSISTÊNCIA E RESPECTIVO IC 95%.....	32
TABELA 4. 8: COMPARAÇÃO DAS FUNÇÕES DE SOBREVIVÊNCIA: VALOR P DOS TESTES LOG-RANK E DE PETO-PETO.	32
TABELA 4. 9: ESTIMAÇÃO DOS QUARTIS DO TEMPO ATÉ À NÃO PERSISTÊNCIA E RESPECTIVO IC 95% DE CADA CATEGORIA DA VARIÁVEL IDADE, VIVE SOZINHO E TERAPÊUTICA.	33
TABELA 4. 10: ANÁLISE UNIVARIADA: <i>HAZARD RATIO</i> E VALOR P DOS TESTES DE WALD.....	35
TABELA 4. 11: ANÁLISE UNIVARIADA: TESTES DA RAZÃO DE VEROSIMILHANÇAS.....	36
TABELA 4. 12: COMPARAÇÃO DE MODELOS MÚLTIPLOS: TESTES DA RAZÃO DE VEROSIMILHANÇAS.	36
TABELA 4. 13: COMPARAÇÃO COM O MODELO OBTIDO NO PASSO ANTERIOR: TESTE DA RAZÃO DE VEROSIMILHANÇAS.....	37
TABELA 4. 14: MODELO MÚLTIPLO FINAL: ESTIMATIVA DE CADA β_j , VALOR P DO TESTE DE WALD, <i>HAZARD RATIO</i> E RESPECTIVO IC95%.....	37
TABELA 4. 15: TESTE DA PROPORCIONALIDADE DAS FUNÇÕES DE RISCO, PROPOSTO POR GRAMBSCH E THERNEAU.	38
TABELA 4. 16: ANÁLISE UNIVARIADA: ESTIMATIVA DOS PARÂMETROS E VALOR P DOS TESTES DE WALD.....	40
TABELA 4. 17: MODELO DE TEMPO DE VIDA ACELERADO DE WEIBULL: ESTIMATIVA DOS COEFICIENTES, PARÂMETROS E TESTE DE WALD.....	42
TABELA 4. 18: MODELO DE RISCOS PROPORCIONAIS DE WEIBULL: ESTIMATIVA DOS COEFICIENTES, PARÂMETROS, <i>HAZARD RATIO</i> E RESPECTIVO IC95% E TESTE DE WALD.	43
TABELA 4. 19: ANÁLISE UNIVARIADA: ESTIMATIVA DOS PARÂMETROS E VALOR P DOS TESTES DE WALD.....	44
TABELA 4. 20: MODELO DE TEMPO DE VIDA ACELERADO LOG-NORMAL: ESTIMATIVA DOS PARÂMETROS E TESTE DE WALD.....	45
TABELA 4. 21: ANÁLISE UNIVARIADA: ESTIMATIVA DOS PARÂMETROS E VALOR P DOS TESTES DE WALD.....	46
TABELA 4. 22: MODELO DE TEMPO DE VIDA ACELERADO LOG-LOGÍSTICO: ESTIMATIVA DOS PARÂMETROS E TESTE DE WALD.....	47
TABELA 4. 23: MODELO DE POSSIBILIDADES PROPORCIONAIS LOG-LOGÍSTICO: ESTIMATIVA DOS PARÂMETROS, <i>ODDS RATIO</i> E RESPECTIVO IC95%.	48
TABELA 4. 24: COMPARAÇÃO DOS MODELOS PARAMÉTRICOS MÚLTIPLOS: AIC E TESTE DA RAZÃO DE VEROSIMILHANÇAS.....	49
TABELA 4. 25: ANÁLISE UNIVARIADA DE MODELOS DE RISCOS PROPORCIONAIS COM M NÓS INTERNOS: ESTIMATIVA DOS PARÂMETROS E VALOR DE AIC.	50
TABELA 4. 26: MODELOS DE RISCOS PROPORCIONAIS COM M NÓS INTERNOS: ESTIMATIVA DOS PARÂMETROS E VALOR DE AIC.	52
TABELA 4. 27: MODELO FLEXÍVEL DE RISCOS PROPORCIONAIS COM 2 NÓS INTERNOS: ESTIMATIVA DOS COEFICIENTES, PARÂMETROS, <i>HAZARD RATIO</i> E RESPECTIVO IC95% E TESTE DE WALD.	53
TABELA 4. 28: ANÁLISE UNIVARIADA DE MODELOS DE POSSIBILIDADES PROPORCIONAIS COM M NÓS INTERNOS: ESTIMATIVA DOS PARÂMETROS E VALOR DE AIC.	54
TABELA 4. 29: MODELOS DE POSSIBILIDADES PROPORCIONAIS COM M NÓS INTERNOS: ESTIMATIVA DOS PARÂMETROS E VALOR DE AIC.....	57
TABELA A. 1: ANÁLISE BIVARIADA: TABELAS DE CONTINGÊNCIA	63
TABELA A. 1: ANÁLISE BIVARIADA: TABELAS DE CONTINGÊNCIA (CONTINUAÇÃO DA PÁGINA ANTERIOR).....	64

Capítulo 1

Introdução

1.1 Enquadramento

Através do aumento do conhecimento científico, a criação e crescente aplicação de novas tecnologias e a melhoria dos cuidados de saúde, os indivíduos tendem a viver cada vez mais tempo.

Em Portugal, no triénio 1980-1982 a esperança média de vida à nascença era de 71,8 anos (68,2 anos para os homens e 75,2 anos para as mulheres). Três décadas depois, no triénio 2010-2012, a esperança média de vida à nascença aumentou para os 79,8 anos (76,7 anos e 82,6 anos, para homens e mulheres respetivamente). A mesma tendência verifica-se na esperança média de vida aos 65 anos, que em Portugal subiu de 14,9 anos para 18,8 anos (nos triénios 1980-1982 e 2010-2012, respetivamente) (Instituto Nacional de Estatística, 2013).

Esta tendência de aumento da esperança média de vida, à nascença e aos 65 anos, tem levado a um envelhecimento demográfico crescente, especialmente nos últimos anos. No entanto, o facto de a população viver mais tempo não significa que tenha mais saúde. Segundo a Organização Mundial de Saúde (OMS), “Saúde é um estado de completo bem-estar físico, mental e social e não somente ausência de doenças ou enfermidades”. (World Health Organization, 1948)

Atualmente, dado o crescente envelhecimento da população, as doenças crónicas são cada vez mais comuns. Segundo dados do Instituto Nacional de Estatística (INE), em 2011, 40,5% da população residente em Portugal, com idades compreendidas entre os 15 e os 64 anos tem algum problema de saúde ou sofre de uma doença crónica. A incidência destes problemas de saúde tende a aumentar com a idade: estima-se que 51,9% da população entre os 45 e os 54 anos e 68,8% da população entre os 55 e os 64 anos tem algum problema de saúde. Além da idade, o sexo também parece estar associado à incidência destes problemas: estima-se que 44,5% das mulheres tenham algum problema de saúde ou doença crónica, já no caso dos homens apenas 36,5% é afetado (Instituto Nacional de Estatística, 2012).

Ainda de acordo com a mesma fonte, cerca de 16,0% da população com algum tipo de problema de saúde apresenta dificuldade em realizar atividades básicas, especialmente atividades que envolvam mobilidade como andar, subir degraus e transportar objetos. Nas pessoas com idade mais avançada, na faixa etária dos 55 aos 64 anos, esta percentagem aumenta para 36,5%.

Pode então perceber-se que, além de associadas ao envelhecimento da população, as doenças crónicas estão intimamente ligadas a uma redução da qualidade de vida dos doentes, tanto física como emocional. Determinadas doenças crónicas são particularmente preocupantes, dada a sua elevada prevalência ou associação com condições debilitantes e fisicamente dolorosas (Ferreira LN, Ferreira PL, Pereira LN, et al., 2014).

Nos países mais desenvolvidos, onde a esperança de vida é mais elevada e a população está mais envelhecida, a qualidade de vida da população, especialmente ligada às doenças crónicas, torna-se cada vez mais uma preocupação e uma questão de saúde pública (World Health Organization, National Institute of Health, National Institute on Aging, et al., 2011).

Vários estudos mostram que a não adesão e a não persistência à terapêutica, aumentam a morbilidade e mortalidade de diversas doenças (Cramer JA, Roy A, Burrell A, et al., 2008). Assim de forma a melhorar a qualidade de vida do doente crónico, a adesão e a persistência à respetiva terapêutica são fundamentais na melhoria dos resultados em saúde, sejam estes demonstrados através do alívio, cura ou até mesmo na prevenção de alguns sinais ou sintomas.

A adesão à terapêutica (*adherence*), refere-se ao ato de agir em conformidade com as recomendações feitas pelo profissional de saúde, isto é, quando o comportamento do doente na toma da medicação, no cumprimento de uma dieta e/ou nas mudanças no estilo de vida, coincide com as recomendações de um prestador de cuidados de saúde (Cramer JA, Roy A, Burrell A, et al., 2008; Brown e Bussell, 2011). A persistência à terapêutica (*persistence*), por sua vez, refere-se ao ato de continuar o tratamento durante o período prescrito, podendo ser definida como “o período de tempo desde a iniciação até à descontinuação da terapêutica” (Cramer JA, Roy A, Burrell A, et al., 2008).

A não adesão e a não persistência são problemas reconhecidos que implicam consequências adversas para a saúde do doente, sendo particularmente significativos em doentes crónicos que requerem uma terapêutica de longa duração. Segundo a OMS, nos países desenvolvidos a adesão em doentes crónicos é em média de 50%. Em países em desenvolvimento, dada a carência e desigualdade no acesso a cuidados de saúde, a adesão será ainda menor (Sabaté E, 2003 ; Brown e Bussell, 2011).

Assim, é essencial compreender os fatores que influenciam a adesão e a persistência dos doentes à terapêutica, o que permitirá aos profissionais de saúde definir quais os doentes de maior risco e atuar em conformidade, por forma a obter melhores resultados clínicos e consequentemente melhorar a qualidade de vida do doente.

1.2 Definição do problema e Objetivos

Na presente dissertação, um dos objetivos é determinar quais são as características sociodemográficas e de saúde associadas ao tempo até à não persistência de doentes no tratamento de uma dada doença crónica, bem como, perceber de que forma estes fatores influenciam o risco de não persistência destes doentes. Para tal, recorrer-se-á à análise de regressão, isto é, ao estudo de modelos que relacionam uma variável resposta e um conjunto de variáveis explicativas. O objetivo principal deste método é encontrar o modelo que melhor descreve a relação entre a variável dependente e as variáveis independentes (covariáveis), que seja o mais parcimonioso e que melhor se ajusta aos dados.

Pretende-se ainda comparar diferentes modelos de regressão utilizados em análise de dados de sobrevivência, por forma a perceber qual o tipo de modelo que mais se adequa a este género de dados relacionados com a persistência no tratamento de doentes crónicos.

Numa abordagem inicial será considerado apenas se, no final do período de *follow-up* (seguimento), os indivíduos são ou não persistentes à terapêutica. Assim sendo, será ajustado um modelo linear generalizado (modelo de regressão logística), permitindo identificar os fatores associados à não persistência.

No entanto, a modelação de dados de sobrevivência é feita, geralmente, com recurso ao modelo de riscos proporcionais de Cox. Contudo, nem sempre os pressupostos deste modelo são verificados, impossibilitando o seu correto uso. Para além disso, o ajustamento de outros modelos poderá trazer vantagens para a análise.

Assim, serão ajustados aos dados de persistência no tratamento de uma doença crónica, o modelo de riscos proporcionais de Cox, três modelos de sobrevivência paramétricos (Weibull, log-logístico, log-normal) e um modelo paramétrico flexível proposto por Royston e Parmar (2002).

Capítulo 2

Metodologia

Trata-se de um estudo observacional prospetivo, com um período de *follow-up* máximo de 18 meses, realizado numa coorte de 360 indivíduos recrutados por diversas farmácias comunitárias portuguesas.

Para o estudo, apenas foram recrutados doentes que estivessem a iniciar um tratamento medicamentoso para uma determinada doença crónica, sendo por isso excluído do estudo qualquer doente que tivesse, nos 6 meses anteriores ao recrutamento, realizado um tratamento medicamentoso para a doença em análise. Foram igualmente excluídos doentes que dependessem de outrem para tomar a medicação.

2.1 Recolha de dados

A informação de base foi recolhida após o recrutamento. Cada doente respondeu a um questionário feito pelo farmacêutico, com questões relativas às suas características sociodemográficas (idade, nível de escolaridade, situação de emprego, agregado familiar), de saúde (prática de exercício, frequência de ida ao médico, consciência de que tem a doença, realização de um exame de diagnóstico, especialidade do médico prescriptor, existência de outras doenças crónicas, toma de outros medicamentos) e tipo de tratamento que iniciou (terapêutica prescrita).

O seguimento dos doentes em estudo foi feito através da base de dados de vendas das farmácias que participaram no estudo e por entrevista telefónica, sempre que o doente fosse identificado como potencialmente não persistente na base de dados, por forma a confirmar o estado de não persistência.

2.2 Variáveis em estudo

Neste trabalho foram consideradas características sociodemográficas, de saúde e tipo de terapêutica, cujos valores foram recolhidos no início do estudo.

2.2.1 Variáveis sociodemográficas

Idade – Indica a faixa etária do indivíduo na altura do recrutamento:

$$\text{Idade} = \begin{cases} 0 & \text{se } < 60 \text{ anos} \\ 1 & \text{se } 60 \text{ a } 70 \text{ anos} \\ 2 & \text{se } \geq 70 \text{ anos} \end{cases}$$

Nível de escolaridade – Variável categórica que representa o nível de escolaridade do indivíduo:

$$\text{Nível de escolaridade} = \begin{cases} 0 & \text{se } \text{sem ensino} \\ 1 & \text{se } \text{ensino básico} \\ 2 & \text{se } \text{ensino secundário ou superior} \end{cases}$$

Situação de emprego – Variável dicotómica que indica a situação do indivíduo face ao emprego:

$$\text{Situação de emprego} = \begin{cases} 0 & \text{se desempregado} \\ 1 & \text{se empregado/reformado} \end{cases}$$

Vive sozinho – Variável associada ao agregado familiar do indivíduo, indicando se vive ou não sozinho:

$$\text{Vive sozinho} = \begin{cases} 0 & \text{se não vive sozinho} \\ 1 & \text{se vive sozinho} \end{cases}$$

2.2.2 Variáveis de saúde

Exercício físico – Variável associada aos hábitos de prática de exercício físico do indivíduo:

$$\text{Exercício físico} = \begin{cases} 0 & \text{se não pratica nenhum tipo de exercício físico} \\ 1 & \text{se pratica algum tipo de exercício físico} \end{cases}$$

Ida ao médico – Representa a frequência com que o doente vai ao médico por semestre:

$$\text{Ida ao médico} = \begin{cases} 0 & \text{se vai ao médico 1 vez ou menos por semestre} \\ 1 & \text{se vai ao médico mais de 1 vez por semestre} \end{cases}$$

Doença crónica – Variável que indica se o doente sabe que tem a doença crónica em estudo:

$$\text{Doença crónica} = \begin{cases} 0 & \text{se não sabe que tem a doença crónica} \\ 1 & \text{se sabe que tem a doença crónica} \end{cases}$$

Exame – Indica se o doente fez um exame de diagnóstico da doença:

$$\text{Exame} = \begin{cases} 0 & \text{se não fez um exame de diagnóstico} \\ 1 & \text{se fez um exame de diagnóstico} \end{cases}$$

Especialidade do médico – Indica se o médico prescriptor é especialista na doença crónica ou se é de outra especialidade:

$$\text{Especialidade do médico} = \begin{cases} 0 & \text{se médico especialista na doença crónica} \\ 1 & \text{se médico de outra especialidade} \end{cases}$$

Outra doença crónica – Variável associada à existência de outra doença crónica:

$$\text{Outra doença crónica} = \begin{cases} 0 & \text{se não tem outra doença crónica} \\ 1 & \text{se tem outra doença crónica} \end{cases}$$

Outros medicamentos – Variável que indica se o doente está a tomar mais medicamentos:

$$\text{Outros medicamentos} = \begin{cases} 0 & \text{se não toma mais medicamentos} \\ 1 & \text{se toma mais medicamentos} \end{cases}$$

2.2.3 Variáveis de tratamento

Terapêutica – Variável dicotómica que indica o tipo de terapêutica que o doente está a fazer:

$$\text{Terapêutica} = \begin{cases} 0 & \text{se a terapêutica é mensal} \\ 1 & \text{se a terapêutica é semanal} \end{cases}$$

2.3 Cálculo do tempo até à não persistência

Para o cálculo do tempo até à não persistência de cada doente crónico, recorreu-se à informação fornecida pela base de dados de vendas da farmácia, onde foi feito o recrutamento e aquisição da medicação pelo doente. Através da base de dados e do respetivo ID do doente, tem-se acesso às datas de dispensa da medicação e à sua posologia.

Tal como já foi referido anteriormente, a persistência é definida como o número de dias desde o início até à descontinuação da terapêutica. Assim sendo, para o cálculo do tempo foi considerada a data de aquisição do medicamento como sendo a data de início da terapêutica (t_0). A descontinuação do tratamento ou não persistência ocorre caso transcorram 30 dias (*gap* máximo) sem aquisição de nova embalagem de medicação, após o último dia coberto pela medicação anteriormente adquirida.

Os doentes em estudo podiam fazer um de dois tipos distintos de tratamento medicamentoso:

2.3.1 Terapêutica mensal

Cada embalagem deste medicamento contém 1 comprimido, para administração única mensal. Assim sendo, a toma de uma embalagem do medicamento corresponde a 30 dias de persistência no tratamento e deverá ser repetida a cada 30 dias. Todavia, por diversas razões, o doente pode adquirir a nova embalagem com atraso, ultrapassando o período no qual a toma deveria ter sido feita.

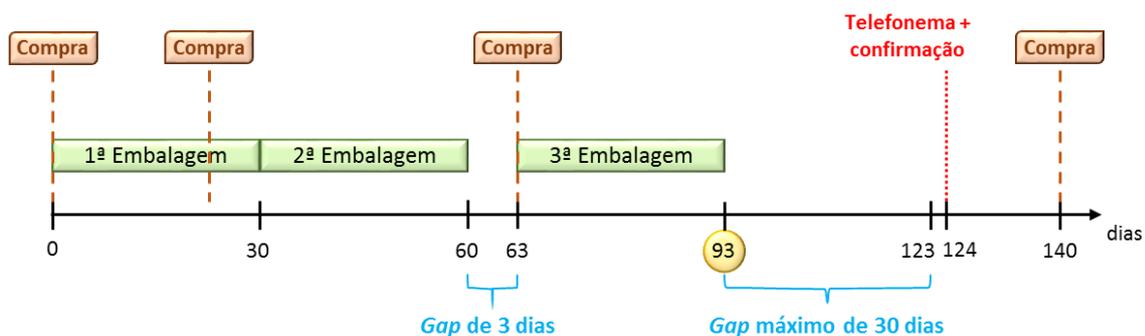


Figura 2. 1: Exemplo do cálculo do tempo até à não persistência, de um doente com terapêutica mensal.

A Figura 2.1 ilustra um exemplo do cálculo do tempo para um indivíduo com terapêutica mensal. Considerando a data de aquisição da 1ª embalagem como o início do tratamento (t_0), visto a terapêutica ser mensal os 30 dias seguintes estão cobertos pela medicação. Assim sendo, após a 1ª toma, o tempo até à não persistência será de pelo menos 30 dias. Se a 2ª embalagem for adquirida antes de terminar o período coberto pela embalagem anterior e o doente tomar o comprimido findo esse período, o tempo até à não persistência será de pelo menos 60 dias.

Por várias razões, o doente pode adquirir a nova embalagem com atraso, ultrapassando o período no qual a toma deveria ter sido feita. Caso a compra e toma da nova embalagem seja feita antes de se atingir o *gap* máximo, o doente mantém-se persistente no tratamento. No exemplo, a 3ª embalagem é adquirida com um atraso de 3 dias e após a sua toma o tempo até à não persistência será pelo menos de 93 dias.

No entanto, se na base de dados de vendas da farmácia não existir informação de dispensa de nova embalagem dentro do *gap* máximo de 30 dias, o doente é identificado como potencialmente não persistente, seguindo-se o contato telefónico.

No caso de o doente ter adquirido a medicação numa farmácia diferente, dentro do *gap* máximo, é recolhido no contato telefónico o dia em que foi feita a compra e o doente mantém-se persistente, continuando a ser acompanhado. Caso contrário, se a não persistência é confirmada, o doente é identificado na base de dados do *follow-up* como não persistente. No exemplo, após confirmação telefónica da não persistência, o doente é identificado como não persistente, sendo o seu tempo até à não persistência igual a 93 dias.

2.3.2 Terapêutica semanal

Cada embalagem deste medicamento contém 4 comprimidos, para administração 1 vez por semana. Desta forma, a toma de 1 comprimido corresponde a uma persistência no tratamento de 7 dias; por sua vez, a toma de uma embalagem corresponde a uma persistência de 28 dias.

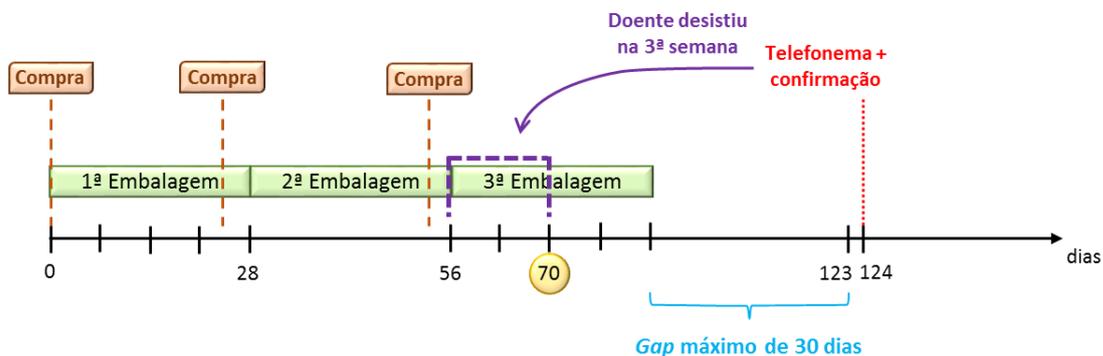


Figura 2. 2: Exemplo do cálculo do tempo até à não persistência, de um doente com terapêutica semanal.

A Figura 2.2 ilustra um exemplo do cálculo do tempo para um indivíduo com terapêutica semanal. Tal como já foi mencionado anteriormente, a data de aquisição da 1ª embalagem é considerada como o início do tratamento (t_0). Uma vez que a terapêutica é semanal, após a 1ª toma os 7 dias seguintes estão cobertos pela medicação, sendo o tempo até à não persistência de pelo menos 7 dias. O 2º comprimido deverá ser tomado na semana seguinte, passando o tempo até à não persistência a ser de pelo menos 14 dias. Desta forma, finda uma embalagem de 4 comprimidos semanais, o tempo do doente é de pelo menos 28 dias.

Se o doente comprar as novas embalagens antes de terminar o período coberto pela medicação anterior e tomar sempre os comprimidos, o seu tempo será aumentado em ciclos de 28 dias. No exemplo, o tempo do doente quando termina a 2ª embalagem é de pelo menos 56 dias e, se terminar a 3ª, será de pelo menos 84 dias.

Novamente, se houver atraso inferior ao *gap* máximo, na compra e toma da nova embalagem, o doente mantém-se persistente no tratamento. Porém, se transcorrerem mais de 30 dias após o fim do período coberto pela última medicação, sem ser adquirida uma nova embalagem, o doente é identificado como potencialmente não persistente, seguindo-se o contacto telefónico.

Se a não persistência é confirmada, o doente é identificado na base de dados do *follow-up* como não persistente. Não obstante, na terapêutica semanal o doente pode ter iniciado a toma da última embalagem sem a terminar. Neste caso, durante o contacto telefónico é necessário perguntar ao doente o número de comprimidos da última embalagem efetivamente tomados, para efeitos do cálculo do tempo até à não persistência. No exemplo, o doente interrompeu o tratamento depois de tomar 2 comprimidos da 3ª embalagem adquirida; assim sendo, o doente é identificado na base de dados de *follow-up* como não persistente, com um tempo até à não persistência igual a 70 dias.

2.4 Análise estatística

2.4.1 Fatores associados à não persistência

Por forma a identificar quais os fatores sociodemográficos e de saúde que contribuem significativamente para a não persistência de doentes, no tratamento de uma determinada doença crónica, foi utilizada a análise de regressão logística.

Foi ajustado um modelo múltiplo com variável resposta binária, correspondendo à persistência ou não persistência na terapêutica pelo doente, no final do estudo. As covariáveis analisadas consistem nas características de base recolhidas aquando do recrutamento.

Para a construção do modelo múltiplo, partiu-se de uma análise de regressão simples, procedendo-se, para cada covariável, ao cálculo do *odds ratio* e respetivo intervalo de confiança a 95%. Em seguida, a seleção de variáveis foi realizada com base no critério de informação de Akaike (AIC) e complementarmente através do método *stepwise* de inclusão progressiva com eliminação regressiva. Para o modelo múltiplo final, foi novamente calculado o *odds ratio* e respetivo intervalo de confiança a 95%, para as covariáveis incluídas no modelo.

