

**UNIVERSIDADE ABERTA**



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

**ANÁLISE DE SOBREVIVÊNCIA AO CANCRO COLO-  
RETAL NOS PERÍODOS 2000-2001 E 2007-2008 NA  
REGIÃO NORTE DE PORTUGAL**

Carlos Manuel Falcão Ferreira Vasconcelos Correia

Mestrado em Bioestatística e Biometria

2016

# UNIVERSIDADE ABERTA



## **ANÁLISE DE SOBREVIVÊNCIA AO CANCRO COLO- RETAL NOS PERÍODOS 2000-2001 E 2007-2008 NA REGIÃO NORTE DE PORTUGAL**

Carlos Manuel Falcão Ferreira Vasconcelos Correia

Mestrado em Bioestatística e Biometria

Orientador:

Professor Doutor Pedro Serranho

Coorientadores:

Professora Doutora Maria do Rosário Ramos

Mestre Luís Antunes

2016

# Resumo

O cancro colo-retal é um dos tipos de cancro mais comuns, em ambos os sexos, tendo em conta os dados existentes nos registos oncológicos nas mais diversas regiões do globo, facto que acontece também na região Norte de Portugal.

O objetivo desta dissertação foi utilizar métodos de análise de sobrevivência tradicional e de análise de sobrevivência relativa, em dados do cancro colo-retal, considerando um período de observação de 5 anos. Os dados analisados são provenientes do Registo Oncológico Regional do Norte (RORENO), obtendo-se, após inspeção e tratamento prévio, um total de 2855 pacientes aos quais foi diagnosticado cancro colo-retal em dois períodos distintos, 2000-2001 e 2007-2008. Pretendeu-se também determinar de que forma as covariáveis influenciam a sobrevivência, e como esta evoluiu entre essas duas coortes. Dado não existir informação sobre as causas de morte, foi objetivo deste estudo estimar que parte da mortalidade observada se deve à doença em questão e que parte se deve a outras causas que afetam a mortalidade da população geral. Foram, também, estimados os coeficientes das covariáveis consideradas para os modelos adotados.

Todos os resultados foram obtidos utilizando como recurso o *software* R e seus pacotes adequados. Utilizaram-se diversos métodos tais como *Kaplan-Meier*, *Cox*, *Ederer II* e *Net Survival*, sendo este último considerado o mais adequado a este tipo de estudo pois é considerado pela comunidade científica o único estimador centrado para análise de sobrevivência relativa e cujas propriedades estatísticas permitem a sua utilização em estudos comparativos entre países. Verificou-se, por este estimador, que não existem diferenças significativas na sobrevivência em função do sexo, idade, distrito ou localização do cancro (cólon ou reto). Por outro lado verificaram-se diferenças significativas para a sobrevivência em função do estadio e das coortes, registando-se um acréscimo global de 8.5% na sobrevivência aos 5 anos de 2000-2001 para 2007-2008. Por fim destacam-se alguns aspetos que merecem atenção em estudos futuros.

**Palavras-chave:** Cancro, Colo-retal, Análise de Sobrevivência Relativa, covariáveis, *Net Survival*

# Abstract

## Colorectal Cancer Survival Analysis in the periods 2000-2001 and 2007-2008 in the North of Portugal

Colorectal cancer is one of the most common types of cancer affecting both genders according to cancer registries all over the world and also at the North Portugal's Cancer Registry.

The objective of this dissertation was to study colorectal cancer survival by application of traditional survival analysis and relative survival analysis, for a 5 year observation period. Analyzed data was obtained from the local cancer registry (RORENO), and after data inspection and treatment, lead to a total of 2855 patients diagnosed with colorectal cancer in two distinct periods, 2000-2001 and 2007-2008.

We intended to determine how covariates affect survival and how has survival evolved between those two cohorts. Given that we do not have any information about death causes, determining what mortality portion was caused by the specific disease and what portion was caused by other causes of the general population, is also an objective. Covariates coefficients were also estimated for the adopted models.

All results were obtained using the R software and appropriate packages. Several methods as Kaplan-Meier, Cox, Ederer II and Net Survival were used, being the latter considered by scientific community the only unbiased estimator for relative survival analysis, and the only eligible for comparisons between countries. Using this estimator there are no significant survival differences stratifying data by sex, age, residence or cancer localization (colon or rectum). On the other hand, those differences are significant stratifying data by cancer stage and by cohorts, increasing 5-year survival by 8.5% from 2000-2001 to 2007-2008. Finally, we highlight some aspects that deserve attention in future studies.

**Keywords:** Cancer, Colorectal, Relative Survival Analysis, covariates, Net Survival

# Agradecimentos

Mais uma etapa. Antes de mais agradeço ao autor desta dissertação pela sua extraordinária capacidade de trabalho e pela sua incessante busca do conhecimento nas suas mais diversificadas áreas, em acumulação com uma atividade profissional exigente.

Aos meus pais pelos valores e princípios que sempre me transmitiram e por terem sido aqueles que sempre e desde sempre me incentivaram e acreditaram em mim.

Às duas meninas que tenho em casa por terem compreendido o facto de não terem podido desfrutar da minha companhia e do salutar convívio em família durante muitas noites, fins de semana e períodos de férias. Por terem que coabitar, inúmeras vezes, com a minha presença ausente.

Ao RORENO do IPO do Porto e à Prof. Doutora Maria José Bento pela sua simpatia e pronta disponibilidade autorizando a utilização da matéria-prima sem a qual esta dissertação não existia, os dados.

Ao Eng<sup>o</sup> Luís Antunes pela materialização da disponibilização dos dados e por todas as suas preciosas indicações e sugestões ao longo de todo este processo.

À Prof. Doutora Maria do Rosário Ramos pela sua visão sempre assertiva e pragmática, pelas suas úteis recomendações, pela sua simpatia e por todo o acompanhamento dado.

Ao Prof. Doutor Pedro Serranho pelo facto de ser um excelente profissional, tendo demonstrado um empenho e dedicação inexcedíveis, tanto na parte curricular do mestrado como durante a orientação da dissertação. Obrigado por me ter acompanhado em todo este trajeto!

Ao anterior acordo ortográfico que não me fazia sentir um estrangeiro na minha própria língua.

Às minhas referências de sempre: ao meu tio Eduardo, ao Prof. Doutor António de Oliveira Salazar e a Charles Maurras.

Aos amigos e a todos aqueles que me querem bem.

# Lista de Abreviaturas

**AJCC** – *American Joint Committee on Cancer.*

**AS** – Análise de Sobrevivência.

**ASR** – Análise de Sobrevivência Relativa.

**ACS** – *American Cancer Society.*

**BB** – *Brownian Bridge*, ponte Browniana.

**CDC** – *Centers for Disease Control and Prevention.*

**CSG** - *Cancer Survival Group da London School of Hygiene.*

**CCR** – Cancro colo-retal.

**GLM** – *Generalized Linear Model* - modelo linear generalizado.

**IARC** – *International Agency for Research on Cancer.*

**IC** – Intervalo de Confiança.

**ICD-O** – *International Classification of Diseases for Oncology.*

**ICSS** - *International Cancer Survival Standard*

**NOS** – *No Other Specification* - sem outra especificação.

**NCI** – *National Cancer Institute*

**NS** – *Net Survival*

**P** – *p-value*

**RORENO** – Registo Oncológico Regional do Norte.

**SE** – *Standard Error* - erro padrão.

**TNM** – *Tumors, Nodes, Metastases* – Tumores, Nódulos, Metástases.

**UICC** – *Union for International Cancer Control.*

**WHO** – *World Health Organization.*

# Índice

Resumo .....	I
Abstract .....	II
Agradecimentos .....	III
Lista de Abreviaturas.....	IV
Índice.....	V
Índice Figuras .....	VIII
Índice Tabelas.....	X
1. Introdução.....	1
2. Enquadramento Teórico .....	3
2.1. Cancro Colo-retal .....	3
2.1.1. Definição .....	3
2.1.2. Localização (Codificação Topográfica) .....	5
2.1.3. Histologia (Codificação Morfológica) .....	7
2.1.4. Estadiamento .....	9
2.1.5. Evolução Científica na área clínica .....	14
2.2. Análise de Sobrevivência.....	16
2.2.1. Definição .....	16
2.2.2. Evento de interesse ou falha.....	16
2.2.3. Período de observação.....	17
2.2.4. Censura .....	17
2.2.5. Função de Sobrevivência.....	18
2.2.6. Função Taxa de Falha .....	19
2.2.7. Estimadores e Modelos.....	21
2.3. Análise de Sobrevivência Relativa.....	23
2.3.1. Tábuas de Mortalidade das Populações .....	24
2.3.2. Sobrevivência Relativa .....	25
2.3.3. Estimadores de Sobrevivência Relativa .....	26
2.3.4. Estimador Net Survival.....	27
2.3.5. Modelos de Regressão .....	28
3. Opções Metodológicas.....	31
3.1. Caracterização dos dados .....	31

3.1.1.	Dados iniciais.....	31
3.1.2.	Cr�terios de inclus�o e exclus�o de indiv�duos.....	33
3.1.3.	Tratamento dos dados para AS.....	34
3.1.4.	Caracter�sticas dos dados tratados .....	36
3.2.	Dados de Mortalidade Geral .....	39
3.2.1.	T�buas de Mortalidade .....	39
3.2.2.	Tratamento das t�buas para ASR.....	40
3.3.	An�lise de Sobreviv�ncia.....	42
3.3.1.	Estimador n�o param�trico de Kaplan-Meier.....	42
3.3.2.	Modelo de Cox de riscos proporcionais.....	43
3.4.	An�lise de Sobreviv�ncia Relativa.....	49
3.4.1.	Estimadores.....	49
3.4.2.	Modelos Aditivos (Excess Mortality).....	54
3.4.3.	Modelo Multiplicativo (Relative Mortality) .....	58
4.	Apresenta�o dos Resultados .....	60
4.1.	Estimativas de Kaplan-Meier .....	60
4.2.	Modelo de Cox de riscos proporcionais.....	68
4.3.	ASR – Estimac�o .....	78
4.3.1.	Sobreviv�ncia Relativa .....	78
4.3.2.	Net Survival .....	87
4.4.	ASR – Modela�o .....	95
4.4.1.	Modelos Aditivos.....	95
4.4.2.	Modelo Multiplicativo.....	104
5.	Considera�es Finais .....	109
5.1.	Principais resultados obtidos .....	109
5.1.1	Estimativas .....	110
5.1.2.	Modelos de regress�o.....	116
5.2.	Aprendizagem e compet�ncias adquiridas.....	119
5.3.	Limita�es do estudo e perspectivas futuras .....	122
5.4.	S�ntese e conclus�es finais.....	125
6.	Anexos.....	127
6.1.	Comandos R utilizados .....	127
6.1.1.	Kaplan-Meier.....	127
6.1.2.	Cox.....	128



6.1.3. ASR .....	128
6.2. Ratetables .....	130
Referências.....	134

# Índice Figuras

Figura 1 Diagrama do cólon e do reto com os códigos topográficos.....	7
Figura 2 Paredes do cólon e reto .....	12
Figura 3 Função Sobrevivência vs Distribuição Acumulada .....	19
Figura 4 Funções de Taxa de Falha .....	20
Figura 5 Exemplo de resultados de sobrevivência observada, esperada relativa .....	26
Figura 6 Estimativas Kaplan-Meier para as duas coortes com IC95%.....	61
Figura 7 Comparação das estimativas Kaplan-Meier para as duas coortes.....	61
Figura 8 Comparação das taxas de mortalidade entre as duas coortes .....	63
Figura 9 Kaplan-Meier – Curvas de sobrevivência por estadio, para as duas coortes.....	64
Figura 10 Kaplan-Meier – Curvas de sobrevivência por grupos de idade, para as duas coortes ....	65
Figura 11 Cox – Resíduos de Schoenfeld por grupos de idade .....	70
Figura 12 Cox – Resíduos de Schoenfeld por sexo .....	70
Figura 13 Cox – Resíduos de Schoenfeld por grupo histológico .....	71
Figura 14 Cox – Resíduos de Schoenfeld por grupo topográfico .....	71
Figura 15 Cox – Resíduos de Schoenfeld por estadio .....	72
Figura 16 Cox – Resíduos de Schoenfeld por coortes .....	73
Figura 17 Cox – Sobrevivência em função da idade.....	75
Figura 18 Cox – Sobrevivência em função do sexo .....	76
Figura 19 Cox – Sobrevivência em função da histologia .....	76
Figura 20 Cox – Sobrevivência em função da topografia.....	77
Figura 21 Cox – Sobrevivência em função do estadio .....	77
Figura 22 Sobrevivência Esperada para as duas coortes .....	78
Figura 23 Histograma – distribuição das idades nas coortes.....	79
Figura 24 Sobrevivência observada, esperada e relativa para as duas coortes.....	81
Figura 25 Sobrevivência Relativa por Sexo .....	82
Figura 26 Sobrevivência Relativa por grupos de idade na coorte 1.....	83
Figura 27 Sobrevivência Relativa por grupos de idade na coorte 2.....	84
Figura 28 Sobrevivência Relativa por estadio na coorte 1.....	85
Figura 29 Sobrevivência Relativa por estadio na coorte 2.....	85
Figura 30 Comparação estimativas Ederer II vs Pohar-Perme por sexos.....	87
Figura 31 Mortalidade acumulada – Causas específica e populacional na coorte 1 .....	93
Figura 32 Mortalidade acumulada – Causas específica e populacional na coorte 2 .....	93
Figura 33 Modelo Aditivo Estève – Brownian Bridge para as covariáveis .....	98
Figura 34 Modelo Aditivo Estève – Brownian Bridge para o modelo global .....	99
Figura 35 Modelo Aditivo Estève – Resíduos de Schoenfeld por sexo .....	101
Figura 36 Modelo Aditivo Estève – Resíduos de Schoenfeld por grupo de idade .....	101
Figura 37 Modelo Aditivo Estève – Resíduos de Schoenfeld por estadio.....	102
Figura 38 Modelo Aditivo Estève – Resíduos de Schoenfeld por coortes.....	102
Figura 39 Mortalidade em Excesso para ambas as coortes.....	103
Figura 40 Mortalidade em Excesso acumulada para ambas as coortes .....	103
Figura 41 Modelo Multiplicativo de Andersen – Brownian Bridge para o modelo global.....	106
Figura 42 Modelo Multiplicativo Andersen – Resíduos de Schoenfeld por sexo.....	106
Figura 43 Modelo Multiplicativo Andersen – Resíduos de Schoenfeld por grupo de idade.....	107

Figura 44 Modelo Multiplicativo Andersen – Resíduos de Schoenfeld por coortes.....	107
Figura 45 Modelo Multiplicativo Andersen – Resíduos de Schoenfeld por estadio .....	108

# Índice Tabelas

Tabela 1 Estadiamento anatómico em função de TNM.....	13
Tabela 2 Variáveis iniciais obtidas do registo .....	31
Tabela 3 Estatísticas descritivas das variáveis iniciais.....	32
Tabela 4 Padrões ICSS por tipo de cancro (Corazziari I and M 2015) .....	35
Tabela 5 Padrões ICSS para grupos de idade para sobrevivência (Corazziari I and M 2015) .....	36
Tabela 6 Estatísticas descritivas relevantes após tratamento dos dados.....	37
Tabela 7 Estrato tábua de mortalidade utilizada para CONCORD-2 (Spika D and Bonaventure A). 39	
Tabela 8 Estimativa Kaplan-Meier para sobrevivência para a Coorte 1 .....	60
Tabela 9 Estimativa Kaplan-Meier para sobrevivência para a Coorte 2 .....	60
Tabela 10 Kaplan-Meier - Teste log-rank para comparação da sobrevivência entre coortes .....	62
Tabela 11 Kaplan-Meier - Tempos medianos de vida e sobrevivência nas duas coortes.....	62
Tabela 12 Kaplan-Meier - Sobrevivência aos 5 anos por estadio .....	64
Tabela 13 Kaplan-Meier - Sobrevivência aos 5 anos por grupo de idade.....	65
Tabela 14 Kaplan-Meier - Sobrevivência aos 5 anos por sexo.....	66
Tabela 15 Kaplan-Meier - Sobrevivência aos 5 anos por distrito.....	66
Tabela 16 Kaplan-Meier - Sobrevivência aos 5 anos por topografia .....	66
Tabela 17 Kaplan-Meier - Sobrevivência aos 5 anos por grupo topográfico .....	67
Tabela 18 Kaplan-Meier - Sobrevivência aos 5 anos por histologia .....	67
Tabela 19 Kaplan-Meier - Sobrevivência aos 5 anos por grupo histológico .....	67
Tabela 20 Cox –Teste Wald com todas as covariáveis.....	68
Tabela 21 Cox –Teste Wald com covariáveis significativas a 5% .....	69
Tabela 22 Cox – Teste de Correlação dos resíduos de Schoenfeld.....	73
Tabela 23 ASR –Sobrevivência esperada da população para as 2 coortes aos 5 anos .....	80
Tabela 24 ASR –Sobrevivência Esperada, Observada e Relativa para as duas coortes – 5 anos ....	81
Tabela 25 ASR –Sobrevivência Relativa por sexo.....	82
Tabela 26 ASR –Sobrevivência Relativa por grupo de idade.....	83
Tabela 27 ASR –Sobrevivência Relativa por estadio .....	84
Tabela 28 ASR –Sobrevivência Relativa por distrito .....	86
Tabela 29 ASR –Sobrevivência Relativa por grupo topográfico.....	86
Tabela 30 ASR –Sobrevivência Relativa por grupo histológico .....	86
Tabela 31 ASR - Sobrevivência por sexo .....	87
Tabela 32 ASR – Net Survival sexo por estadio .....	88
Tabela 33 ASR - Sobrevivência por grupo de idade .....	88
Tabela 34 ASR – Net Survival Idade por estadio .....	89
Tabela 35 ASR - Sobrevivência por estadio.....	89
Tabela 36 ASR - Sobrevivência por distrito .....	90
Tabela 37 ASR – Net Survival distrito por estadio.....	90
Tabela 38 ASR - Sobrevivência por grupo topográfico.....	91
Tabela 39 ASR – Net Survival grupo topográfico por estadio .....	91
Tabela 40 ASR - Sobrevivência por grupo histológico.....	92
Tabela 41 ASR – Net Survival grupo histológico por estadio .....	92
Tabela 42 ASR – Taxa de mortalidade acumulada – Causas específica e populacional.....	93
Tabela 43 ASR – Resultados globais sobrevivência e nº de mortes.....	94

Tabela 44 ASR – Modelo Aditivo Estève com todas as covariáveis .....	95
Tabela 45 ASR – Brownian Bridge para Modelo Aditivo Estève com todas as covariáveis.....	96
Tabela 46 ASR – Modelo Aditivo Estève ajustado.....	97
Tabela 47 ASR – Brownian Bridge para Modelo Aditivo Estève ajustado.....	98
Tabela 48 ASR – Modelos Aditivos – comparação entre coeficientes e seus erros padrão .....	100
Tabela 49 ASR – Modelo Multiplicativo de Andersen com todas as covariáveis.....	104
Tabela 50 ASR – Modelo Multiplicativo de Andersen ajustado.....	105
Tabela 51 ASR – Brownian Bridge para Modelo Multiplicativo Anderson ajustado.....	105
Tabela 52 Resultados gerais Net Survival .....	110
Tabela 53 Comparação entre razões de taxa de mortalidade .....	117

# 1. Introdução

Neste estudo pretendeu-se aplicar métodos de Análise de Sobrevida a dados de doentes oncológicos, por forma a estimar a sua função de sobrevivência, verificar qual a importância das covariáveis para essa mesma função, e qual a interpretação dos respetivos coeficientes em modelos de regressão.

Em Análise de Sobrevida a variável resposta é o tempo decorrido até à ocorrência de um evento de interesse. As técnicas estatísticas clássicas de análise de regressão não conseguem tratar dados com a presença de censura, isto é, quando existe uma observação parcial da resposta motivada pela interrupção da observação para determinado objeto de estudo ou paciente, pela não ocorrência do evento de interesse ou falha durante o período de observação ou pela ocorrência de falha devido a causas externas ou eventos competitivos, (Colosimo E. A. 2006). Através da Análise de Sobrevida podemos estimar as funções de sobrevivência (probabilidade de uma observação não falhar até um certo tempo  $t$ ) e de taxa de falha (probabilidade da falha ocorrer num determinado intervalo ínfimo de tempo).

Neste estudo aplicaram-se técnicas de Análise de Sobrevida em dados relativos a cancro colo-retal do Registo Oncológico Regional do Norte (RORENO) sediado no Instituto Português de Oncologia do Porto (IPO-Porto).

Foram analisadas duas coortes, isto é, dois conjuntos de sujeitos em observação, com as mesmas características demográficas e que partilham a mesma experiência, evento ou condição (C.R. Rao 2007), sendo que no presente estudo essa condição comum, em cada uma das coortes, se refere à data de diagnóstico do cancro colo-retal nos períodos entre 2000-2001 e 2007-2008.

Tendo em conta o tipo de dados disponibilizados pelo RORENO e aqui analisados, o evento de falha considerado é a morte do doente, sendo o tempo de falha o tempo que decorre desde o diagnóstico até à morte.

Dado que, neste tipo de dados, a informação sobre as causas de morte está muitas vezes indisponível ou não é fiável, foram utilizados métodos de Análise de Sobrevida Relativa, sendo apresentadas as vantagens desses métodos relativamente aos tradicionais para o caso específico em estudo. A Análise de Sobrevida Relativa, compara a sobrevivência observada com a sobrevivência

esperada num grupo, da população geral, com as mesmas características demográficas e permitiu-nos estimar a mortalidade associada ao diagnóstico do cancro na ausência de informação sobre a causa de morte. Essa informação, a existir, deveria levar a que as mortes por causas terças fossem tratadas como censuras (Lambert P. 2009).