O diagnóstico ao ajustamento do modelo múltiplo incluí o teste de Hosmer e Lemeshow e o teste de Qui-quadrado com base nos resíduos de Pearson e nos resíduos *deviance*. Foi ainda feita uma análise da sensibilidade e especificidade do modelo, recorrendo-se à curva ROC (“*Receiver Operating Characteristic*”), de modo a avaliar a capacidade discriminatória do modelo.

2.4.2 Fatores associados ao tempo até à não persistência

Com a finalidade de identificar os fatores que têm influência significativa no tempo até à não persistência no tratamento de uma doença crónica, recorreu-se à análise de sobrevivência.

A variável resposta considerada é o tempo até à não persistência no tratamento, definido como o tempo desde o início até à descontinuação da terapêutica. As covariáveis consideradas foram todos os fatores sociodemográficos e de saúde relativos a cada indivíduo, cujos valores foram recolhidos no início do estudo.

Foram ajustados diferentes tipos de modelos, de forma a identificar qual se adequa mais a este género de dados de sobrevivência. Foram considerados cinco modelos diferentes: o modelo de Cox, três modelos de sobrevivência paramétricos (Weibull, log-logístico, log-normal) e o modelo paramétrico flexível (Royston e Parmar), mais recente e ainda pouco conhecido.

Toda a análise estatística foi feita recorrendo ao *software R versão 3.0.1*.

Capítulo 3

Fundamentos teóricos

3.1 Regressão logística

A regressão logística é um método estatístico, muito utilizado para modelar e descrever a relação entre uma variável resposta binária e uma ou mais variáveis explicativas.

Segundo Hosmer, Lemeshow e Sturdivant (2013), este método pode ser utilizado em estudos onde a variável de interesse é o tempo até à ocorrência de um acontecimento. No entanto, nestes casos a variável resposta é transformada numa variável dicotómica, sendo apenas modelada a ocorrência ou não ocorrência do acontecimento, ficando por isso excluído da análise o tempo até ao acontecimento, pelo que há perda de informação.

No passado, a regressão logística tornou-se um método comum na análise de dados de sobrevivência, dada a sua fácil implementação no *software* estatístico disponível. No entanto, visto que este método não considera o tempo até à ocorrência do acontecimento e visto que atualmente estão disponíveis, com igual facilidade de implementação, outros métodos mais adequados à análise de dados de sobrevivência, a utilização da regressão logística é desaconselhada. Ainda assim, numa primeira abordagem da análise dos dados, nesta dissertação será ajustado um modelo de regressão logística.

3.1.1 Ajustamento do modelo

Na regressão logística com variável resposta (Y) binária ou dicotómica, esta pode tomar um de dois valores possíveis, $Y = 0$ se o acontecimento não ocorre ou $Y = 1$ se o acontecimento ocorre. A probabilidade de sucesso é dada por $P(Y = 1|\mathbf{x}) = \pi(\mathbf{x})$.

No âmbito do presente trabalho, o acontecimento de interesse é a não persistência no tratamento pelo doente crónico. Desta forma, $Y = 0$ corresponde a um doente persistente no final do período de seguimento (insucesso) e $Y = 1$ corresponde a um doente não persistente no final do período de seguimento (sucesso).

O valor médio condicional $E(Y|\mathbf{x})$, isto é, o valor médio da variável resposta Y condicional ao valor das covariáveis X_j , $j = 1 \dots p$. Quando é utilizada a distribuição logística, este valor é dado por:

$$E(Y|\mathbf{x}) = \pi(\mathbf{x}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \quad (3.1)$$

A relação entre a variável resposta e as covariáveis pode ser descrita com recurso a diferentes funções de ligação. Na presente dissertação será utilizada a função de ligação logit, dada pela seguinte transformação de (3.1):

$$\ln\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (3.2)$$

Para ajustar o modelo é necessário estimar os parâmetros $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$ sendo, para tal, utilizado o método da máxima verossimilhança. Considerando uma amostra de n observações independentes $(x_i, y_i), i = 1 \dots n$, é possível obter as estimativas de $\boldsymbol{\beta}$ encontrando os valores que maximizam a função de verossimilhança:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (3.3)$$

Para a seleção de variáveis a incluir no modelo final foram seguidas duas abordagens, por forma a fazer uma seleção mais correta. Foi utilizado o critério de informação de Akaike (AIC) e o método *stepwise* de inclusão progressiva (valor p de entrada $p_E = 0,1$) com eliminação regressiva (valor p de eliminação $p_R = 0,15$).

Assim, as variáveis cuja inserção no modelo diminui o valor de AIC e gera um valor p associado ao teste da razão de verossimilhanças inferior a 0,1 são incluídas no modelo. Após a inclusão de cada variável, verificou-se se as variáveis anteriormente incluídas no modelo se mantinham significativas, através da análise do respetivo valor p .

Obtido o modelo final, é possível analisar os *odds ratios* associados a cada variável. Para cada variável x_i , a possibilidade de sucesso (*odds*) dos indivíduos com $x_i = 1$ é dada por $\pi(1)/[1 - \pi(1)]$, da mesma forma, a possibilidade de sucesso dos indivíduos com $x_i = 0$ é dada por $\pi(0)/[1 - \pi(0)]$. A razão entre as possibilidades de sucesso de um indivíduo com $x_i = 1$ e de um indivíduo com $x_i = 0$ é dada pelo *odds ratio*:

$$\text{Odds ratio} = \frac{\pi(1)/[1 - \pi(1)]}{\pi(0)/[1 - \pi(0)]} = e^{\beta_1} \quad (3.4)$$

3.1.2 Diagnóstico do modelo

Por fim, o ajustamento do modelo deverá ser avaliado recorrendo-se ao teste de Hosmer e Lemeshow e à análise de resíduos de Pearson e resíduos *deviance*. É ainda aconselhável verificar a capacidade discriminatória do modelo, através da análise da área abaixo da Curva ROC (“*Receiver Operating Characteristic*”).

3.2 Análise de sobrevivência

A análise de sobrevivência, ou análise estatística de dados de sobrevivência, engloba um conjunto de procedimentos estatísticos, onde a variável dependente é o tempo desde um instante inicial bem definido até à ocorrência de um determinado acontecimento de interesse. O tempo desde o início do *follow-up* de um doente até à ocorrência do acontecimento é denominado “tempo de vida”. O acontecimento, por sua vez, é uma qualquer experiência de interesse que possa ocorrer a um indivíduo, seja morte, recaída ou no caso do presente estudo, o doente crónico deixar de ser persistente no tratamento.

Em estudos onde existem longos períodos de *follow-up*, os participantes por vezes não atingem o tempo total de seguimento previsto. Em geral, o procedimento mais comum é excluir da análise os dados destes participantes, visto que se desconhece o seu verdadeiro tempo até à ocorrência do acontecimento em causa. A grande vantagem da análise de sobrevivência reside na possibilidade de incluir na análise estes dados incompletos. Os indivíduos para os quais não se conhece o tempo de sobrevivência exato, porque abandonaram o estudo, foram perdidos para o *follow-up* ou o tempo de seguimento chegou ao fim antes de ocorrer o acontecimento, dão origem a observações “censuradas”.

Assim sendo, o tempo de sobrevivência de um indivíduo é representado pela variável aleatória contínua e não negativa T e a indicação de censura ou ocorrência do acontecimento é dada pela variável indicatriz δ , sendo $\delta = 1$ caso seja observado o acontecimento e $\delta = 0$ caso o acontecimento não seja observado, isto é quando ocorre censura.

Na análise de sobrevivência existem duas funções fundamentais, que caracterizam a distribuição do tempo de sobrevivência (T):

- Função de sobrevivência (*survival function*), descreve a probabilidade de um indivíduo sobreviver para além do instante t . É uma função monótona decrescente e contínua.

$$S(t) = P(T > t), \quad t > 0 \quad (3.5)$$

- Função de risco (*hazard function*), representa a probabilidade de morte no instante t , sabendo que o indivíduo sobreviveu até esse instante. É uma função não negativa, que pode tomar várias formas (monótona crescente, monótona decrescente, constante, *bathtub-shaped* ou *hump-shaped*).

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (3.6)$$

Uma característica importante destas duas funções é o facto de estarem relacionadas entre si e também com a função densidade de probabilidade $f(t)$. Assim conhecendo apenas uma das funções é possível chegar facilmente às restantes.

$$S(t) = \exp \left[- \int_0^t h(u) du \right] \quad (3.7)$$

$$h(t) = - \frac{d \ln S(t)}{dt} \quad (3.8)$$

$$f(t) = h(t) S(t) \quad (3.9)$$

A função de risco cumulativa (*cumulative hazard function*) mede o risco de ocorrência do acontecimento de interesse até ao instante t . Esta função define-se como:

$$H(t) = \int_0^t h(u)du \quad \text{ou de forma equivalente} \quad H(t) = -\ln S(t) \quad (3.10)$$

3.2.1 Estimação não paramétrica

Estimador de Kaplan-Meier

Tal como já foi mencionado, em estudos com longos períodos de *follow-up* por vezes o acontecimento não é observado em certos indivíduos, para os quais o verdadeiro tempo de sobrevivência não é conhecido, gerando dados censurados. Neste caso, é habitual utilizar o estimador não-paramétrico da função de sobrevivência, designado por estimador de Kaplan-Meier.

Considere-se:

- Uma amostra de dimensão n , sendo $t_{(1)}, \dots, t_{(r)}$ $r \leq n$ os tempos distintos até à ocorrência do acontecimento, na amostra
- d_i o número de ocorrências num determinado $t_{(i)}$ $i = 1, \dots, r$
- n_i o número de indivíduos em risco imediatamente antes de $t_{(i)}$ $i = 1, \dots, r$

Então, o estimador de Kaplan-Meier é dado por:

$$\hat{S}(t) = \prod_{i: t_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right) \quad (3.11)$$

Este estimador considera tantos intervalos quanto o número de acontecimentos distintos que ocorreram, sendo a estimativa $\hat{S}(t)$ uma função em escada com saltos nos instantes de morte observados. O gráfico da função $\hat{S}(t)$ ou curva de Kaplan-Meier, permite avaliar o ritmo a que os acontecimentos vão ocorrendo em diferentes grupos, apenas atingindo o valor zero quando o maior tempo observado não é censurado.

Os limites do intervalo de confiança *pointwise*, isto é, o intervalo de $100(1 - \alpha)\%$ de confiança para a função de sobrevivência em cada instante t , são dados pela expressão:

$$[\hat{S}(t) \pm z_{1-\alpha/2} \widehat{se}(\hat{S}(t))] \quad (3.12)$$

Ainda assim, quando a estimativa da função de sobrevivência se aproxima de 0 ou 1, podem ser obtidos limites fora do intervalo (0,1). Por forma a ultrapassar este problema, o intervalo de confiança pode ser obtido para uma transformação da função de sobrevivência, como por exemplo a transformação $\ln\{-\ln[S(t)]\}$. Nestas condições os limites do intervalo de confiança são dado por:

$$[\ln\{-\ln[\hat{S}(t)]\} \pm z_{1-\alpha/2} \widehat{se}\{\ln[-\ln(\hat{S}(t))]\}] \quad (3.13)$$

Tal como proposto por Hall e Wellner (1980), é também possível obter bandas de confiança para toda a função de sobrevivência.

Para verificar se existem diferenças estatisticamente significativas entre as funções de sobrevivência de dois ou mais grupos, é necessário recorrer a testes não paramétricos. Existe uma classe de testes cuja estatística de teste é definida por:

$$\frac{[\sum_{j=1}^r w_j (d_{1j} - e_{1j})]^2}{\sum_{j=1}^r w_j^2 v_{1j}} \quad (3.14)$$

onde $(d_{1j} - e_{1j})$ é a diferença entre o número de mortes observadas e o número de mortes esperadas no grupo 1 no instante t_j com $j = 1 \dots k$ e w_j é uma constante conhecida.

Consoante o peso atribuído à constante w_j , diferentes testes podem ser obtidos:

- $w_j = 1$ teste log-rank
- $w_j = n_j$ teste de Gehan
- $w_j = \sqrt{n_j}$ teste de Tarone-Ware
- $w_j = \sum_{i:t(i) \leq t(j)} \left(1 - \frac{d_i}{n_{i+1}}\right)$ teste de Peto-Peto

Na presente dissertação recorreu-se aos testes não paramétricos log-rank e Peto-Peto. O teste log-rank é o mais potente quando as funções de risco dos grupos são proporcionais. Caso não haja proporcionalidade de riscos, este teste é ainda bastante potente desde que as funções de risco não se cruzem.

A hipótese de riscos proporcionais pode ser avaliada informalmente com base na representação gráfica das funções $\ln[-\ln \hat{S}(t)]$ de cada grupo, que se deverão manter razoavelmente paralelas.

Estimação dos percentis do tempo de vida

Por forma a caracterizar a localização da distribuição do tempo de vida é aconselhável recorrer à estimativa da mediana. Esta medida de localização corresponde ao tempo no qual se estima, que para 50% dos indivíduos em estudo, ainda não tenha ocorrido o acontecimento de interesse. Após a estimação da função de sobrevivência através do estimador de Kaplan-Meier, a estimativa da mediana é facilmente obtida através da expressão:

$$\hat{\chi}_{50} = \min\{t_{(i)}: \hat{S}(t_{(i)}) \leq 0,5\} \quad (3.15)$$

onde $t_{(i)}$ é o i -ésimo instante de morte, $i = 1 \dots k$

Em estudos onde a estimativa da função de sobrevivência é sempre superior a 0,5 não é possível estimar a mediana do tempo de vida. Além da mediana pode ser conveniente estimar outro percentil de probabilidade p :

$$\hat{\chi}_p = \min\{t_{(i)}: \hat{S}(t_{(i)}) \leq 1 - p\} \quad (3.16)$$

3.3 Modelo de regressão de Cox

O modelo de riscos proporcionais de Cox é atualmente o procedimento mais utilizado para modelar a relação entre um conjunto de covariáveis e uma variável resposta do tipo tempo de vida. A utilização deste modelo permite não só avaliar o efeito de diferentes tratamentos no tempo de vida dos indivíduos em estudo, como também o efeito de outras características de base dos indivíduos.

O modelo de regressão de Cox apresenta-se geralmente sob a forma de uma função de risco dada pelo produto de dois termos: a função de risco subjacente $h_0(t)$, que corresponde à função de risco de um indivíduo com vetor de covariáveis associado $\mathbf{x} = 0$ (indivíduo padrão) e o fator $e^{\boldsymbol{\beta}'\mathbf{x}}$, denominado risco relativo, que representa o efeito multiplicativo das covariáveis na função de risco. Assim, para um determinado indivíduo a que está associado o vetor de covariáveis $\mathbf{x} = (x_1, \dots, x_p)'$, a função de risco no instante t é definida por:

$$h(t; \mathbf{x}) = h_0(t)e^{\boldsymbol{\beta}'\mathbf{x}} \quad (3.17)$$

onde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ é o vetor dos coeficientes de regressão.

Com base na função de risco, é possível comparar indivíduos que difiram apenas no valor de uma covariável, através do risco relativo (*hazard ratio*). Considere-se dois indivíduos com vetor de p covariáveis \mathbf{x}_1 e \mathbf{x}_2 , respetivamente, que diferem apenas no valor da variável x_j . O risco relativo é dado pela razão das respetivas funções de risco:

$$\frac{h(t; \mathbf{x}_1)}{h(t; \mathbf{x}_2)} = \frac{h_0(t)\exp(\beta_1 x_{11} + \dots + \beta_j x_{1j} + \dots + \beta_p x_{1p})}{h_0(t)\exp(\beta_1 x_{21} + \dots + \beta_j x_{2j} + \dots + \beta_p x_{2p})} = \exp[\beta_j(x_{1j} - x_{2j})] \quad (3.18)$$

Assim, $\exp(\beta_j)$ representa o risco relativo de ocorrência do acontecimento para dois indivíduos que diferem uma unidade no valor da covariável x_j , sendo iguais os valores das restantes covariáveis.

Uma característica importante do modelo é o facto do risco relativo ($e^{\boldsymbol{\beta}'\mathbf{x}}$) envolver as covariáveis e não depender do tempo. Por conseguinte, o modelo pressupõe riscos proporcionais, isto é, pressupõe que o efeito das covariáveis no tempo de vida não sofre alterações durante o período de seguimento do indivíduo. Além disso, a forma exponencial do risco relativo assegura que as estimativas dos riscos sejam não negativas, o que torna o modelo de Cox muito apelativo. Por outro lado, uma vez que a função de risco subjacente não é especificada, o modelo de Cox é semiparamétrico.

Tal como referido por Kleinbaum (2005), embora possa ser preferível usar um modelo paramétrico, na maioria das vezes não temos a certeza de qual é o modelo mais apropriado. E apesar do modelo de Cox ser semiparamétrico, este é bastante robusto gerando resultados muito próximos do modelo paramétrico mais correto. Estas características contribuem para a sua grande popularidade.

3.3.1 Ajustamento do modelo

Geralmente, em estudos de sobrevivência é recolhida informação relativa a um conjunto de variáveis, cuja influência sobre o tempo de sobrevivência dos indivíduos em estudo pode ser desconhecida. A inclusão no modelo de variáveis desnecessárias não só diminui a eficácia do modelo como aumenta valores de p e margem de erro dos intervalos de confiança.

Assim, por forma a obter um modelo parcimonioso, é importante identificar quais as variáveis explicativas com influência significativa na sobrevivência dos indivíduos. Vários métodos de seleção de variáveis podem ser utilizados. Na presente dissertação far-se-á uso do método sugerido por Collett (2003), que compara vários modelos alternativos.

O primeiro passo deste método consiste na construção de modelos contendo apenas uma variável. Para tal é necessário estimar o vetor dos parâmetros β_j , pela maximização da função de verosimilhança parcial proposta por Cox (1975):

$$L(\beta) = \prod_{i=1}^k \frac{\exp(\beta' X_i)}{\sum_{l \in R_i} \exp(\beta' X_l)} \quad (3.19)$$

Em seguida, é calculado o valor da estatística $-2 \ln \hat{L}$ para cada um dos modelos, que é comparado, através do teste da razão de verosimilhanças, com o valor da estatística para o modelo nulo (sem variáveis explicativas). A estatística do teste da razão de verosimilhanças é dada pela expressão:

$$G = -2 \ln[\hat{L}_{nulo} / \hat{L}_{modelo}] \quad (3.20)$$

Se uma variável explicativa por si só reduzir significativamente o valor da estatística, é considerada potencialmente importante, influenciando o tempo de sobrevivência.

Posteriormente, é construído um modelo que inclui todas as variáveis consideradas potencialmente importantes e calcula-se novamente o valor da estatística $-2 \ln \hat{L}$. Na presença de certas variáveis, outras podem deixar de ser significativas. Desta forma, as variáveis são omitidas do modelo, uma de cada vez, e o valor da estatística é calculado de novo para os vários modelos. Consequentemente, variáveis cuja omissão não aumenta significativamente o valor de $-2 \ln \hat{L}$, são removidas do modelo.

As variáveis que isoladamente não foram consideradas potencialmente importantes, podem na presença de outras variáveis revelar influência sobre o tempo de sobrevivência. Assim, estas variáveis são incluídas no modelo obtido anteriormente, uma de cada vez, sendo mantidas as que levarem a uma redução significativa de $-2 \ln \hat{L}$.

Por último, procede-se a uma verificação final, por forma a confirmar que nenhuma variável incluída no modelo pode ser omitida sem aumentar significativamente o valor da estatística $-2 \ln \hat{L}$, mas também que nenhuma variável não incluída reduz de forma significativa o valor da estatística.

Nesta dissertação, o nível de significância considerado para a inclusão ou omissão de variáveis, pelo método sugerido por Collett (2003), será de 0,1 tal como recomenda o autor.

3.3.2 Diagnóstico do modelo

Por forma a avaliar a qualidade do ajustamento do modelo final, é fundamental realizar uma análise de resíduos. Existem vários tipos de resíduos de interesse que permitem analisar diferentes aspetos do ajustamento do modelo: resíduos de Cox-Snell, de Schoenfeld, Martingala, *deviance* e *score*.

Nesta dissertação serão utilizados os resíduos propostos por Schoenfeld (1982), que permitem avaliar a hipótese de riscos proporcionais após ajustado o modelo.

Contrariamente a outros tipos de resíduos, a cada indivíduo corresponde um conjunto de resíduos de Schoenfeld, isto é, os resíduos são definidos para o i -ésimo indivíduo e para cada covariável x_j incluída no modelo, sendo dados por:

$$r_{ji} = \delta_i \{x_{ji} - a_{ji}\} \quad \text{com} \quad a_{ji} = \frac{\sum_{l \in R_i} x_{jl} \exp(\hat{\beta}' x_l)}{\sum_{l \in R_i} \exp(\hat{\beta}' x_l)} \quad (3.21)$$

onde R_i é o conjunto de indivíduos em risco no instante t_i .

Grambsch e Therneau (1994) recomendam a utilização dos resíduos de Schoenfeld padronizados (*scaled Schoenfeld residuals*), que permitem avaliar de forma mais eficaz o pressuposto de riscos proporcionais. É aconselhado pelos autores fazer uma análise gráfica e recorrer a um teste estatístico, baseado nestes resíduos e numa função do tempo, $g(t)$.

3.4 Modelos de sobrevivência paramétricos

Tal como referido na secção 3.3, o modelo de Cox é bastante utilizado dada a sua facilidade de implementação e interpretação. Visto ser um modelo de regressão semi-paramétrico, não é necessário assumir uma distribuição particular para o tempo de vida, tornando este modelo bastante flexível.

Por outro lado, nos modelos de sobrevivência paramétricos assume-se que o tempo de vida segue uma determinada família de distribuições com parâmetros desconhecidos. Quando existem fortes indícios de que uma distribuição do tempo de vida é adequada, é preferível utilizar estes modelos, uma vez que se tornam mais eficientes do que o modelo de Cox. Além disso, estes modelos permitem incluir covariáveis cujo efeito não é proporcional em termos de risco, mas que pode ser proporcional noutra escala, por exemplo em termos de possibilidade (*odds*).

3.4.1 Modelos de tempo de vida acelerado

Nos modelos de tempo de vida acelerado (AFT), as covariáveis têm um efeito multiplicativo no tempo de sobrevivência, acelerando ou travando o tempo até ao acontecimento de interesse. A representação log-linear destes modelos é dada pela expressão:

$$\ln T = \mu - \alpha'x + \sigma\varepsilon \quad (3.22)$$

onde μ é o termo independente, α é o vetor de parâmetros de regressão e σ é o parâmetro de escala. A distribuição da variável aleatória ε não depende das covariáveis.

Considere-se o indivíduo padrão com tempo de vida $T_0 = \exp(\mu + \sigma\varepsilon)$ e cuja função de sobrevivência é $S_0(t) = P[\exp(\mu + \sigma\varepsilon) > t]$, então:

$$S(t; \mathbf{x}) = S_0(t \exp(\alpha' \mathbf{x})) \quad (3.23)$$

Assim, o efeito das covariáveis consiste numa modificação da escala do tempo através do fator de aceleração $\exp(\boldsymbol{\alpha}'\mathbf{x})$, pelo que se o fator for:

- Inferior a 1, o tempo até ao acontecimento é acelerado pelas covariáveis
- Superior a 1, o tempo até ao acontecimento é travado pelas covariáveis

Modelo de regressão de Weibull

É simultaneamente um modelo de riscos proporcionais e um modelo de tempo de vida acelerado. Considere-se uma distribuição de Weibull com parâmetro de escala $\lambda > 0$ e parâmetro de forma $\gamma > 0$. As funções de sobrevivência, de risco e densidade de probabilidade são dadas respetivamente por:

$$S(t) = \exp(-\lambda t^\gamma) \quad (3.24)$$

$$h(t) = \lambda \gamma t^{\gamma-1} \quad (3.25)$$

$$f(t) = \lambda \gamma t^{\gamma-1} \exp(-\lambda t^\gamma) \quad (3.26)$$

Os pressupostos do modelo de regressão de Weibull podem ser verificados através de uma análise do gráfico da função $\ln[-\ln(\hat{S}(t))]$ vs $\ln(t)$:

- Se o gráfico for razoavelmente linear, então pode assumir-se que o tempo de sobrevivência tem distribuição de Weibull.
- Se as funções correspondentes a vários grupos forem razoavelmente paralelas, pode assumir-se que os riscos são proporcionais.
- Se é razoável aceitar ambos os pressupostos anteriores, então também é plausível aceitar o pressuposto de tempo de vida acelerado.

Sob a forma de modelo de riscos proporcionais, o modelo de regressão de Weibull tem uma aparência semelhante ao modelo de Cox. No entanto, uma vez que se assume que o tempo de sobrevivência segue uma distribuição de Weibull, a função de risco do indivíduo padrão ($h_0(t)$) é especificada. Assim, a função de risco de um indivíduo com vetor de covariáveis \mathbf{x} é:

$$h(t; \mathbf{x}) = h_0(t) \exp(\boldsymbol{\beta}'\mathbf{x}) = \lambda \gamma t^{\gamma-1} \exp(\boldsymbol{\beta}'\mathbf{x}) \quad (3.27)$$

onde $\lambda \exp(\boldsymbol{\beta}'\mathbf{x})$ é o parâmetro de escala e γ é o parâmetro de forma.