Após esta introdução, o enquadramento teórico que suporta esta dissertação é apresentado no capítulo 2. Introduce-se o conceito de Análise de Sobrevivência e as suas abordagens tradicionais não paramétricas e semiparamétricas. Introduce-se, também o conceito de Análise de Sobrevivência Relativa e a sua pertinência para o estudo em questão. Aqui foram abordados alguns dos métodos de estimação e modelação mais utilizados neste tipo de análise, nomeadamente o estimador *Net Survival* e os modelos de regressão aditivos e multiplicativos.

O capítulo 3 que se segue apresenta o método, começando por introduzir o problema de estudo que consiste na estimação e modelação da sobrevivência, tendo como base registos oncológicos relativos a diagnósticos de cancro colo-retal em duas coortes constituídos por pacientes da região norte de Portugal cujo diagnóstico foi registado em dois períodos: 2000-2001 e 2007-2008. São apresentados os procedimentos de recolha, processamento e tratamento estatístico, bem como a descrição das variáveis em equação. Aqui são apresentados os métodos de análise de sobrevivência introduzidos no capítulo de Enquadramento Teórico e utilizados no restante trabalho.

No capítulo 4 são apresentados os resultados deste estudo, sendo comparados os resultados das duas coortes em questão. São, também, comparados os resultados obtidos através de diferentes métodos. É utilizada a ferramenta computacional R e os seus pacotes mais adequados para o estudo em questão para a determinação de todos os resultados, incluindo a componente gráfica.

Termina-se, no capítulo 5 com a síntese do estudo e dos principais resultados obtidos, com as limitações do mesmo, com a aprendizagem e competências adquiridas e com as conclusões finais.

## **2. Enquadramento Teórico**

### **2.1. Cancro Colo-retal**

#### **2.1.1. Definição**

O cancro resulta de efeitos acumulados de alterações genéticas, somáticas ou hereditárias, e mudanças fisiológicas na célula que se traduzem na manifestação da doença. Hanahan e Weinberg, (Hanahan D. 2011) enunciam um conjunto de características essenciais na fisiologia celular para que seja possível o crescimento tumoral: autossuficiência em sinais de crescimento, insensibilidade a sinais de inibição de crescimento, evasão da apoptose, potencial replicativo ilimitado, indução da angiogénese, invasão de tecidos e metástase, reprogramação do metabolismo energético e, por último, evasão à imuno-destruição. Estas alterações implicam num funcionamento inadequado de moléculas responsáveis pelo controlo da multiplicação celular.

O advento da biologia molecular e celular e das técnicas em análise patológica permitem hoje uma compreensão muito mais profunda da doença, em comparação com o que se sabia sobre a mesma algumas décadas atrás, nomeadamente com a sequenciação do genoma humano, e em particular do genoma ligado ao cancro. Quando as células apresentam alterações genéticas no seu DNA, motivadas por um acontecimento raro e imprevisível, de entre os acima enunciados, que as modifica e dota de capacidade de se multiplicarem muito rapidamente, poderá ser formado um tumor. Quando este conjunto de células permanece localizado estamos perante um tumor benigno. Pelo contrário, quando não permanece localizado e adquire capacidade de mobilidade a qual lhe permite invadir outros tecidos, através dos sistemas sanguíneo ou linfático, estamos perante um tumor maligno ou cancro. Sempre que essa invasão ocorre estamos perante a ocorrência de metástases (Silva 2011).

Independentemente dos locais por onde se espalha o cancro, este é sempre classificado de acordo com o local onde se inicia.



O Cancro Colo-retal, doravante descrito como CCR é um cancro que se inicia no cólon ou no reto. Estes tipos de cancro podem ser referidos separadamente como cancros do cólon ou cancros do reto, no entanto são normalmente associados devido ao facto de apresentarem muitas características comuns, nomeadamente a nível topográfico, ou seja, a nível da sua localização de origem.

O cólon e o reto fazem parte do sistema digestivo. Neste sistema, o estômago e o intestino delgado processam a energia a partir da alimentação, enquanto que o cólon e o reto, situados no intestino grosso, absorvem os fluídos de forma a gerar desperdícios sólidos, os quais são libertados pelo organismo.

O cólon constitui a maior parte do intestino grosso e é composto por 4 secções principais:

- Ascendente, inicia no cego, o qual faz a ligação ao intestino delgado e se estende para cima do lado direito do abdómen.
- Transverso, assim designado pois atravessa do lado direito para o lado esquerdo na parte superior do abdomen,
- Descendente, estende-se para baixo do lado esquerdo.
- Sigmoide, assim designado devido à sua forma em “S”.

Em termos de localização topográfica serão consideradas em termos de codificação, para além destas 4 secções principais, algumas subsecções tal como veremos adiante.

Quanto ao reto, este recebe os desperdícios resultantes do processamento no cólon, as fezes, e armazena-as até serem libertadas pelo corpo.

As paredes do cólon são compostas por algumas camadas e o CCR tem origem na camada mais interna e pode crescer e estender-se para algumas ou todas as camadas mais externas. O estadiamento do CCR depende em grande parte do grau de crescimento do mesmo para as camadas externas.

Tipicamente os CCR desenvolvem-se lentamente ao longo dos anos, mas antes do seu aparecimento formam-se normalmente tumores benignos designados por pólipos. Esses pólipos podem transformar-se em cancro, sendo que essa probabilidade depende do tipo de pólipo em questão. Os pólipos adenomatosos ou adenomas apresentam maior probabilidade de se tornarem CCR que os pólipos

inflamatórios, o que se leve a considerar na sua presença que o paciente se encontra em condição pré cancerígena.

Quando o cancro se forma, este cresce para as camadas externas do cólon, e aí pode chegar aos vasos sanguíneos e linfáticos e daí para os gânglios linfáticos mais próximos ou para partes mais distantes do corpo, metastizando.

A maior consciencialização para os processos preventivos permite que em cada vez mais casos seja possível detetar os pólipos antes que estes se transformem, eventualmente, em cancro (ACS 2014).

Quanto aos fatores de risco que aumentam, normalmente, a probabilidade de desenvolver CCR, para além da idade (a partir dos 50 anos o risco aumenta significativamente), outros fatores têm vindo a ser apontados como de risco, tais como inatividade física, má alimentação, obesidade, álcool e tabaco (ACS 2014).

### **2.1.2. Localização (Codificação Topográfica)**

A *International Agency for Research on Cancer* (IARC) é a agência especializada em cancro da *World Health Organization* (WHO).

O objetivo da IARC, (<http://codes.iarc.fr/>), é promover a colaboração internacional na pesquisa do cancro recorrendo às suas valências nas áreas de epidemiologia, ciências laboratoriais e bioestatística, de forma a identificar as causas do cancro e fazer com que sejam adotadas medidas preventivas, tendo como objetivo reduzir a sua incidência.

A WHO criou através desta agência codificações para doenças oncológicas, tanto a nível de localização topográfica como de histologia, as quais têm sido utilizadas desde 1976, ano em que foi editada a primeira edição da *International Classification of Diseases for Oncology* (ICD-O).

Estas codificações vão sendo atualizadas, acompanhando a evolução científica e as suas novas descobertas, levando a que tivesse sido editada a segunda edição em 1990, a terceira em 2000 e uma revisão em 2011.

O dígito 3 na nomenclatura ICD-O-3 refere-se ao facto da codificação utilizada atualmente ser relativa à terceira edição do ICD-O.

A codificação topográfica, constituindo uma das formas de caracterização dos cancros, dá-nos a indicação do local onde é detetado o CCR de entre as várias secções do cólon e do reto,

O código tem sempre um prefixo C, seguido por dois dígitos que indicam a localização, e separado por um ponto, um dígito que indica a sublocalização.

Abaixo é apresentada a atual codificação topográfica ICD-O-3 para a zona colorretal. De notar que o código C18.9 se refere a um CCR com origem no cólon mas para o qual não foi possível determinar a secção específica do mesmo, sendo que a sigla NOS significa *No Other Specification* (sem outra especificação).

Tendo em conta as mais recentes atualizações, a codificação topográfica para CCR é apresentada abaixo e é ilustrada na Figura 1.

## **C18 CÓLON**

- C18.0 Ceco
- C18.1 Apêndice
- C18.2 Cólon ascendente
- C18.3 Ângulo hepático
- C18.4 Cólon transverso
- C18.5 Ângulo esplénico
- C18.6 Cólon descendente
- C18.7 Cólon sigmoide
- C18.8 Cólon com lesão invasiva
- C18.9 Cólon, NOS

## **C19 JUNÇÃO RETOSIGMOIDE**

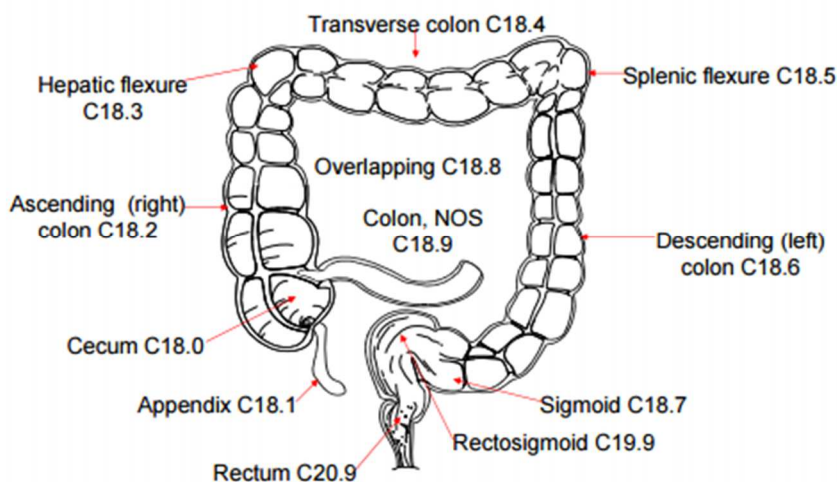
- C19.9 Junção retosigmoide

## **C20 RETO**

- C20.9 Reto, NOS

Figura 1 Diagrama do cólon e do reto com os códigos topográficos

(CDC 2012)



### 2.1.3. Histologia (Codificação Morfológica)

Quanto à sua histologia, ou seja, a caracterização do seu tecido biológico a nível de formação, estrutura e função, o CCR pode ser de vários tipos.

Tal como para a localização, a IARC criou codificações ICD-O para classificar os diferentes tipos de cancro quanto à sua histologia.

O código morfológico descreve as características do tumor em si, incluindo o seu tipo de célula e a sua atividade biológica. Este código é composto por quatro dígitos que indicam o tipo de célula ou histologia e, separado por uma barra (/), um dígito que indica o comportamento.

Esse dígito pode apresentar os seguintes valores:

- 0 – benigno
- 1 – comportamento incerto
- 2 – tumor in situ (células não passam da camada superficial dos tecidos, não existindo rutura da membrana basal)
- 3 – maligno na localização primária
- 6 – maligno nas metástases distantes
- 9 – maligno, em localização incerta

Consultando o website <http://codes.iarc.fr/> , conseguimos verificar, à data de hoje, que existe um número de códigos histológicos um pouco superior a mil e trezentos, tendo em conta todas as localizações classificadas atualmente desde o código 8000/0 até ao código 9992/3 (de notar que existem muitos códigos dentro desses limites aos quais não estão associadas quaisquer morfologias, tais como os pertencentes aos intervalos 9568-9649 ou 9770-9799). Se tivermos em conta, apenas, as localizações relativas ao CCR, C18-C20, restringimos o número de códigos às seguintes gamas, segundo o AJCC Cancer Staging Handbook, (AJCC 2010):

8000-8152, 8154-8231, 8243-8245, 8247-8248, 8250-8576, 8940-8950 e 8980-8981.

De acordo com a publicação (ACS 2005) da American Cancer Society, mais de 95% dos CCR são do tipo Adenocarcinoma, os quais se formam nas glândulas que geram o muco que lubrifica o interior do cólon e reto. Existem outros tipos de CCR, menos comuns, que podem começar no cólon e no reto, tais como os que surgem a partir dos tumores carcinóides (com origem em células produtoras de hormonas), tumores estromais gastrointestinais (com origem em células intersticiais da parede do cólon), linfomas (com origem em células do sistema imunitário) e sarcomas (com origem nos vasos sanguíneos, músculo e tecido conjuntivo da parede do cólon e reto), (ACS 2014).

De acordo com o ICD-O-3, histologias com os três primeiros dígitos de 814 até 848 são classificadas como adenomas e adenocarcinomas (excluindo nesse intervalo os códigos 824x – tumores neuroendócrinos ou carcinóides).

Os códigos mais comuns (CDC 2012), neste contexto, para CCR são:

**8140/3** adenocarcinoma, NOS

**8210/3** adenocarcinoma em polipose adenomatosa

**8220/3** adenocarcinoma em polipose adenomatosa do cólon

**8261/3** adenocarcinoma em adenoma viloso

**8480/3** adenocarcinoma mucinoso

## 2.1.4. Estadiamento

O estadiamento do cancro descreve o grau de severidade do cancro baseado na magnitude do tumor original e na dispersão que possa ter ocorrido pelo corpo (AJCC 2015).

O *American Joint Committee on Cancer*, AJCC publica, nas suas edições, *Cancer Staging Manual* e *Cancer Staging Atlas*, os dados mais detalhados sobre estadiamento de cancro. Essas publicações são reconhecidas como referência pelas comunidades científicas. O estadiamento da AJCC é utilizado por médicos e profissionais da saúde em todo o mundo de forma a facilitar uma uniformização na descrição das doenças neoplásicas. Uma classificação e estadiamento apropriados permitem aos médicos determinar os tratamentos mais adequados, avaliar resultados com maior fiabilidade e comparar estatísticas reportadas por várias instituições de índole local, regional ou nacional.

O *Cancer Staging Manual* (AJCC 2010) atualmente na sétima edição (2010) é constantemente revisto e atualizado de forma a incorporar novos conhecimentos sobre etiologia e patologia do cancro. À medida que mais conhecimento temos, o estadiamento deve adaptar-se para acomodar essa nova informação. O sistema de estadiamento adotado pelo AJCC é a classificação TNM, a qual é aceite como padrão e referência pela comunidade científica.

### Estadiamento TNM

A classificação TNM foi criada pela *Union for International Cancer Control*, UICC e é desenvolvida e mantida por um Comité que inclui representantes da *International Federation of Gynecology and Obstetrics*, FIGO, *International Association for the Study of Lung Cancer*, IASLC e o AJCC, de forma a manter essa mesma classificação unificada ((UICC 2015).

A classificação TNM do cancro por extensão anatómica da doença, constituindo uma linguagem comum para os profissionais da oncologia, passou a ser considerada como o principal determinante para o tratamento e prognóstico

adequados, bem como de base para a investigação e troca de informação uniformizada entre centros de tratamento, (ACS 2015)

A classificação TNM, sendo o padrão adotado no meio, vai para além da prática clínica, constituindo informação vital para as entidades responsáveis pelo desenvolvimento de políticas de controlo e prevenção do cancro. Por esse motivo é crucial a sua inclusão nos sistemas de registo de cancro.

A classificação TNM é um sistema baseado na anatomia que regista a região primária do tumor, eventuais extensões nos nódulos linfáticos, bem como a presença ou ausência de metástases (AJCC 2012).

Cada aspeto individual da TNM é denominado como uma categoria:

A categoria **T** dá informação, relativamente ao tumor primário, sobre o seu tamanho e o seu grau de desenvolvimento nos tecidos no local onde se iniciou.

**TX** significa que o tumor não pode ser medido

**T0** significa que não existe evidência do tumor primário (não pode ser encontrado)

**Tis** significa que as células do cancro só crescem na camada mais superficial do tecido sem invadir os tecidos mais internos, podendo ser chamado como *in situ* cancer ou *pre-cancer*.

**T1, T2, T3 e T4** descrevem o tamanho do tumor e/ou o grau de invasão dos tecidos. Quanto maior o número, maior é o tumor e/ou o grau de invasão dos tecidos

A categoria **N** descreve o comprometimento dos nódulos linfáticos próximos do local de origem

**NX** significa que os nódulos linfáticos próximos não podem ser avaliados

**N0** significa que os nódulos linfáticos próximos não contêm cancro

**N1, N2 e N3** podem descrever o tamanho, localização e/ou número de nódulos afetados pelo cancro. Quanto maior o número, maior é a invasão dos nódulos pelo cancro

A categoria **M** descreve a presença ou não de metástases em locais mais distantes do corpo

**M0** significa que não foram encontrados vestígios de dispersão do cancro para localizações distantes

**M1** significa que o cancro alcançou órgãos ou tecidos distantes (foram encontradas metástases)

Estas categorias são gerais e existem algumas especificidades para alguns tipos de cancro, nuns o T pode, por exemplo descrever o tamanho, noutros o grau de invasão dos tecidos. Podem, também, para alguns cancros, as classificações recorrerem a subcategorias específicas, como por exemplo T3a ou T3b.

### **Estadiamento anatómico**

Depois de determinados os valores de T, N e M, os quais constituem os grupos de prognóstico, estes podem ser combinados de forma a que seja atribuído um estadiamento anatómico em função dos mesmos. Normalmente esse estadiamento é representado por um número romano de I a IV, onde o estadio IV é o que representa o cancro no estado mais avançado. Assim, quando estamos perante a presença de metástases, M assume o valor M1. O estadio 0 indica, na maior parte dos casos, *pre-cancer*, isto é, num estado de desenvolvimento muito embrionário, normalmente associado a T igual a Tis.

### **Especificidades CCR**

A classificação TNM para CCR oferece um nível de detalhe superior aos outros sistemas de estadiamento, adicionando maior precisão na identificação de subgrupos de diagnóstico (American Joint Committee on Cancer 2010). Este estadiamento é baseado na profundidade da invasão do interior ou para além das paredes do cólon e do reto e invasão ou aderência para órgãos e estruturas adjacentes (T), o número de nódulos linfáticos envolvidos (N) e a presença ou ausência de metástases distantes (M).

A definição das categorias T, N, M foi evoluindo, de forma a poder fornecer a informação o mais pertinente e precisa possível para uma melhor identificação e



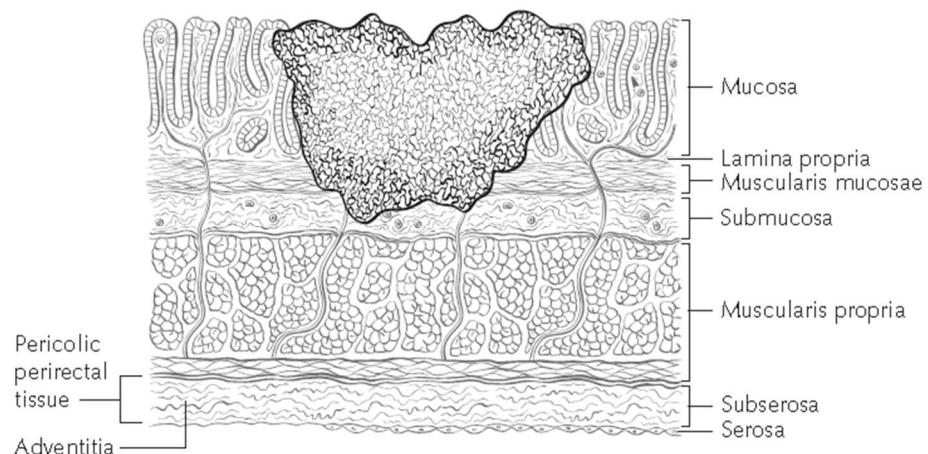
classificação do diagnóstico, e na sétima, e atual, edição do AJCC poderemos encontrar subdivisões para T4, N1, N2 e M1.

No entanto, os dados objeto deste estudo foram classificados de acordo com a 6ª edição do AJCC, de 2002, dado que essa era a edição vigente aquando da recolha desses mesmos dados.

Assim, considerando CCR, de acordo com a 6ª edição do AJCC, a classificação TNM pode ser especificada da forma que se segue.

Figura 2 Paredes do cólon e reto

(AJCC 2010)



O Tumor Primário (**T**), tendo em conta a Figura 2, relativa à localização do tumor primário nos tecidos das paredes do cólon apresenta os seguintes valores:

**TX** significa que o tumor não pode ser avaliado

**T0** significa que não existe evidência do tumor primário (não pode ser encontrado)

**Tis** *Tumor in situ*, intraepitelial ou invasão da Lamina Própria, sem extensão para a *Muscularis Mucosa*

**T1** Tumor invade a Submucosa

**T2** Tumor invade a *Muscularis* própria

**T3** Tumor invade os tecidos pericoretais ou a Subserosa através da *Muscularis Própria*

**T4** Tumor invade diretamente ou adere a outros órgãos ou estruturas e/ou penetra na superfície do peritoneu visceral

Os nódulos linfáticos regionais (**N**) apresentam os seguintes valores:

**NX** significa que os nódulos linfáticos próximos não podem ser avaliados

**N0** significa que os nódulos linfáticos próximos não foram invadidos

**N1** Invasão de 1 a 3 nódulos

**N2** Invasão de 4 ou mais nódulos

As metástases distantes (**M**) apresentam os seguintes valores:

**MX** metástases distantes não podem ser avaliadas

**M0** significa que não foram encontradas metástases distantes

**M1** foram encontradas metástases distantes

Em função destas classificações e subclassificações TNM, para o CCR o Estadiamento Anatómico adotado encontra-se na tabela abaixo, adaptada do AJCC Handbook, 6ª edição (AJCC 2002)

*Tabela 1 Estadiamento anatómico em função de TNM*

(AJCC 2002)

Estadiamento Anatómico /TNM AJCC 6ed				
Estadio	T	N	M	
<b>0</b>	Tis	N0	M0	
<b>I</b>	T1	N0	M0	
	T2	N0	M0	
<b>II</b>	<b>IIA</b>	T3	N0	M0
	<b>IIB</b>	T4	N0	M0
<b>III</b>	<b>IIIA</b>	T1-T2	N1	M0
	<b>IIIB</b>	T3-T4	N1	M0
	<b>IIIC</b>	Any T	N2	M0
<b>IV</b>	Any T	Any N	M1	

Sempre tendo como objetivo a melhoria constante da identificação dos subgrupos de prognóstico, o estadiamento anatómico em função das classificações TNM vem sofrendo alterações e adaptações, sendo que na 7ª edição foram criadas diversas subclassificações não só nas componentes TNM, no caso T4, N1, N2 e M1, mas também nos estadios II (adicionado o IIC) e IV (subdividido em IVA e IVB).

### **2.1.5. Evolução Científica na área clínica**

A pesquisa científica na área do CCR é contínua e visa prevenir a doença, bem como melhorar os tratamentos. Esta evolução no campo dos tratamentos tem como objetivo prolongar a vida dos pacientes após o diagnóstico, tendendo a influenciar, assim, positivamente a sobrevivência dos mesmos.

No campo da genética o desenvolvimento de testes como o Oncotype Dx® Colon Cancer Assay, ColoPrint® e ColDx™ tem como objetivo ajudar a prever a atividade de diferentes genes do CCR, de forma a conseguir prever o risco de alastramento do mesmo.

Foram, também, desenvolvidos novos testes através do reconhecimento dum tipo de RNA específico do CCR, os quais permitem identificar nos gânglios linfáticos mais próximos a presença da doença, levando a um estadiamento superior ao previsto sem a utilização deste método. Esta melhor identificação do estadio poderá dar origem a diferentes tipo de tratamentos.

A nível de cirurgia vai existindo um melhor entendimento sobre o que pode aumentar o sucesso da mesma, tal como o assegurar de que os gânglios linfáticos são suficientemente removidos durante a operação. Abordagens de cirurgia laparoscópica e robótica têm sido, também, avaliadas, no tratamento do CCR.

A nível de quimioterapia vem sendo introduzidas novas drogas, novas formas de as combinar e outras que têm sido utilizadas noutros cancros, tais como *cisplatin* ou *gemcitabine*.

Terapias direcionadas têm sido utilizadas para tratar o CCR, incluindo *bevacizumab* (*Avastin*), *cetuximab* (*Erbitux*), e *panitumumab* (*Vectibix*). Estas terapias inicialmente vocacionadas para os cancros mais avançados têm sido testadas em

combinação com a quimioterapia em cânceres em estágios de menor evolução, de forma a tentar diminuir os riscos de recidiva (ACS 2014).

A nível de imunoterapia têm sido estudados alguns tipos de vacinas de forma a tratar o CCR ou impedir o reaparecimento do mesmo.

A pesquisa tem ocorrido também de forma a se tentar obter as melhores formas de combinar as diferentes terapias: quimioterapia, radioterapia, imunoterapia e terapia direcionada.

## **2.2. Análise de Sobrevivência**

### **2.2.1. Definição**

A Análise de Sobrevivência (AS) é geralmente definida como um conjunto de métodos estatísticos utilizados para descrever, explicar ou prever o tempo decorrido até à ocorrência de eventos. A origem do nome *Survival Analysis* deriva do facto desses métodos terem sido originalmente desenvolvidos por bioestatísticos para analisar tempos de vida até à ocorrência de mortes. Contudo, estes mesmos métodos poderão ser utilizados para um vasto leque de fenómenos sociais tais como nascimentos, divórcios, detenções, migrações e outros (Allison 2010).

Em AS os sujeitos são observados durante um determinado período de tempo e o foco situa-se no tempo que demora até que um determinado evento de interesse (falha) suceda, a partir do início do período de observação. Nos estudos de AS é normal existirem observações censuradas, ou seja, em que o evento de interesse não ocorreu durante esse período de observação. A estatística clássica não está preparada para lidar com estes fenómenos de censura, podendo ser utilizada apenas se estes forem eliminados do estudo, levando a que se perca informação relevante. A AS utiliza métodos próprios para lidar com dados censurados, motivo pelo qual deverá ser utilizada na presença de censura, em detrimento dos métodos estatísticos baseados em análise de regressão.

### **2.2.2. Evento de interesse ou falha**

O ponto de partida numa AS deverá ser a definição do evento que irá ser estudado. Esse evento deverá ser uma alteração qualitativa que ocorre (ou não) num instante de tempo observado, dentro do período de observação.

No presente estudo o evento de falha é a morte de pacientes a quem foi diagnosticado CCR.

### **2.2.3. Período de observação**

A AS requer que os indivíduos sejam observados durante um período de tempo definido. Durante este intervalo de tempo poderá ou não ser observada a falha. Os indivíduos devem ser comparáveis no início desse período, excetuando as diferenças medidas pelas covariáveis consideradas. A escala de medida para este período de observação é tipicamente uma escala temporal, mas poderão ser utilizadas outras escalas tais como, por exemplo, em engenharia, o número de ciclos ou a quilometragem (Colosimo E. A. 2006).

A observação ocorre desde o início do período até à ocorrência do evento de falha, se este não for repetível. Caso não ocorra o evento de falha a observação decorre até ao final definido para esse período.

No presente estudo o período de observação corresponde ao tempo de *follow-up* (seguimento) aos pacientes a quem foi diagnosticado CCR. Esse período tem início na data de diagnóstico, a duração máxima de 5 anos e é apresentado em dias.

### **2.2.4. Censura**

Mesmo que os períodos de observação sejam longos, normalmente terminam antes que ocorra o evento de interesse em todos os indivíduos, o que é usual acontecer especialmente em estudos clínicos. Outra possibilidade é a morte ser provocada por outro motivo que não o de interesse (CCR), não podendo também ser considerada uma falha. Assim, estamos na presença de observações incompletas ou parciais. Estas observações são denominadas censuras e podem, também, ser causadas por perda de acompanhamento do indivíduo no decorrer do período de estudo. A censura mais comum (Allison 2010) é a censura à direita, na qual a falha não ocorreu antes do final do período de observação, mas poderá ocorrer censura à esquerda (quando a falha ocorre antes do início do período de observação) ou censura aleatória quando, por exemplo, um paciente é retirado do estudo sem ter ocorrido a falha ou quando morre por uma causa diferente da estudada.

Todas as observações provenientes de um estudo de sobrevivência devem ser utilizadas na análise estatística, pois mesmo sendo incompletas, as observações

censuradas fornecem informações relevantes sobre o tempo de vida dos pacientes e, conseqüentemente, a sua omissão poderá levar a conclusões enviesadas (Colosimo E. A. 2006).

No presente estudo é efetuada censura à direita para todos os indivíduos para os quais não ocorreu o evento de falha (morte) durante o período de 5 anos após diagnóstico.

### **2.2.5. Função de Sobrevivência**

O Tempo de Falha, representado pela variável aleatória não negativa  $T$ , é usualmente caracterizado em AS pelas suas funções de Taxa de Falha e de Sobrevivência. Esta é a variável de interesse em estudos de AS.

A Função de Sobrevivência é definida como a probabilidade de uma observação de determinado indivíduo num estudo não falhar até um certo tempo  $t$ , ou seja a probabilidade do sujeito sob observação sobreviver até esse instante. Em termos e probabilidades essa função é representada por

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

onde  $t$  é um determinado instante de tempo e  $T$  é uma variável aleatória que representa o instante no qual ocorre o instante de falha.

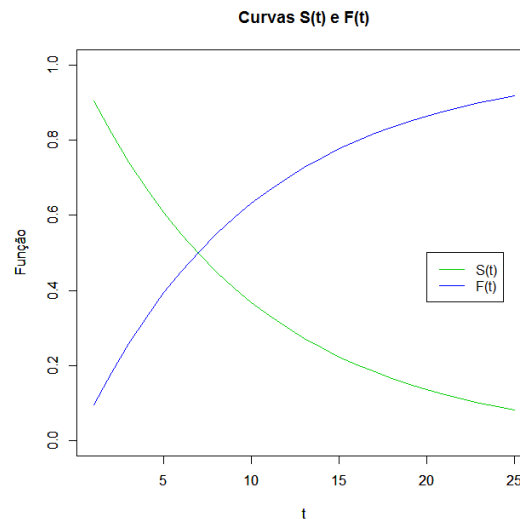
Esta função é não crescente e geralmente em estudos biológicos tende para zero quando o tempo aumenta, isto é:  $S(t) \rightarrow 0$  quando  $t \rightarrow \infty$ .

Por outro lado, a probabilidade de uma observação não sobreviver ao tempo  $t$ , é-nos dada pela Função de distribuição de probabilidade acumulada da variável  $T$

$$F(t) = P(T \leq t) = 1 - S(t). \quad (2)$$

A figura abaixo pretende mostrar graficamente a relação entre estas duas funções, exemplificando para uma função de sobrevivência com o modelo exponencial subjacente  $S(t) = e^{-0.1t}$ , e conseqüentemente  $F(t) = 1 - e^{-0.1t}$ . De notar que F é sempre crescente e S é sempre decrescente, em sentido lato.

Figura 3 Função Sobrevivência vs Distribuição Acumulada



## 2.2.6. Função Taxa de Falha

A probabilidade da falha ocorrer entre os instantes de tempo  $t_1$  e  $t_2$  pode ser expressa como

$$S(t_1) - S(t_2), \quad (3)$$

assim, a função taxa de falha  $\lambda(t)$  para o intervalo  $(t_1, t_2)$ , tendo em conta que a falha não ocorreu antes de  $t_1$  é dada por

$$\lambda = \frac{S(t_1) - S(t_2)}{(t_2 - t_1)S(t_1)}. \quad (4)$$

De forma a tentarmos representar a taxa de falha instantânea no tempo t, redefinimos o intervalo como  $(t, t + \Delta t)$ , em que  $\Delta t \approx 0$ , assumindo a expressão (4) a seguinte forma



$$\lambda = \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}, \quad (5)$$

assumindo que o  $\Delta t$  tende para zero,  $\lambda(t)$  representa a taxa de falha instantânea no tempo  $t$  condicional à sobrevivência até esse instante. Esta função descreve a forma como a taxa de instantânea de falha muda com o tempo

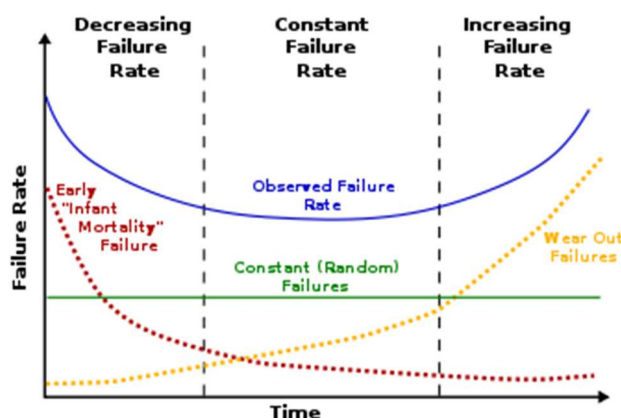
$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \quad (6)$$

A função taxa de falha, não tendo que ser monótona, poderá ser constituída por períodos de tempo em que é crescente, constante ou decrescente ao longo do tempo.

A sobrevivência dos seres humanos apresenta, geralmente, uma curva que é uma combinação desses três tipos de inclinação. A curva resultante é conhecida como curva da Banheira e indica que após uma taxa mais elevada inicial devido à mortalidade infantil, há um decrescimento ao qual se segue uma taxa quase constante. Com o envelhecimento dos seres humanos verifica-se um aumento da taxa de falha, sendo a sua curva crescente. Estes tipos de taxa de falha estão representados na figura abaixo. Esta curva é, também, utilizada na engenharia para caracterizar a fiabilidade e durabilidade dos produtos.

Figura 4 Funções de Taxa de Falha

(Wikipedia 2016)



A função de taxa de falha é mais informativa que a função de sobrevivência, pois diferentes funções de sobrevivência podem ter formas semelhantes apesar das respectivas taxas de falha poderem diferir drasticamente (Colosimo E. A. 2006).

No presente estudo, e dado que o evento de interesse ou falha é a morte, passaremos doravante a denominar a taxa de falha como taxa de mortalidade.

### **2.2.7. Estimadores e Modelos**

Tendo em conta que em AS estamos na presença de algo muito particular, como as observações censuradas, as técnicas de análise de regressão não possuem ferramentas que permitam o seu tratamento. Em AS são utilizados estimadores não paramétricos adequados para a função de sobrevivência, bem como modelos de regressão paramétricos e semiparamétricos e respetiva estimação dos coeficientes das covariáveis incluídas nesses mesmos modelos.

#### **Estimação**

Embora em AS possa não se verificar a presença de observações censuradas, essa situação na prática é muito pouco usual. Na presença de censura são necessárias técnicas estatísticas especializadas que permitam acomodar a informação contida em todas as observações, mesmo as censuradas, ou seja naquelas em que a falha não ocorreu até ao final do período de observação.

No presente estudo o primeiro estimador para a função de sobrevivência utilizado é o estimador de *Kaplan-Meier*, o qual é, sem dúvida, o estimador não paramétrico mais utilizado em estudos clínicos, embora existam outros tais como o estimador de Nelson-Aalen e o estimador da tabela de vida ou atuarial (Colosimo E. A. 2006).

A AS pode ser aplicada a um vasto domínio de aplicações, e não apenas ao mundo clínico. Por exemplo, em engenharia são comuns os estudos sobre a durabilidade de componentes ou produtos e em ciências sociais poderão ser estudados casamentos, nascimentos, divórcios, migrações. O estimador de *Kaplan-Meier* é,

assim, o mais adotado em AS, para toda uma vasta panóplia de estudos com eventos de interesse de diversa natureza, tais como os utilizados nos estudos acima citados.

No entanto, na situação específica em que o evento de interesse é a morte humana (tal como acontece no presente estudo), outras abordagens mais adequadas deverão ser utilizadas, como iremos introduzir no capítulo 2.3.

### **Modelos de Regressão**

Em AS, o método mais utilizado para análise de regressão de dados de sobrevivência é, de longe, o modelo de regressão de Cox (Allison 2010), o qual combina o modelo de taxas de risco proporcionais com o método de estimação de verosimilhança parcial. Este modelo é semiparamétrico porque inclui uma componente não paramétrica como a taxa de falha basal e uma componente paramétrica em função dos coeficientes das covariáveis, não fazendo qualquer assunção sobre a distribuição da probabilidade do tempo dos eventos. Os modelos paramétricos, assumindo distribuições de probabilidade tais como a exponencial, Weibull, Gompertz, log-normal, log-logística ou gama, assentam em hipóteses que os tornam algo restritivos (Allison 2010), não tendo sido utilizados no presente estudo.

Tal como na estimação da função de sobrevivência, no caso específico deste estudo em que o evento de interesse é a morte humana, outras abordagens mais adequadas deverão ser utilizadas, como iremos introduzir, também no capítulo 2.3., no qual introduziremos a Análise de Sobrevivência Relativa.

## 2.3. Análise de Sobrevivência Relativa

Nos casos em que o evento de interesse é a morte,  $S(t)$  representa a probabilidade de estar vivo no instante de tempo  $t$ . Se a esta premissa juntarmos o facto do evento de interesse ser a morte causada por uma certa doença, e não sendo as causas de morte conhecidas, não é possível estimar diretamente a proporção de mortes causadas pela doença em questão. Nestes casos devemos recorrer a métodos de AS Relativa (Pohar M. 2006), os quais comparam a morte causada por sujeitos com a doença, com a morte da população em geral.

Com os avanços no conhecimento que levam à melhoria contínua dos prognósticos e tratamentos, os tempos de *follow-up* tendem a tornar-se mais longos, tornando comum o facto da proporção de mortes devidas a outras causas que não a doença em questão, se vá tornando cada vez mais significativa. A AS Relativa desenvolveu métodos que permitem comparar as mortes observadas no estudo com as mortes esperadas para conjunto similar de indivíduos da população geral, durante esse mesmo período de observação, retiradas das tábuas de mortalidade (Perme M.P. 2011).

Assim, no presente estudo verificamos que sendo o evento de interesse a morte, que o período de observação adotado é elevado (aumentando a probabilidade de morte causada por outros riscos), que nos interessa estudar a morte causada pelo CCR e que nos registos fonte do estudo não temos informação sobre as causas de morte, aqui confirma-se a necessidade da aplicação de métodos de AS Relativa.

A Sobrevivência relativa, pode então, ser definida como a razão entre a sobrevivência observada  $S_o(t)$  e a sobrevivência esperada da população  $S_p(t)$

$$S_r(t) = \frac{S_o(t)}{S_p(t)}. \quad (7)$$

A sobrevivência observada para um conjunto de indivíduos pode ser estimada pelo estimador de *Kaplan-Meier* tendo em conta os dados observados e a natureza

desconhecida das causas de morte, as quais não estão especificadas nos dados do registo.

Para o cálculo da sobrevivência esperada são considerados os dados obtidos das tabelas de população, para um grupo com as mesmas características demográficas do grupo de observação, a nível de sexo, idade (no diagnóstico) e ano de calendário (no evento).

### **2.3.1. Tábuas de Mortalidade das Populações**

As tábuas de mortalidade das populações permitem, assim, comparar a sobrevivência observada num conjunto de pacientes, com a de um conjunto semelhante da população durante o mesmo período de tempo. O grupo populacional ideal será aquele que tem exatamente as mesmas características em todos os aspetos que possam afetar a sobrevivência, excetuando a doença em questão. Desta forma a sobrevivência relativa permitir-nos-á produzir informação sobre a redução da sobrevivência devida à doença. Estas condições ideais não são geralmente conseguidas, pois das tabelas populacionais conseguimos ter apenas informação sobre a idade, sexo e ano da coorte. Essa informação é bastante importante, mas poderá ser insuficiente em alguns casos, para se poder estimar corretamente a sobrevivência relativa. Como exemplo poderemos ter o caso de termos uma percentagem muito mais elevada de fumadores num grupo de pacientes com cancro de pulmão do que na população em geral (Perme 2007). Como o facto de fumar afeta a sobrevivência, a sobrevivência de fumadores sem cancro de pulmão é menor que a de não fumadores na população em geral. Neste caso, se ignorarmos a componente relativa ao tabagismo, a sobrevivência relativa não será corretamente estimada pois irá incluir informação não só sobre o cancro de pulmão, mas também, as consequências de fumar, as quais serão eventualmente manifestadas por outras doenças (Blakely T. 2012).

No entanto, mesmo que algumas variáveis importantes não existam nas tábuas de mortalidade populacionais, os métodos de sobrevivência relativa conseguem retirar

informação importante a partir dos dados, desde que se tenham em mente a estrutura dessas mesmas tabelas, aquando da interpretação dos dados.

Para além do sexo, idade e ano de calendário, outras características como região (incluída no nosso estudo), a raça e a classe social podem ser incluídas nessas tabelas, o que não acontece em Portugal, no caso das duas últimas.

Exemplificando, no Reino Unido, incluindo Inglaterra, País de Gales, Escócia e Irlanda do Norte, as tábuas de mortalidades incluem, também, informação sobre a classe social organizada em cinco grupos, a que chamam “*Quintile of deprivation score*”, (CSG 2016).

Também como exemplo, nos Estados Unidos da América, um dos elementos a ter em conta nessas tábuas de mortalidade é a raça, denominada nas mesmas como “*Race/Ethnicity*”, (Spika D. 2015)

### **2.3.2. Sobrevivência Relativa**

A Sobrevivência Relativa pode ser interpretada como sendo função da mortalidade causado apenas e especificamente pela doença em questão. A curva observada é calculada tomando todas as mortes como eventos e ignorando as causas de morte (Perme M.P. 2007).

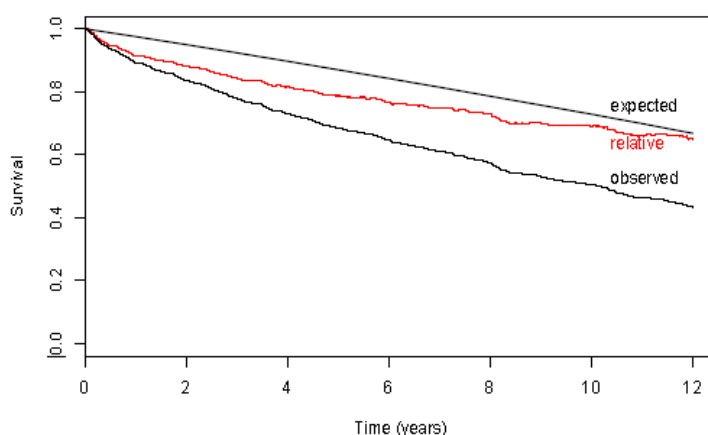
É importante notar o facto de que a Curva de Sobrevivência Relativa, não é propriamente uma curva de sobrevivência, dado que poderá não ser monótona decrescente. Por outro lado, e dependendo do estudo, o seu valor, apesar de não negativo, poderá ser superior a 1. Esta situação ocorre quando a sobrevivência do grupo de observação é superior ao da mortalidade geral, por exemplo quando a natureza dos estudos implica que o grupo de observação esteja clinicamente mais bem controlado e está sujeito a um modo de vida mais saudável (Perme M.P. 2007).