A função de sobrevivência do modelo de Weibull sob a forma de modelo de riscos proporcionais é, por sua vez, dada por:

$$S(t; \mathbf{x}) = S_0(t) \exp(\boldsymbol{\beta}'\mathbf{x}) = \exp(-\lambda t^\gamma \exp(\boldsymbol{\beta}'\mathbf{x})) \quad (3.28)$$

Como modelo de tempo de vida acelerado, o modelo de Weibull pode ser escrito na forma log-linear:

$$\ln T = \mu + \boldsymbol{\alpha}'\mathbf{x} + \sigma \varepsilon \quad (3.29)$$

Uma vez que o tempo de vida segue uma distribuição de Weibull, a variável aleatória ε segue uma distribuição de Gumbel. Assim, a função de sobrevivência do modelo de Weibull de um indivíduo com vetor de covariáveis \mathbf{x} é dada por:

$$S(t; \mathbf{x}) = \exp \left[-\exp \left(\frac{\ln t - \mu - \boldsymbol{\alpha}'\mathbf{x}}{\sigma} \right) \right] \quad (3.30)$$

Comparando as funções de sobrevivência obtidas a partir do modelo de riscos proporcionais e da representação log-linear do modelo de tempo de vida acelerado, é possível chegar às seguintes relações entre os parâmetros:

$$\lambda = \exp \left(-\frac{\mu}{\sigma} \right) \quad \gamma = \frac{1}{\sigma} \quad \beta_j = -\frac{\alpha_j}{\sigma} \quad (3.31)$$

Modelo de regressão log-normal

A variável T segue uma distribuição log-normal com parâmetros μ e σ , se $\ln T$ segue uma distribuição normal de parâmetros μ e σ^2 . As funções de sobrevivência, de risco e densidade de probabilidade de T são dadas por:

$$S(t) = 1 - \Phi \left(\frac{\ln t - \mu}{\sigma} \right) \quad (3.32)$$

$$h(t) = \frac{\phi \left(\frac{\ln t - \mu}{\sigma} \right)}{t\sigma \left[1 - \Phi \left(\frac{\ln t - \mu}{\sigma} \right) \right]} \quad (3.33)$$

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} t^{-1} \exp \left[-\frac{1}{2} \left(\frac{\ln t - \mu}{\sigma} \right)^2 \right] \quad (3.34)$$

para $t > 0$ e onde $\Phi(\cdot)$ é a função de distribuição da normal padrão, dada por $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-u^2/2) du$.

A função de risco da distribuição log-normal tem uma forma bastante característica: é igual a zero quando $t = 0$, cresce até um máximo que depende do valor de σ e, em seguida, decresce até zero à medida que t tende para infinito.

O modelo de regressão log-normal é, tal como o modelo de Weibull, um modelo de tempo de vida acelerado, pelo que pode ser escrito na forma log-linear:

$$\ln T = \mu + \boldsymbol{\alpha}'\mathbf{x} + \sigma\varepsilon \quad (3.35)$$

Uma vez que o tempo de vida segue uma distribuição log-normal, a variável aleatória ε segue uma distribuição normal de parâmetros $\mu = 0$ e $\sigma = 1$. Assim, a função de sobrevivência do modelo de regressão log-normal de um indivíduo com vetor de covariáveis \mathbf{x} é dada por:

$$S(t; \mathbf{x}) = 1 - \Phi \left(\frac{\ln t - \mu - \boldsymbol{\alpha}'\mathbf{x}}{\sigma} \right) \quad (3.36)$$

Note-se ainda que a distribuição de Weibull e a distribuição log-normal são casos particulares da distribuição gama generalizada.

3.4.2 Modelos de possibilidades proporcionais

Nos modelos de possibilidades proporcionais, assume-se que a razão de possibilidades (*odds ratio*) se mantém constante ao longo do tempo. Além disso, as covariáveis têm um efeito multiplicativo na possibilidade (*odds*) de um indivíduo sobreviver para além do instante t (sucesso). Assim, a possibilidade de sobrevivência de um indivíduo com vetor de covariáveis \mathbf{x} é dado por:

$$\frac{S(t; \mathbf{x})}{1 - S(t; \mathbf{x})} = e^{\eta} \frac{S_0(t)}{1 - S_0(t)} \quad (3.37)$$

onde $\eta = \boldsymbol{\beta}'\mathbf{x}$ e $S_0(t)$ é a função de sobrevivência do indivíduo padrão.

O logaritmo da razão de possibilidades de sobrevivência para além de t , entre o i -ésimo indivíduo com vetor de covariáveis \mathbf{x}_i e um indivíduo padrão, é apenas η_i . Assim o modelo é linear para o logaritmo da razão de possibilidades.

Modelo de regressão log-logístico

É simultaneamente um modelo de tempo de vida acelerado e um modelo de possibilidades proporcionais. Considere-se uma distribuição log-logística com parâmetro de escala $\lambda > 0$ e parâmetro de forma $\gamma > 0$. As funções de sobrevivência, de risco e densidade de probabilidade são dadas por:

$$S(t) = \frac{1}{1 + \lambda t^\gamma} \quad (3.38)$$

$$h(t) = \frac{\lambda \gamma t^{\gamma-1}}{1 + \lambda t^\gamma} \quad (3.39)$$

$$f(t) = \frac{\lambda \gamma t^{\gamma-1}}{(1 + \lambda t^\gamma)^2} \quad (3.40)$$

Para esta família de distribuições o logaritmo da possibilidade de um indivíduo não sobreviver para além de t (insucesso) é uma função linear do logaritmo do tempo e é dado pela seguinte expressão:

$$\ln \left[\frac{1 - S(t)}{S(t)} \right] = \ln \lambda + \gamma \ln t \quad (3.41)$$

onde γ é o declive e $\ln \lambda$ é a ordenada na origem.

Recorrendo a uma análise do gráfico da estimativa desta função *versus* o logaritmo do tempo, é possível verificar os pressupostos do modelo de regressão log-logístico:

- Se o gráfico for razoavelmente linear, então pode assumir-se que o tempo de sobrevivência tem distribuição log-logística.
- Se as funções correspondentes a vários grupos forem razoavelmente paralelas, pode assumir-se que as possibilidades são proporcionais.
- Se é razoável aceitar ambos os pressupostos anteriores, então também é plausível aceitar o pressuposto de tempo de vida acelerado.

Como modelo de possibilidades proporcionais, a função de sobrevivência e de risco do modelo log-logístico são dadas, respetivamente, por:

$$S(t; \mathbf{x}) = \frac{1}{1 + \exp(\boldsymbol{\beta}'\mathbf{x})\lambda t^\gamma} = \frac{1}{1 + \exp(\boldsymbol{\beta}'\mathbf{x} + \ln \lambda + \gamma \ln t)} \quad (3.42)$$

$$h(t; \mathbf{x}) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x})\lambda \gamma t^{\gamma-1}}{1 + \exp(\boldsymbol{\beta}'\mathbf{x})\lambda t^\gamma} \quad (3.43)$$

Através da função de sobrevivência pode obter-se facilmente a possibilidade de sobrevivência. Assim, o modelo de regressão log-logístico escrito sob a forma de modelo de possibilidades proporcionais, é dado por:

$$\frac{S(t; \mathbf{x})}{1 - S(t; \mathbf{x})} = \exp(-\boldsymbol{\beta}'\mathbf{x} - \ln \lambda - \ln t) = e^{-\boldsymbol{\beta}'\mathbf{x}} \frac{S_0(t)}{1 - S_0(t)} \quad (3.44)$$

À semelhança dos restantes modelos de tempo de vida acelerado, o modelo log-logístico pode ser escrito na forma log-linear:

$$\ln T = \mu + \boldsymbol{\alpha}'\mathbf{x} + \sigma \varepsilon \quad (3.45)$$

Uma vez que o tempo de vida segue uma distribuição log-logística, a variável aleatória ε segue uma distribuição logística. Assim, a função de sobrevivência do modelo log-logístico de um indivíduo com vetor de covariáveis \mathbf{x} é dada por:

$$S(t; \mathbf{x}) = \frac{1}{1 + \exp\left(\frac{\ln t - \mu - \boldsymbol{\alpha}'\mathbf{x}}{\sigma}\right)} \quad (3.46)$$

Através da comparação das funções de sobrevivência obtidas a partir do modelo de possibilidades proporcionais e da representação log-linear do modelo de tempo de vida acelerado, é possível chegar às seguintes relações entre os parâmetros:

$$\lambda = \exp\left(-\frac{\mu}{\sigma}\right) \quad \gamma = \frac{1}{\sigma} \quad \beta_j = -\frac{\alpha_j}{\sigma} \quad (3.47)$$

Note-se que a distribuição log-logística e a distribuição log-normal apresentam formas idênticas produzindo resultados muito semelhantes. No entanto, o modelo log-normal apesar de ser um modelo do tempo de vida acelerado, contrariamente ao modelo log-logístico, não é um modelo de possibilidades proporcionais.

Com base no *odds* de insucesso, é ainda possível comparar indivíduos que diferem apenas no valor de uma covariável, através do *odds ratio*. Para tal, considere-se dois indivíduos com vetores de p covariáveis \mathbf{x}_1 e \mathbf{x}_2 , respetivamente, que diferem apenas no valor da variável x_j . O *odds ratio* é dado pela razão das respetivas *odds* de insucesso:

$$\frac{\frac{1 - S(t; \mathbf{x}_1)}{S(t; \mathbf{x}_1)}}{\frac{1 - S(t; \mathbf{x}_2)}{S(t; \mathbf{x}_2)}} = \frac{\exp(\beta_1 x_{11} + \dots + \beta_j x_{1j} + \dots + \beta_p x_{1p})}{\exp(\beta_1 x_{21} + \dots + \beta_j x_{2j} + \dots + \beta_p x_{2p})} = \exp[\beta_j (x_{1j} - x_{2j})] \quad (3.48)$$

Desta forma, $\exp(\beta_j)$ representa o *odds ratio* de ocorrência do acontecimento até ao instante t , para dois indivíduos que diferem de uma unidade no valor da covariável x_j , sendo iguais os valores das restantes covariáveis.

3.5 Modelos Flexíveis de Royston e Parmar

Apesar da utilização de modelos paramétricos ter vantagens comparativamente ao modelo de Cox, por vezes os modelos paramétricos mais simples podem não ser suficientemente flexíveis para representar adequadamente a função de risco associada a um grupo de indivíduos. Este facto contribuiu para a popularidade do modelo de Cox, cuja função de risco subjacente não necessita de ser definida, tal como mencionado anteriormente, pelo que não influencia as estimativas dos riscos relativos.

Os modelos propostos por Royston e Parmar (2002), por sua vez, são generalizações dos modelos paramétricos anteriormente referidos, que introduzem maior flexibilidade relativamente à forma da distribuição de sobrevivência que podem modelar.

Para obter modelos mais flexíveis a abordagem seguida pelos autores consiste em modelar uma transformação de $S(t)$ como uma função *spline* cúbica restrita do logaritmo do tempo, em vez de simplesmente uma função linear do logaritmo do tempo:

$$g[S(t; \mathbf{x})] = s(\ln t, \gamma) + \boldsymbol{\beta}' \mathbf{x} \quad (3.49)$$

Um *spline* cúbico natural ou restrito é um *spline* definido por funções polinomiais cúbicas, que é restrito à linearidade para além dos pontos que o delimitam, isto é, para além dos nós limite (k_{min} e $k_{máx}$). Os nós limite são geralmente, mas não necessariamente, determinados pelo primeiro e pelo último tempo de vida observado. Além destes nós, existem m nós internos (k_1, \dots, k_m com $k_1 > k_{min}$ e $k_m < k_{máx}$) cujo número e localização pode ser definido. É de notar, no entanto que, segundo os autores, desde que existam nós suficientes o aumento do número de nós e a sua localização não é particularmente importante, uma vez que não altera a curva ajustada.

Por forma a ajustar um *spline* cúbico natural para uma variável z , é necessário criar novas variáveis $v_1(z) \dots v_m(z)$. Estas novas variáveis são transformações de z , obtidas a partir dos nós internos $(k_j, j = 1 \dots m)$ em que:

$$v_j(z) = (z - k_j)_+^3 - \lambda_j(z - k_{min})_+^3 - (1 - \lambda_j)(z - k_{máx})_+^3 \quad (3.50)$$

onde $\lambda_j = \frac{k_{máx} - k_j}{k_{máx} - k_{min}}$ e $(z - a)_+ = \max(0, z - a)$.

Assim, o *spline* cúbico natural pode ser escrito sob a forma de uma função de parâmetros $\boldsymbol{\gamma}$ e variáveis $v_1(z) \dots v_m(z)$:

$$s(z, \boldsymbol{\gamma}) = \gamma_0 + \gamma_1 z + \gamma_2 v_1(z) + \dots + \gamma_{m+1} v_m(z) \quad (3.51)$$

A complexidade da curva depende do número de nós incluídos e consequentemente do número de graus de liberdade ($df = m + 1$). Os autores aconselham a utilização de modelos com $m < 4$, uma vez que curvas com mais nós internos podem tornar-se instáveis.

Além disso, Royston e Lambert (2011) recomendam usar modelos com *spline* cúbico do logaritmo do tempo, em vez da escala não transformada do tempo, uma vez que resulta numa menor variação entre curvas com diferentes números de nós internos.

A posição dos nós internos pode ser obtida com base nos percentis empíricos da distribuição do logaritmo dos tempos observados, conforme tabela abaixo:

Tabela 3. 1: Posição dos nós internos em modelos *spline*, para diferentes graus de liberdade.

Nós internos	<i>d.f.</i>	Percentis		
1	2	50		
2	3	33	67	
3	4	25	50	75

Os modelos flexíveis são extensões dos modelos de Weibull e log-logístico, sendo o efeito das covariáveis proporcional na escala adequada (*hazard* e *odds*, respetivamente).

3.5.1 Modelo flexível de riscos proporcionais

Por forma a obter um modelo flexível de riscos proporcionais, a abordagem proposta por Royston e Parmar (2002) é modelar o logaritmo da função de risco cumulativa subjacente, como uma função *spline* cúbica do logaritmo do tempo:

$$\ln H(t; \mathbf{x}) = \ln H_0(t) + \boldsymbol{\beta}' \mathbf{x} = s(\ln t; \boldsymbol{\gamma}) + \boldsymbol{\beta}' \mathbf{x} \quad (3.52)$$

Sendo o correspondente logaritmo da função de risco dado por:

$$\ln h(t; \mathbf{x}) = \ln \left\{ \frac{d}{dt} s(\ln t; \boldsymbol{\gamma}) \right\} + s(\ln t; \boldsymbol{\gamma}) + \boldsymbol{\beta}' \mathbf{x} \quad (3.53)$$

Quando não são considerados nós internos no *spline* cúbico, ($m = 0$ e $df = 1$), a distribuição de base é a distribuição de Weibull:

$$\ln H(t) = \gamma_0 + \gamma_1 \ln t \quad (3.54)$$

Comparando os logaritmos da função de risco cumulativa, obtidos a partir do modelo flexível com $m = 0$ e do modelo de riscos proporcionais, é possível chegar às seguintes relações entre os parâmetros:

$$\gamma_0 = \ln \lambda \quad e \quad \gamma_1 = \gamma \quad (3.55)$$

3.5.2 Modelo flexível de possibilidades proporcionais

Analogamente, para obter um modelo flexível de possibilidades proporcionais, a abordagem proposta pelos autores consiste em modelar o logaritmo do *odds* de insucesso subjacente, como uma função *spline* cubica do logaritmo do tempo:

$$\ln \left(\frac{1 - S(t; \mathbf{x})}{S(t; \mathbf{x})} \right) = \ln \left(\frac{1 - S_0(t; \mathbf{x})}{S_0(t; \mathbf{x})} \right) + \boldsymbol{\beta}' \mathbf{x} = s(\ln t; \gamma) + \boldsymbol{\beta}' \mathbf{x} \quad (3.56)$$

Sendo a correspondente função de sobrevivência dada por:

$$S(t; \mathbf{x}) = \frac{1}{1 + \exp(s(\ln t; \gamma) + \boldsymbol{\beta}' \mathbf{x})} \quad (3.57)$$

Quando não são considerados nós internos no *spline* cúbico, ($m = 0$ e $df = 1$), a distribuição de base é a distribuição log-logística:

$$S(t) = \frac{1}{1 + \exp(\gamma_0 + \gamma_1 \ln t + \boldsymbol{\beta}' \mathbf{x})} \quad (3.58)$$

Comparando os logaritmos da função de risco cumulativa, obtidos a partir do modelo flexível com $m = 0$ e do modelo de possibilidades proporcionais, é possível chegar às seguintes relações entre os parâmetros:

$$\gamma_0 = \ln \lambda \quad e \quad \gamma_1 = \gamma \quad (3.59)$$

3.5.3 Ajustamento do modelo

À semelhança de outros modelos de regressão, a seleção de variáveis a incluir no modelo pode ser feita através de um dos vários métodos disponíveis (por exemplo método *stepwise*, entre outros).

Segundo Royston e Parmar (2002), é de esperar que as estimativas dos coeficientes sejam semelhantes entre o modelo de Weibull, o modelo de Cox e os modelos flexíveis de riscos proporcionais. De forma análoga, as estimativas dos coeficientes obtidas no modelo log-logístico e nos modelos flexíveis de possibilidades proporcionais deverão ser semelhantes.

Assim, as variáveis a incluir nos modelos flexíveis de riscos proporcionais e flexíveis de possibilidades proporcionais, podem ser selecionadas através dos modelos de Weibull ou de Cox e log-logístico, respetivamente.

É ainda sugerido utilizar o valor de AIC como fonte de informação para a escolha do número de nós internos a incluir no modelo flexível.

3.5.4 Diagnóstico do modelo

Para avaliar a qualidade do ajustamento do preditor linear do modelo flexível, pode realizar-se uma análise dos resíduos. Para tal, Royston e Lambert (2011) sugerem usar um tipo de resíduo idêntico ao resíduo martingala. Assim, o resíduo para um determinado indivíduo i é dado por:

$$r_i = \delta_i - \hat{H}_i(t_i) \quad (3.60)$$

onde δ_i é o indicador de censura e $\hat{H}_i(t_i)$ a estimativa da função de risco cumulativa no tempo de censura ou morte do indivíduo (t_i).

Se o modelo estiver bem ajustado então $E(r_i | \mathbf{x}_i \hat{\beta}) = 0$, além disso para qualquer covariável x incluída no modelo $E(r_i | x_i) = 0$. Para uma melhor interpretação dos resíduos, é ainda recomendado pelos autores recorrer à representação gráfica destes resíduos *versus* cada covariável x , incluindo uma curva de suavização. Caso não exista um padrão de afastamento sistemático da curva de suavização em relação ao eixo horizontal ($y = 0$), o modelo apresenta um bom ajustamento.

Capítulo 4

Resultados

Foi realizada uma análise descritiva dos valores de cada variável independente, por forma a descrever e resumir as características inerentes da amostra em estudo.

Do total de indivíduos em estudo ($n=360$), verificou-se que a maioria (63,3%) tinha um nível de escolaridade básico, em geral viviam acompanhados (77,7%) e 23,1% dos indivíduos encontravam-se desempregados. Relativamente à idade dos indivíduos, verificou-se que a idade mínima e máxima foi de 50 e 87 anos, respetivamente. Após categorizada a variável idade, constatou-se que 35,3% dos indivíduos em estudo tinha idade igual ou inferior a 60 anos e 27,8% tinha idade superior a 70 anos.

Os indivíduos foram também questionados relativamente a alguns dos seus hábitos e cuidados com a saúde. Cerca de metade (50,6%) referiu não praticar qualquer tipo de exercício físico e 61,7% referiu ir ao médico mais de uma vez a cada seis meses. Apenas 57,8% dos indivíduos referiu ter tido conhecimento por parte do médico, de que tinha a doença crónica e a maioria (77,8%) diz ter realizado um exame de diagnóstico.

O tipo de terapêutica predominante (67,2%) na amostra em estudo é a toma semanal, sendo que, para apenas 22,2% dos indivíduos o médico prescriptor da terapêutica medicamentosa, era um médico da especialidade. Foi também registado para todos os indivíduos, se estes tinham ou não mais alguma doença crónica, tendo a maioria (68,1%) referido que sim e se estariam ou não a tomar outros medicamentos, além da medicação para a doença em análise, sendo que 86,1% da amostra afirmou estar a tomar mais medicamentos.

No final do estudo, foi observada a ocorrência do acontecimento em 76,4% dos indivíduos, isto é, 275 indivíduos interromperam a toma do medicamento antes do final do período de seguimento, tornando-se não persistentes. Dos 85 indivíduos para os quais não se observou a não persistência, 6 foram perdidos para *follow-up*, permitindo-nos afirmar que 79 indivíduos se mantiveram persistentes até ao final do estudo.

4.1 Modelo de regressão logística

Para proceder à análise de regressão logística, a variável resposta foi a Persistência. Esta é uma variável binária, que toma o valor zero ($Y = 0$) caso o indivíduo seja “Persistente” no final do estudo, ou o valor um ($Y = 1$) caso o indivíduo seja “Não Persistente” no final do estudo.

Além de ser importante analisar a distribuição dos valores de cada variável independente, através de uma análise descritiva univariada, é também pertinente averiguar se há evidência de associação com a variável resposta, através de uma análise bivariada. Para tal, foram criadas as respetivas tabelas de contingência e aplicado o teste de qui-quadrado de independência a cada uma.

As tabelas de contingência criadas encontram-se em anexo (Anexo A), sendo, no entanto, importante referir a seu respeito, que não foram obtidas células em branco, nem células com frequência esperada inferior a 5.

Tabela 4. 1: Testes do qui-quadrado de independência

Variável	χ^2	valor p
Idade	9,1963	0,0101
Nível de escolaridade	0,8378	0,6578
Situações de emprego	0,9871	0,3204
Vive sozinho	7,6941	0,0055
Exercício físico	0,1581	0,6909
Ida ao médico	0,0421	0,8374
Doença crónica	0,3057	0,5803
Exame	0,8135	0,3671
Especialidade do médico	0,1045	0,7464
Outra doença crónica	0,5584	0,4549
Outros medicamentos	0,2939	0,5877
Terapêutica	2,3127	0,1283

A Tabela 4.1 apresenta o resultado do teste de qui-quadrado de independência para cada covariável. Através dos resultados desta análise, quando cada variável é considerada isoladamente, pode afirmar-se que a idade e o facto de o doente viver ou não sozinho, são as únicas variáveis que revelam influência significativa na não persistência dos doentes no tratamento.

A Figura 4.1 apresenta a distribuição dos indivíduos por estado de persistência e segundo as classes das variáveis Idade e Vive sozinho. Como se pode ver pela Figura 4.1 A quando comparadas as classes etárias, é na classe dos 60 aos 70 anos que se encontra a maior percentagem de indivíduos persistentes no final do estudo (30,1%). Da mesma forma, é para idades acima de 70 anos que se verifica a maior percentagem de não persistência (86,0%), por classe etária.

No caso dos doentes que vivem acompanhados (Figura 4.1 B), cerca de 25,4% manteve-se persistente até ao final do seguimento; por outro lado é no grupo de doentes que vivem sozinhos que se verifica a maior percentagem de não persistência no final do seguimento (90%).

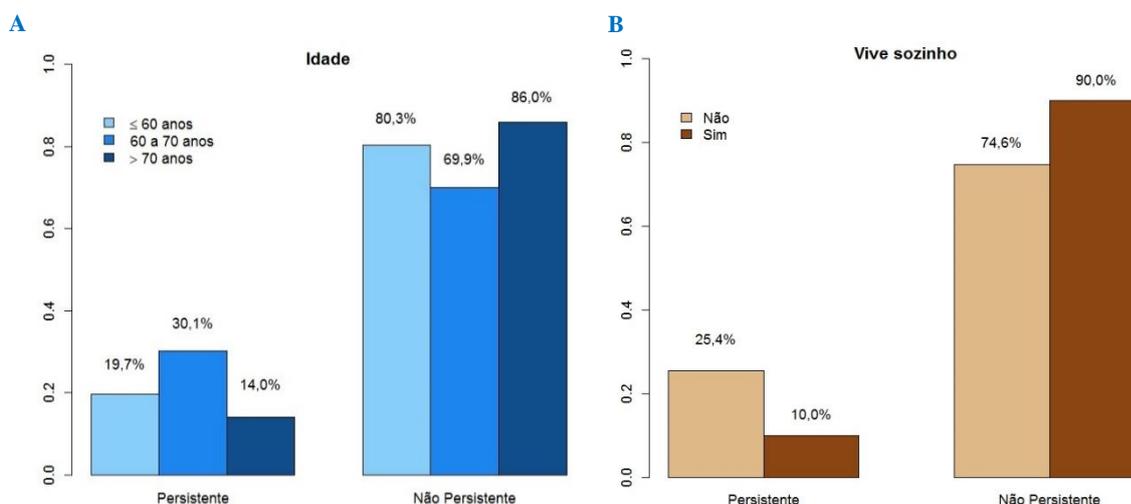


Figura 4. 1: Distribuição dos indivíduos, por estado de persistência no final do estudo segundo a classe de Idade (A) e segundo vive ou não sozinho (B).