Retirado da tese de doutoramento de Maja Pohar Perme (Perme M.P. 2007), abaixo está representado um exemplo da estimação da curva de sobrevivência, a partir de dados dum estudo do University Clinical Centre in Ljubljana, sobre enfarte de miocárdio. Foram seguidos 1017 pacientes que sofreram esse episódio entre 1982 e 1986, tendo sido seguidos durante 14 anos, havendo registos de determinado

número de variáveis efetuados durante a admissão. O evento de interesse considerado foi a morte, dado que a informação sobre a causa específica de morte era impossível de determinar. O objetivos desse estudo era estimar de que forma o enfarte afetou a sobrevivência após a sua ocorrência. Dado o longo período de *follow-up* e o facto da mortalidade causada pelo enfarte não ser muito substancial, será de esperar muitas mortes devidas a outras causas durante esse período.

Figura 5 Exemplo de resultados de sobrevivência observada, esperada relativa

(Perme M.P. 2007)



Aqui as curvas observadas e esperadas para um período de 12 anos são comparadas. Conseguimos observar, aqui, que durante o primeiro ano a sobrevivência observada decresce muito mais rapidamente que a sobrevivência da população.

A curva observada estima a sobrevivência esperada devida a todas as causas de morte e é muito menor que a curva de sobrevivência relativa, fornecendo-nos, essa diferença, a informação sobre o excesso de mortalidade causada pelo enfarte.

### 2.3.3. Estimadores de Sobrevivência Relativa

Tal como já definido em capítulos anteriores, caso o evento morte esteja inequivocamente associado à doença em questão, utilizamos tipicamente o

estimador de *Kaplan-Meier* ou o modelo de *Cox* e a nossa análise limitar-se-á à sobrevivência observada,  $S_o(t)$ .

Quando as causas de morte não são conhecidas, tal como acontece nos dados deste estudo e tipicamente em todos os registos, os estimadores que têm sido, tradicionalmente utilizados, são o *Ederer I* (Ederer 1961), *Ederer II* (Ederer F. 1959) e *Hakulinen* (Hakulinen 1982). Assim, estes estimadores permitem-nos comparar a sobrevivência observada com a sobrevivência de um grupo da população em geral com as mesmas características daquele grupo em estudo.

### **2.3.4. Estimador Net Survival**

Todavia, os estimadores tradicionais acima enunciados não fornecem informação independente da mortalidade das populações nacionais, não se tornando adequados quando é necessário efetuar comparações entre países, como acontece em estudos mas alargados como o EUROCARE e o CONCORD. Para resolver esta questão surgiu, em 2012, o estimador Pohar-Perme (Perme M.P. 2012). A partir do aparecimento deste estimador, este passou a ser considerado o único não enviesado, tendo Roche et al (Roche L. 2012) afirmado que os registos de cancro devem abandonar todos os métodos clássicos (*Ederer I*, *Ederer II* e *Hakulinen*) e adotar este novo de Pohar-Perme. Afirmações similares poderão ser encontradas em outros artigos tais como os publicados no *International Journal of Cancer* em 2013, (Jooste V. 2013) e (Monnereau A. 2013).

Este estimador estima a sobrevivência quando estamos na situação hipotética em que a doença sob estudo é a única causa possível de morte. Isto é tornado possível decompondo a taxa de mortalidade observada em mortes devidas à doença e em mortes devidas a outras causas. É assumido que a taxa de mortalidade devida a outras causas nos é dada pelas tábuas de mortalidade da população e que a taxa de mortalidade observada é superior à taxa de mortalidade da população (Perme M.P. 2011). Tendo esta assunção em mente, utilizamos o termo taxa de mortalidade em excesso (Excess Hazard) e temos a relação definida por

$$\lambda_O = \lambda_P + \lambda_E \quad (8)$$



em que  $\lambda_o$  representa a taxa de mortalidade observada,  $\lambda_p$  a taxa de mortalidade da população e  $\lambda_E$  a taxa de mortalidade em excesso.

A função de sobrevivência obtida a partir da taxa de mortalidade em excesso é intitulada *Net Survival*.

Assim, apesar de anteriormente os estimadores *Ederer I*, *Ederer II* e *Hakulinen* serem considerados adequados, eles estimam a sobrevivência relativa como resultado de diferentes formas de estimar a sobrevivência esperada. Assim, a taxa de sobrevivência relativa resultante e dada pela equação (7), faz com que o resultado seja dependente da mortalidade das populações. Por sua vez, a *Net Survival*, sendo calculada a partir da taxa de mortalidade em excesso torna-se o seu resultado independente da população, permitindo assim a comparabilidade entre países, e é estimada pelo estimador de Pohar-Perme (Perme M.P. 2011).

Paul Dickman e Paul Lambert (Lambert P. 2014) afirmam que o *Ederer II* pode, também, ser utilizado como estimador de *Net Survival*, sem grandes enviesamentos.

### **2.3.5. Modelos de Regressão**

Têm sido várias as abordagens para os modelos de regressão, sendo que em todas essas abordagens a resposta é modelada através da função taxa de mortalidade (Pohar M. 2006), sendo que deverá ser assumida uma relação entre a taxa de mortalidade observada e a taxa de mortalidade da população. Essas taxas de mortalidade tal como poderemos ver adiante no capítulo 3.4.2 são função das variáveis consideradas nas tábuas de mortalidade e de outras variáveis que se pretenda incluir na regressão, estas últimas apenas para  $\lambda_o$  e  $\lambda_E$ , os quais representam, respetivamente, a taxa de mortalidade observada, estimada a partir dos nossos dados e a taxa de mortalidade em excesso, ou seja a taxa de mortalidade resultante de causa específica, neste caso específico o CCR.

A taxa de mortalidade da população obtida diretamente das tábuas de mortalidade para um grupo com as mesmas características do grupo em observação em termos de idade, sexo e coorte é representada por  $\lambda_p$ .

Basicamente conseguimos distinguir dois tipos de modelos. Os aditivos podem ser representados de forma simplificada, pela equação (8).

Por este motivo estes modelos aditivos são também conhecidos como modelos de Excesso de Mortalidade.

Os modelos multiplicativos podem ser representados de forma simplificada pela seguinte equação

$$\lambda_o = \lambda_p \times \lambda_R \quad (9)$$

em que  $\lambda_o$  e  $\lambda_p$  apresentam o mesmo significado que na equação anterior e  $\lambda_R$  representa a taxa de mortalidade relativa, pois tal como se deduz da equação acima,  $\lambda_R = \lambda_o/\lambda_E$ , ou seja é obtida pela razão entre a mortalidade observada e a mortalidade da população.

Por este motivo estes modelos multiplicativos são também conhecidos como modelos de Mortalidade Relativa.

Os modelos usualmente considerados (Pohar M. 2006) e utilizados neste estudo são os modelos aditivos de *Hakulinen-Tenkanen* (Hakulinen T. 1987), *Poisson* (Dickman P.W. 2003) e *Estève* (Estève J. 1990) e o modelo multiplicativo de *Andersen* (Andersen P.K. 1985).

Qualquer uma destas quatro abordagens para a estimação de modelos de sobrevivência relativa produz resultados bastante semelhantes, embora o modelo de *Estève* venha a ser reconhecido como teoricamente superior (Dickman P.W. 2003) ao de *Hakulinen-Tenkanen* e ao de *Poisson*, pois utiliza informação de tempos de sobrevivência exatos, não dependendo da aproximação a um modelo binomial ou de *Poisson*, respetivamente.

A literatura sobre sobrevivência relativa tem considerado os modelos aditivos como mais adequados para a análise do cancro (Pohar M. 2006). Mais recentemente (Elie C. 2011) é, também, referido que os modelos aditivos são quase exclusivamente a primeira escolha para estudos de cancro baseados em populações, parecendo o modelo multiplicativo ser adotado mais frequentemente no estudo de outras

doenças crónicas. Opiniões semelhantes são manifestadas em (Dickman P.W. 2003) e em (Buckley 1984), cujos autores consideram os modelos aditivos mais apropriados biologicamente.

Mais detalhe sobre os estimadores e modelos utilizados no presente estudo serão apresentados mais adiante nas opções metodológicas, nos capítulos 3.4.1, 3.4.2 e 3.4.3.

## 3. Opções Metodológicas

### 3.1. Caracterização dos dados

#### 3.1.1. Dados iniciais

Foi efetuada uma extração das bases de dados do RORENO, de dados relativos a pacientes a quem foi diagnosticado CCR em duas coortes distintas, cujas datas de diagnóstico correspondem, respetivamente, aos períodos 2000-2001 e 2007-2008.

Apresentamos abaixo a tabela com os campos extraídos, o tipo de variáveis que os representam e o seu significado:

*Tabela 2 Variáveis iniciais obtidas do registo*

Variável	Tipo	Significado/Observações
RORENO_ID	Qualitativa – nominal	Identidade do paciente. Não utilizada para o estudo, podendo no entanto ser útil para alguma eventual necessidade de esclarecimento
Idade	Quantitativa	Idade do paciente à data de diagnóstico (em anos)
Estado Vital	Qualitativa – nominal	Estado à data de final <i>follow-up</i> (0 – vivo, 1 – morto, 9 – desconhecido). Variável utilizada como evento de interesse (morte) ou censura em Análise de Sobrevida
Data de <i>follow-up</i>	Quantitativa	Data de final de <i>follow-up</i> (dd/mm/aa).
Residência - Distrito	Qualitativa - nominal	Distrito do paciente (região Norte)
Sexo	Qualitativa – nominal	Masculino/Feminino
Localização Topográfica	Qualitativa – nominal	Classificação topográfica do CCR
Estudo Histológico	Qualitativa – ordinal	Classificação Histológica do CCR. De notar que todos os registos correspondem ao comportamento identificado como xxxx/3, o que indica que são malignos na localização primária
T	Qualitativa – ordinal	Tumor – Estadiamento TNM
N	Qualitativa – ordinal	Nódulo – Estadiamento TNM
M	Qualitativa – ordinal	Metástase – Estadiamento TNM
Estadio	Qualitativa – ordinal	Estadiamento anatómico AJCC 6ªed. Esta variável é resultado da combinação das variáveis T, N e M
Data do Diagnóstico	Quantitativa	Data em que foi feito o diagnóstico CCR (dd/mm/aa)

Ano do Diagnóstico	Qualitativa – ordinal	Ano em que foi feito o diagnóstico CCR (ano)
--------------------	-----------------------	--

Foram, inicialmente, extraídos 7536 registos, sendo que 2928 correspondem à coorte 2000-2001 e 4608 à coorte 2007-2008.

A estes dados correspondem as seguintes estatísticas descritivas mais relevantes:

*Tabela 3 Estatísticas descritivas das variáveis iniciais*

<b>Variável</b>	<b>Estatísticas</b>
Idade	Mínimo: 16 1º Quartil: 60 Mediana: 69 Média: 67.42 3º Quartil: 76 Maximo: 99 Registos Incompletos\desconhecidos: 107
Estado Vital	0 (vivo) – 3094 1 (morto) – 4200 Registos Incompletos\desconhecidos: 242
Data_de follow-up	03/02/2000 a 04/12/2014
Residencia - Distrito	Moda: Porto com 4123 registos Registos Incompletos\desconhecidos: 149
Sexo	Feminino: 3180 Masculino: 4356
Localização Topográfica	Mínimo: 18.0 Máximo: 20.9 Moda: 20.9 com 2328 registos
Estudo Histológico	Mínimo: 8000 Máximo: 8936 Moda: 8140 com 6510 registos
T	Mínimo: 0 Máximo: 4 Moda: 3 com 2359 registos Registos Incompletos\desconhecidos: 3914
N	Mínimo: 0 Máximo: 3 Moda: 0 com 1777 registos Registos Incompletos\desconhecidos: 4153

M	Mínimo: 0 Máximo: 1 Moda: 0 com 2212 registos Registos Incompletos\desconhecidos: 4579
Estadio	Mínimo: 0 Máximo: IV Moda: IIA com 744 registos Registos Incompletos\desconhecidos: 4907
Data do Diagnóstico	01/01/2000 a 31/12/2001 – Coorte 1 01/01/2007 a 31/12/2008 – Coorte 2
Ano do Diagnóstico	2000 ou 2001 – 2928 2007 ou 2008 – 4608

### 3.1.2. Critérios de inclusão e exclusão de indivíduos

- Relativamente às variáveis estado vital e data de *follow-up*, constituiu critério de eliminação:
  - data de *follow-up* em branco, com estado vital = 1 (1 registo)
  - Estado vital desconhecido (242 registos)
- Relativamente às variáveis T, N e M foram eliminados 4430 registos nos casos em que T, N ou M apresentam valores em branco ou X, o que corresponde a casos em que não podem ser medidos). Excetuam-se, aqui, os casos:
  - Em que M=1, pois isto significa que independentemente dos valores de T ou N, o estadio resultante é sempre o IV.
  - Em que N=2 e M=0, sendo desconhecido T, pois o estadio resultante é sempre o IIIC
- Relativamente à variável estadio, considerando-a como resultante das variáveis TNM, foram efetuadas as seguintes operações:
  - Eliminados estadios 0, pois para o estudo em questão interessam os diagnósticos positivos (estadio de I a IV). 7 registos (2 T0N0M0 e 5 TisN0M0).

- Acertados estadios IIC, III e Iva para os valores corretos de acordo com a tabela AJCC 6<sup>a</sup>ed, segundo a qual os valores válidos são: 0, I, IIA, IIB, IIIA, IIIB, IIIC e IV

Finalizado este tratamento o qual visa a validação e completude de todas as variáveis envolvidas em todos os registos, foram eliminados 4681 registos. No entanto, o tratamento dos dados não está terminado aqui, pois para que os mesmos possam ser utilizados em Análise de Sobrevida necessitam de outros tratamentos mais específicos, tal como indicado na secção seguinte.

### **3.1.3. Tratamento dos dados para AS**

Em Análise de Sobrevida interessa-nos, para além da ocorrência ou não do evento de interesse (morte) durante o período de observação, o registo dos tempos de sobrevivência.

Assim, e também para facilitar o tratamento dos dados, foi criada uma variável que representa a diferença, em dias, entre a data de final de *follow-up* e a data de diagnóstico, considerando que a data de final de *follow-up* é a data de óbito ou de fim de estudo e a data de diagnóstico é a data em que foi feito o diagnóstico CCR. Este valor terá que obrigatoriamente ser não negativo, e é expresso em dias. Na sequência da criação desta variável foi detetado um registo com data de *follow-up* (20/01/2001) inferior à data de diagnóstico (09/01/2007), originando um valor negativo para a variável Tempos (-2180). Esse registo foi, assim, eliminado.

Interessando-nos o estudo de Análise de Sobrevida ao CCR a 5 anos, e tendo em conta que em ambas as coortes existem, no máximo, observações em 2 anos bissextos, iremos considerar um tempo máximo de observação de 1827 dias.

Assim, para ambas as coortes, os dados foram censurados ao fim de 5 anos, isto é, sempre que o tempo decorrido entre a data de diagnóstico e a data de *follow-up* era superior a 1827, o valor da variável acima criada foi igualado a 1827 e foi considerado como censura.

Assim, efetuámos censura à direita sempre que o evento de interesse (morte) não ocorre durante o período de observação de 5 anos, utilizando neste caso a escala de medida em dias.

O programa Surveillance, Epidemiology, and End Results (SEER) do National Cancer Institute (NCI) fornece padrões para o estudo das populações adultas com cancro. Esta padronização é utilizada para comparar a sobrevivência entre diferentes populações com diferentes distribuições de idades. Esses padrões são denominados International Cancer Survival Standard (ICSS) e classificam os cancros em 3 grupos com padrões similares de incidência em função da idade. (Corazziari I and M 2015).

*Tabela 4 Padrões ICSS por tipo de cancro (Corazziari I and M 2015)*

<b>Padrão ICSS</b>	<b>Localizações do Cancro</b>
1	Lábio, língua, glândulas salivares, cavidade oral, orofaringe, hipofaringe, cabeça & pescoço, esôfago, estômago, intestino delgado, cólon, reto, fígado, trato biliar, pâncreas, cavidades nasais, laringe, pulmão, pleura, mama, corpos úteros, ovário, vagina & vulva, pénis, bexiga, rim, melanoma coróide, linfoma não-Hodgkin, mieloma múltiplo, leucemia linfática crónica, leucemia mieloide aguda, leucemia mieloide crónica, leucemia, próstata
2	Nasofaringe, tecidos macios, melanoma, colo do útero, cérebro, glândula tiróide, osso
3	Testículo, doença de Hodgkin, leucemia linfática aguda

Assim, o CCR está englobado no padrão ICSS 1 relativo a localizações de cancro com aumento de incidência em função da idade. Os padrões ICSS 2 e ICSS 3 são relativos a localizações com incidência constante em função da idade e com especial incidência nos adultos jovens, respetivamente.

São considerados, segundo estes padrões, 5 grupos de idades, estando representados na tabela abaixo conjuntamente com os relativos pesos na incidência nas populações.



Tabela 5 Padrões ICSS para grupos de idade para sobrevivência (Corazziari I and M 2015)

Grupos de idade padrão para sobrevivência	Peso na população (%)		
	ICSS 1	ICSS 2	ICSS 3
15-44	7	28	60
45-54	12	17	10
55-64	23	21	10
65-74	29	20	10
75+	29	14	10

Criamos, então, uma variável para estratificação das idades, com os seguintes grupos: 15-44, 45-54, 55-64, 65-74 e 75+.

Foram, também, criados grupos para as variáveis topografia e histologia.

Quanto à topografia foram criados dois grupos, seguindo as classificações adotadas em (Bento M.J. 2014) e (RORENO 2015) :

- 1 - Cólon (topografias C18.0 a C18.9)
- 2 - Reto (C20.9), o qual normalmente engloba a Junção Retosigmoide (C19.9).

Foi criada também uma variável para estratificar os grupos histológicos, incluindo os seguintes dois grupos, seguindo também as classificações adotadas pelo Registo Oncológico Regional do Norte (RORENO 2015):

1 – Adenomas/Adenocarcinomas (histologias 814x, 821x, 822x, 826x, 844x e 848x)

2 – Outras histologias (800x, 801x, 804x, 805x, 807x, 812x, 824x, 849x, 856x, 874x, 889x e 893x).

Cerca de 97% dos registos correspondem ao primeiro grupo (adenomas e adenocarcinomas).

### 3.1.4. Características dos dados tratados

As variáveis T, N e M não serão consideradas no estudo pois as mesmas constituem *inputs* da variável estadio, de acordo com a Tabela 1.

Assim, as variáveis que consideraremos para o estudo são:

- Idade
- Grupo de Idade
- Estado à data de final de *follow-up*
- Distrito
- Sexo
- Data de diagnóstico
- Data de final de *follow-up*
- Localização Topográfica
- Grupo Topográfico
- Grupo Histológico
- Estadio
- Dias decorridos entre as datas de diagnóstico e de final de *follow-up*
- Coortes

Assim, depois de tratados chegámos ao valor total de 2855 registos, 843 para a coorte 1 e 2012 para a coorte 2.

*Tabela 6 Estatísticas descritivas relevantes após tratamento dos dados*

<b>Variável</b>	<b>Coorte 1</b>	<b>Coorte 2</b>
Idade	Mínimo: 22 1º Quartil: 58 Mediana: 66 Média: 64.54 3º Quartil: 73 Maximo: 90	Mínimo: 21 1º Quartil: 58 Mediana: 68 Média: 66.05 3º Quartil: 75 Maximo: 98
Grupo de idade	1(15-44): n=53, média=38.6, mediana=41.0 (6.3%) 2(45-54): n=106, média=50.1, mediana=50.5 (12.6%) 3(55-64): n=221, média=60.1, mediana=60 (26.2%) 4(65-74): n=312, média=69.7, mediana=70 (37.0%) 5(75+): n=151, média=79.6, mediana=79 (17.9%)	1(15-44): n=91, média=38.3, mediana=40 (4.5%) 2(45-54): n=254, média=50, mediana=50 (12.6%) 3(55-64): n=500, média=59.8, mediana=60 (24.9%) 4(65-74): n=638, média=69.7, mediana=70 (31.7%) 5(75+): n=529, média=80, mediana=79 (26.3%)
Estado à data de final de <i>follow-up</i>	0 (vivo) – 398 (47.2%) 1 (morto) – 445 (52.8%)	0 (vivo) – 1079 (53.6%) 1 (morto) – 933 (46.4%)

Residencia –	Porto: 508 (60.3%)	Porto: 1274 (63.3%)
Distrito	Braga: 212 (25.1%)	Braga: 330 (16.4%)
	Bragança: 49 (5.8%)	Viana do Castelo: 158 (7.9%)
	Vila Real: 48 (5.7%)	Vila Real: 128 (6.4%)
	Viana do Castelo: 26 (3.1%)	Bragança: 122 (6.1%)
Sexo	2 (Feminino): 364 (43.2%)	2 (Feminino): 821 (40.8%)
	1 (Masculino): 479 (56.8%)	1 (Masculino): 1191 (59.2%)
Grupo	1 (cólón): 509 (60.4%)	1 (cólón): 1137 (56.5%)
Topográfico	2 (reto): 334 (39.6%)	2 (reto): 875 (43.5%)
Grupo	1 (adenomas/adenocarcinomas): 817 (96.9%)	1 (adenomas/adenocarcinomas): 1952 (97.0%)
Histológico	2 (outras histologias): 26 (3.1%)	2 (outras histologias): 60 (3.0%)
Estadio	I: 160 (19.0%)	I: 291 (14.5%)
	IIA: 195 (23.1%)	IIA: 548 (27.2%)
	IIB: 24 (2.8%)	IIB: 60 (3.0%)
	IIIA: 30 (3.6%)	IIIA: 53 (2.6%)
	IIIB: 131 (15.5%)	IIIB: 404 (20.1%)
	IIIC: 58 (6.9%)	IIIC: 156 (7.8%)
	IV: 245 (29.1%)	IV: 500 (24.9%)
Tempos (dias)	Mínimo: 0	Mínimo: 0
	1º Quartil: 507	1º Quartil: 642.5
	Mediana: 1501	Mediana: 1827
	Média: 1191	Média: 1283.4
	3º Quartil: 1827	3º Quartil: 1827
	Maximo: 1827	Maximo: 1827
Coortes	843	2012

Com o tratamento de dados o número inicial de registos passou de 7536 para 2855. Apesar dos dados tratados representarem 38% dos dados iniciais, os pacientes inicialmente identificados como estando no estadio IV aumentaram de 543 para 745 (passando a representar 26% dos registos, enquanto que nos dados iniciais representavam apenas 7%). O aumento no número de pacientes agora identificados com esse estadio tem a ver com o facto dos mesmos terem, inicialmente, o seu valor não preenchido, o que foi corrigido pelo facto da variável M=1 tornar inequívoco o estadio IV.

## 3.2. Dados de Mortalidade Geral

### 3.2.1. Tábuas de Mortalidade

O programa CONCORD desenvolvido pelo grupo Cancer Survival Group (CSG) da London School of Hygiene and Tropical Medicine <http://csg.lshtm.ac.uk/> tem como objetivo descrever e explicar as variações internacionais da sobrevivência ao cancro. Depois de uma primeira fase em que foram recolhidos dados de mortalidade, a nível internacional, para o período entre 1990 e 1994, o estudo CONCORD-2, inclui dados de mais de 270 registos em 61 países recolhidos a partir de 1995 até 2010.

Esta pesquisa pretende explicar tendências e desigualdades a nível regional, nacional e internacional. Os seus resultados deverão ajudar as autoridades de saúde a desenvolver políticas na área do tratamento e prevenção do cancro que permitam aumentar a sobrevivência e diminuir as desigualdades àqueles níveis.

De forma a tornar possível essa comparabilidade, o CSG disponibiliza as tábuas de mortalidade utilizadas por esse estudo, sendo essas tabelas organizadas por ano civil, idade e sexo, fornecendo informação sobre a taxa de mortalidade  $m_x$ , probabilidade de morte  $q_x$  e probabilidade de sobrevivência  $p_x = 1 - q_x$  (Spika D 2015).

As tábuas de mortalidade aí disponibilizadas e utilizadas neste estudo no âmbito de AS relativa são referentes ao Norte de Portugal (área coincidente com a dos dados observados do ROENO) e têm como origem o Instituto Nacional de Estatística.

*Tabela 7 Estrato tábua de mortalidade utilizada para CONCORD-2 (Spika D and Bonaventure A)*

Continent	Country	Registry	Year	Sex	Age (Years)	Mortality rate (mx)	Probability of death (qx)	Probability of survival (px)
Europe	Portugal	Portugal North	1995	Male	0	0.006692	0.006670	0.993330
Europe	Portugal	Portugal North	1995	Male	1	0.000795	0.000795	0.999205
Europe	Portugal	Portugal North	1995	Male	2	0.000396	0.000396	0.999605

Europe	Portugal	Portugal North	1995	Male	3	0.000300	0.000300	0.999700
Europe	Portugal	Portugal North	1995	Male	4	0.000246	0.000246	0.999754
Europe	Portugal	Portugal North	1995	Male	5	0.000215	0.000215	0.999785
Europe	Portugal	Portugal North	1995	Male	6	0.000201	0.000201	0.999799
Europe	Portugal	Portugal North	1995	Male	7	0.000198	0.000198	0.999802
Europe	Portugal	Portugal North	1995	Male	8	0.000204	0.000204	0.999796
Europe	Portugal	Portugal North	1995	Male	9	0.000220	0.000220	0.999780
Europe	Portugal	Portugal North	1995	Male	10	0.000246	0.000246	0.999754

### 3.2.2. Tratamento das tábuas para ASR

Relativamente às coortes em observação, as mortes podem ocorrer durante o período 2000-2006 para a coorte 1 e 2007-2013 para a coorte 2.

As tábuas de mortalidade disponibilizadas abrangem o período que decorre desde 1995 e 2010. Assim, na ausência de dados relativos a 2011, 2012 e 2013, assumimos para esses anos os valores de  $m_x$ ,  $q_x$  e  $p_x$  de 2010, completando a tabela. Foram, também, retirados os registos relativos ao período 1995-1999, dado que o mesmo não coincide com os períodos de observação das coortes estudados. Como resultado final obtivemos tábuas relativas ao período 2000-2013 (o qual será utilizado em ambas as coortes, dado que todo o período temporal é abrangido).

Das tábuas de mortalidade iniciais foram também retirados os campos *Continent*, *Country* e *Registry*, por não serem necessários, dado que o presente estudo se foca apenas na região Norte de Portugal e não inclui no seu âmbito comparabilidade da sobrevivência CCR a nível regional, nacional ou internacional.

A partir destes dados foi necessário criar duas matrizes, uma para cada sexo, com as probabilidades  $p_x$  condicionais de sobrevivência a 1 ano, cujas linhas representam a idade e cujas colunas representam os anos relativos aos períodos de observação do estudo (2000-2013).

Tendo como *input* estas duas matrizes é utilizado o comando *transrate* do pacote *ReIsurv* (Perme M.P. 2015), para criar uma tabela com as dimensões idade, sexo e ano, denominada como *ratetable*, utilizada para ASR em todas as funções do pacote R referido.

O comando *transrate* transforma as probabilidades condicionais  $p_x$  de sobrevivência a 1 ano em taxas médias de mortalidade diárias  $m_x$ , pois o dia é a unidade utilizada em todas as variáveis do referido pacote.

Assim, a transformação efetuada é conseguida pela fórmula

$$m_x(\text{diária}) = -\log(p_x)/365.24. \quad (10)$$

Por exemplo, se um paciente do sexo feminino chegar aos 99 anos em 2013, a probabilidade expetável de conseguir chegar aos 100 anos é de 59.14% e a sua taxa de mortalidade diária é de 0.144% durante esse período,

$$-\log(0.5914)/365.24 = 0.00144. \quad (11)$$

A *ratetable* criada está, assim, estruturada por anos, tendo em cada ano as taxas de mortalidade diárias organizadas por idade sequencial (0-99 anos), para ambos os sexos. Esta estrutura pode ser visualizada no anexo 6.2.

Tal como anteriormente referido, dado que replicámos os dados das tábuas de mortalidade de 2010 para 2011, 2012 e 2013, os dados da *ratetable* destes últimos anos são iguais aos de 2010. Conseguimos, também, constatar que as taxas de mortalidade geral na região norte de Portugal foram diminuindo ao longo deste período (2000-2010).

### 3.3. Análise de Sobrevivência

Utilizando, então, um tempo de observação de 5 anos após o diagnóstico, vamos considerar métodos não paramétricos, através do estimador de *Kaplan-Meier* e métodos semiparamétricos (com inclusão de covariáveis) através do Modelo de Cox de riscos proporcionais.

Iremos recorrer ao estimador de *Kaplan-Meier* pois o mesmo é, “*sem dúvida, o mais utilizado em estudos clínicos e vem ganhando cada vez mais espaço em estudos de confiabilidade*” (Colosimo E. A. 2006). Os estimadores não paramétricos, como o de *Kaplan-Meier*, não permitem a inclusão direta de covariáveis quantitativas na análise. Através destes estimadores conseguimos, de forma simples, dividir os dados em estratos, no entanto se estiverem envolvidas várias covariáveis, o método mostra-se limitado pois obriga à criação de inúmeros estratos, alguns com dimensões de amostra eventualmente pequenas, logo com uma mais difícil comparabilidade. Para além disso, as variáveis contínuas têm de ser transformadas em discretas, agrupando os dados em classes, perdendo informação nesse processo.

De forma a podermos incluir covariáveis quantitativas diretamente na análise teremos que utilizar modelos de regressão paramétricos como o Exponencial ou o de Weibull ou semiparamétricos como o modelo de Cox. Este modelo é bastante utilizado em estudos médicos dada a sua flexibilidade, fruto da presença de uma componente não paramétrica tal como se descreve nos pontos abaixo.

#### 3.3.1. Estimador não paramétrico de Kaplan-Meier

Para este estimador não paramétrico proposto por Kaplan e Meier (Kaplan, Meier 1958),  $\hat{S}(t)$ , é uma função escada com degraus nos tempos relativos aos eventos de interesse, ordenados e não uniformes, e considera um número de intervalos de tempo igual ao número de eventos de interesse distintos.

Considerando que:

$n$  é a dimensão da amostra observada;

$t_j$  são os instantes onde ocorrem os eventos de interesse, distintos e ordenados;

$n_j$  são o número de indivíduos em risco no instante  $t_j$ , os quais não foram censurados nem morreram até ao instante imediatamente anterior a  $t_j$ ;

$d_j$  são o número de eventos de interesse no instante  $t_j$ ;

o estimador de *Kaplan-Meier* é, então, definido como

$$\hat{S}(t) = \prod_{j: t_j < t} \left( \frac{n_j - d_j}{n_j} \right) = \prod_{j: t_j < t} \left( 1 - \frac{d_j}{n_j} \right). \quad (12)$$

Este estimador, tal como outros, apresenta uma variabilidade associada à amostragem, sendo a sua variância assintótica estimada pela fórmula de Greenwood

$$\widehat{var}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{j: t_j < t} \frac{d_j}{n_j(n_j - d_j)} \quad (13)$$

em que o intervalo de confiança (IC) aproximado a  $100(1 - \alpha)\%$  de confiança para o estimador de *Kaplan-Meier* é dado por

$$\hat{S}(t) \mp Z_{\alpha/2} \sqrt{\widehat{var}(\hat{S}(t))} \quad (14)$$

tendo em conta que  $\alpha/2$  indica o percentil da distribuição normal padrão.

### 3.3.2. Modelo de Cox de riscos proporcionais

Ao contrário dos modelos paramétricos, tais como o exponencial ou Weibull, que não são utilizados neste estudo, o modelo de Cox não exige a existência duma distribuição estatística para o tempo de sobrevivência. Este modelo, proposto por Cox (Cox 1972), é denominado como modelo de riscos proporcionais.

O modelo de Cox parte do princípio que existe proporcionalidade entre a taxa de risco de grupos distintos. Assim, este modelo é dividido numa componente paramétrica (constituída pelos coeficientes de proporcionalidade entre grupos) e



numa componente não paramétrica (taxa de risco de base). Esta componente não paramétrica dota o modelo de grande versatilidade, o que leva a que seja bastante utilizado em estudos clínicos (Colosimo E. A. 2006).

Tendo em conta que o princípio básico para estimar o efeito das covariáveis é a proporcionalidade dos riscos ou taxas de mortalidade, neste caso, ao longo do período de observação, vamos considerar a sua forma mais simples.

Ainda seguindo a linha de raciocínio em (Colosimo E. A. 2006), se considerarmos uma única covariável, a qual representa um indicador de grupo, e dois grupos aleatórios de pacientes os quais recebem um tratamento padrão (grupo 0) e um tratamento novo (grupo 1), assumimos proporcionalidade entre as respetivas taxas de mortalidade,  $\lambda_0(t)$  e  $\lambda_1(t)$  se

$$\frac{\lambda_1(t)}{\lambda_0(t)} = K \quad (15)$$

em que  $K$  é a razão das taxas de mortalidade, razão essa constante ao longo do tempo  $t$ .

Assim, se a função taxa de mortalidade do grupo com o tratamento padrão é  $\lambda_0(t)$ , a função taxa de mortalidade para o grupo com o novo tratamento  $\lambda_1(t)$  é dada por  $K\lambda_0(t)$ .

Neste caso se  $K > 1$  o novo tratamento não parece ser eficaz, sendo-o apenas para valores de  $K$  inferiores a  $1$ , o que significaria que a taxa de mortalidade para o grupo com o novo tratamento é menor.

Como a razão de taxas de mortalidade  $K$  é não negativa, é conveniente considerar  $K = \exp(\beta x)$ , sendo  $\beta$  o coeficiente da covariável  $x$  no modelo e  $x = 0$  para o grupo  $0$  e  $x = 1$  para o grupo  $1$ .

Assim, a forma mais simples para o modelo de Cox para uma única covariável é nos dada por

$$\lambda(t|x) = \begin{cases} \lambda_0(t)\exp(\beta), & \text{se } x = 1 \\ \lambda_0(t) & \text{se } x = 0 \end{cases} \quad (16)$$

De forma genérica, considerando  $p$  covariáveis, de forma a que  $x$  seja um vetor com as componentes  $x = (x_1, x_2, \dots, x_p)'$  e  $\beta$  o vetor com os coeficientes  $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ , a expressão geral do modelo de Cox é-nos dado por

$$\lambda(t|x) = \lambda_0(t)\exp(\beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p) = \lambda_0(t)\exp(x'\beta). \quad (17)$$

Assim, confirmamos que o modelo é composto pelo produto de duas componentes, uma não paramétrica e outra paramétrica. A componente não paramétrica  $\lambda_0(t)$  é uma função não negativa no tempo e é normalmente denominada de taxa de mortalidade, pois  $\lambda(t|x) = \lambda_0(t)$  quando  $x$  é nulo.

O modelo acima expresso em termos de função de mortalidade, pode ser expresso em termos de função de sobrevivência

$$S(t|x) = [S_0(t)]^{\exp(x'\beta)}, \quad (18)$$

este modelo é também denominado de modelo de riscos proporcionais pois a razão das taxas de mortalidade é constante no tempo. Isto é, a razão das taxas de mortalidade para os indivíduos  $i$  e  $j$  pode ser representada por

$$\frac{\lambda(t|x_i)}{\lambda(t|x_j)} = \frac{\lambda_0(t)\exp(x_i'\beta)}{\lambda_0(t)\exp(x_j'\beta)} = \exp(x_i'\beta - x_j'\beta). \quad (19)$$

A estimação dos coeficientes  $\beta$  do modelo de Cox pelo método da máxima verosimilhança é inapropriada, dada a presença e dependência da componente não paramétrica  $\lambda_0(t)$ , como se pode verificar pela função de máxima verosimilhança

$$L(\beta) = \prod_{i=1}^n [f(t_i | X_i)]^{\delta_i} [S(t_i | X_i)]^{1-\delta_i} = \prod_{i=1}^n [\lambda(t_i | X_i)]^{\delta_i} [S(t_i | X_i)]. \quad (20)$$

No modelo de Cox, a função de verosimilhança é dada por

$$S(t_i | X_i) = \exp\left\{-\int_0^{t_i} \lambda_0(u)\exp\{X_i'\beta\}du\right\} = [S_0(t_i)]^{\exp\{X_i'\beta\}} \quad (21)$$

logo, aplicando este resultado em (20) obtemos

$$L(\beta) = \prod_{i=1}^n [\lambda_0(t_i) \exp\{X_i' \beta\}]^{\delta_i} [S_0(t_i)]^{\exp\{X_i' \beta\}} \quad (22)$$

que é função da componente não paramétrica  $\lambda_0(t)$ , como queríamos demonstrar.

### Método da verosimilhança parcial

De forma a evitar essa dependência, Cox formalizou no seu artigo original (Cox 1975), o método da máxima verosimilhança parcial, condicionando a construção da função de verosimilhança ao conhecimento da história passada de falhas e censuras (Colosimo E. A. 2006).

Considerando-se que para uma amostra de  $n$  indivíduos, existem  $k \leq n$  falhas distintas nos tempos  $t_1, t_2, \dots, t_k$  considera-se a probabilidade condicional da  $i$ -ésima observação falhar no tempo  $t_i$ , conhecendo as observações sob risco nesse período

$$P[\text{indivíduo falhar em } t_i \mid \text{uma falha em } t_i \text{ e história até } t_i] = \quad (23)$$

$$\frac{P[\text{indivíduo falhar em } t_i \mid \text{sobrevive a } t_i \text{ e história até } t_i]}{P[\text{uma falha em } t_i \mid \text{história até } t_i]} =$$

$$\frac{\lambda_i(t|X_i)}{\sum_{j \in R(t_i)} \lambda_j(t|X_j)} = \frac{\lambda_0(t) \exp\{X_i' \beta\}}{\sum_{j \in R(t_i)} \lambda_0(t) \exp\{X_j' \beta\}} = \frac{\exp\{X_i' \beta\}}{\sum_{j \in R(t_i)} \exp\{X_j' \beta\}}$$

em que  $R(t_i)$  é o conjunto dos índices das observações sob risco em  $t_i$ . Assim, condicionando ao historial de falhas e censuras até ao tempo  $t_i$ , verificamos que a componente não paramétrica desaparece da equação (23).

A função de verosimilhança parcial a ser utilizada para inferir os parâmetros do modelo é formada pelo produto de todos os termos representados na equação (23), considerando os distintos tempos de falha,

$$L(\beta) = \prod_{i=1}^k \frac{\exp\{X'_i\beta\}}{\sum_{j \in R(t_i)} \exp\{X'_j\beta\}} = \prod_{i=1}^n \left( \frac{\exp\{X'_i\beta\}}{\sum_{j \in R(t_i)} \exp\{X'_j\beta\}} \right)^{\delta_i} \quad (24)$$

em que  $\delta_i$  é o indicador de falha.

Mais tarde, (Andersen 1982) provaram a consistência desses estimadores de máxima verosimilhança parcial, tornando possível a aplicação de estatísticas como as de Wald, razão das verosimilhanças e *score* (Colosimo E. A. 2006).

### Interpretação dos coeficientes

O teste de Wald é um método para testar a significância de determinadas covariáveis num modelo estatístico, constituindo uma das formas de testar se os parâmetros associados a um grupo de covariáveis é nulo. Se o teste para esse conjunto de covariáveis for significativo, podemos concluir que os coeficientes  $\beta$  associados a essas covariáveis é diferente de zero. Caso contrário, se o teste de Wald não for significativo para determinadas covariáveis, as mesmas podem ser omitidas do modelo (Kyngas 2001).

### Adequação do modelo

Sendo o modelo de Cox baseado no pressuposto de riscos proporcionais, mesmo que a magnitude do risco varie para diferentes indivíduos, como função dos diferentes valores das suas covariáveis, esse mesmo efeito é constante ao longo do tempo, ou seja, não há interações com o mesmo (Allison 2010).

Assim sendo, e de modo a verificar a adequabilidade do modelo, recorreremos aos resíduos de Schoenfeld (Schoenfeld 1982), os quais nos permitem avaliar o pressuposto dos riscos proporcionais, após o ajustamento do modelo de Cox.

Considerando que se o  $i$ -ésimo indivíduo com o vetor de covariáveis  $x_i = (x_{i1}, \dots, x_{ip})$  falhar, temos para esse indivíduo um vetor de resíduos  $r_i = (r_{i1}, \dots, r_{ip})$ , em que cada componente  $r_{iq}$ , para  $q = (1, \dots, p)$ , condicional a uma falha no conjunto de risco  $R(t_i)$  é definida por (Colosimo E. A. 2006)

$$r_{iq} = x_{iq} - \frac{\sum_{j \in R(t_i)} x_{jq} \exp\{x'_j \hat{\beta}\}}{\sum_{j \in R(t_i)} \exp\{x'_j \hat{\beta}\}} \quad (25)$$

Assim, os resíduos são definidos para cada falha e não para as censuras. O conjunto dos resíduos de Schoenfeld constitui uma matriz com  $d$  linhas (representando o número de falhas) e  $p$  colunas (representando o número de covariáveis consideradas). Assim cada linha representa um tempo de falha e cada coluna representa uma covariável, sendo a  $i$ -ésima linha dessa matriz obtida pela equação (25). Para os resíduos é normal que  $\sum_i r_i = 0$ . De modo a que seja possível considerar a estrutura de correlação entre os resíduos, é normalmente utilizada (Colosimo E. A. 2006) a forma padronizada dos resíduos de Schoenfeld definida por

$$S_i^* = [\mathfrak{I}(\hat{\beta})]^{-1} \times r_i \quad (26)$$

em que  $\mathfrak{I}(\hat{\beta})$  é a matriz da informação observada (Therneau 2000).

De forma a que exista proporcionalidade entre as taxas de falha, deveremos ter em conta a restrição  $\beta(t) = \beta$ , assim se  $\beta(t)$  não for constante, o impacto de uma ou mais covariáveis na taxa de falha pode variar com o tempo (Grambsch 1994). Assim, o gráfico de  $\beta(t)$  versus  $t$  deverá ser uma linha horizontal.

Com dados reais, é muito pouco provável que a proporcionalidade das taxas de falha seja totalmente constante, mas poderá ou não sê-lo de forma aproximada, o que leva a que a interpretação gráfica possa levar a conclusões subjetivas. Assim, poderá ser útil a obtenção de medidas estatísticas tais como o coeficiente de correlação de Pearson ( $\rho$ ), o qual poderá ou não verificar se existem correlações lineares entre os resíduos padronizados de Schoenfeld e o tempo. Valores de  $\rho$  próximos de zero tendem a mostrar não haver evidências para rejeitarmos a suposição de taxas de falha proporcionais (Colosimo E. A. 2006).