4.1.1 Ajustamento do modelo logístico

Na presença de várias variáveis independentes que podem potencialmente ser incluídas no modelo logístico, o primeiro passo do ajustamento é a seleção de variáveis que influenciam a não persistência dos doentes. Para tal começou-se por ajustar modelos com uma covariável, sendo de seguida estimado para cada coeficiente o *odds ratio* (OR) e o respetivo intervalo de confiança a 95%. Recorreu-se também ao teste de Wald e ao teste da razão de verosimilhanças, por forma a verificar a influência de cada covariável na não persistência. Os resultados obtidos encontram-se na Tabela 4.2 e na Tabela 4.3.

Tabela 4. 2: Análise univariada: *odds ratio* e valor p dos testes de Wald

Variável		OR[IC95%]	valor p
Idade	≤ 60 anos	Ref.	-
	Entre 60 e 70 anos	0,570 [0,321 ; 1,011]	0,0545
	> 70 anos	1,506 [0,737 ; 3,076]	0,2616
Nível de escolaridade	Sem ensino	Ref.	-
	Básico	1,386 [0,646 ; 2,971]	0,4019
	Secundário ou Superior	1,457 [0,618 ; 3,438]	0,3897
Situação de emprego	Desempregado	Ref.	-
	Empregado/Reformado	1,392 [0,789 ; 2,456]	0,2540
Vive sozinho	Não	Ref.	-
	Sim	3,057 [1,404 ; 6,659]	0,0049
Exercício físico	Não	Ref.	-
	Sim	1,143 [0,693 ; 1,885]	0,6000
Ida ao médico	1 vez ou menos por semestre	Ref.	-
	Mais de uma vez por semestre	0,915 [0,546 ; 1,534]	0,7370
Doença crónica	Não	Ref.	-
	Sim	1,191 [0,721 ; 1,968]	0,4960
Exame	Não	Ref.	-
	Sim	1,362 [0,766 ; 2,423]	0,2930
Especialidade do médico	Médico da especialidade	Ref.	-
	Outra	0,861 [0,465 ; 1,593]	0,6340
Outra doença crónica	Não	Ref.	-
	Sim	0,780 [0,450 ; 1,354]	0,3770
Outros medicamentos	Não	Ref.	-
	Sim	0,753 [0,349 ; 1,624]	0,4690
Terapêutica	Mensal	Ref.	-
	Semanal	1,544 [0,922 ; 2,586]	0,0990

Tabela 4. 3: Análise univariada: Testes da razão de verossimilhanças.

Variável	$-2\ln[\hat{L}_{nulo}/\hat{L}_{ajustado}]$	valor p
Idade	9,23	0,0099
Nível de escolaridade	0,80	0,6700
Situações de emprego	1,27	0,2598
Vive sozinho	9,77	0,0018
Exercício físico	0,28	0,5995
Ida ao médico	0,11	0,7363
Doença crónica	0,46	0,4963
Exame	1,08	0,2987
Especialidade do médico	0,23	0,6310
Outra doença crónica	0,79	0,3725
Outros medicamentos	0,55	0,4589
Terapêutica	2,68	0,1017

Segundo Hosmer, Lemeshow e Sturdivant (2013), as variáveis independentes cujo valor p do teste da razão de verossimilhanças for inferior a 0,25, são variáveis candidatas para inclusão no modelo múltiplo. Assim sendo, são consideradas variáveis candidatas a Idade (valor $p=0,0099$), Vive sozinho (valor $p=0,0018$) e a Terapêutica (valor $p=0,1017$).

Após completar a análise univariada, segue-se a seleção e inclusão das variáveis no modelo múltiplo, através do critério de Akaike e do método *stepwise* de inclusão progressiva com eliminação regressiva. O ponto de partida é o modelo nulo, isto é, o modelo que inclui apenas o parâmetro constante (β_0).

Visto que os modelos com uma covariável já foram ajustados anteriormente, a seleção inicia-se com a comparação através do teste da razão de verossimilhanças, de cada um destes modelos com o modelo nulo. Tendo o modelo que inclui a variável Vive sozinho, obtido o menor valor de AIC (373,10) e o menor valor p (0,0018), esta será a primeira variável a ser incluída no modelo múltiplo.

No passo seguinte, foi verificado se mais alguma variável seria importante, tendo em conta que a variável Vive sozinho já se encontrava no modelo. Para tal, foram ajustados modelos incluindo a variável Vive sozinho e cada uma das restantes, sendo de seguida comparados, através do teste da razão de verossimilhanças, com o modelo anteriormente obtido. A Idade é a variável cuja inclusão no modelo origina o menor valor de AIC (368,62) e o menor valor p (0,0144). Visto este valor p ser inferior ao p_E e o valor de AIC ser menor do que o valor do modelo anterior, a variável foi incluída no modelo múltiplo.

Como se pode ver na Tabela 4.4, ao comparar o modelo obtido neste passo da seleção de variáveis com o modelo nulo, a estatística de teste tem um valor $G=18,25$ e valor $p=0,0004$, o que indica uma diferença significativa entre os modelos e como tal, o modelo ajustado neste passo fornece-nos mais informação.

Prosseguindo a análise, foram ajustados os modelos que incluem as variáveis selecionadas e cada uma das restantes variáveis. Constatou-se que o modelo que inclui a Terapêutica produz o menor valor de AIC e valor p inferior ao p_E (valor $p=0,0962$), sendo por isso a variável incluída no modelo. No passo seguinte, verificou-se que nenhuma das restantes variáveis diminuía o valor de AIC ao ser introduzida no modelo. O modelo múltiplo final inclui as variáveis Vive sozinho, Idade e Terapêutica, com um valor de AIC de 367,85.

Tabela 4. 4: Valor de AIC e Teste da razão de verosimilhanças, em cada passo da seleção de variáveis.

Modelo	Residual deviance	AIC	Teste da razão de verosimilhanças	
			Estatística G	valor p
Nulo	378,87	380,87	-	-
Vive sozinho	369,10	373,10	9,77 ⁽¹⁾	0,0018 ⁽¹⁾
Vive sozinho + Idade	360,62	368,62	18,25 ⁽¹⁾	0,0004 ⁽¹⁾
			8,48 ⁽²⁾	0,0144 ⁽²⁾
Vive sozinho + Idade + Terapêutica	357,85	367,85	21,02 ⁽¹⁾	0,0003 ⁽¹⁾
			2,77 ⁽³⁾	0,0962 ⁽³⁾

⁽¹⁾ Comparação face ao modelo nulo; ⁽²⁾ Face ao modelo Vive sozinho; ⁽³⁾ Face ao modelo Vive sozinho + Idade

4.1.2 Interpretação dos coeficientes do modelo final

Os resultados correspondentes ao modelo múltiplo final encontram-se na Tabela 4.5. É de notar que as estimativas dos coeficientes são positivas com exceção da estimativa correspondente à classe etária dos 60 aos 70 anos.

Tabela 4. 5: Modelo múltiplo final: estimativa de cada β_j , respetivo teste de Wald e *odds ratio*.

Modelo final	$\hat{\beta}_j$	valor p	$\widehat{OR}[IC95\%]$
Constante	1,062	8,59e-05	-
Vive sozinho	1,156	0,0048	3,176 [1,423 ; 7,088]
Idade – “60 a 70 anos”	-0,728	0,0156	0,483 [0,268 ; 0,871]
Idade – “> 70 anos”	0,086	0,8197	1,090 [0,519 ; 2,292]
Terapêutica	0,458	0,0942	1,581 [0,925 ; 2,704]

O coeficiente da covariável que indica se o doente vive ou não sozinho foi estimado em $\hat{\beta}_1 = 1,156$, o que leva a um valor de $\widehat{OR} = 3,176$. Este valor de *odds ratio* significa que um indivíduo que vive sozinho tem 3 vezes mais possibilidade de se tornar não persistente do que um indivíduo que não vive sozinho.

À covariável Idade correspondem dois coeficientes, um para cada variável *dummy*. O primeiro coeficiente foi estimado em $\hat{\beta}_2 = -0,728$, o que origina $\widehat{OR} = 0,483$, pelo que um doente com idade compreendida entre os 60 e os 70 anos tem metade da possibilidade de descontinuar a medicação do que um doente com 60 anos ou menos. A estimativa do coeficiente da segunda variável *dummy*, $\hat{\beta}_3 = 0,086$, leva a um valor de $\widehat{OR} = 1,090$. Este valor sugere que a possibilidade de um indivíduo com mais de 70 anos se tornar não persistente, é idêntica à de um indivíduo com 60 anos ou menos. Além disso, note-se que o intervalo de confiança a 95% contém o valor 1, pelo que não há evidência de que as possibilidades sejam diferentes entre as duas faixas etárias.

Para a covariável Terapêutica, o coeficiente foi estimado em $\hat{\beta}_4 = 0,458$, o que leva a $\widehat{OR} = 1,581$. Este valor significa que um indivíduo que se encontra a fazer uma terapêutica semanal tem mais 58% de possibilidade de interromper o tratamento, do que um indivíduo que está a fazer uma terapêutica mensal. No entanto com 95% de confiança não se pode concluir que a possibilidade é diferente, uma vez que o intervalo de confiança contém o valor 1.

4.1.3 Diagnóstico do modelo

Tanto quanto sabemos, o modelo múltiplo obtido é o mais parcimonioso e as variáveis incluídas influenciam a não persistência dos doentes no tratamento da doença crónica. No entanto, é necessário avaliar a qualidade de ajustamento do modelo e a sua capacidade de discriminação dos indivíduos em cada um dos grupos definidos pela variável resposta.

Tal como se pode ver pela Tabela 4.4, a diferença entre os valores da *deviance* do modelo final e do modelo nulo é de 21,02, o que no teste da razão de verosimilhanças corresponde a um valor $p=0,0003$, existindo assim evidência estatisticamente significativa de que o modelo ajustado fornece mais informação do que o modelo nulo.

O teste de Hosmer-Lemeshow e os testes aos resíduos de Pearson e aos resíduos da *deviance* permitem tirar conclusões relativamente à qualidade de ajustamento do modelo final.

Na Tabela 4.6 são apresentados os valores da estatística de diagnóstico destes testes, bem como o respetivo valor p . Em todos os testes os valores p obtidos são elevados, indicando que o ajustamento global do modelo é bom.

Tabela 4. 6: Testes para avaliação da qualidade do ajustamento.

Teste	Estatística de teste	valor p
Resíduos de Pearson	362,20	0,4135
Resíduos da <i>deviance</i>	357,85	0,4774
Hosmer-Lemeshow	2,84	0,9440

Em seguida, por forma a avaliar a capacidade discriminatória do modelo recorreu-se a outro indicador, a curva ROC, mais especificamente à área abaixo da curva. Para o modelo ajustado, a área obtida e respetivo intervalo de confiança foi de 0,66 [0,59 ; 0,73]. Este valor significa que apesar de existir discriminação entre os indivíduos persistentes e não persistentes, esta é bastante fraca.

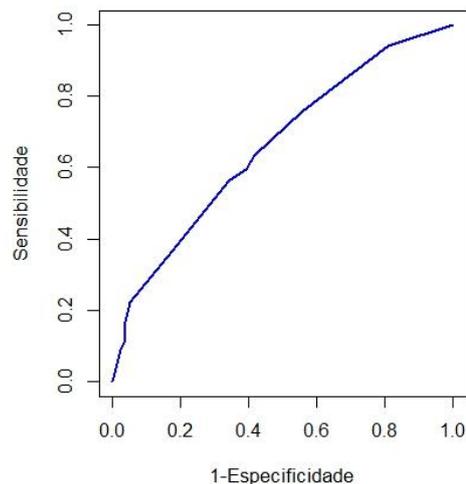


Figura 4. 2: Curva ROC.

4.2 Análise de sobrevivência

Nesta secção pretende-se identificar os fatores que influenciam o tempo até à não persistência de indivíduos no tratamento de uma dada doença crónica. A forma como estes fatores influenciam o risco de interrupção indevida do tratamento é também quantificada e analisada. À semelhança da regressão logística, as covariáveis consideradas incluem os fatores sociodemográficos e de saúde recolhidos aquando do recrutamento.

4.2.1 Estimação não paramétrica

O período durante o qual o estudo decorreu foi de 18 meses (545 dias). Pelo Figura 4.3, pode verificar-se que a probabilidade de um indivíduo sobreviver, isto é, a probabilidade de um doente se manter persistente para além do 525º dia (tempo do último acontecimento observado), foi estimada em $\hat{S}(525) = 0,226$. Este valor indica que, sem informação de potenciais fatores que influenciem o tempo até à não persistência dos indivíduos, qualquer doente tem uma probabilidade estimada de 0,226, de se manter persistente no tratamento da doença crónica por mais de 525 dias. É de notar ainda que $\hat{S}(t) = 0,226$ para $525 < t < 545$, não sendo possível obter uma estimativa da probabilidade de persistência para além de 545 dias (18 meses).

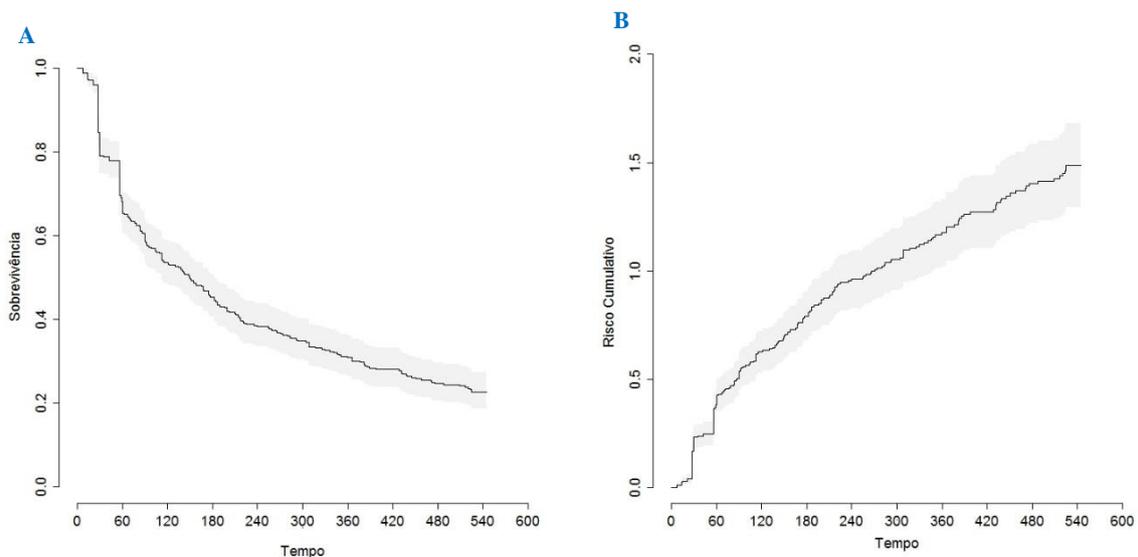


Figura 4. 3: Curva de sobrevivência (A) e estimativa da função de risco cumulativa (B).

Através da estimação dos quartis do tempo até à não persistência (Tabela 4.7), estima-se que após o 56º dia, pelo menos 25% da população interrompa a toma da medicação. Por sua vez, a mediana do tempo até à não persistência foi estimada em 149 dias, ou seja, após cerca de 5 meses metade dos doentes tornar-se-á não persistente. A estimação do 3º quartil mostra que, até 472 dias 75% dos indivíduos interrompe indevidamente a toma da medicação.

Tabela 4. 7: Estimação dos quartis do tempo de persistência e respetivo IC 95%.

Estimativas dos quartis	[IC 95%]
$\hat{\chi}_{0,25} = 56$	[35 ; 56]
$\hat{\chi}_{0,50} = 149$	[112 ; 182]
$\hat{\chi}_{0,75} = 472$	[382 ; $+\infty$ [

Posteriormente, obteve-se a estimativa da função de sobrevivência para cada grupo definido pelos valores de cada variável e recorreu-se ao teste log-rank e ao teste de Peto-Peto, para avaliar a existência de diferenças significativas entre os grupos, no que diz respeito ao tempo até à não persistência (Tabela 4.8).

Tabela 4. 8: Comparação das funções de sobrevivência: valor p dos testes log-rank e de Peto-Peto.

Variável	Teste log-rank	Teste de Peto-Peto
Idade	0,0127	0,0530
Nível de escolaridade	0,2634	0,1150
Situação de emprego	0,1370	0,0899
Vive sozinho	0,0081	0,0717
Exercício físico	0,6970	0,9026
Ida ao médico	0,6204	0,5061
Doença crónica	0,5665	0,7916
Exame	0,7503	0,6364
Especialidade do médico	0,5016	0,1351
Outra doença crónica	0,8231	0,8056
Outros medicamentos	0,7518	0,2628
Terapêutica	0,0324	0,0049

Para um nível de significância $\alpha = 0,05$, pelo teste log-rank há evidência de diferenças estatisticamente significativas entre as funções de sobrevivência, para as variáveis Idade (valor $p=0,0127$), Vive sozinho (valor $p=0,0081$) e Terapêutica (valor $p=0,0324$). Pelo teste de Peto-Peto apenas se verificaram diferenças significativas entre os grupos definidos por estas variáveis, quando considerado $\alpha = 0,1$. É razoável esperar que estas variáveis tenham influência sobre o tempo até à não persistência e que sejam incluídas nos modelos de sobrevivência.

Como se pode verificar pelo Gráfico A da Figura 4.4, até aos 8 meses de toma (240 dias), as curvas dos grupos “ ≤ 60 anos” e “ > 70 anos” estão sobrepostas, enquanto a curva do grupo “60 a 70 anos” se mantém acima das anteriores. Após os 8 meses, verifica-se um afastamento das duas primeiras curvas que se mantêm paralelas até ao final do estudo e um afastamento gradual da curva de sobrevivência do grupo “60 a 70 anos”, sugerindo um melhor prognóstico de persistência desta faixa etária face às restantes, à medida que o tempo passa.

A Tabela 4.9 contém as estimativas dos quartis do tempo de vida para cada categoria das variáveis consideradas e os respetivos intervalos de 95% de confiança. Para a variável Idade, a estimativa mais elevada da mediana do tempo de vida foi obtida para a categoria “60 a 70 anos”, sendo aproximadamente 6 meses ($\hat{\chi}_{0,5} = 185$). Por sua vez, para as categorias “ ≤ 60 anos” e “ > 70 anos” estima-se que metade dos indivíduos interrompa o tratamento por volta do 4º mês ($\hat{\chi}_{0,5} = 122$, $\hat{\chi}_{0,5} =$

112 respectivamente). Ademais, para a faixa etária “60 a 70 anos” não foi possível estimar o 3º quartil, uma vez que no final do período de seguimento mais de 25% dos doentes, com idade compreendida nesta faixa, se manteve persistente. Este conjunto de resultados também sugere que indivíduos com idade entre 60 e 70 anos têm um melhor prognóstico de persistência do que os restantes indivíduos.

Para a variável Vive sozinho (Gráfico B da Figura 4.4), as curvas de sobrevivência coincidem nos primeiros 3 meses (90 dias), observando-se um afastamento gradual nos meses seguintes. Note-se que a curva correspondente aos doentes que vivem acompanhados está sempre acima da curva dos doentes que vivem sozinhos, sugerindo um melhor prognóstico de persistência para os doentes que vivem acompanhados.

Estima-se que até ao 56º dia, 25% dos indivíduos da população venha a interromper o tratamento medicamentoso, independentemente de viver sozinho ou acompanhado (Tabela 4.9). A estimação da mediana do tempo de vida, mostra a diferença na persistência entre os dois grupos, sendo estimada em $\hat{\chi}_{0,5} = 152$ para o grupo que vive acompanhado e em $\hat{\chi}_{0,5} = 130$ para o grupo que vive sozinho. Além disso, estima-se que 75% dos indivíduos que vivem sozinhos interrompa a medicação por volta do 8º mês ($\hat{\chi}_{0,75} = 240$). No entanto, não é possível estimar este quartil para o grupo de indivíduos que vivem acompanhados, uma vez que no fim do seguimento mais de 25% dos indivíduos ainda se mantinha persistente. Estes resultados voltam a sugerir que indivíduos que vivem acompanhados têm um melhor prognóstico de persistência no tratamento, do que os indivíduos que vivem sozinhos.

O Gráfico C da Figura 4.4 mostra um decréscimo acentuado na persistência, em ambos os grupos terapêuticos, entre o início do seguimento e o 2º mês, a partir do qual a curva da terapêutica mensal se mantém constantemente acima da curva da terapêutica semanal.

Como se pode ver pela Tabela 4.9, a mediana do grupo com terapêutica semanal foi estimada em $\hat{\chi}_{0,5} = 122$ (cerca de 4 meses) e o 3º quartil em $\hat{\chi}_{0,75} = 431$ (cerca de 14 meses). Para o grupo com terapêutica mensal, a mediana foi estimada em $\hat{\chi}_{0,5} = 187$ (cerca de 6 meses); no entanto, visto no fim do seguimento mais de 25% dos indivíduos se manterem persistentes, não foi possível estimar o 3º quartil do tempo até à não persistência no tratamento. Os resultados obtidos sugerem que o grupo de doentes com terapêutica mensal tem um melhor prognóstico de persistência do que o grupo de doentes com terapêutica semanal.

Tabela 4. 9: Estimação dos quartis do tempo até à não persistência e respetivo IC 95% de cada categoria da variável Idade, Vive sozinho e Terapêutica.

Variável	$\hat{\chi}_{0,25}$ [IC95%]	$\hat{\chi}_{0,5}$ [IC95%]	$\hat{\chi}_{0,75}$ [IC95%]
Idade			
≤ 60 anos	56 [30 ; 60]	122 [90 ; 190]	398 [308 ; +∞ [
60 a 70 anos	56 [56 ; 84]	185 [143 ; 346]	-
> 70 anos	32,5 [28 ; 60]	112 [84 ; 196]	317 [254 ; 488]
Vive sozinho			
Não	56 [35 ; 59]	152 [112 ; 216]	-
Sim	56 [30 ; 61]	130 [86 ; 185]	240 [199 ; 432]
Terapêutica			
Mensal	61 [58 ; 90]	187 [143 ; 325]	-
Semanal	56 [28 ; 56]	122 [87 ; 169]	431 [309 ; +∞ [

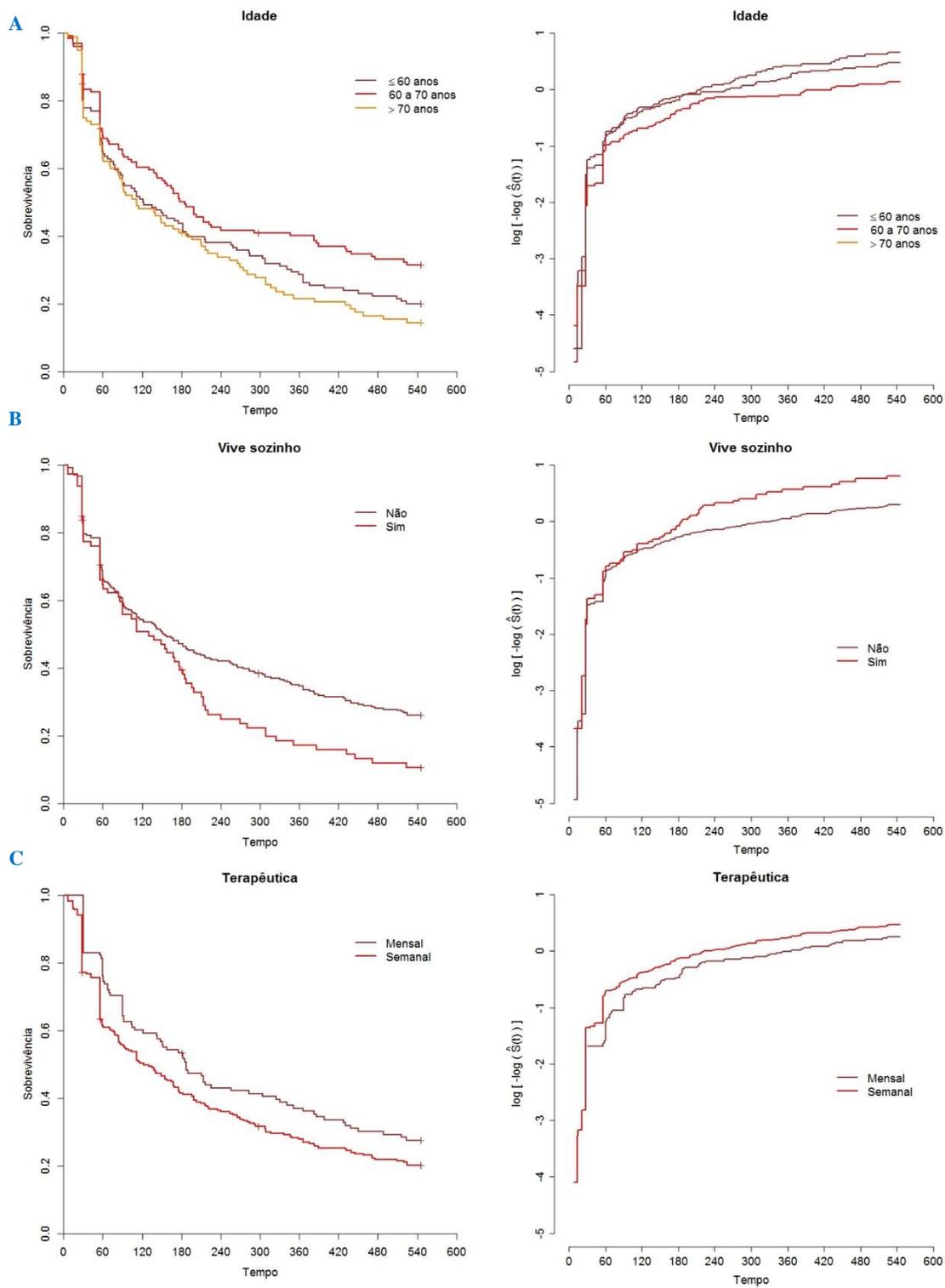


Figura 4. 4: Gráficos das estimativas $\hat{S}(t)$ (à esquerda) e $\ln[-\ln \hat{S}(t)]$ (à direita), para as variáveis Idade (A), Vive sozinho (B) e Terapêutica (C).