## 3.4. Análise de Sobrevivência Relativa

Os estimadores de sobrevivência relativa *Ederer I*, *Ederer II*, *Hakulinen* e *Net Survival*, introduzidos nos cap. 2.3.3 e 2.3.4, bem como os modelos de regressão introduzidos no cap. 2.3.5, são descritos nos subcapítulos abaixo.

### 3.4.1. Estimadores

#### **Ederer I**

Este estimador *Ederer I* (Ederer 1961) calcula a sobrevivência individual esperada, isto é, a ocorrência ou não de morte, para cada paciente da população com características semelhantes às de cada paciente do grupo em estudo. A soma desses eventos de interesse dá-nos o número esperado de pacientes que deverão morrer num determinado período, assumindo que eles seguem as taxas de mortalidade obtidas a partir das tábuas de mortalidade. A média desses valores dá-nos um estimador não enviesado da sobrevivência esperada para um subconjunto da população de características similares às do grupo de pacientes e pode ser utilizada como denominador na equação (7).

Todavia, podem surgir problemas com este método, pois, por exemplo, se um grupo de pacientes recentemente admitido em determinado estudo não tiver uma distribuição semelhante em termos de idade e sexo, relativamente a outro grupo introduzido nas primeiras fases desse mesmo estudo, como os tempos de censura não são independentes de idade e sexo, este fator poderá causar um enviesamento da sobrevivência observada. Nestes casos seria favorável incluir informação sobre censura nos cálculos da sobrevivência esperada (Perme M.P. 2007).

#### **Ederer II**

Ederer e Heise (Ederer F. 1959), sugeriram um método para estimar a sobrevivência esperada, que permite tempos de *follow-up* heterógeneos. Contrariamente ao método *Ederer I*, que calcula a média da sobrevivência esperada de entre todos os pacientes no estudo, o método *Ederer II* apenas inclui cada paciente na condição de que o mesmo esteja dentro do período de *follow-up*.

Assim, a sobrevivência esperada depende da mortalidade observada, dependência essa que faz aumentar a variância da sobrevivência esperada, tornando-a não negligenciável comparada com a sobrevivência observada (Perme M. P. 2007).

### **Hakulinen**

O método de *Hakulinen* toma em consideração os tempos de *follow-up* potencialmente heterógeneos mas ignora os instantes de tempo de ocorrência dos eventos de interesse, ou seja, de morte. Aqui o tempo de *follow-up* é substituído pelo potencial tempo de *follow-up*, mantendo-se as censuras mas substituindo os instantes de tempo onde ocorre a morte pelo potencial máximo período de *follow-up* do estudo, o qual corresponde à diferença entre a data de término do estudo e a data de diagnóstico. O intervalo de tempo é dividido de acordo com esses tempos e a probabilidade da sobrevivência esperada condicional é calculada em cada intervalo, baseada nos pacientes em risco nesse período, de acordo com os seus tempos de *follow-up* potenciais. Dessa forma a sobrevivência esperada não depende da mortalidade observada, mas toma em consideração os tempos de censura possivelmente heterogêneos (Perme M. P. 2007).

Estabeleceu-se um consenso entre as comunidades de registo de cancro, especialmente na Europa, destacando o estimador *Ederer II* como preferível, de entre os estimadores acima referidos, passando a ser o método por defeito a ser utilizado pela Eurocare a partir de 2011, tal como referido no relatório técnico do NCI (*National Cancer Institute*, (Hyunsoon C. 2011).

### **Net Survival**

Pohar-Perme demonstrou que a sobrevivência relativa tem sido calculada com um estimador enviesado de *Net Survival* (Perme M.P. 2012). No entanto esse enviesamento na, prática, poderá ser diminuto, pelo menos na abordagem do estimador *Ederer II* (Lambert P. 2014).

De acordo com os autores do estimador (Perme M.P. 2012), será, abaixo descrito o método utilizado. Previamente e assumindo o modelo aditivo (8), para cada indivíduo  $i$  no grupo de estudo, introduzimos as seguintes notações:

$T_{Ei}$  – Tempo até à morte causada pela doença em causa;

$T_{Pi}$  – Tempo até à morte causada por outras causas na população em geral;

$C_i$  – Tempo até à censura;

Dado que a ocorrência de morte devido à causa específica impede a observação do tempo até à morte devida a outras causas, nunca poderemos observar em simultâneo  $T_{Ei}$  e  $T_{Pi}$ , apenas podemos observar o que tem o menor desses tempos, o que poderemos considerar:

$T_i = \min\{T_{Ei}, T_{Pi}\}$ , que está sujeito à censura  $C_i$ ;

$U_i = \min\{T_i, C_i\}$  – tempo de *follow-up* do indivíduo  $i$  que define o indicador de falha  $\delta_i$  que é igual a 0 se o evento for censurado ( $T_i > C_i$ ), e 1 caso contrário;

$X_i$  – Valor das covariáveis para o sujeito  $i$ ;

$D_i$  – Subgrupo de  $X_i$  contendo as variáveis demográficas (tipicamente idade, sexo e ano);

Assim, os dados observados no indivíduo  $i$  são  $(U_i, \delta_i, X_i)$  e as funções de sobrevivência são dadas por  $S_{Ei}(t) = P(T_{Ei} > t | X_i)$  e  $S_{Pi}(t) = P(T_{Pi} > t | D_i)$ , que geram as taxas de risco  $\lambda_{Ei}(t)$  e  $\lambda_{Pi}(t)$ , respetivamente. Utilizamos, também, o símbolo  $\Lambda$  para representar a função taxa de risco acumulada. Assumimos, ainda, que a população é finita de dimensão  $N$  e que o grupo em estudo tem dimensão  $n$ .

Num contexto em que desconhecemos as causas de morte, “*relative survival setting*” temos dois conjuntos de dados: os do grupo de indivíduos com a doença de interesse e dos estratos da população em geral obtidos a partir das tábuas de mortalidade. Assumimos, também, que os indivíduos do grupo em observação são comparáveis à população em geral em relação a fatores de risco externos. O valor  $\lambda_{Pi}(t)$  é obtido a partir das tábuas de mortalidade e assumimos que a causa de morte específica em estudo é rara e tem um efeito negligenciável naquele valor.

Para enquadrar o estimador *Net Survival* de Pohar-Perme, partimos dum contexto em que as causas de morte são conhecidas e perfeitamente identificadas “*cause-*



*specific setting*”, e posteriormente adaptamos a um contexto de sobrevivência relativa, “*relative survival setting*”, sendo este último o considerado no presente estudo, dado não existir informação sobre a causa de morte nos registos em análise.

### ***cause specific setting***

Partindo, então, do princípio que as causas de morte estão bem identificadas, os dados observados podem ser representados por um processo de contagem.

O enquadramento teórico da Teoria dos Processos de Contagem, dada a sua complexidade e distanciamento matemático do âmbito deste estudo, é descrito no apêndice A.2 do documento (Perme M.P. 2007).

Assim os dados observados podem ser representados pelo processo de contagem  $N_{Ei}(t) = I\{T_i \leq t, T_i \leq C_i, T_{Ei} \leq T_{Pi}\}$  e pelo processo de risco  $Y(t)$ .

De forma a obter uma estimativa não enviesada de  $\Lambda_E(t)$ , pondera-se tanto o processo de contagem como o de risco de cada indivíduo utilizando a distribuição no tempo de sobrevivência na população, definindo

$$N_{Ei}^\omega = \frac{N_{Ei}(t)}{S_{Pi}(t_i)}, Y_i^\omega = \frac{Y_i(t)}{S_{Pi}(t_i)} \quad (27)$$

e propõe-se o estimador  $\hat{\Lambda}_E^\omega(t) = \int_0^t dN_E^\omega(u)/Y^\omega(u)$ , que é centrado apesar da censura devida aos riscos da população  $N_E^\omega = \sum N_{Ei}^\omega$  e  $Y^\omega = \sum Y_i^\omega$ . Foi considerado este estimador, de Nelson-Aalen, porque foi apontado como um estimador consistente para a mortalidade acumulada (Perme M.P. 2012).

O peso de  $Y$  implica o aumento do número de indivíduos ainda em risco com influência na proporção de pacientes que morreram devido aos riscos da população. O peso de  $N_{Ei}$  implica o aumento do número de eventos, da mesma forma. Uma vez que a informação requerida sobre os pesos pode ser obtida diretamente das tábuas de mortalidade, não é necessário recorrer a modelação nem às respetivas hipóteses assumidas.

**relative survival setting**

Desconhecendo as causas de morte, uma forma de estimar a *Net Survival*  $S_E(t) = \exp\left\{-\int_0^t \lambda_E(u)du\right\}$ , é utilizar um modelo de sobrevivência relativa multivariado. Assumindo que o risco populacional é totalmente determinado pelas variáveis demográficas, a assunção de independência condicional implica que a censura devida aos riscos da população é condicionalmente independente e a *Net Survival* pode ser estimada como a média dos valores previstos para cada indivíduo

$$P(t < T_E \leq t + \Delta t | T_P > t, T_E > t, D) = P(t < T_E \leq t + \Delta t | T_E > t, D). \quad (28)$$

Para estimar  $\lambda_E(t)$ , seguindo a mesma lógica da equação (27), Maja Pohar propõe ponderar o estimador *Ederer II*. Assumindo  $N_i^\omega(t) = N_i(t)/S_{Pi}$  e  $N^\omega(t) = \sum N_i^\omega(t)$ , o estimador é definido como

$$\hat{\Lambda}_E(t) = \int_0^t \frac{dN^\omega(u)}{Y^\omega(u)} - \int_0^t \frac{\sum_{i=1}^n Y_i^\omega(u) d\Lambda_{Pi}(u)}{Y^\omega(u)}, \quad (29)$$

para encontrar a estimativa da população pelo estimador acima (quando  $S_{Ci} = S_C$  para todo o  $i$ ), notamos que

$$E\{Y_i^\omega(t)\} = \frac{S_C(t)S_{Oi}(t)}{S_{Pi}(t)} = S_C(t)S_{Ei}(t) \quad (30)$$

e

$$E\{dN_i^\omega(t) | \mathcal{F}_t\} = \frac{E\{dN_i(t) | \mathcal{F}_t\}}{S_{Pi}(t)} = Y_i(t)d\Lambda_{Oi}(t)/S_{Pi}(t) \quad (31)$$

onde  $\mathcal{F}_t$  representa toda a informação disponível até ao instante de tempo  $t$ .

Como

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{Y_i(t)d\Lambda_{Oi}(t)}{S_{Pi}(t)} &\rightarrow \frac{S_C(t)}{N} \sum_{i=1}^N \frac{S_{Oi}(t)d\Lambda_{Oi}(t)}{S_{Pi}(t)} \\ &= \frac{S_C(t)}{N} \sum_{i=1}^N S_{Ei}(t)d\Lambda_{Oi}(t) \end{aligned} \quad (32)$$

o valor de  $\widehat{d\Lambda}_E(t)$  na equação (29) é equivalente a

$$\begin{aligned} & \frac{\sum_{i=1}^N S_{Ei}(t) d\Lambda_{Oi}(t)}{\sum_{i=1}^N S_{Ei}(t)} - \frac{\sum_{i=1}^N S_{Ei}(t) d\Lambda_{Pi}(t)}{\sum_{i=1}^N S_{Ei}(t)} \\ &= \frac{\sum_{i=1}^N S_{Ei}(t) d\Lambda_{Ei}(t)}{\sum_{i=1}^N S_{Ei}(t)} = d\Lambda_E(t) \end{aligned} \quad (33)$$

ficando, assim demonstrado que o estimador *Net Survival* de Maja Pohar é um método consistente para estimar  $\lambda_E(t)$ . Ao contrário dos estimadores tradicionais, este método não pode ser expresso como uma razão entre a sobrevivência observada e a sobrevivência da população, mas sim como resultante de funções definidas, apenas, a partir da taxa de mortalidade em excesso, o que leva a que seja possível a comparação de sobrevivência entre países, tal como referido no Cap.2.3.4. O facto de ter sido considerado, por diversos autores, o único estimador não enviesado, leva também, a que este estimador de *Net Survival* de Pohar-Perme, seja o estimador de eleição para análise de sobrevivência para os registos de cancro. A este facto acresce a viabilidade técnica, pois a criação do pacote *relsurv* do R, permite uma simples aplicação do método, mesmo em grandes conjuntos de dados (Roche L 2012).

### 3.4.2. Modelos Aditivos (Excess Mortality)

No modelo aditivo, a taxa de mortalidade observada (O) é a soma da taxa de mortalidade da população (P) e da taxa de mortalidade em excesso (E)

$$\lambda_o(t, z) = \lambda_p(t, c) + \lambda_E(t, z) \quad (34)$$

onde  $z = (c, x)$  e  $c$  o vetor das variáveis pelas quais as tábuas de mortalidade da população estão estratificadas e  $x$  o vetor de covariáveis adicionais que se pretendam incluir na análise de regressão.

A taxa de mortalidade em excesso é, usualmente (Pohar M. 2007), modelada como

$$\lambda_E(t, z) = h_0(t) \exp(\beta z), \quad (35)$$

assim sendo, a equação (34) torna-se

$$\lambda_O(t, z) = \lambda_P(t, c) + h_0(t) \exp(\beta z) \quad (36)$$

sendo  $h_0$  a taxa de mortalidade em excesso de base. A equação (34) dá-nos uma relação multiplicativa entre as funções de sobrevivência

$$S_{O(t,z)} = S_P(t, c) S_R(t, z), \quad (37)$$

equação esta que é equivalente à equação (7).

De notar, da equação (34) que se assume que  $\lambda_O(t, z) \geq \lambda_P(t, c)$  durante todo o período  $t$ , quaisquer que sejam os valores das covariáveis. Normalmente em estudos relativos ao cancro, esta assunção tende a ser verdadeira. Caso não o seja, a escolha destes modelos poderá não ser adequada.

Os modelos aditivos usualmente utilizados são o de *Hakulinen-Tenkanen* (T. Hakulinen T. 1987), *Poisson* (Dickman P.W. 2003) e *Estève* (Estève J. 1990), os quais podem ser descritos da forma abaixo apresentada.

### **Modelo aditivo Hakulinen-Tenkanen**

Com este modelo (Hakulinen T. 1987) os pacientes devem ser agrupados em  $K$  estratos, indexados por  $k$ , com um estrato para cada combinação das covariáveis relevantes tais como idade, sexo, estadio e ano da coorte. Uma tábua de mortalidade deverá ser estimada para cada estrato, com intervalos de tempo de *follow-up* indexados por  $i$ . A taxa  $\lambda_E(t)$  é assumida como uma função multiplicativa das covariáveis e constante em cada intervalo de tempo  $i$ .

Esta taxa de mortalidade em excesso pode, assim, assumir a forma

$$\lambda_E(t) = \exp\{\beta'z + \gamma'fu(t)\} \quad (38)$$

na qual  $z$  é o vetor das covariáveis,  $fu$  o vetor dos indicadores de intervalo de *follow-up*, função do tempo e  $\beta$  e  $\gamma$  os vetores dos coeficientes de regressão. Se considerarmos as proporções específicas de sobrevivência observada e esperada para cada intervalo  $i$  como  $p_{ki}$  e  $p_{ki}^*$  respectivamente, as equações (34) e (7) levam-nos a

$$\ln\left(-\ln\left(\frac{p_{ki}}{p_{ki}^*}\right)\right) = \beta'z + \gamma_i, \quad (39)$$

o que constitui um modelo linear generalizado (GLM) com estrutura de erro binomial e com uma função de ligação log-log combinada com a razão entre as proporções específicas de sobrevivência observada e esperada para cada intervalo.

### **Modelo aditivo Poisson**

Este modelo sendo um GLM é similar ao modelo de *Hakulinen-Tenkanen*, requerendo também o mesmo agrupamento dos dados, tendo no entanto uma estrutura de erro de *Poisson* (Dickman P.W. 2003). Aqui  $\lambda_E(t)$  assume, também, a forma da equação (38). Assume-se que o número de mortes observadas  $d_{ki}$  do estrato  $k$  no tempo  $i$  segue uma distribuição de *Poisson*,  $\mu_{ki}$ , na qual  $\mu_{ki} = \lambda_{E,ki} \times y_{ki}$ , sendo  $y_{ki}$  a combinação indivíduo-tempo em risco para o grupo  $k$  no intervalo de tempo  $i$ .

Sendo o número de mortes esperadas, na população,  $d_{ki}^*$ , a equação (34) pode ser escrita como

$$\frac{\mu_{ki}}{y_{ki}} = \frac{d_{ki}^*}{y_{ki}} + \exp(\beta'z + \gamma_i), \quad (40)$$

ou seja

$$\ln(\mu_{ki} - d_{ki}^*) = \beta'z + \gamma_i + \ln(y_{ki}). \quad (41)$$

Assim, temos um GLM com estrutura de erro de *Poisson*, função de ligação  $\ln(\mu_{ki} - d_{ki}^*)$ , offset  $\ln(y_{ki})$  e resultado  $d_{ki}$ . Neste modelo  $\frac{d_{ki}^*}{y_{ki}}$  é considerado como uma média das taxas de mortalidade esperadas da população  $\lambda_p$ .

### Modelo aditivo Estève

O modelo aditivo de *Estève* (Estève J.1990) utiliza os dados individuais e estima os coeficientes utilizando a abordagem da máxima verosimilhança. A função de verosimilhança para este modelo aditivo é-nos dada por

$$L(\beta) = \prod_{j=1}^n (\lambda_o(t_j))^{d_j} \exp\left(-\int_0^{t_j} \lambda_o(s) ds\right) \quad (42)$$

$$= \prod_{j=1}^n \left( \lambda_p(t_j) + \exp(\beta'z + \gamma'fu(t_j)) \right) d_j \exp\left(-\int_0^{t_j} (\lambda_p(s) + \exp(\beta'z + \gamma'fu(s))) ds\right)$$

onde  $t_j$  é o tempo de sobrevivência para a pessoa  $j$  e  $d_j$  o indicador do seu estado, o qual representa 1 se houver morte ou 0 se existir censura em  $t_j$ , isto aplicado a cada um dos  $j = 1, \dots, n$  indivíduos.

Considerando a mortalidade total como a soma da mortalidade da população e da mortalidade em excesso, função log-verosimilhança é então

$$l(\beta) = \sum_{j=1}^n d_j \ln\left(\lambda_p(t_j) + \exp(\beta'z + \gamma'fu(t_j))\right) - \sum_{j=1}^n \int_0^{t_j} \exp(\beta'z + \gamma'fu(s)) ds - \sum_{j=1}^n \int_0^{t_j} \lambda_p(s) ds. \quad (43)$$

O último termo da equação acima não depende de  $\beta$  e  $\gamma$  e pode ser omitido, o que constitui uma vantagem a nível computacional, pois apenas é necessária a informação da mortalidade esperada no momento  $t_j$  para cada um dos indivíduos (Dickman P.W. 2003).

Apesar do modelo ser especificado em tempo contínuo, é assumido, tal como nos outros modelos aditivos, que a mortalidade é constante em intervalos de tempo predefinidos.

A hipótese de proporcionalidade das taxas de risco para os modelos de sobrevivência relativa pode ser testado através dos resíduos parciais de Schoenfeld, tal como para o modelo de Cox (Perme M.P. 2015) de forma a garantir que a razão das taxas de mortalidade, neste caso da mortalidade em excesso, se mantêm constantes ao longo do período de observação.

Outro método para testar o ajuste dos modelos, testando a assunção dos riscos proporcionais é através da formação de uma *Brownian Bridge* (BB) (Perme M.P. 2015). O enquadramento teórico da BB, dada a sua complexidade e distanciamento matemático do âmbito deste estudo é descrito no apêndice A.1 do documento (Perme M.P. 2007).

Na interpretação do resultado desta BB interessará que o seu valor máximo absoluto não ultrapasse os valores críticos de 1.36 e 1.63, para níveis de significância de 1 e 5%, respetivamente (Perme M.P. 2007). São ideais os valores mais próximos de zero pois indicam uma maior constância do efeito das covariáveis ao longo do tempo.

### 3.4.3. Modelo Multiplicativo (Relative Mortality)

O modelo multiplicativo pode ser dado por

$$\lambda_o(t, z) = \lambda_p(t, c)v(t, z) \quad (44)$$

em que  $v(t, z)$  pode ser visto como a taxa de mortalidade relativa pois é igual à razão  $\lambda_o/\lambda_p$ .

Este modelo tem menos restrições matemáticas (Pohar M. 2007) que o modelo aditivo (34), mas tem sido considerado que os modelos aditivos são mais adequados, pelo menos em estudos relativos ao cancro (Dickman P.W. 2003), (Buckley 1984).

A taxa de mortalidade relativa é, usualmente (Andersen P.K. 1985), modelada como

$$v(t, z) = v_0(t)\exp(\beta'z) \quad (45)$$

onde  $v_0(t)$  é uma taxa de mortalidade de base.

Assim sendo, a equação (44) torna-se

$$\lambda_o(t, z) = \lambda_p(t, c)v_0(t) \exp(\beta'z). \quad (46)$$

### **Modelo Multiplicativo Andersen**

O modelo multiplicativo de *Andersen* (Andersen P.K. 1985) assume a forma

$$\lambda_o(t, z) = v_0(t, z) \exp\left(\beta'z + \ln(\lambda_p(t, c))\right) \quad (47)$$

o que o torna um modelo de *Cox* com uma variável adicional dependente do tempo com o coeficiente fixado em 1. Apesar de  $\lambda_p(t)$  ser um processo contínuo, as tábuas de mortalidade das populações são normalmente divididas em intervalos anuais, com alterações dos valores de  $\lambda_p(t)$  de intervalo em intervalo.

Tal como para os modelos aditivos, também para o modelo multiplicativo poderemos utilizar os métodos para teste de ajustamento recorrendo aos resíduos parciais de Schoenfeld e BB (Perme M.P. 2015).



## 4. Apresentação dos Resultados

Nos resultados abaixo apresentados iremos considerar um nível de significância de 5% e um período de observação de 5 anos.

### 4.1. Estimativas de Kaplan-Meier

Considerando que  $\hat{S}(t_j+)$  é a estimativa de *Kaplan-Meier* para a função de sobrevivência no intervalo entre o instante  $t_j$  e o instante onde ocorre o evento de interesse seguinte, são apresentadas de seguida as estimativas para ambas as coortes.

São, apresentados também, os intervalos de confiança a 95% para a função de sobrevivência baseados na fórmula de Greenwood. Nas tabelas abaixo é apresentado um extrato das tabelas completas.

Tabela 8 Estimativa Kaplan-Meier para sobrevivência para a Coorte 1

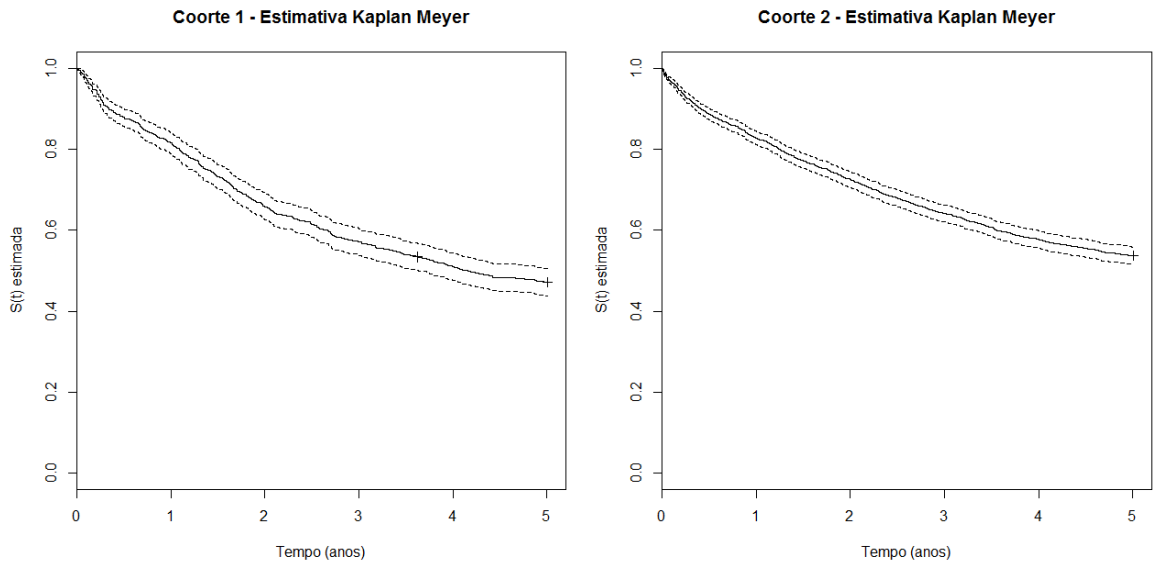
$t_j$	$n_j$	$d_j$	$\hat{S}_1(t_j+)$	Lower 95 %CI	Upper 95 %CI
0	843	1	0.999	0.996	1.000
368 ( $\approx 1$ ano)	687	1	0.814	0.787	0.840
730 ( $\approx 2$ anos)	555	1	0.657	0.625	0.689
1095 ( $\approx 3$ anos)	482	1	0.571	0.537	0.604
1465 ( $\approx 4$ anos)	429	1	0.509	0.475	0.543
1813 ( $\approx 5$ anos)	398	1	0.472	0.438	0.506

Tabela 9 Estimativa Kaplan-Meier para sobrevivência para a Coorte 2

$t_j$	$n_j$	$d_j$	$\hat{S}_1(t_j+)$	Lower 95 %CI	Upper 95 %CI
0	2012	2	0.999	0.998	1.000
365 ( $\approx 1$ ano)	1666	1	0.828	0.811	0.844
731 ( $\approx 2$ anos)	1459	1	0.725	0.705	0.744
1095 ( $\approx 3$ anos)	1291	1	0.641	0.620	0.662
1463 ( $\approx 4$ anos)	1161	1	0.577	0.555	0.598
1825 ( $\approx 5$ anos)	1080	1	0.536	0.514	0.558

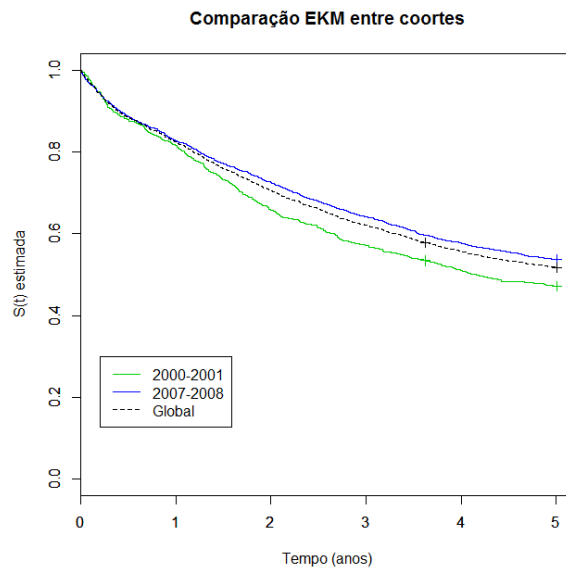
Essas estimativas de *Kaplan-Meier* para a função de sobrevivência dos dois coortes estão representadas nas figuras abaixo

Figura 6 Estimativas Kaplan-Meier para as duas coortes com IC95%



A comparação entre as duas funções de sobrevivência pode ser observada na seguinte figura.

Figura 7 Comparação das estimativas Kaplan-Meier para as duas coortes



Aplicando o teste de log-rank de comparação das duas estimativas de *Kaplan-Meier* das funções de sobrevivência por coorte obtemos um *p-value* (*p*) de 0.001.

Para cada coorte são apresentados o número de eventos de interesse ocorridos (Observado) e também o número de eventos de interesse esperados (Esperado). O valor da estatística de teste foi de 10.8 com 1 grau de liberdade. Sabendo que o teste log-rank é utilizado para testar a hipótese nula de que não há diferença entre os grupos (coortes 1 e 2) e sendo o  $p$  inferior a 0.05, podemos, para uma significância de 5%, concluir que existem diferenças estatisticamente significativas na sobrevivência observada entre as duas coortes, isto é, existe efeito de grupo, se bem que esta diferença se deve também a outras covariáveis que diferem entre as coortes, tal como veremos adiante.

Conseguimos, também, verificar que a sobrevivência aos 5 anos global, de 51.7% é mais distante da sobrevivência da coorte 1 (+4.5%) que da sobrevivência da coorte 2 (-1.9%), o que é esperado dado ser a dimensão da coorte 2 cerca de 2.4 vezes superior à da coorte 1.

Da mesma forma se verificam diferenças, para as coortes 1 e 2, da sobrevivência global a 1 ano de 82.4% (+1%,-0.4%), a 2 anos 70.5% (+4.8%,-2%), a 3 anos 62% (+4.9%,-2.1%) e a 4 anos 55.7% (+4.8%,-2%).

*Tabela 10 Kaplan-Meier - Teste log-rank para comparação da sobrevivência entre coortes*

Coorte	N	Observado	Esperado	$(O-E)^2/E$	$(O-E)^2/V$
1	843	445	390	7.73	10.8
2	2012	933	988	3.05	10.8
Qui-quadrado 10.8 com 1 grau de liberdade, $p = 0.00102$					

Comparando estatísticas das duas curvas de sobrevivência conseguimos obter os tempos medianos e percentagem de sobrevivência aos 5 anos na tabela abaixo.

*Tabela 11 Kaplan-Meier - Tempos medianos de vida e sobrevivência nas duas coortes*

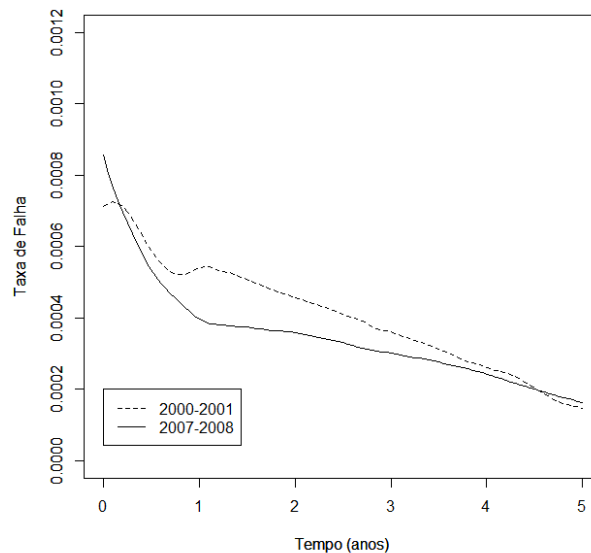
Coorte	Tempo Mediano de Vida	Sobrevivência aos 5 anos
1	4.13 anos	0.472
2	Sup. a 5 anos	0.536
Global	Sup. a 5 anos	0.517

Tal como se pode verificar graficamente conseguimos perceber que, tanto a sobrevivência como o tempo mediano de vida são superiores na coorte 2. A

estimativa de tempo em que 50% dos pacientes permanece vivo é de cerca de 4.13 anos (1509 dias) para a coorte 1. Para a coorte 2 este tempo mediano é superior ao período de observação de 5 anos.

Observa-se que existe uma relação entre a função de sobrevivência e a taxa de mortalidade, sendo que a uma maior taxa de mortalidade nos 2 primeiros anos, aproximadamente, corresponde um decréscimo mais acentuado na função de sobrevivência tanto para a coorte 1 como para a coorte 2.

Figura 8 Comparação das taxas de mortalidade entre as duas coortes



Conseguimos, também, demonstrar pelo teste log-rank, que existem curvas de sobrevivência com diferenças estatisticamente significativas por estadio em ambas as coortes, com valores para a estatística qui-quadrado de 385 e 995, respetivamente com 6 graus de liberdade e  $p < 0.001$ . Para o número total de pacientes verifica-se a mesma tendência, com  $p < 0.001$ , também.

Aplicando, também, o mesmo teste log-rank para as curvas 2 a 2 (procedimento adotado, da mesma forma, para as covariáveis grupo de idade e distrito) verificamos que na coorte 1 as diferenças são estatisticamente significativas entre o estadio IIIC e os restantes e entre o estadio IV e os restantes, não o sendo entre os estadios (I, IIA, IIB, IIIA) e (IIB, IIIA, IIIB).

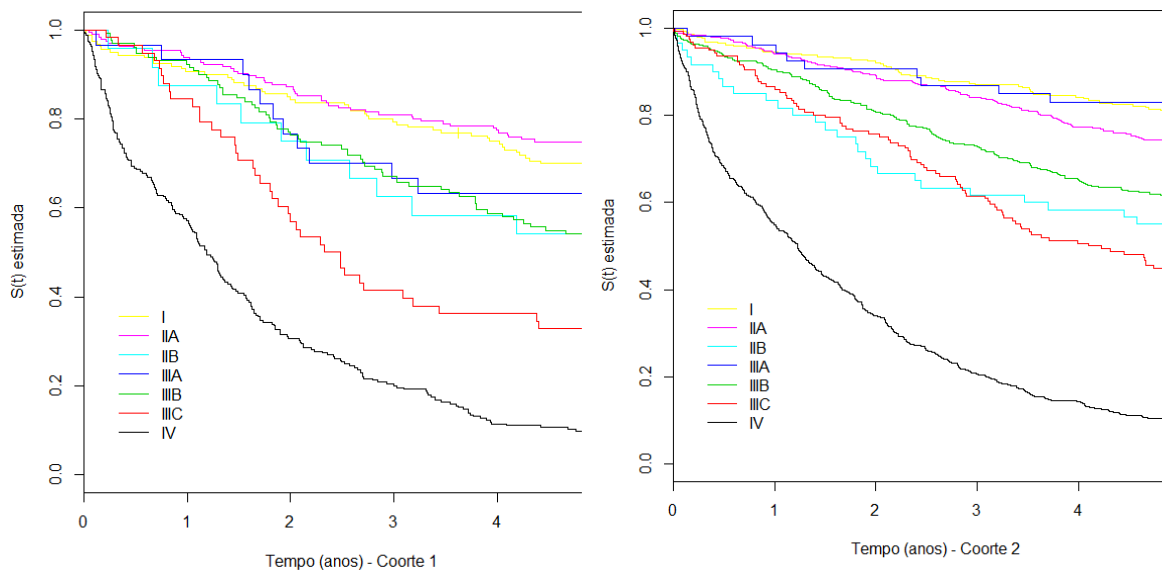
Na coorte 2 as diferenças são estatisticamente significativas entre o estadios IV e os restantes e entre os estadios IIIB e IIIC e os restantes excetuando o estadio IIB.

Tabela 12 Kaplan-Meier - Sobrevivência aos 5 anos por estadio

Sobrevivência aos 5 anos - Estimativa Kaplan-Meier						
Estadio	Coorte 1	n <sub>1</sub>	Coorte 2	n <sub>2</sub>	Total	n <sub>T</sub>
I	0.699	160	0.801	291	0.765	451
IIA	0.738	195	0.734	548	0.735	743
IIB	0.542	24	0.550	60	0.548	84
IIIA	0.633	30	0.830	53	0.759	83
IIIB	0.527	131	0.611	404	0.591	535
IIIC	0.310	58	0.449	156	0.411	214
IV	0.094	245	0.100	500	0.098	745

Torna-se perceptível que os estadios IIIC e IV têm bastante influência na sobrevivência, na coorte 1, sendo que na coorte 2 apenas se destaca o estadio IV, tal como também podemos verificar graficamente.

Figura 9 Kaplan-Meier – Curvas de sobrevivência por estadio, para as duas coortes



Pelo teste log-rank conseguimos perceber existirem curvas de sobrevivência com diferenças significativas em função da idade ( $p < 0.001$  em ambas as coortes e no conjunto global).

Na coorte 1 existem diferenças estatisticamente significativas entre o grupo etário 75+ e todos os outros excetuando o grupo 65-74. Existem também diferenças estatisticamente significativas entre os grupos 15-44 e 65-74.

Na coorte 2 existem diferenças estatisticamente significativas entre o grupo etário 75+ e todos os outros. Existem, também, diferenças estatisticamente significativas entre o grupo etário 45-54 e os grupos etários superiores.

Figura 10 Kaplan-Meier – Curvas de sobrevivência por grupos de idade, para as duas coortes

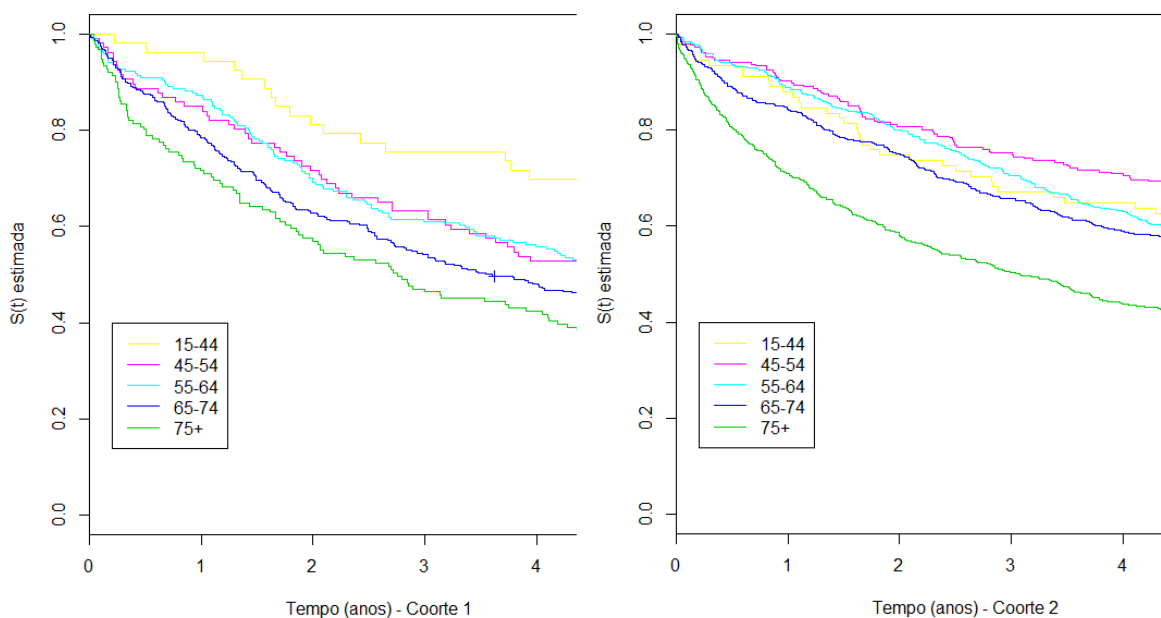


Tabela 13 Kaplan-Meier - Sobrevivência aos 5 anos por grupo de idade

Sobrevivência aos 5 anos - Estimativa Kaplan-Meier						
Idade	Coorte 1	n <sub>1</sub>	Coorte 2	n <sub>2</sub>	Total	n <sub>T</sub>
15-44	0.679	53	0.615	91	0.639	144
45-54	0.528	106	0.665	254	0.625	360
55-64	0.511	221	0.566	500	0.549	721
65-74	0.442	312	0.555	638	0.518	950
75+	0.364	151	0.410	529	0.400	680

Quanto à covariável sexo, não existem diferenças significativas para as curvas de sobrevivência ( $p = 0.276$  e  $p = 0.69$  nas coortes 1 e 2, respectivamente). O mesmo se verifica para o conjunto global com  $p = 0.394$ .

Tabela 14 Kaplan-Meier - Sobrevivência aos 5 anos por sexo

Sobrevivência aos 5 anos - Estimativa Kaplan-Meier						
Sexo	Coorte 1	n <sub>1</sub>	Coorte 2	n <sub>2</sub>	Total	n <sub>T</sub>
Masculino	0.455	479	0.531	1191	0.509	1670
Feminino	0.494	364	0.544	821	0.529	1185

O mesmo teste permitiu verificar não existirem diferenças para as curvas de sobrevivência em função do distrito para a coorte 2 ( $p = 0.455$ ), embora na coorte 1 pareçam existir curvas com essas diferenças ( $p = 0.0031$ ), sendo possível, por exemplo, verificar, nesta coorte, que no distrito de Vila Real (29.2%) a sobrevivência aos 5 anos é muito inferior à dos outros distritos: Viana do Castelo 57.7%, Bragança 57.1%, Braga 54.2% e Porto 44.5%. No conjunto global das duas coortes a diferença também não é significativa com  $p = 0.13$ .

Verificou-se que na coorte 1 existem diferenças estatisticamente significativas entre o distrito de Vila Real e os restantes excetuando o Porto, e entre o distrito do Porto e Viana ou Braga, não existindo essas diferenças entre os distritos de Braga, Bragança e Viana do Castelo.

Tabela 15 Kaplan-Meier - Sobrevivência aos 5 anos por distrito

Sobrevivência aos 5 anos - Estimativa Kaplan-Meier						
Distrito	Coorte 1	n <sub>1</sub>	Coorte 2	n <sub>2</sub>	Total	n <sub>T</sub>
Braga	0.542	212	0.558	330	0.552	542
Bragança	0.571	49	0.525	122	0.538	171
Porto	0.445	508	0.542	1274	0.514	1782
V. Castelo	0.557	26	0.481	158	0.495	184
V. Real	0.292	48	0.508	128	0.449	176

Na tabela seguinte, e dado existirem inúmeras topografias com um número muito pequeno de casos, apenas são apresentadas na tabela os códigos histológicos com 10 ou mais casos, pois não parece fazer sentido calcular a sobrevivência para números menores de casos. Da mesma forma serão tratados os *inputs* de todas as tabelas abaixo, sempre que se justifique.

Tabela 16 Kaplan-Meier - Sobrevivência aos 5 anos por topografia

Sobrevivência aos 5 anos - Estimativa Kaplan-Meier						
Topografia	Coorte 1	n <sub>1</sub>	Coorte 2	n <sub>2</sub>	Total	n <sub>T</sub>
C18.0	0.500	28	0.475	101	0.481	129
C18.1	-	-	-	9	-	9
C18.2	0.577	26	0.531	162	0.537	188
C18.3	0.458	24	0.500	62	0.488	86

C18.4	0.333	24	0.522	69	0.473	93
C18.5	0.455	11	0.500	40	0.490	51
C18.6	0.250	24	0.594	64	0.500	88
C18.7	0.442	156	0.587	482	0.552	638
C18.8	-	4	0.462	13	0.529	17
C18.9	0.509	212	0.496	135	0.504	347
C19.9	0.531	92	0.418	110	0.470	202
C20.9	0.455	242	0.540	765	0.519	1007

Verificou-se também não existirem diferenças significativas para as curvas de sobrevivência em função dos grupos topográficos sendo  $p = 0.425$  para a coorte 1,  $p = 0.709$  para a coorte 2 e  $p = 0.835$  para o conjunto global.