4.3 Modelo de Riscos Proporcionais de Cox

O modelo de Cox é um modelo de regressão amplamente utilizado na análise de sobrevivência. Na presença de um grande número de variáveis explicativas, este modelo pode ser usado para identificar o subconjunto de variáveis que influencia significativamente o tempo de sobrevivência dos indivíduos.

4.3.1 Ajustamento do modelo

À semelhança da regressão logística e de acordo com o método proposto por Collett (2003), a seleção de variáveis para o modelo de Cox inicia-se com o ajustamento de modelos com uma covariável. Os resultados obtidos encontram-se na Tabela 4.10.

Tabela 4. 10: Análise univariada: *hazard ratio* e valor p dos testes de Wald.

Variável		$\hat{\beta}_j$	$\widehat{HR}[IC95\%]$	valor p
Idade	≤ 60 anos	Ref.	-	-
	60 a 70 anos	-0,293	0,746 [0,561 ; 0,992]	0,0439
	> 70 anos	0,136	1,146 [0,859 ; 1,530]	0,3541
Nível de escolaridade	Sem ensino	Ref.	-	-
	Básico	0,216	1,241 [0,833 ; 1,849]	0,2890
	Secundário ou Superior	0,357	1,429 [0,922 ; 2,213]	0,1100
Situação de emprego	Desempregado	Ref.	-	-
	Empregado/Reformado	0,216	1,241 [0,934 ; 1,650]	0,1370
Vive sozinho	Não	Ref.	-	-
	Sim	0,362	1,436 [1,093 ; 1,886]	0,0094
Exercício físico	Não	Ref.	-	-
	Sim	0,047	1,048 [0,827 ; 1,328]	0,6970
Ida ao médico	1 vez ou menos por semestre	Ref.	-	-
	Mais de uma vez por semestre	-0,060	0,941 [0,739 ; 1,200]	0,6260
Doença crónica	Não	Ref.	-	-
	Sim	0,067	1,069 [0,840 ; 1,361]	0,5850
Exame	Não	Ref.	-	-
	Sim	0,039	1,040 [0,778 ; 1,390]	0,7920
Especialidade do médico	Médico da especialidade	Ref.	-	-
	Outra	0,100	1,106 [0,833 ; 1,467]	0,4870
Outra doença crónica	Não	Ref.	-	-
	Sim	-0,028	0,973 [0,757 ; 1,250]	0,8290
Outros medicamentos	Não	Ref.	-	-
	Sim	0,057	1,059 [0,760 ; 1,476]	0,7360
Terapêutica	Mensal	Ref.	-	-
	Semanal	0,288	1,334 [1,033 ; 1,723]	0,0272

Com o objetivo de encontrar as variáveis que podem influenciar o tempo até à não persistência, recorreu-se ao teste da razão de verosimilhanças, para comparar cada um dos modelos ajustados anteriormente com o modelo nulo.

Tabela 4. 11: Análise univariada: Testes da razão de verosimilhanças.

Variável	Estatística G	valor p
Idade	8,61	0,0135
Nível de escolaridade	2,72	0,2571
Situações de emprego	2,30	0,1293
Vive sozinho	6,35	0,0117
Exercício físico	0,15	0,6974
Ida ao médico	0,24	0,6266
Doença crónica	0,30	0,5844
Exame	0,07	0,7913
Especialidade do médico	0,49	0,4827
Outra doença crónica	0,05	0,8292
Outros medicamentos	0,12	0,7340
Terapêutica	5,05	0,0247

Como se pode ver pela Tabela 4.11, as únicas variáveis para as quais o valor p no teste da razão de verosimilhanças foi inferior a 0,1 foram: a Idade (valor p=0,0135), Vive sozinho (valor p=0,0117) e a Terapêutica (valor p=0,0247). De acordo com Collett (2003), estas variáveis são consideradas potencialmente importantes e devem ser todas incluídas num modelo múltiplo.

O segundo passo da seleção de variáveis é a exclusão das variáveis incluídas no modelo múltiplo, uma de cada vez, por forma a verificar se a omissão de alguma variável aumenta significativamente o valor da estatística $-2 \ln \hat{L}$. Tal como se pode ver pela Tabela 4.12, houve um aumento significativo da estatística $-2 \ln \hat{L}$ para todas as variáveis, pelo que nenhuma delas deixou de ser significativa na presença das restantes.

Tabela 4. 12: Comparação de modelos múltiplos: Testes da razão de verosimilhanças.

Modelo	Teste da razão de verosimilhanças	
	Estatística G	valor p
Idade + Terapêutica + Vive sozinho	-	-
Idade + Vive sozinho	4,77	0,0289
Idade + Terapêutica	6,81	0,0090
Terapêutica + Vive sozinho	7,70	0,0212

De seguida, é necessário verificar se, na presença das covariáveis já incluídas no modelo, alguma das variáveis que isoladamente não se revelou importante influencia agora significativamente o tempo até à não persistência. Os resultados encontram-se na Tabela 4.13 onde se pode ver que nenhuma inclusão de outra variável no modelo anterior diminuí significativamente o respetivo valor da estatística $-2 \ln \hat{L}$.

Tabela 4. 13: Comparação com o modelo obtido no passo anterior: Teste da razão de verossimilhanças.

Modelo	Teste da razão de verossimilhanças	
	Estatística G	valor p
Modelo múltiplo ⁽¹⁾	-	-
Modelo múltiplo ⁽¹⁾ + Nível de escolaridade	4,29	0,1173
Modelo múltiplo ⁽¹⁾ + Situações de emprego	1,64	0,2001
Modelo múltiplo ⁽¹⁾ + Exercício físico	0,001	0,9769
Modelo múltiplo ⁽¹⁾ + Ida ao médico	0,63	0,4289
Modelo múltiplo ⁽¹⁾ + Doença crónica	0,32	0,5733
Modelo múltiplo ⁽¹⁾ + Exame	0,07	0,7948
Modelo múltiplo ⁽¹⁾ + Especialidade do médico	0,49	0,4835
Modelo múltiplo ⁽¹⁾ + Outra doença crónica	0,45	0,5044
Modelo múltiplo ⁽¹⁾ + Outros medicamentos	0,03	0,8528

⁽¹⁾Modelo que inclui as covariáveis Idade, Terapêutica e Vive sozinho

Assim, obtemos um modelo final que contém apenas as variáveis Idade, Vive sozinho e Terapêutica (Tabela 4.14), à semelhança do modelo de regressão logística.

Tabela 4. 14: Modelo múltiplo final: estimativa de cada β_j , valor p do teste de Wald, *hazard ratio* e respetivo IC95%.

Modelo final	$\hat{\beta}_j$	$se(\hat{\beta}_j)$	valor p	\widehat{HR} [IC95%]
Idade – “60 a 70 anos”	-0,362	0,147	0,0141	0,697 [0,522 ; 0,930]
Idade – “> 70 anos”	-0,021	0,155	0,8919	0,979 [0,722 ; 1,327]
Vive sozinho	0,388	0,145	0,0073	1,474 [1,110 ; 1,957]
Terapêutica	0,286	0,133	0,0314	1,331 [1,026 ; 1,727]

4.3.2 Interpretação dos coeficientes do modelo final

Note-se que apenas as estimativas dos coeficientes associados à covariável idade são negativas, originando estimativas do risco relativo inferiores a 1.

O coeficiente correspondente à classe etária “60 a 70 anos” foi estimado em $\hat{\beta}_1 = -0,362$, o que leva a um $\widehat{HR} = 0,697$. Assim, para um doente com idade compreendida entre os 60 e os 70 anos, há um decréscimo estimado de 30,3% no risco de descontinuar indevidamente a toma da medicação relativamente a um doente com 60 anos ou menos, para doentes com os mesmos valores das restantes variáveis. A estimativa do coeficiente correspondente à classe etária “> 70 anos”, $\hat{\beta}_2 = -0,021$, leva a

um valor de $\widehat{HR} = 0,979$. Este valor significa que um doente com mais de 70 anos tem um risco estimado de se tornar não persistente idêntico ao de um doente com 60 anos ou menos, para doentes com iguais valores nas restantes variáveis. Note-se ainda que o intervalo de 95% de confiança inclui o valor 1, pelo que não há evidência de que os riscos sejam diferentes, entre indivíduos das duas classes etárias.

Para a covariável que indica se o doente vive ou não sozinho, o coeficiente foi estimado em $\hat{\beta}_3 = 0,388$, o que leva a um valor de $\widehat{HR} = 1,474$. Com base neste valor estima-se que um indivíduo que vive sozinho tem um acréscimo de 47,4% no risco de interromper o tratamento do relativamente a um indivíduo que vive acompanhado, para indivíduos na mesma faixa etária e que estão a fazer a mesma terapêutica.

A estimativa do coeficiente da covariável Terapêutica, $\hat{\beta}_4 = 0,286$, dá origem a $\widehat{HR} = 1,331$. Portanto, estima-se que um indivíduo que está a fazer uma terapêutica semanal tem mais 33,1% de risco de se tornar não persistente, do que um indivíduo que está a fazer uma terapêutica mensal, para indivíduos com os mesmos valores das variáveis Idade e Vive sozinho.

4.3.3 Diagnóstico do modelo

A qualidade do ajustamento do modelo final foi avaliada através da análise de resíduos. Tal como sugerido por Grambsch e Therneau (1994), por forma a verificar a proporcionalidade das funções de risco correspondentes às variáveis incluídas no modelo, a análise focou-se nos resíduos de Schoenfeld padronizados, fazendo-se uso do gráfico destes resíduos *versus* $g(t)$. Complementarmente, recorreu-se ao teste proposto pelos autores, tendo sido utilizada como função do tempo $g(t) = t$ e $g(t) = \ln t$.

Tabela 4. 15: Teste da proporcionalidade das funções de risco, proposto por Grambsch e Therneau.

Variável	Transformação t		Transformação $\ln t$	
	χ^2	valor p	χ^2	valor p
Idade – “60 a 70 anos”	0,51	0,4733	0,47	0,4939
Idade – “> 70 anos”	0,08	0,7714	0,09	0,7667
Vive sozinho	2,87	0,0900	2,65	0,1033
Terapêutica	5,96	0,0147	2,50	0,1136
Global	10,19	0,0372	6,43	0,1696

De acordo com os resultados obtidos (Tabela 4.15), quando considerada a transformação logarítmica do tempo e um nível de significância de 0,05, não existe evidência para afirmar que as funções de risco não sejam proporcionais. Note-se ainda que foi obtido um resultado semelhante para o modelo global (valor $p=0,1696$).

No entanto, considerando a escala do tempo não transformada, existe evidência significativa para afirmar que as funções de risco associadas à variável Terapêutica não são proporcionais (valor $p=0,0147$), a um nível de significância de 0,05.

De acordo com a análise dos gráficos dos resíduos (Figura 4.5), não parece haver uma tendência definida (crescente ou decrescente) da curva de suavização, que indique não proporcionalidade das funções de risco associadas às variáveis incluídas no modelo final.

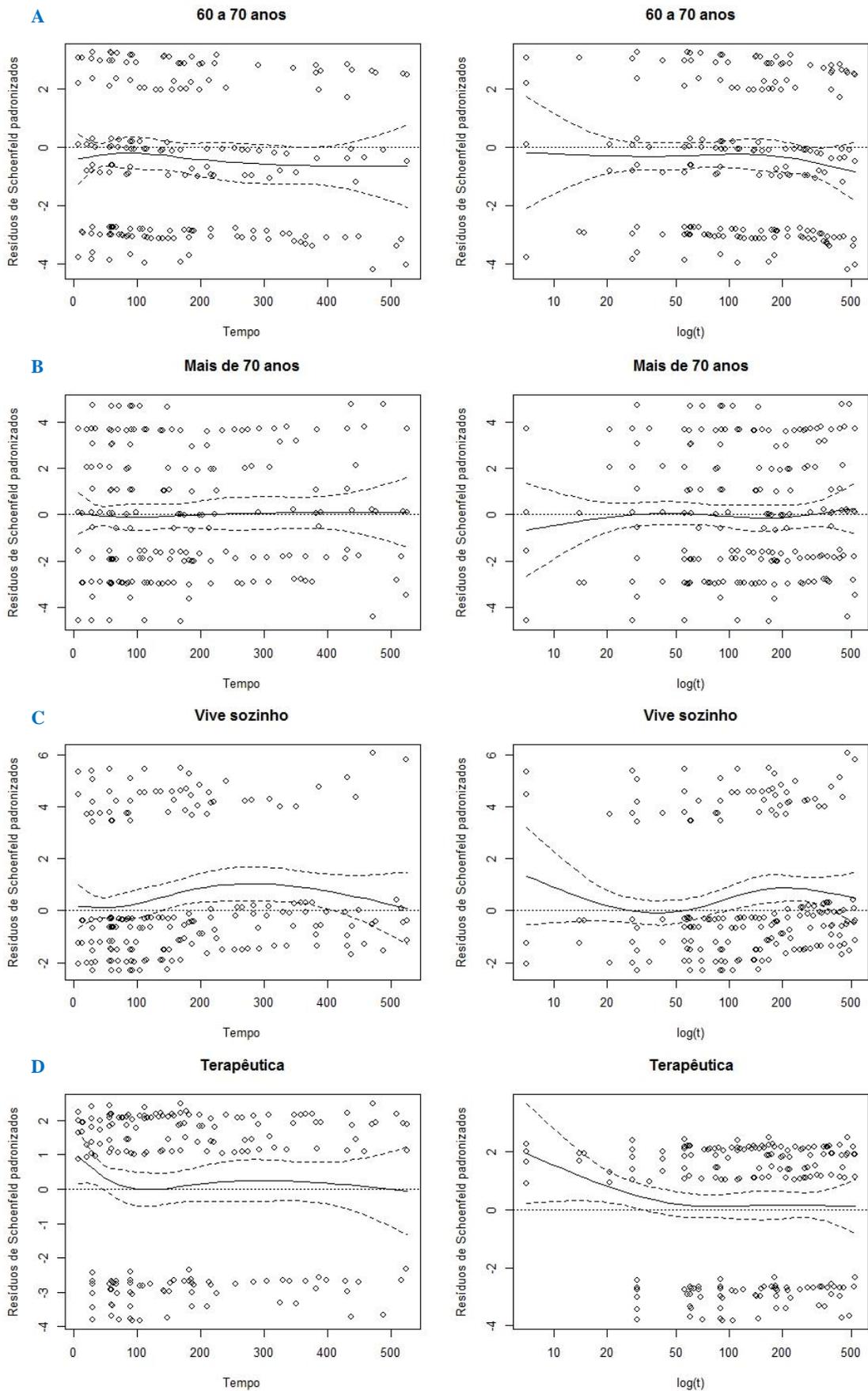


Figura 4. 5: Gráficos dos resíduos de Schoenfeld padronizados versus o tempo (à esquerda) e versus o logaritmo do tempo (à direita), para as variáveis Idade – “60 a 70 anos” (A), Idade – “Mais de 70 anos” (B), Vive sozinho (C) e Terapêutica (D).

4.4 Modelos Paramétricos

Tal como no modelo de riscos proporcionais de Cox, a variável resposta nos modelos paramétricos é o tempo até à não persistência no tratamento de uma determinada doença crónica. As covariáveis que vão ser incluídas nos modelos paramétricos são a Idade, Vive sozinho e Terapêutica, anteriormente identificadas como relevantes tanto no ajustamento do modelo de regressão logística como no modelo de Cox. Estas foram também as únicas variáveis que se mostraram significativas na análise não paramétrica.

4.4.1 Modelos de tempo de vida acelerado

Geralmente, recorre-se à função *survreg* disponível no *package* “*survival*” no *software* R, para ajustar modelos de regressão paramétricos, sob a forma de modelos de tempo de vida acelerado. É de salientar que, caso o modelo possa ser escrito sob a forma de modelo de riscos proporcionais, esta função não devolve diretamente as estimativas dos parâmetros do modelo sendo por isso necessário calculá-las posteriormente.

Modelo de regressão de Weibull

Tal como mencionado anteriormente, o modelo de regressão de Weibull é simultaneamente um modelo de riscos proporcionais e um modelo de tempo de vida acelerado.

Primeiramente, foram ajustados modelos de regressão de Weibull com uma covariável sob a forma de tempo de vida acelerado, com o auxílio da função *survreg*. Tal como se pode ver pela Tabela 4.16, para um nível de significância de 0,05, apenas para o grupo de indivíduos com mais de 70 anos não se verificam diferenças significativas quando comparado com o grupo de referência da respetiva variável ($p=0,3000$). De notar ainda que $\exp(\hat{\alpha}_j) > 1$ apenas para o grupo de indivíduos com idade compreendida entre os 60 e 70 anos, pelo que apenas neste grupo se pode inferir um aumento do tempo até à não persistência comparativamente ao grupo de referência.

Tabela 4. 16: Análise univariada: estimativa dos parâmetros e valor p dos testes de Wald.

Variável		$\hat{\alpha}_j$	$\exp(\hat{\alpha}_j)$ [IC95%]	valor p
Idade	≤ 60 anos	Ref.	-	-
	60 a 70 anos	0,391	1,478 [1,038 ; 2,105]	0,0303
	> 70 anos	-0,189	0,828 [0,579 ; 1,183]	0,3000
	Scale (σ)	1,238	-	-
	Intercept (μ)	5,590	-	-
Vive sozinho	Não	Ref.	-	-
	Sim	-0,521	0,594 [0,425 ; 0,831]	0,0024
	Scale (σ)	1,236	-	-
	Intercept (μ)	5,795	-	-
Terapêutica	Mensal	Ref.	-	-
	Semanal	-0,335	0,715 [0,520 ; 0,984]	0,0395
	Scale (σ)	1,246	-	-
	Intercept (μ)	5,903	-	-

As estimativas, com base no modelo de Weibull, das funções de sobrevivência associadas a cada variável podem ser comparadas graficamente com as estimativas de Kaplan-Meier dessas funções, permitindo averiguar a adequabilidade do modelo de Weibull (Figura 4.6).

Para as classes etárias “ ≤ 60 anos” e “ > 70 anos” verifica-se uma sobreposição razoável das curvas de sobrevivência estimadas pelo método de Kaplan-Meier e pelo modelo de Weibull após cerca de 8 e 6 meses de seguimento, respetivamente, até ao final do estudo. Contudo, na fase inicial do seguimento o modelo de Weibull sobrestima a sobrevivência em todos os grupos da variável Idade, com particular ênfase no grupo com idades compreendidas entre os 60 e 70 anos (Figura 4.6 A). Note-se que foi apenas para este grupo que se obteve $exp(\hat{\alpha}_j) > 1$, podendo este valor estar sobrestimado.

Para o grupo de doentes que vive sozinho, verifica-se uma sobreposição dos gráficos, o que indica que para este grupo o modelo parece ajustar-se razoavelmente bem. No entanto, o mesmo não acontece para o grupo de doentes que vive acompanhado (Figura 4.6 B).

Para os grupos da variável Terapêutica observa-se, à semelhança dos grupos das restantes variáveis, uma sobrestimação das funções de sobrevivência, verificando-se alguma sobreposição das curvas de Kaplan-Meier e do modelo de Weibull apenas após cerca de 10 meses (Figura 4.6 C).

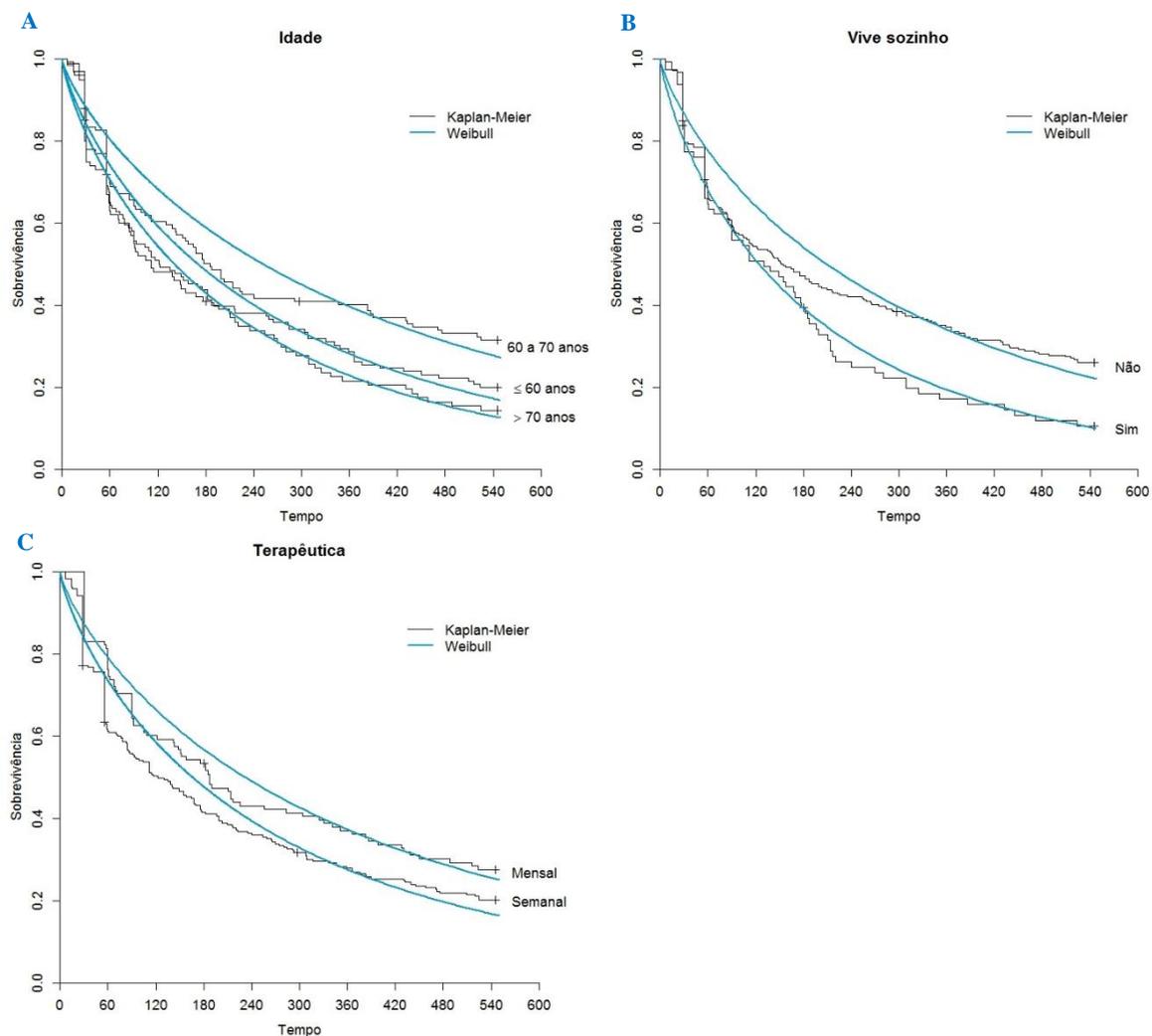


Figura 4. 6: Curvas de sobrevivência: comparação entre as estimativas de Kaplan-Meier e as curvas estimadas pelo modelo de Weibull para as variáveis Idade (A), Vive sozinho (B) e Terapêutica (C).

Fazendo-se novamente uso da função *survreg*, ajustou-se o modelo de regressão de Weibull múltiplo, que inclui as três variáveis analisadas previamente e cujos resultados se encontram na tabela 4.17.

Tabela 4. 17: Modelo de tempo de vida acelerado de Weibull: estimativa dos coeficientes, parâmetros e teste de Wald.

Variável	$\hat{\alpha}_j$	$exp(\hat{\alpha}_j)$ [IC95%]	$se(\hat{\alpha}_j)$	valor p
Idade – “60 a 70 anos”	0,466	1,594 [1,121 ; 2,266]	0,180	0,0094
Idade – “> 70 anos”	0,016	1,016 [0,701 ; 1,473]	0,189	0,9320
Vive sozinho	-0,537	0,585 [0,414 ; 0,825]	0,176	0,0022
Terapêutica	-0,320	0,726 [0,528 ; 0,999]	0,163	0,0493
Parâmetro				
Scale (σ)	1,220	-	-	-
Intercept (μ)	5,834	-	-	-

De acordo com as estimativas da Tabela 4.17, o modelo múltiplo de Weibull sob a forma de tempo de vida acelerado é dado por:

$$\widehat{\ln T} = 5,834 + 0,466 \times \text{"60 a 70 anos"} + 0,016 \times \text{"> 70 anos"} - 0,537 \times \text{"Vive sozinho"} - 0,320 \times \text{"Terapêutica"} + 1,220 \times \varepsilon \quad (3.61)$$

Assim, o fator de aceleração da classe etária “60 a 70 anos” foi estimado em $exp(\hat{\alpha}_1) = 1,594$. Este valor traduz-se num aumento de 1,6 vezes do tempo até à não persistência num doente com idade entre os 60 e os 70 anos, comparativamente a um doente com idade igual ou inferior a 60 anos, em doentes com os mesmos valores nas restantes variáveis. Por sua vez, o valor do fator de aceleração para a classe etária com mais de 70 anos foi estimado em $exp(\hat{\alpha}_2) = 1,016$, indicando que para doentes com iguais valores nas restantes variáveis, o tempo até à descontinuação do tratamento é idêntico para um doente com mais de 70 anos e um doente com 60 anos ou menos.

Para a covariável que especifica se cada indivíduo vive sozinho ou acompanhado, o fator de aceleração foi estimado em $exp(\hat{\alpha}_3) = 0,585$. Esta estimativa indica que o tempo até à interrupção indevida do tratamento de um indivíduo que vive sozinho é aproximadamente metade do tempo correspondente a um indivíduo que vive acompanhado, quando comparados indivíduos na mesma faixa etária e a fazer a mesma terapêutica.

Estima-se que o fator de aceleração para a covariável terapêutica seja $exp(\hat{\alpha}_4) = 0,726$, o que significa que o tempo até à não persistência de um doente que está a fazer uma terapêutica semanal corresponde a 72,6% do tempo de um doente que está a fazer uma terapêutica mensal, para doentes com os mesmos valores na variável Idade e Vive sozinho.

A Tabela 4.18 apresenta os valores estimados dos parâmetros do modelo de Weibull na forma de modelo de riscos proporcionais, visto que tal é necessário para poder comparar os coeficientes estimados no modelo de Weibull com os do modelo de Cox.

Tabela 4. 18: Modelo de riscos proporcionais de Weibull: estimativa dos coeficientes, parâmetros, *hazard ratio* e respetivo IC95% e teste de Wald.

Variável	$\hat{\beta}_j$	\widehat{HR} [IC95%]	$se(\hat{\beta}_j)$
Idade – “60 a 70 anos”	-0,382	0,682 [0,511 ; 0,911]	0,147
Idade – “> 70 anos”	-0,013	0,987 [0,728 ; 1,338]	0,155
Vive sozinho	0,440	1,553 [1,171 ; 2,059]	0,144
Terapêutica	0,262	1,300 [1,001 ; 1,688]	0,133
Parâmetro			
γ	0,820	-	-
λ	0,008	-	-

Assim, a estimativa da função de sobrevivência de um indivíduo com vetor de covariáveis $\mathbf{x} =$ (Idade – “60 a 70 anos”, Idade – “> 70 anos”, Vive sozinho, Terapêutica) é dada por:

$$\hat{S}(t; \mathbf{x}) = \exp[-0,008 \times t^{0,820} \exp(-0,382 \times \text{"60 a 70 anos"} - 0,013 \times \text{">70 anos"} + 0,440 \times \text{"Vive sozinho"} + 0,262 \times \text{"Terapêutica"})] \quad (3.62)$$

Note-se que as estimativas do risco relativo obtidas com o modelo de Weibull (Tabela 4.18) são idênticas às obtidas com o modelo de Cox (Tabela 4.14). O risco relativo correspondente à classe etária “60 a 70 anos” foi estimado em $\widehat{HR} = 0,682$ ($\widehat{HR} = 0,697$ no modelo de Cox), o que se traduz num decréscimo estimado de 31,8% no risco de um doente com idade compreendida entre os 60 e os 70 anos descontinuar indevidamente a toma da medicação, comparativamente a um doente com 60 anos ou menos, em doentes com os mesmos valores nas restantes variáveis. Para a classe etária “Mais de 70 anos”, o valor estimado do risco relativo foi de $\widehat{HR} = 0,987$ ($\widehat{HR} = 0,979$ no modelo de Cox). Assim, para doentes com iguais valores nas restantes variáveis, um doente com mais de 70 anos e um doente com 60 anos ou menos têm um risco idêntico de se tornar não persistentes. De forma semelhante ao observado no modelo de Cox, no modelo de Weibull o intervalo de 95% de confiança deste risco relativo contém o valor 1, não existindo evidência de que os riscos sejam diferentes para indivíduos das duas classes etárias.

Para a covariável que indica se o doente vive ou não sozinho, o risco relativo foi estimado em $\widehat{HR} = 1,553$ ($\widehat{HR} = 1,474$ no modelo de Cox), pelo que um indivíduo que vive sozinho tem um acréscimo estimado de cerca de 55,3% no risco de interromper o tratamento do que um indivíduo que vive acompanhado, para indivíduos na mesma faixa etária e a fazer a mesma terapêutica.

A estimativa do risco relativo para a covariável terapêutica é de $\widehat{HR} = 1,300$ ($\widehat{HR} = 1,331$ no modelo de Cox). Portanto, estima-se que um indivíduo que está a fazer uma terapêutica semanal tem mais 30,0% de risco de se tornar não persistente, do que um indivíduo que está a fazer uma terapêutica mensal, para indivíduos com os mesmos valores nas variáveis Idade e Vive sozinho.

Modelo de regressão log-normal

Sendo o modelo de regressão log-normal um modelo de tempo de vida acelerado, pode ser ajustado no *software* R através da função *survreg*, à semelhança do modelo anterior.

A análise iniciou-se com o ajustamento de modelos com uma covariável (Tabela 4.19). Comparativamente aos modelos de Weibull, verificou-se um aumento do valor p para os vários grupos com exceção do grupo de indivíduos que se encontra a fazer a terapêutica semanal. De notar ainda que este é o único grupo para o qual o intervalo de 95% de confiança não inclui o valor 1, havendo evidência de diferença no tempo de vida de um indivíduo que faz o tratamento semanal comparativamente a um indivíduo que faz o tratamento mensal.

Tabela 4. 19: Análise univariada: estimativa dos parâmetros e valor p dos testes de Wald.

Variável		$\hat{\alpha}_j$	$exp(\hat{\alpha}_j)$ [IC95%]	valor p
Idade	≤ 60 anos	Ref.	-	-
	60 a 70 anos	0,335	1,398 [0,972 ; 2,012]	0,0709
	> 70 anos	-0,140	0,869 [0,591 ; 1,278]	0,4760
	Scale (σ)	1,443	-	-
	Intercept (μ)	4,991	-	-
Vive sozinho	Não	Ref.	-	-
	Sim	-0,368	0,692 [0,479 ; 0,999]	0,0491
	Scale (σ)	1,446	-	-
	Intercept (μ)	5,156	-	-
Terapêutica	Mensal	Ref.	-	-
	Semanal	-0,404	0,667 [0,481 ; 0,926]	0,0155
	Scale (σ)	1,439	-	-
	Intercept (μ)	5,343	-	-

Em seguida comparou-se novamente as estimativas das funções de sobrevivência associadas ao grupo definido por cada variável com as estimativas de Kaplan-Meier dessas funções, por forma a avaliar a adequabilidade do modelo log-normal (Figura 4.7).

Para a variável Idade, apenas para o grupo de doentes com mais de 70 anos, se verifica alguma sobreposição entre a estimativa de Kaplan-Meier e a curva estimada pelo modelo log-normal. A função de sobrevivência da classe etária “60 a 70 anos” estimada pelo modelo log-normal está acima da estimativa de Kaplan-Meier até aos 10 meses, e abaixo da mesma até ao final do seguimento (Figura 4.7 A). Estes resultados sugerem que o modelo log-normal não será o mais adequado.

Para a variável Vive sozinho e Terapêutica, não se verifica sobreposição das respetivas curvas de sobrevivência, voltando a sugerir que o modelo log-normal não será o mais adequado (Figura 4.7 B e 4.7 C).

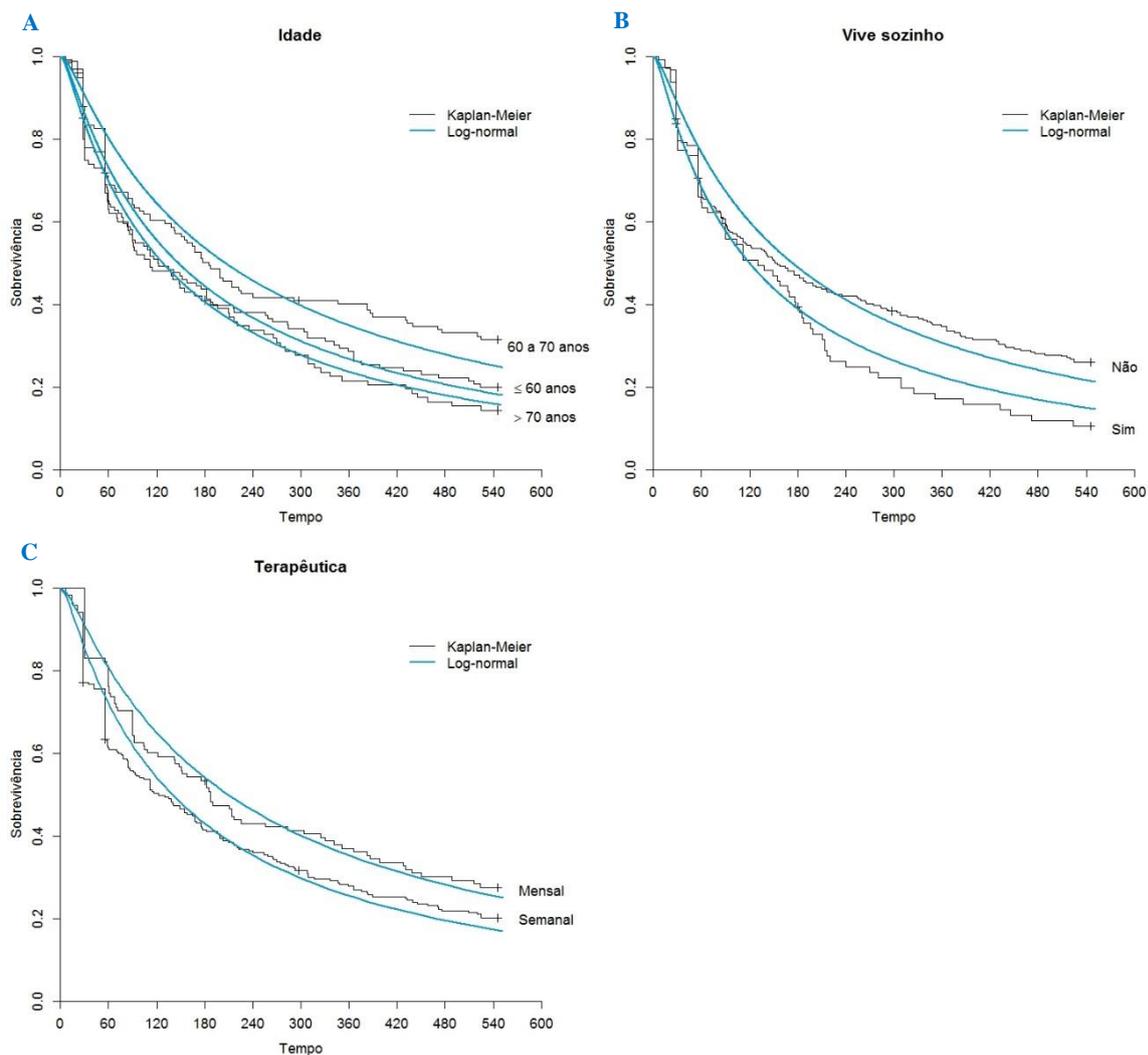


Figura 4. 7: Curvas de sobrevivência: comparação entre as estimativas de Kaplan-Meier e as curvas estimadas pelo modelo log-normal para as variáveis Idade (A), Vive sozinho (B) e Terapêutica (C).

Ainda assim, optou-se por proceder ao ajustamento do modelo log-normal cujos resultados se encontram na Tabela 4.20.

Tabela 4. 20: Modelo de tempo de vida acelerado log-normal: estimativa dos parâmetros e teste de Wald.

Modelo final	$\hat{\alpha}_j$	$exp(\hat{\alpha}_j)$ [IC95%]	$se(\hat{\alpha}_j)$	valor p
Idade – “60 a 70 anos”	0,411	1,508 [1,048 ; 2,170]	0,186	0,0268
Idade – “> 70 anos”	0,033	1,033 [0,694 ; 1,539]	0,203	0,8720
Vive sozinho	-0,386	0,680 [0,467 ; 0,988]	0,191	0,0433
Terapêutica	-0,406	0,666 [0,480 ; 0,925]	0,167	0,0153
Scale (σ)	1,424	-	-	-
Intercept (μ)	5,273	-	-	-

De um modo geral, as estimativas do factor de aceleração são da mesma ordem de grandeza das estimativas obtidas com base no modelo de regressão de Weibull.

4.4.2 Modelos de possibilidades proporcionais

Modelo de regressão log-logístico

Como já foi referido antes, o modelo de regressão log-logístico é simultaneamente um modelo de possibilidades proporcionais e um modelo de tempo de vida acelerado, pelo que pode ser ajustado recorrendo-se à função *survreg*.

Tal como no modelo de regressão de Weibull, começou-se por ajustar modelos com uma covariável sob a forma de modelos de tempo de vida acelerado. Na Tabela 4.21 é possível ver que os resultados obtidos são idênticos aos do modelo log-normal. Note-se que o único modelo no qual o intervalo de confiança a 95% do fator de aceleração não contém o valor 1 é o modelo que inclui apenas a variável Terapêutica. Além disso, para $\alpha = 0,05$ apenas no grupo de indivíduos que está a fazer a Terapêutica semanal se verificaram diferenças significativas no tempo até à não persistência, quando comparado com o grupo de referência respetivo ($p=0,0192$).

Tabela 4. 21: Análise univariada: estimativa dos parâmetros e valor p dos testes de Wald.

Variável		$\hat{\alpha}_j$	$exp(\hat{\alpha}_j)$ [IC95%]	valor p
Idade	≤ 60 anos	Ref.	-	-
	60 a 70 anos	0,362	1,436 [0,976 ; 2,111]	0,0661
	> 70 anos	-0,136	0,873 [0,585 ; 1,303]	0,5060
	Scale (σ)	0,879	-	-
	Intercept (μ)	4,942	-	-
Vive sozinho	Não	Ref.	-	-
	Sim	-0,341	0,711 [0,489 ; 1,035]	0,0751
	Scale (σ)	0,883	-	-
	Intercept (μ)	5,112	-	-
Terapêutica	Mensal	Ref.	-	-
	Semanal	-0,410	0,664 [0,471 ; 0,935]	0,0192
	Scale (σ)	0,878	-	-
	Intercept (μ)	5,301	-	-

À semelhança dos modelos anteriores, as funções de sobrevivência estimadas com base nos modelos com uma covariável foram comparadas com as estimativas de Kaplan-Meier (Figura 4.8).

Como se pode observar, as curvas são muito idênticas às obtidas com o modelo log-normal (Figura 4.7), verificando-se sobreposição razoável das curvas associadas à variável Idade apenas para o grupo de doentes com mais de 70 anos. Existe uma subestimação acentuada da função de sobrevivência para a classe etária “60 a 70 anos” obtida pelo modelo log-logístico a partir dos 10 meses de seguimento (Figura 4.8 A).

Na variável Vive sozinho e Terapêutica, volta a não se verificar sobreposição das respetivas curvas de sobrevivência, o que sugere que o modelo log-logístico não será o mais adequado (Figura 4.8 B e 4.8 C).

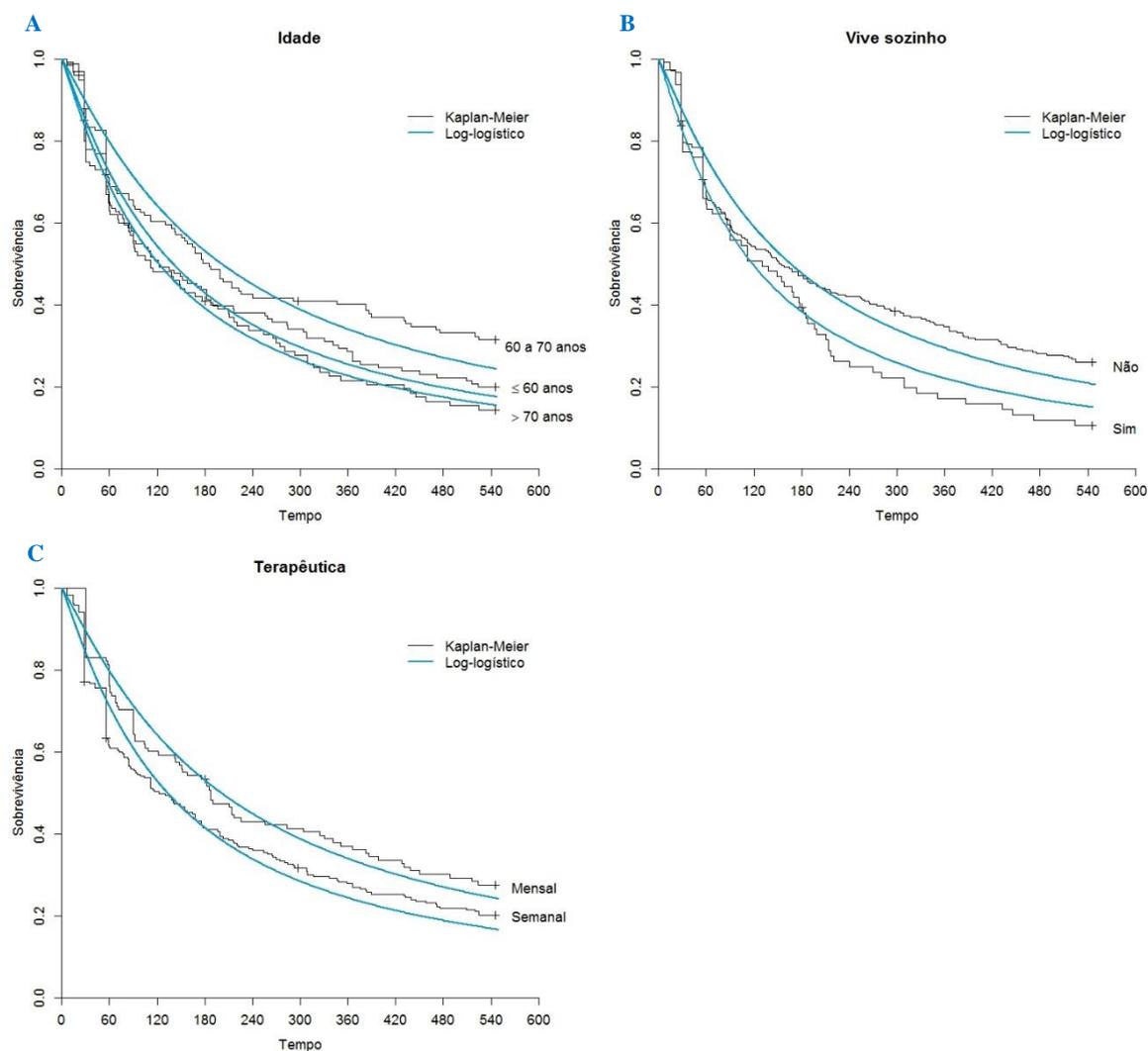


Figura 4. 8: Curvas de sobrevivência: comparação entre as estimativas de Kaplan-Meier e as curvas estimadas com o modelo log-logístico para as variáveis Idade (A), Vive sozinho (B) e Terapêutica (C).

No entanto, à semelhança do modelo log-normal, optou-se por proceder ao ajustamento do modelo log-logístico múltiplo cujos resultados se encontram na Tabela 4.22.

As estimativas dos coeficientes de regressão são muito próximas dos valores obtidos para o modelo log-normal. É de notar, no entanto que a estimativa do parâmetro de escala ($\hat{\sigma} = 0,878$) é bastante inferior à do modelo anterior.

Tabela 4. 22: Modelo de tempo de vida acelerado log-logístico: estimativa dos parâmetros e teste de Wald.

Variável	$\hat{\alpha}_j$	$exp(\hat{\alpha}_j)$ [IC95%]	$se(\hat{\alpha}_j)$	valor p
Idade – “60 a 70 anos”	0,460	1,584 [1,074 ; 2,334]	0,198	0,0202
Idade – “> 70 anos”	0,042	1,043 [0,688 ; 1,581]	0,212	0,8430
Vive sozinho	-0,379	0,684 [0,466 ; 1,004]	0,196	0,0525
Terapêutica	-0,415	0,660 [0,468 ; 0,931]	0,175	0,0178
<i>Scale</i>	0,867	-	-	-
<i>Intercept</i>	5,225	-	-	-

À semelhança do modelo de Weibull, é possível a partir das estimativas anteriormente obtidas chegar aos parâmetros estimados do modelo log-logístico sob a forma de modelo de possibilidades proporcionais (Tabela 4.23).

Tabela 4. 23: Modelo de possibilidades proporcionais log-logístico: estimativa dos parâmetros, *odds ratio* e respetivo IC95%.

Variável	$\hat{\beta}_j$	$se(\hat{\beta}_j)$	\widehat{OR} [IC95%]
Idade – “60 a 70 anos”	-0,530	0,228	0,589 [0,376 ; 0,921]
Idade – “> 70 anos”	-0,049	0,245	0,953 [0,590 ; 1,539]
Vive sozinho	0,437	0,226	1,549 [0,995 ; 2,410]
Terapêutica	0,479	0,202	1,614 [1,086 ; 2,398]
γ	1,153	-	-
λ	0,002	-	-

Apesar das dúvidas relativas à adequabilidade deste modelo, de acordo com o mesmo para a classe etária “60 a 70 anos” o *odds ratio* foi estimado em $\widehat{OR} = 0,589$, o que indica um decréscimo estimado de 41,1% na possibilidade de um doente nesta classe etária descontinuar indevidamente a toma da medicação, comparativamente a um doente com 60 anos ou menos, em doentes com os mesmos valores nas variáveis Vive sozinho e Terapêutica. Para o grupo de doentes com mais de 70 anos, o valor estimado do *odds ratio* foi de $\widehat{OR} = 0,953$. Assim, para doentes com iguais valores nas restantes variáveis, um doente com mais de 70 anos tem uma possibilidade de se tornar não persistente ligeiramente menor (decréscimo de apenas 4,7%), quando comparado a um doente com 60 anos ou menos.

O *odds ratio* correspondente à covariável que indica se o doente vive ou não sozinho, foi estimado em $\widehat{OR} = 1,549$, pelo que um indivíduo que vive sozinho tem um acréscimo estimado de cerca de 54,9% na possibilidade de interromper o tratamento do que um indivíduo que vive acompanhado, quando comparados indivíduos na mesma classe etária e a fazer a mesma terapêutica.

A estimativa do *odds ratio* para a covariável terapêutica é de $\widehat{OR} = 1,614$. Este valor traduz-se num aumento estimado de 61,4% na possibilidade de um indivíduo que está a fazer uma terapêutica semanal se tornar não persistente, relativamente a um indivíduo que está a fazer uma terapêutica mensal, para indivíduos com os mesmos valores nas restantes variáveis.

4.4.3 Comparação dos modelos paramétricos

Através da análise univariada com base em cada um dos modelos paramétricos anteriores, foi possível perceber que as estimativas da função de sobrevivência que mais se aproximaram da estimativa de Kaplan-Meier, foram obtidas pelos modelos de Weibull com uma covariável.

No entanto, analisando o valor de AIC dos modelos paramétricos múltiplos obtidos anteriormente, verifica-se que o modelo de regressão log-normal origina o menor valor desta estatística, sugerindo que este é o modelo que melhor se ajusta de entre os 3 considerados. Contudo, é de notar que o menor valor p do teste da razão de verosimilhanças é obtido para o modelo de regressão de Weibull múltiplo.

Tabela 4. 24: Comparação dos modelos paramétricos múltiplos: AIC e Teste da razão de verosimilhanças.

Modelo	AIC	Teste da razão de verosimilhanças	
		Estatística G	valor p
Modelo de regressão de Weibull	3653,99	22,18	0,0002
Modelo de regressão log-normal	3608,18	15,86	0,0032
Modelo de regressão log-logístico	3623,74	15,29	0,0041

4.5 Modelos Flexíveis de Royston e Parmar

Os modelos flexíveis propostos por Royston e Parmar podem ser ajustados no *software R* através da função *flexsurvspline* do *package* “*flexsurv*”. A complexidade do modelo depende do número de nós internos (m) incluídos na função *spline*, definidos na função *flexsurvspline* pelo argumento k , cuja localização predefinida é dada pelos quantis da distribuição de $\ln(T)$.

Além do número de nós internos, é ainda possível escolher a transformação da função de sobrevivência através do argumento *scale*:

- *scale* = “*hazard*” obtém-se um modelo de riscos proporcionais
- *scale* = “*odds*” obtém-se um modelo de possibilidades proporcionais

4.5.1 Modelo flexível de riscos proporcionais

Tal como mencionado anteriormente, o modelo flexível de riscos proporcionais é obtido modelando o logaritmo da função de risco cumulativa, $\ln H(t, \mathbf{x})$, que pode ser obtido através da correspondente transformação da função de sobrevivência, $g[S(t, \mathbf{x})] = \ln(-\ln(S(t, \mathbf{x})))$.

A análise desta família de modelos iniciou-se com o ajustamento de modelos incluindo cada uma das três variáveis selecionadas no modelo de Cox, fazendo-se variar o número de nós internos (Tabela 4.25).

Note-se que quando não são incluídos nós internos ($m = 0$), obtém-se o modelo mais simples apenas com dois parâmetros, os nós limite, e que coincide com o modelo de Weibull. De facto, comparando os valores dos parâmetros estimados pelo modelo flexível sem nós internos com os valores dos parâmetros estimados pelo modelo de riscos proporcionais de Weibull (que podem ser obtidos a partir da Tabela 4.16), é possível verificar que são coincidentes.

Analisando o critério de Akaike para cada modelo, é possível ver que o valor diminuí com o aumento do número de nós internos até $m = 3$. Em todos os modelos com 4 nós internos o valor de AIC volta a crescer, o que indica que para além de $m = 3$ se verifica um aumento de complexidade da curva que é desnecessário, tal como sugerido por Royston e Parmar (2002).

É ainda de notar que as estimativas dos coeficientes de regressão não se alteram muito com o aumento do número de nós internos.

Tabela 4. 25: Análise univariada de modelos de riscos proporcionais com m nós internos: estimativa dos parâmetros e valor de AIC.

Modelos Flexíveis de Riscos Proporcionais com uma covariável						
Variável	Estimativas	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$
Idade	$\hat{\beta}_1$	-0,316	-0,294	-0,297	-0,296	-0,296
	$\hat{\beta}_2$	0,153	0,130	0,138	0,136	0,136
	γ_0	-4,514	-8,001	-10,933	-13,967	-14,819
	γ_1	0,808	1,788	2,723	3,775	4,078
	γ_2	-	0,098	0,368	0,582	0,806
	γ_3	-	-	-0,268	-0,257	-0,365
	γ_4	-	-	-	-0,057	-0,089
	γ_5	-	-	-	-	-0,035
	AIC	3661,98	3598,65	3582,77	3580,73	3582,11
Vive sozinho	$\hat{\beta}_1$	0,421	0,349	0,366	0,363	0,362
	γ_0	-4,688	-8,129	-11,074	-14,107	-14,935
	γ_1	0,809	1,783	2,721	3,773	4,067
	γ_2	-	0,098	0,369	0,583	0,798
	γ_3	-	-	-0,270	-0,258	-0,350
	γ_4	-	-	-	-0,058	-0,102
	γ_5	-	-	-	-	-0,029
	AIC	3661,62	3599,17	3583,07	3581,05	3582,43
Terapêutica	$\hat{\beta}_1$	0,269	0,264	0,265	0,264	0,263
	γ_0	-4,736	-8,256	-11,160	-14,231	-15,046
	γ_1	0,802	1,793	2,719	3,783	4,074
	γ_2	-	0,099	0,365	0,585	0,804
	γ_3	-	-	-0,263	-0,261	-0,362
	γ_4	-	-	-	-0,053	-0,094
	γ_5	-	-	-	-	-0,027
	AIC	3665,78	3600,86	3585,32	3583,25	3584,61

O efeito da inclusão de nós internos na forma das estimativas das funções de sobrevivência, associadas a cada variável, pode ser analisado através da comparação dos gráficos destas estimativas e das obtidas pelo estimador de Kaplan-Meier (Figura 4.9).

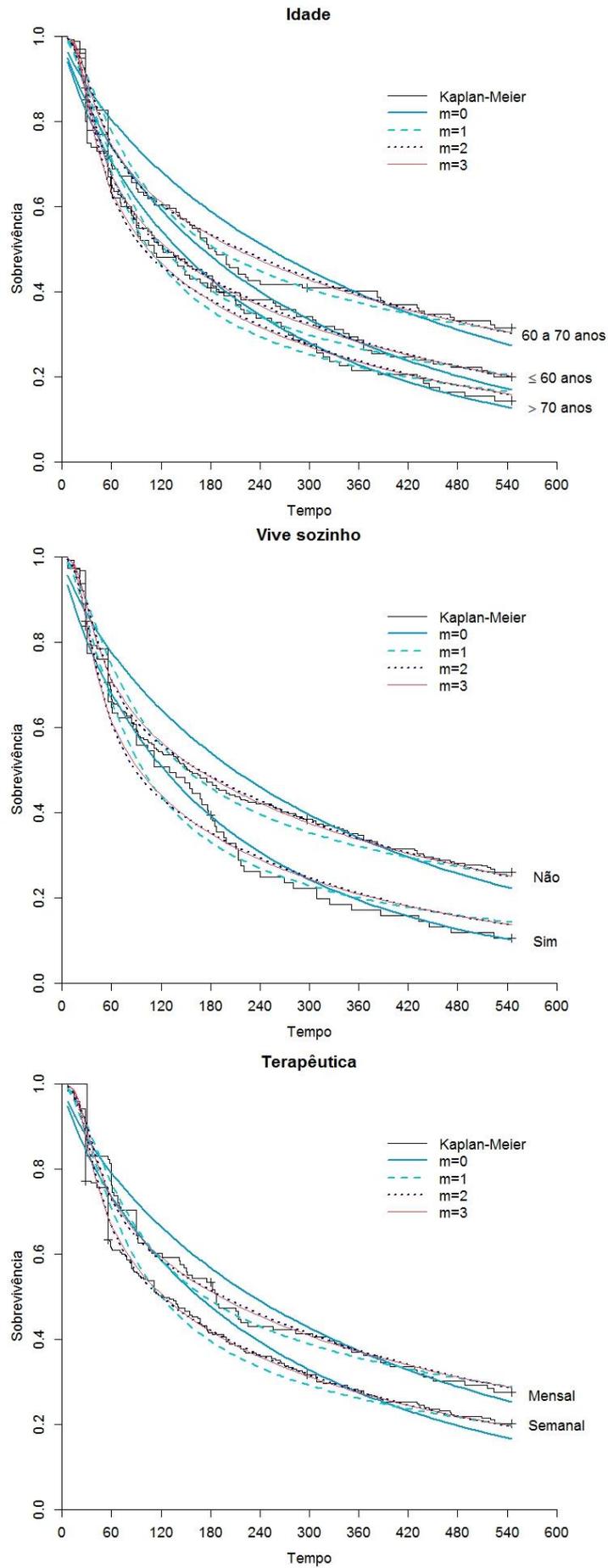


Figura 4. 9: Curvas de sobrevivência: comparação entre as estimativas de Kaplan-Meier e as curvas estimadas pelos modelos flexíveis de riscos proporcionais com m nós para as variáveis Idade (A), Vive sozinho (B) e Terapêutica (C).

De um modo geral, pela Figura 4.9 é possível perceber que para os vários grupos das três variáveis em análise, a inclusão de 1 nó interno ($m = 1$) produz uma clara alteração na estimativa da curva de sobrevivência, que se torna mais flexível no início do seguimento aproximando-se da estimativa de Kaplan-Meier. A adição de um 2º nó produz um efeito na curvatura das estimativas das funções que apesar de menos evidente, melhora o ajustamento face aos modelos com 1 nó interno, levando à sobreposição das curvas à respetiva estimativa de Kaplan-Meier. Os modelos com 3 nós internos, por sua vez, não alteram significativamente a forma da função de sobrevivência estimada face aos modelos com $m = 2$, sugerindo que a presença do 3º nó interno possa ser dispensável, visto acrescentar complexidade desnecessária aos modelos.

Pela Figura 4.9 A é possível perceber que para todos os grupos da variável idade, a inclusão de nós internos no modelo flexível diminui a sobrestimação das funções de sobrevivência no início do seguimento, face ao modelo de Weibull ($m = 0$). Para o grupo de indivíduos com 60 anos ou menos, a estimativa da função de sobrevivência obtida com o modelo com $m = 2$ coincide com a estimativa de Kaplan-Meier ao longo de todo o seguimento. A inclusão de 2 nós internos também melhora a estimativa da função associada aos restantes grupos da variável idade, verificando-se apenas uma pequena sobrestimação da sobrevivência entre os 6 e os 10 meses no grupo etário “60 a 70 anos” e uma pequena subestimação entre os 4 e os 8 meses no grupo etário “> 70 anos”.

Para o grupo de doentes que vive acompanhado, a estimativa da função de sobrevivência com base no modelo flexível com 2 nós internos está sobreposta à estimativa de Kaplan-Meier, ao longo de todo o seguimento. Este facto sugere que o modelo revela um bom ajustamento para este grupo de indivíduos. No entanto, para o grupo de doentes que vive sozinho a inclusão de nós internos não melhorou a estimativa da função de sobrevivência face ao modelo de Weibull, verificando-se uma subestimação da sobrevivência até aos 6 meses e uma sobrestimação após essa altura (Figura 4.9 B).

No caso da variável Terapêutica, as estimativas da função de sobrevivência associadas a cada grupo de doentes, obtidas pelo modelo flexível com 2 nós internos, estão totalmente sobrepostas às respetivas estimativas de Kaplan-Meier, verificando-se por isso uma grande melhoria face ao modelo de Weibull (Figura 4.9 C).

Em seguida procedeu-se ao ajustamento dos modelos flexíveis múltiplos de riscos proporcionais com m nós internos ($m = 0, 1, 2, 3$), cujos resultados se encontram na tabela abaixo:

Tabela 4. 26: Modelos de riscos proporcionais com m nós internos: estimativa dos parâmetros e valor de AIC.

Modelo Flexível de Riscos Proporcionais				
Variável	$m = 0$	$m = 1$	$m = 2$	$m = 3$
Idade – “60 a 70 anos”	-0,382	-0,360	-0,364	-0,363
Idade – “> 70 anos”	-0,013	-0,022	-0,015	-0,016
Vive sozinho	0,440	0,373	0,390	0,387
Terapêutica	0,262	0,261	0,260	0,259
γ_0	-4,784	-8,205	-11,189	-14,184
γ_1	0,820	1,788	2,738	3,778
γ_2	-	0,097	0,374	0,580
γ_3	-	-	-0,276	-0,250
γ_4	-	-	-	-0,067
AIC	3653,99	3593,00	3576,56	3574,67

De acordo com os resultados da Tabela 4.26, é possível perceber que a introdução de nós no modelo flexível leva a uma diminuição do valor de AIC associado ao modelo, sugerindo uma melhoria no ajustamento do modelo com a inclusão dos nós.

O valor de AIC dos modelos com 2 e com 3 nós internos é muito próximo, pelo que ambos revelam um bom ajustamento. Este resultado e a análise gráfica univariada anterior sugerem que a inclusão do 3º nó interno poderá não ser necessária, pelo que na presente dissertação se optou por focar a análise no modelo mais parcimonioso.

Tabela 4. 27: Modelo flexível de riscos proporcionais com 2 nós internos: estimativa dos coeficientes, parâmetros, *hazard ratio* e respetivo IC95% e teste de Wald.

Variável	$\hat{\beta}_j$	$se(\hat{\beta}_j)$	$\widehat{HR}[IC95\%]$
Idade – “60 a 70 anos”	-0,364	0,147	0,695 [0,521 ; 0,928]
Idade – “> 70 anos”	-0,015	0,155	0,985 [0,727 ; 1,335]
Vive sozinho	0,390	0,145	1,478 [1,113 ; 1,962]
Terapêutica	0,260	0,133	1,297 [1,000 ; 1,683]
γ_0	-11,189	1,041	-
γ_1	2,738	0,298	-
γ_2	0,374	0,075	-
γ_3	-0,276	0,074	-

Assim, para o modelo flexível com 2 nós internos, a estimativa do logaritmo da função de risco cumulativa é dada por:

$$\ln \widehat{H}(t, \mathbf{x}) = -11,189 + 2,738 \times \ln t + 0,374 \times v_1(\ln t) - 0,276 \times v_2(\ln t) - 0,364 \times \text{"60 a 70 anos"} - 0,015 \times \text{">70 anos"} + 0,390 \times \text{"Vive sozinho"} + 0,259 \times \text{"Terapêutica"} \quad (3.63)$$

com,

$$v_1(\ln t) = (\ln t - 0,364)_+^3 - \lambda_1 (\ln t - (-11,189))_+^3 - (1 - \lambda_1) (\ln t - 2,738)_+^3$$

e

$$v_2(\ln t) = (\ln t - (-0,276))_+^3 - \lambda_2 (\ln t - (-11,189))_+^3 - (1 - \lambda_2) (\ln t - 2,738)_+^3$$

onde,

$$\lambda_1 = \frac{2,738 - 0,364}{2,738 - (-11,189)} = 0,170 \quad \text{e} \quad \lambda_2 = \frac{2,738 - (-0,276)}{2,738 - (-11,189)} = 0,216$$

4.5.2 Modelo flexível de possibilidades proporcionais

O modelo flexível de possibilidades proporcionais é obtido modelando o logaritmo da função de possibilidade cumulativa, $\ln O(t, \mathbf{x})$, que pode ser obtido através da correspondente transformação da função de sobrevivência, $g[S(t, \mathbf{x})] = \ln(S(t, \mathbf{x})^{-1} - 1)$.

De forma semelhante aos modelos flexíveis de riscos proporcionais, a análise desta família de modelos iniciou-se com o ajustamento de modelos com uma covariável, fazendo-se variar o número de nós internos (Tabela 4.27).

Repare-se que o modelo flexível mais simples, quando não são incluídos nós internos ($m = 0$), coincide com o modelo log-logístico. Comparando os valores dos parâmetros estimados pelo modelo de possibilidades proporcionais log-logístico (que podem ser obtidos a partir da Tabela 4.21), com os valores dos parâmetros estimados pelo modelo flexível sem nós internos (Tabela 4.27), este facto torna-se evidente.

Tabela 4. 28: Análise univariada de modelos de possibilidades proporcionais com m nós internos: estimativa dos parâmetros e valor de AIC.

Modelos Flexíveis de Possibilidades Proporcionais com uma covariável						
Variável	Estimativas	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$
Idade	$\hat{\beta}_1$	-0,411	-0,396	-0,390	-0,389	-0,389
	$\hat{\beta}_2$	0,155	0,152	0,162	0,159	0,157
	γ_0	-5,623	-8,496	-11,643	-14,102	-15,053
	γ_1	1,138	1,957	2,968	3,826	4,166
	γ_2	-	0,089	0,430	0,536	0,800
	γ_3	-	-	-0,360	-0,164	-0,390
	γ_4	-	-	-	-0,167	0,003
	γ_5	-	-	-	-	-0,167
	AIC		3628,66	3599,01	3583,67	3583,32
Vive sozinho	$\hat{\beta}_1$	0,386	0,345	0,370	0,366	0,365
	γ_0	-5,788	-8,622	-11,781	-14,232	-15,192
	γ_1	1,132	1,944	2,959	3,814	4,157
	γ_2	-	0,088	0,431	0,536	0,799
	γ_3	-	-	-0,362	-0,163	-0,387
	γ_4	-	-	-	-0,170	-0,003
	γ_5	-	-	-	-	-0,165
	AIC		3629,87	3600,53	3584,88	3584,54
Terapêutica	$\hat{\beta}_1$	0,467	0,469	0,459	0,458	0,456
	γ_0	-6,036	-8,930	-12,038	-14,516	-15,474
	γ_1	1,139	1,965	2,967	3,833	4,175
	γ_2	-	0,089	0,427	0,541	0,809
	γ_3	-	-	-0,356	-0,173	-0,406
	γ_4	-	-	-	-0,160	0,012
	γ_5	-	-	-	-	-0,166
	AIC		3627,56	3597,41	3582,34	3581,99

Pelo critério de Akaike associado a cada modelo, pode verificar-se que o valor diminuí com o aumento do número de nós internos até $m = 3$, no entanto, nos modelos com 4 nós internos o valor de AIC volta a aumentar. Este resultado, de forma análoga aos modelos flexíveis de riscos proporcionais, indica que devem ser utilizados modelos com $m < 4$. Uma vez mais, com o aumento do número de nós internos não se verificam grandes alterações nas estimativas dos coeficientes de regressão.

Em seguida procedeu-se à comparação dos gráficos das estimativas obtidas pelos modelos flexíveis de possibilidades proporcionais com uma covariável e dos gráficos das estimativas de Kaplan-Meier (Figura 4.10). Apesar do efeito sobre as estimativas das curvas de sobrevivência ser menos evidente nesta família de modelos flexíveis do que na anterior, como se pode ver pelos vários gráficos da Figura 4.10, em geral, a inclusão de nós internos melhora as estimativas associadas aos vários grupos, aproximando-as das respetivas curvas de Kaplan-Meier.

Como se pode ver na Figura 4.10 A, para o grupo de indivíduos com 60 anos ou menos as estimativas da função de sobrevivência obtidas com os modelos com $m = 2$ e com $m = 3$ coincidem com a estimativa de Kaplan-Meier ao longo de todo o seguimento. A inclusão de 2 ou 3 nós internos também melhora a estimativa da função associada ao grupo etário “> 70anos”, verificando-se apenas uma pequena sobrestimação da sobrevivência após 10 meses de seguimento ($t = 300$). No entanto, no grupo etário “60 a 70 anos” apesar da inclusão dos nós internos melhorar a estimação da curva de sobrevivência face ao modelo log-logístico ($m = 0$), apenas existe sobreposição das curvas até aos 6 meses de seguimento ($t = 180$).

A função de sobrevivência do grupo de doentes que vive acompanhado, estimada com base no modelo flexível com 2 nós internos, está sobreposta à estimativa de Kaplan-Meier ao longo de todo o seguimento, indicando um bom ajustamento do modelo para este grupo de doentes. Contudo, para o grupo de doentes que vive sozinho, a inclusão de nós internos não melhorou a estimativa da função de sobrevivência face ao modelo log-logístico (Figura 4.10 B).

No caso da variável Terapêutica, de forma semelhante ao observado na família de modelos flexíveis anterior, as funções de sobrevivência associadas a cada grupo de indivíduos, estimadas pelo modelo flexível com 2 nós internos, estão sobrepostas às respetivas estimativas de Kaplan-Meier durante todo o período de seguimento. Este resulta indica uma grande melhoria face ao modelo log-logístico (Figura 4.10 C).

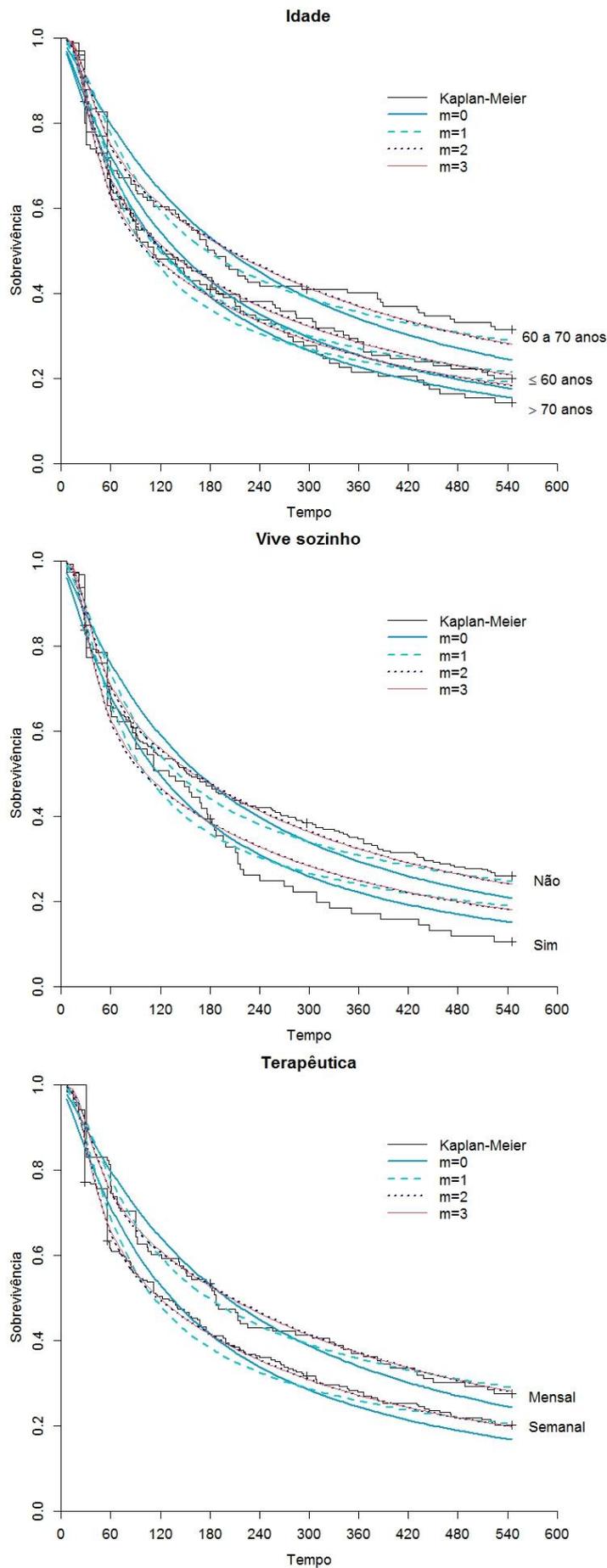


Figura 4. 10: Curvas de sobrevivência: comparação entre as estimativas de Kaplan-Meier e as curvas estimadas pelos modelos flexíveis de possibilidades proporcionais com m nós para as variáveis Idade (A), Vive sozinho (B) e Terapêutica (C).

Posteriormente procedeu-se ao ajustamento de modelos flexíveis múltiplos de possibilidades proporcionais com m nós internos no *spline* cúbico. Uma vez mais, de acordo com os resultados anteriormente obtidos e com a sugestão de Royston e Parmar (2002), nesta fase da análise não foram ajustados modelos com mais de 3 nós internos.

Tabela 4. 29: Modelos de possibilidades proporcionais com m nós internos: estimativa dos parâmetros e valor de AIC.

Modelo Flexível de Possibilidades Proporcionais				
Estimativa	$m = 0$	$m = 1$	$m = 2$	$m = 3$
Idade – “60 a 70 anos”	-0,530	-0,507	-0,506	-0,505
Idade – “> 70 anos”	-0,049	-0,038	-0,034	-0,036
Vive sozinho	0,437	0,396	0,421	0,417
Terapêutica	0,479	0,480	0,470	0,468
γ_0	-6,024	-8,884	-12,065	-14,481
γ_1	1,153	1,971	2,995	3,840
γ_2	-	0,089	0,436	0,533
γ_3	-	-	-0,368	-0,154
γ_4	-	-	-	-0,179
AIC	3623,74	3594,60	3578,99	3578,73

Segundo os resultados da Tabela 4.29 e tal como era de esperar, a introdução de nós internos origina uma diminuição do valor de AIC associado ao modelo, o que indica uma melhoria do ajustamento do mesmo. Note-se que o valor de AIC dos modelos com 2 e com 3 nós internos é idêntico, pelo que estes modelos estão igualmente bem ajustados. Uma vez mais, nestas condições optou-se por prosseguir com a análise do modelo mais simples.

Para o modelo flexível com 2 nós internos, a estimativa do logaritmo da função de possibilidade cumulativa de um indivíduo com vetor de covariáveis $\mathbf{x} =$ (Idade – “60 a 70 anos”, Idade – “> 70 anos”, Vive sozinho, Terapêutica) é dado por:

$$\ln \widehat{O}(t, \mathbf{x}) = -12,065 + 2,995 \times \ln t + 0,436 \times v_1(\ln t) - 0,368 \times v_2(\ln t) - 0,506 \times \text{"60 a 70 anos"} - 0,034 \times \text{">70 anos"} + 0,421 \times \text{"Vive sozinho"} + 0,470 \times \text{"Terapêutica"} \quad (3.64)$$

com,

$$v_1(\ln t) = (\ln t - 0,436)_+^3 - \lambda_1 (\ln t - (-12,065))_+^3 - (1 - \lambda_1) (\ln t - 2,995)_+^3$$

e

$$v_2(\ln t) = (\ln t - (-0,368))_+^3 - \lambda_2 (\ln t - (-12,065))_+^3 - (1 - \lambda_2) (\ln t - 2,995)_+^3$$

onde,

$$\lambda_1 = \frac{2,995 - 0,436}{2,995 - (-12,065)} = 0,170 \quad \text{e} \quad \lambda_2 = \frac{2,995 - (-0,368)}{2,995 - (-12,065)} = 0,223$$

Capítulo 5

Discussão e Conclusões

Um dos objetivos do presente trabalho de dissertação era identificar quais as características sociodemográficas e de saúde associadas ao tempo até à não persistência de doentes no tratamento de uma dada doença crónica, e perceber de que forma influenciam o risco de não persistência destes doentes.

Pretendeu-se comparar diferentes famílias de modelos, por forma a identificar qual é a que mais se adequa ao tipo de dados em análise, permitindo fazer inferências o mais próximo da realidade possível.

Assim sendo, numa fase inicial da análise estatística recorreu-se ao modelo de regressão logística, um método frequentemente utilizado em estudos longitudinais devido à sua fácil implementação. Através deste método foram selecionados três fatores que influenciam significativamente a não persistência de doentes no tratamento de uma doença crónica, obtendo-se um modelo múltiplo que inclui as variáveis Idade, Vive sozinho e Terapêutica.

Numa segunda fase da análise recorreu-se ao modelo semiparamétrico de Cox, atualmente um dos mais conhecidos e utilizados em análise de sobrevivência. Com base neste modelo verificou-se que as únicas variáveis com influência significativa sobre o tempo até à não persistência são: a Idade, Vive sozinho e a Terapêutica. Este resultado vai ao encontro do obtido na regressão logística, bem como de alguns estudos que identificam a Idade e a frequência da Terapêutica como fatores associados ao tempo até à não persistência (Bender BG, Pedan A, Varasteh LT, 2006; Casula M, Catapano AL, Piccinelli R, et al., 2014).

Apesar de se ter chegado a resultados concordantes, no que diz respeito à identificação dos fatores, é importante ter em conta que a regressão logística é um método que considera apenas se no final do seguimento os doentes são ou não persistentes e ignora o tempo até à não persistência, existindo por isso perda de informação. A análise de sobrevivência, por sua vez, é um método mais adequado para analisar este tipo de dados, uma vez que a variável resposta aqui considerada é o tempo até à ocorrência do acontecimento de interesse. Além disso, é de salientar que este método permite incluir dados censurados na análise.

Numa etapa seguinte do estudo, foram ajustados aos dados modelos de análise de sobrevivência mais complexos que o modelo de Cox. Assim, a análise prosseguiu com o ajustamento de três modelos paramétricos bem conhecidos: modelo de Weibull, log-normal e log-logístico. Estes modelos podem ser escritos sob diferentes formas, podendo as estimativas dos parâmetros ser comparadas entre si quando são representados na forma de modelos de tempo de vida acelerado.

Na análise univariada dos modelos paramétricos verificou-se que o modelo de Weibull origina as estimativas das funções de sobrevivência, associadas a cada variável, mais próximas das estimativas de Kaplan-Meier, sendo por isso a nível univariado a família de modelos que melhor parece descrever o tempo até à não persistência. Por sua vez, o modelo log-normal e o log-logístico são modelos que produzem resultados análogos, o que é evidente tanto a nível univariado (os gráficos das estimativas das curvas de sobrevivência são praticamente coincidentes), como a nível multivariado (os coeficientes estimados são muito próximos). No entanto, de todos os modelos de sobrevivência ajustados até esta etapa do estudo, o modelo log-normal tem associado o menor valor de AIC, sendo deste ponto de vista o modelo que melhor se ajusta aos dados analisados, de entre os modelos considerados.

A fácil implementação e interpretação do modelo de Cox motivam fortemente a sua ampla utilização. No entanto, é importante investir na especificação da função de risco subjacente, porque um

modelo paramétrico que seja adequado levará à obtenção de estimativas mais precisas do risco e da probabilidade de sobrevivência.

Nos últimos anos foi proposta e desenvolvida uma família de modelos paramétricos (Royston e Parmar, 2002; Royston e Lambert, 2011) que visa não só aumentar a flexibilidade face aos modelos paramétricos já conhecidos, como também manter a fácil interpretação do modelo de Cox.

De um modo geral, em ambas as famílias de modelos flexíveis (riscos proporcionais e possibilidades proporcionais), a nível univariado a inclusão de nós internos melhora o processo de estimação das funções de sobrevivência face ao modelo de Weibull e ao modelo log-logístico. O aumento do número de nós internos presentes no modelo (até um máximo de 3 nós) contribui também para esta melhoria fazendo diminuir o valor de AIC associado a cada modelo.

No entanto, são os modelos flexíveis de riscos proporcionais com 2 e 3 nós internos que originam as estimativas da função de sobrevivência mais próximas das estimativas de Kaplan-Meier. Na análise de regressão múltipla, estes são também os modelos que apresentam o menor valor de AIC, sendo por isso os modelos mais adequados e que melhor descrevem a função de risco e a função de sobrevivência. Uma vez que as estimativas dos coeficientes dos dois modelos são muito próximas e ambos os modelos estão igualmente bem ajustados aos dados, é preferível optar pelo modelo mais parcimonioso, que é o modelo com 2 nós internos.

Assim, de acordo com este modelo, para indivíduos com o mesmo valor nas restantes variáveis, um doente com idade entre os 60 e 70 anos tem um risco estimado de se tornar não persistente inferior ao de um doente com 60 anos ou menos, enquanto um doente com mais de 70 anos tem um risco estimado idêntico. Por sua vez, um doente que vive sozinho tem um risco estimado de interromper o tratamento superior ao de um doente que vive acompanhado, para doentes na mesma faixa etária e a fazer a mesma terapêutica. E por fim, um doente que está a fazer uma terapêutica semanal tem um risco estimado de se tornar não persistente superior ao de um doente que está a fazer uma terapêutica mensal, para doentes com os mesmos valores nas variáveis Idade e Vive sozinho.

Em suma, o modelo de Cox é muito utilizado uma vez que não é necessário especificar uma função de risco subjacente, tornando-o de fácil implementação e interpretação. Na presença de fortes indícios de que uma dada distribuição do tempo de vida é adequada, é preferível utilizar modelos paramétricos, visto serem mais eficientes. Por outro lado, os modelos de Royston e Parmar, além de serem modelos paramétricos, têm maior flexibilidade esperando-se por isso que representem de forma mais fidedigna a função de risco associada a cada grupo de indivíduos.

No presente estudo, a inclusão de nós internos no modelo flexível de riscos proporcionais melhorou bastante a estimativa da função de sobrevivência, especialmente face ao modelo de Weibull. O modelo flexível de riscos proporcionais com 2 nós revelou-se o modelo mais adequado e parcimonioso. De um modo geral, a utilização de modelos flexíveis na análise do efeito de covariáveis no tempo até à não persistência no tratamento de uma dada doença crónica, origina estimativas do risco relativo mais precisas do que utilizando outras famílias de modelos, contribuindo assim para um melhor conhecimento do fenómeno em estudo.

No futuro, seria interessante replicar o presente estudo para uma amostra de dimensão superior, uma vez que no modelo múltiplo alguns grupos continham poucos indivíduos. Igualmente importante seria que o tempo de seguimento fosse mais longo. Além disso, no presente estudo não foi explorada a não proporcionalidade nos modelos flexíveis, pelo que um próximo passo seria incluir variáveis que levem à existência de riscos ou possibilidades não proporcionais. Por fim, seria de interesse explorar diferentes doenças crónicas, por forma a verificar a adequabilidade dos modelos paramétricos flexíveis.

Bibliografia

- Bender BG, Pedan A, Varasteh LT. Adherence and persistence with fluticasone propionate/salmeterol combination therapy. *Journal of Allergy and Clinical Immunology* 2006; **118**: 899–904.
- Brown MT, Bussell JK. Medication adherence: WHO Cares? *Mayo Clinic Proceedings* 2011; **86**: 304–314.
- Casula M, Catapano AL, Piccinelli R, et al. Assessment and Potential Determinants of Compliance and Persistence to Antiosteoporosis Therapy in Italy. *American Journal of Managed Care* 2014; **20**: 138–145.
- Collett D. *Modelling Survival Data in Medical Research*. 2nd ed. Boca Raton: Chapman and Hall/CRC, 2003.
- Cox DR. Partial likelihood. *Biometrika* 1975; **62**: 269–276.
- Cramer JA, Roy A, Burrell A, et al. Medication Compliance and Persistence: Terminology and Definitions. *Value in Health* 2008; **11**: 44–47.
- Ferreira LN, Ferreira PL, Pereira LN, et al. EQ-5D Portuguese population norms. *Quality of Life Research* 2014; **23**: 425–430.
- Grambsch PM, Therneau TM. Proportional Hazards Test and Diagnostics Based on Weighted Residuals. *Biometrika* 1994; **81**: 515–526.
- Hall WJ, Wellner JA. Confidence Bands for a Survival Curve from Censored Data. *Biometrika* 1980; **67**: 133–143.
- Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. 3rd ed. New Jersey: John Wiley & Sons, 2013.
- Instituto Nacional de Estatística. *Estatísticas Demográficas 2012*. Lisboa: INE, 2013.
- Instituto Nacional de Estatística. *Saúde e Incapacidade em Portugal 2011*. Lisboa: INE, 2012.
- Kleinbaum DG, Klein M. *Survival Analysis: A Self-Learning Text*. 2nd ed. New York: Springer, 2005.
- Royston P, Lambert PC. *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. 1st ed. Texas: Stata Press, 2011.
- Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* 2002; **21**: 2175–2197.
- Sabaté E. *Adherence to long-term therapies: Evidence for action*. Switzerland: World Health Organization, 2003.
- Schoenfeld D. Partial Residuals for The Proportionnal Hazards Regression Model. *Biometrika* 1982; **69**: 239–241.
- World Health Organization. Constitution of WHO: principles, www.who.int/about/mission/en/ (1948, acedido 16 de Novembro de 2015).
- World Health Organization, National Institute of Health, National Institute on Aging, et al. *Global Health and Aging*. 11-7737, 2011.

Anexo A

Regressão Logística

Tabela A. 1: Análise Bivariada: Tabelas de contingência

		Persistente	Não persistente
Idade	≤ 60 anos	25	102
	Entre 60 e 70 anos	40	93
	> 70 anos	14	86
		Persistente	Não persistente
Nível de escolaridade	Sem ensino	11	29
	Básico	49	179
	Secundário ou Superior	19	73
		Persistente	Não persistente
Situação de emprego	Desempregado	22	61
	Empregado/Reformado	57	220
		Persistente	Não persistente
Vive sozinho	Não	71	209
	Sim	8	72
		Persistente	Não persistente
Exercício físico	Não	42	140
	Sim	37	141
		Persistente	Não persistente
Ida ao médico	1 vez ou menos por semestre	29	109
	Mais de uma vez por semestre	50	172
		Persistente	Não persistente
Doença crónica	Não	36	116
	Sim	43	165
		Persistente	Não persistente
Exame	Não	21	59
	Sim	58	222
		Persistente	Não persistente
Especialidade do médico	Médico da especialidade	16	64
	Outra	63	217

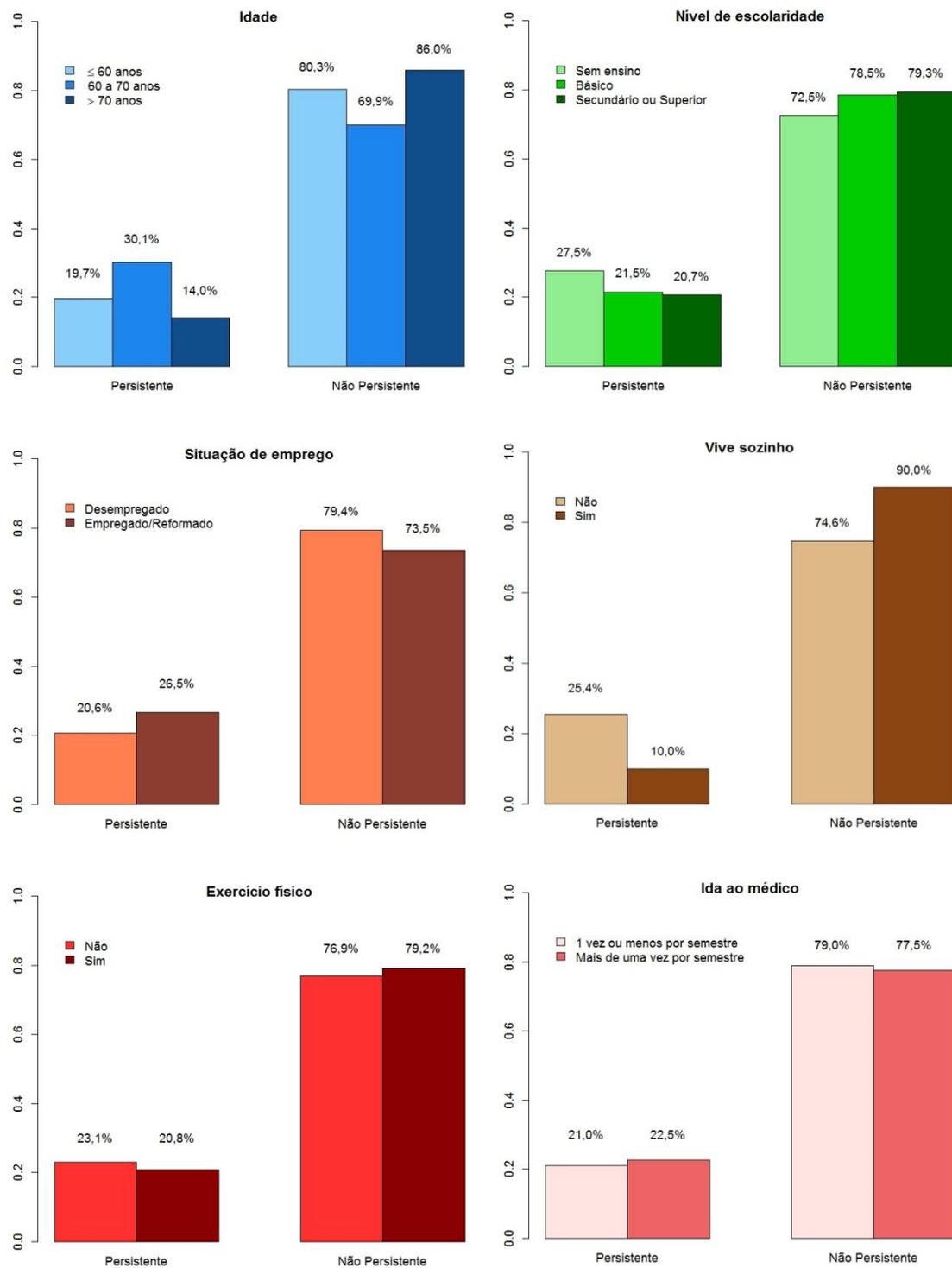


Figura A. 1: Análise Bivariada: Distribuição dos indivíduos, por estado de persistência no final do estudo segundo a classe de cada variável independente em análise.

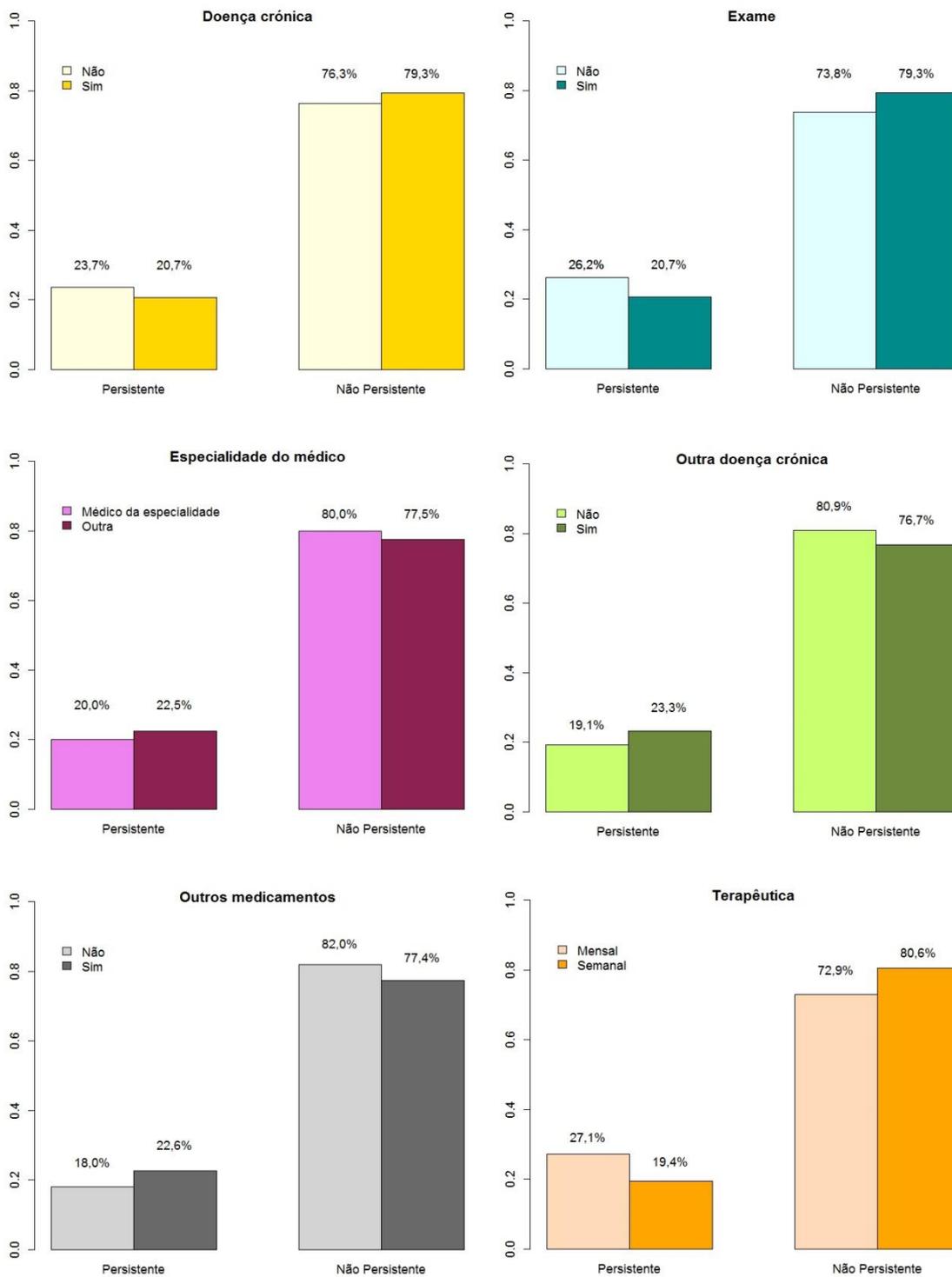


Figura A. 1: Análise Bivariada: Distribuição dos indivíduos, por estado de persistência no final do estudo segundo a classe de cada variável independente em análise (continuação da página anterior).

Anexo B

Análise de sobrevivência

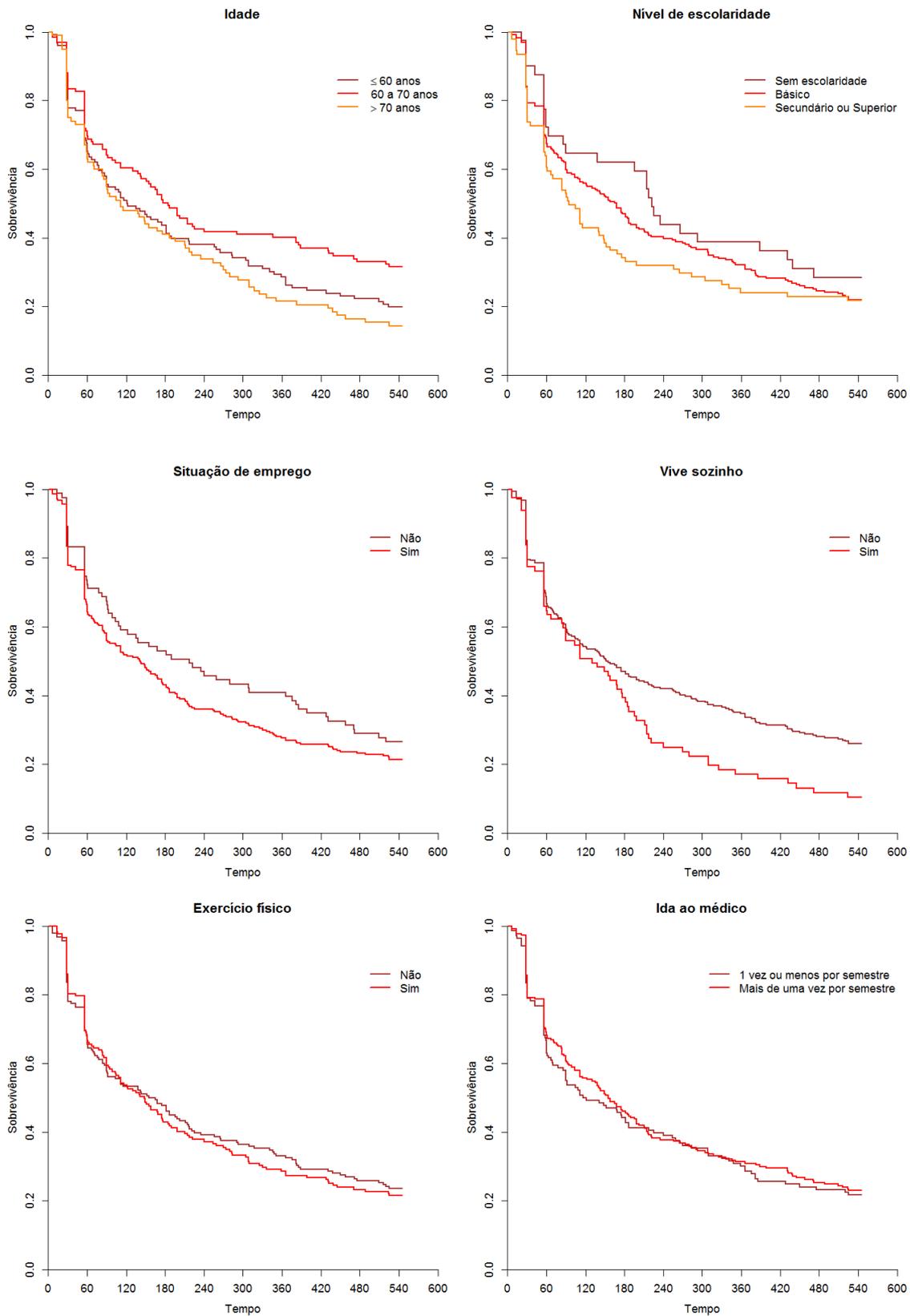


Figura B. 1: Gráficos das estimativas $\hat{S}(t)$ para cada variável independente em análise.

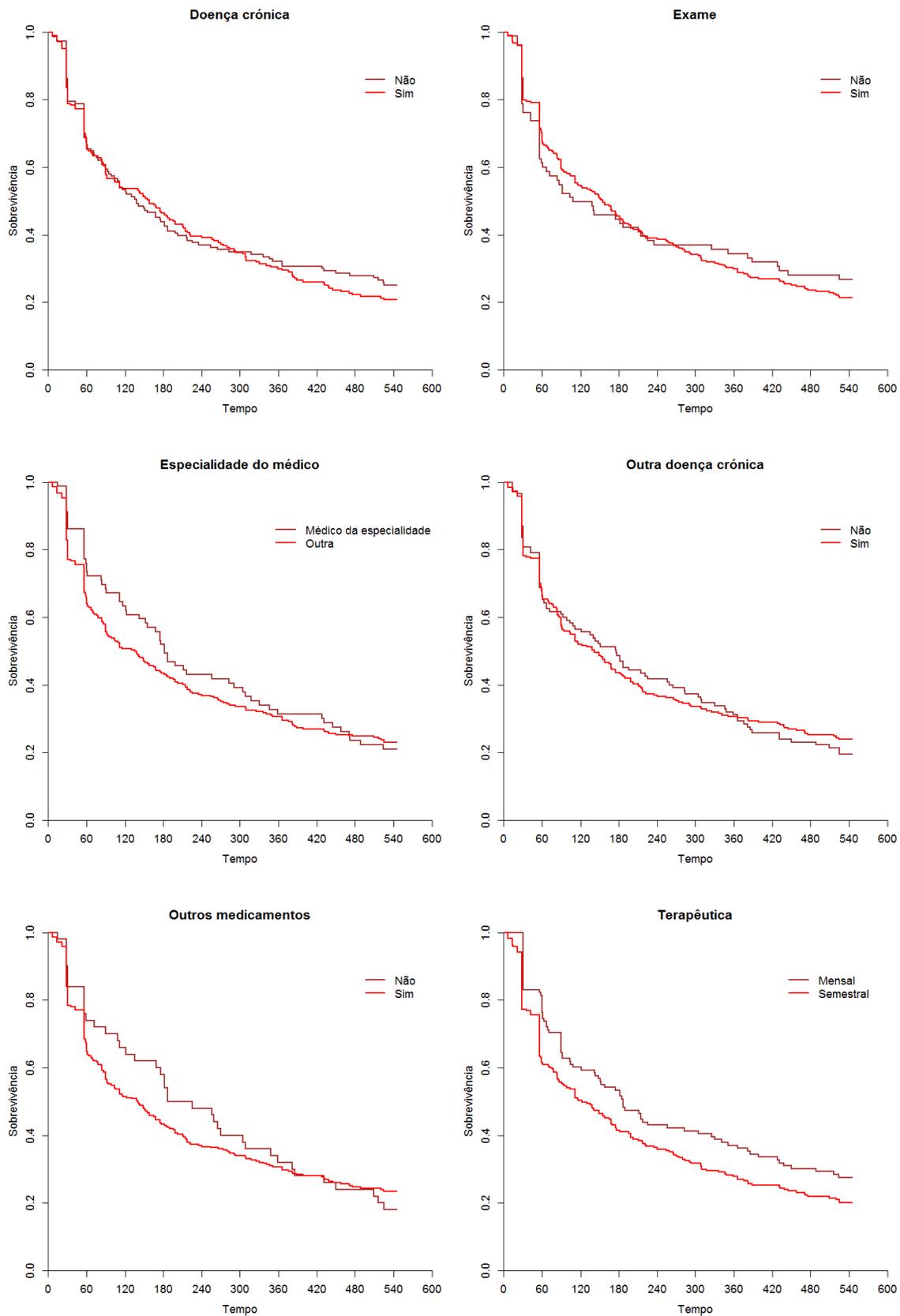


Figura B. 1: Gráficos das estimativas $\hat{S}(t)$ para cada variável independente em análise (continuação da página anterior).