*Tabela 17 Kaplan-Meier - Sobrevivência aos 5 anos por grupo topográfico*

Sobrevivência aos 5 anos - Estimativa Kaplan-Meier						
Topografia (Grupos)	Coorte 1	$n_1$	Coorte 2	$n_2$	Total	$n_T$
Cólon	0.470	509	0.545	1137	0.522	1646
Reto	0.476	334	0.525	875	0.511	1007

*Tabela 18 Kaplan-Meier - Sobrevivência aos 5 anos por histologia*

Sobrevivência aos 5 anos - Estimativa Kaplan-Meier							
Histologia	Histgrp	Coorte 1	$n_1$	Coorte 2	$n_2$	Total	$n_T$
8140	Adenocarcinoma, NOS	1	0.479	759	0.540	1754	0.522 2513
8210	Adenocarcinoma em polipose adenomatosa	1	-	5	0.771	48	0.774 53
8246	Carcinoma neuroendócrino, NOS	2	-	-	0.545	11	0.545 11
8480	Adenocarcinoma mucinoso	1	0.444	27	0.464	112	0.460 139
8490	Carcinoma de células em anel de sinete	2	-	4	0.385	13	0.353 17

Verificou-se também não existirem diferenças significativas para as curvas de sobrevivência em função dos grupos histológicos na coorte 1 sendo  $p = 0.057$ . Na coorte 2,  $p < 0.001$  e no conjunto global,  $p < 0.001$  as diferenças são significativas.

*Tabela 19 Kaplan-Meier - Sobrevivência aos 5 anos por grupo histológico*

Sobrevivência aos 5 anos - Estimativa Kaplan-Meier						
Histologia (Grupos)	Coorte 1	$n_1$	Coorte 2	$n_2$	Total	$n_T$
Adenomas/Adenocarcinomas	0.476	817	0.542	1952	0.523	2769
Outras histologias	0.346	26	0.350	60	0.349	86



## 4.2. Modelo de Cox de riscos proporcionais

De forma a determinar quais as covariáveis significativas para o Modelo de Cox, consideramos, numa primeira fase, a inclusão de todas as covariáveis:

- Grupo de Idades
- Sexo
- Distrito
- Grupo Histológico
- Grupo Topográfico
- Estadio
- Coortes

Obtemos a seguinte tabela de resultados dos testes de Wald aos parâmetros:

*Tabela 20 Cox –Teste Wald com todas as covariáveis*

Covariável	$\hat{\beta}_i$	P
Idade (45-54)	0.16600	0.31074
Idade (55-64)	0.36988	0.01388
Idade (65-74)	0.57673	9.39e-05
Idade (75+)	1.15942	7.22e-15
Sexo(feminino)	-0.13294	0.01736
Distrito(Bragança)	0.01011	> 0.05
Distrito(Porto)	0.09863	> 0.05
Distrito(V. Castelo)	0.25665	0.03784
Distrito(V. Real)	0.08144	> 0.05
Hstgrp(outras)	0.42357	0.00223
Topgrp(reto)	0.15344	0.00579
Estadio (IIA)	0.16624	> 0.05
Estadio (IIB)	0.95067	5.51e-07
Estadio (IIIA)	0.14663	> 0.05
Estadio (IIIB)	0.76875	1.13e-10
Estadio (IIIC)	1.30757	< 2e-16
Estadio (IV)	2.34341	< 2e-16
Coortes(2007-2008)	-0.23574	6.43e-05

Na tabela acima foram consideradas como valores referência para a taxa base (*baseline*) o sexo masculino, o grupo de idades 15-44, o estadio I, o distrito de Braga, as histologias referentes a adenomas e adenocarcinomas, a localização topográfica cólon e a coorte 2000-2001.

Excluimos do modelo a covariável Distrito, cujos coeficientes dos seus fatores não são, no geral, significativos a 5%. Assim voltaremos a aplicar o teste de Wald agora apenas ao modelo com as covariáveis Grupo de Idade, Sexo, Estadio, Grupo Topográfico, Grupo Histológico e Coortes.

Considerando, então, o modelo com essas covariáveis obtemos a tabela de resultados dos testes de Wald, na qual incluímos os coeficientes dessas mesmas covariáveis, quando significativos, as respectivas taxas de risco e os IC de 95% dessas mesmas taxas:

*Tabela 21 Cox – Teste Wald com covariáveis significativas a 5%*

Covariável	$\widehat{\beta}_i$	$\text{Exp}(\widehat{\beta}_i)$	Lower 95 %CI	Upper 95 %CI	<i>p</i>
Idade (55-64)	0.37529	1.4554	1.0842	1.9538	0.012498
Idade (65-74)	0.58704	1.7986	1.3472	2.4014	6.86e-05
Idade (75+)	1.16131	3.1941	2.3853	4.2772	6.44e-15
Sexo(feminino)	-0.13080	0.8774	0.7868	0.9784	0.018664
Hstgrp(outras)	0.42738	1.5332	1.1696	2.0100	0.001977
Topgrp(reto)	0.15144	1.1635	1.0436	1.2972	0.006355
Estadio (IIB)	0.95688	2.6036	1.7957	3.7748	4.45e-07
Estadio (IIIB)	0.75738	2.1327	1.6889	2.6931	1.98e-10
Estadio (IIIC)	1.28441	3.6125	2.7860	4.6843	<2e-16
Estadio (IV)	2.33835	10.3641	8.4141	12.7661	<2e-16
Coortes(2007-2008)	-0.21894	0.8034	0.7166	0.9007	0.000175

Para verificar se o princípio da proporcionalidade entre taxas de mortalidade é satisfeito analisaremos os gráficos de resíduos de Schoenfeld para essas covariáveis e fatores na Figura 11.

Figura 11 Cox – Resíduos de Schoenfeld por grupos de idade

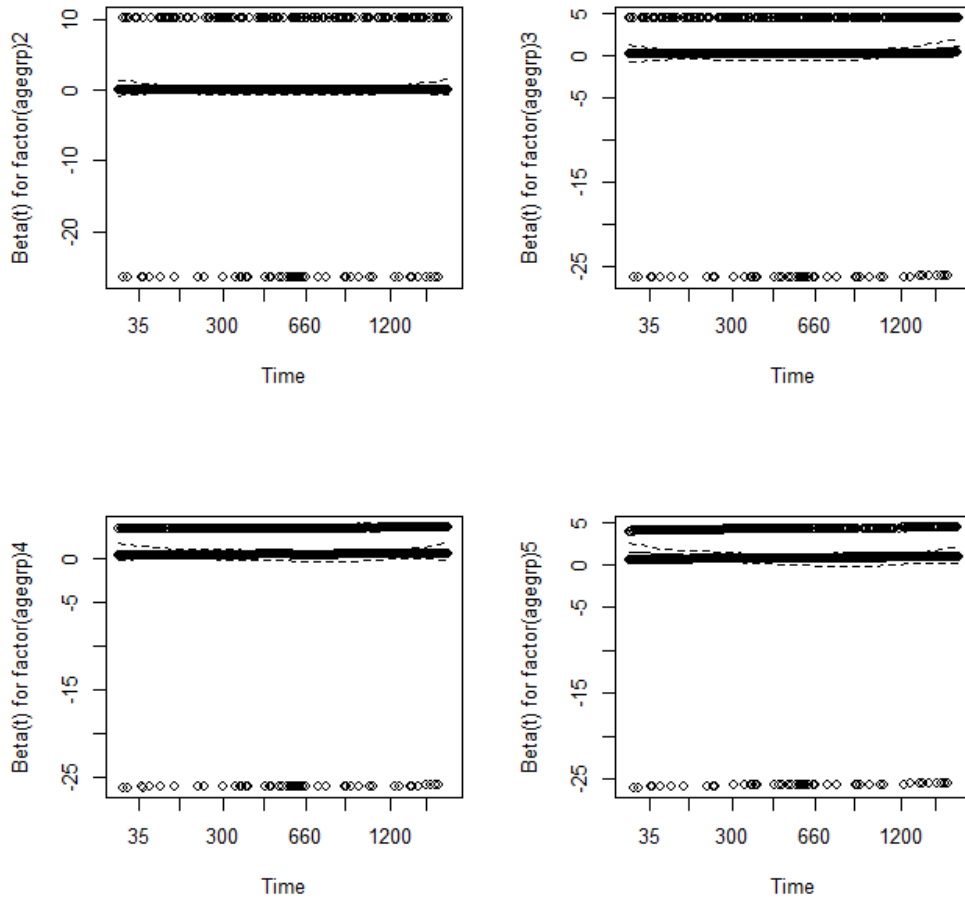


Figura 12 Cox – Resíduos de Schoenfeld por sexo

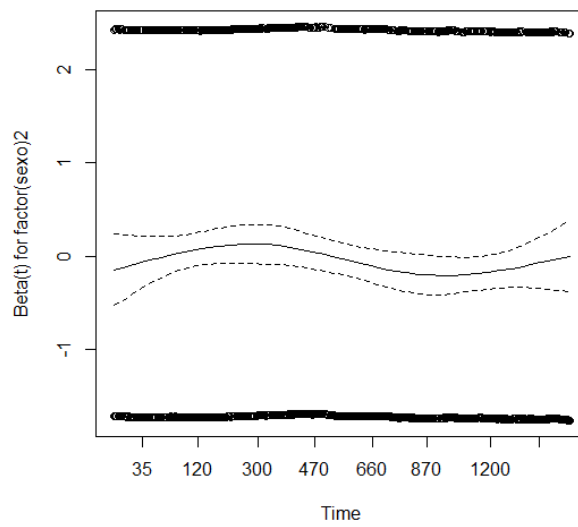


Figura 13 Cox – Residuos de Schoenfeld por grupo histológico

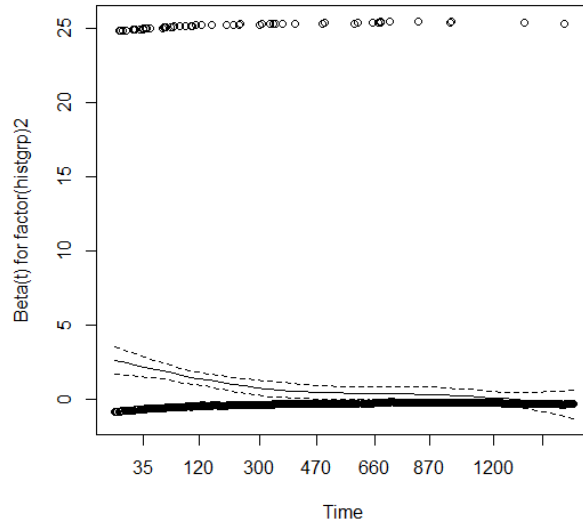


Figura 14 Cox – Residuos de Schoenfeld por grupo topográfico

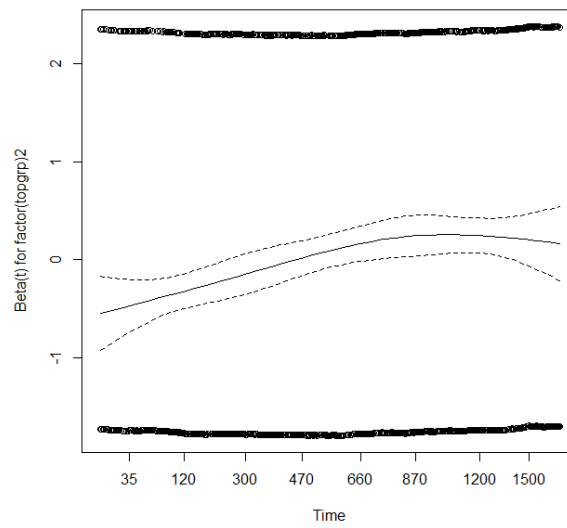


Figura 15 Cox – Residuos de Schoenfeld por estadio

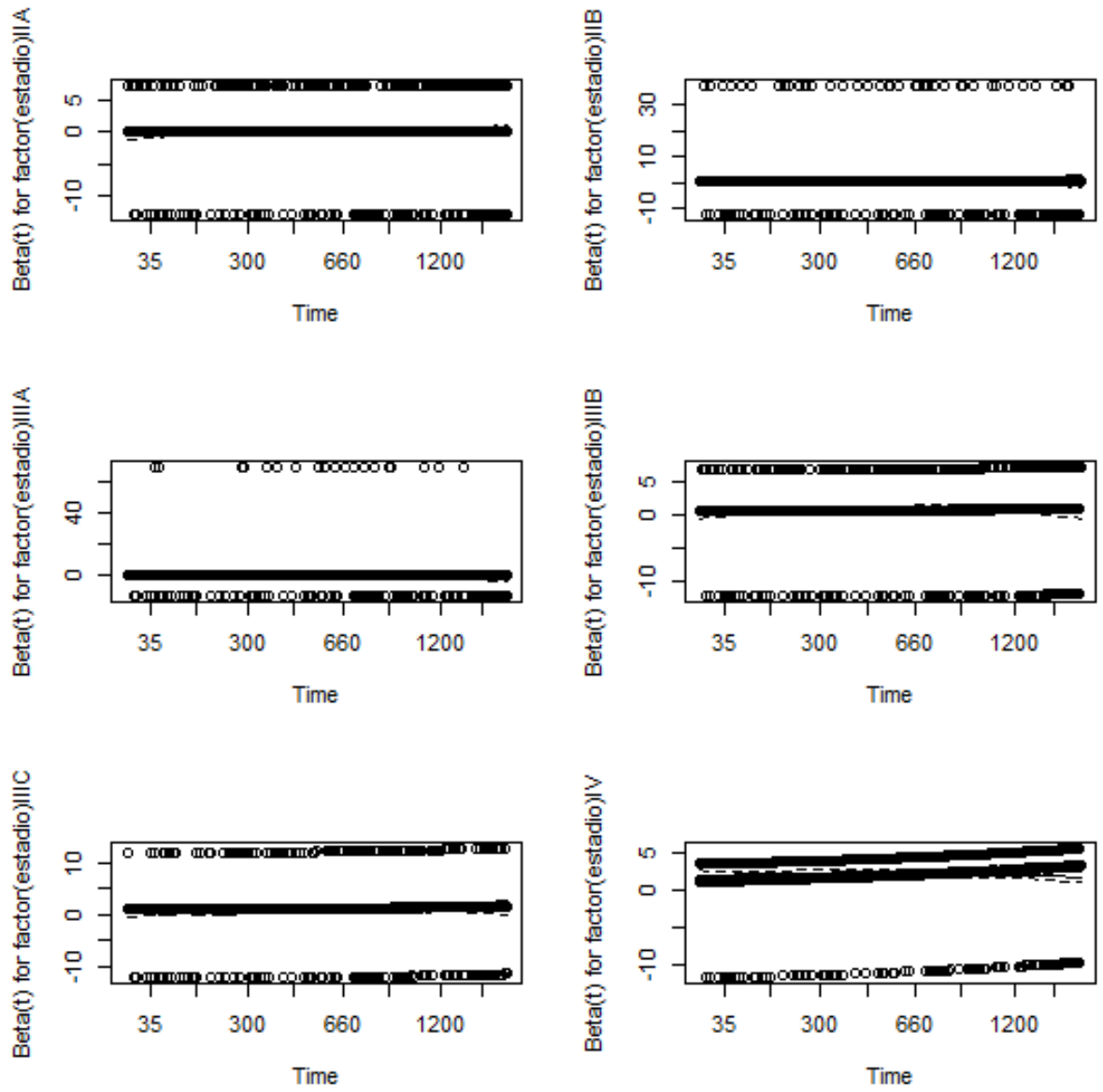
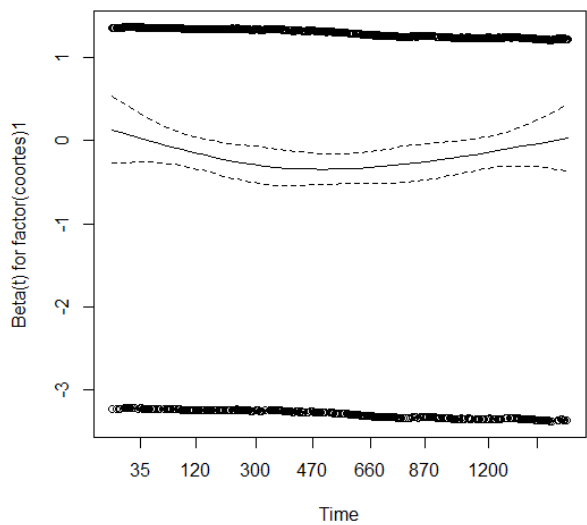


Figura 16 Cox – Resíduos de Schoenfeld por coortes



Pela análise dos gráficos, uma vez que as curvas adaptadas não se distanciam muito de uma reta horizontal, deduz-se que o modelo de Cox é adequado, pois aparenta haver proporcionalidade entre as taxas de mortalidade.

Podemos verificar, também, que não existe correlação linear, ou é muito fraca, entre os resíduos de Schoenfeld e o tempo.

Tabela 22 Cox – Teste de Correlação dos resíduos de Schoenfeld

Covariável	$\rho$	P
Idade (55-64)	0.0403	1.35e-1
Idade (65-74)	-0.0016	9.51e-1
Idade (75+)	-0.0209	4.38e-1
Sexo(feminino)	-0.0159	5.54e-1
Hstgrp(outras)	-0.0957	3.82e-4
Topgrp(reto)	0.0948	4.04e-4
Estadio (IIB)	-0.0368	1.74e-1
Estadio (IIIB)	-0.0041	8.80e-1
Estadio (IIIC)	0.0168	5.33e-1
Estadio (IV)	-0.0278	3.09e-1
Coortes(2007-2008)	-0.0024	9.28e-1

Tendo em conta a análise efetuada à proporcionalidade entre taxas de mortalidade e o teste de correlação dos resíduos efetuada, podemos aplicar o modelo de Cox.

Assim, dados os coeficientes calculados e apresentados na Tabela 21, cujos efeitos são ajustados entre todas as covariáveis e o cumprimento do pressuposto da proporcionalidade das taxas de mortalidade estamos em condições de afirmar que:

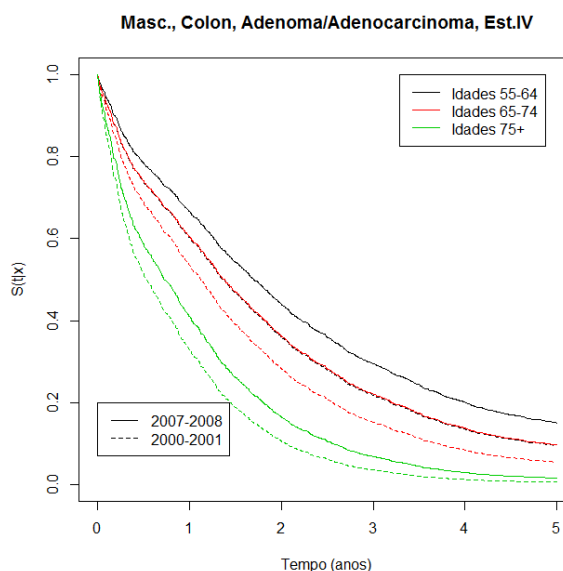
- Sendo os coeficientes das covariável relativa aos grupos de idade (55-64)  $\widehat{\beta}_1=0.37529$ , (65-74)  $\widehat{\beta}_2=0.58704$  e (75+)  $\widehat{\beta}_3=1.16131$ , podemos esperar que as taxas de mortalidade dos grupos de idade (55-64), (65,74) e (75+) aumentem, respetivamente 1.46 vezes (de 1.08 a 1.95 com um IC de 95%), 1.80 vezes (de 1.35 a 2.40 com um IC de 95%) e 3.19 vezes (de 2.39 a 4.28 com um IC de 95%) em relação ao grupo de idade (15-44). O coeficiente do grupo (45-54) não foi considerado significativo, pelo que este grupo foi agregado ao grupo de *baseline* (15-44).;
- Sendo o coeficiente da covariável sexo(feminino)  $\widehat{\beta}_4=-0.13080$ , podemos esperar que a taxa de mortalidade diminua cerca 0.88 vezes para o sexo feminino em relação ao sexo masculino (de 0.79 a a 0.98 vezes com um IC de 95%);
- Sendo o coeficiente da covariável histgrp(outras)  $\widehat{\beta}_5=0.42738$ , podemos esperar que a taxa de mortalidade de todas as outras histologias aumente 1.53 vezes em relação aos adenomas e adenocarcinomas (de 1.17 a a 2.01 vezes com um IC de 95%);
- Sendo o coeficiente da covariável topgrp(reto)  $\widehat{\beta}_6=0.15144$ , podemos esperar que a taxa de mortalidade de CCR localizados no reto (incluindo a junção retosigmoide) aumente 1.16 vezes em relação aos CCR localizados no cólon (de 1.04 a a 1.30 vezes com um IC de 95%);
- Sendo os coeficientes das covariável estadio(IIB)  $\widehat{\beta}_7=0.95688$ , estadio(IIIB)  $\widehat{\beta}_8=0.75738$ , estadio(IIIC)  $\widehat{\beta}_9=1.28441$  e estadio(IV)  $\widehat{\beta}_{10}=-2.33835$ , podemos esperar que as taxas de mortalidade dos estadios IIB, IIIB, IIIC e IV aumentem, respetivamente 2.60 vezes (de 1.80 a 3.77 com um IC de 95%), 2.13 (de 1.69 a 2.69 com um IC de 95%), 3.61 vezes (de 2.79 a 4.68 com um IC de 95%) e 10.36 vezes (de 8.41 a 12.77 com um IC de 95%), em

relação ao estadio I. Os coeficientes dos estadios IIA e IIIA não foram considerados significativos, pelo que foram agregados ao grupo de baseline (Estadio I);

- Sendo o coeficiente da covariável coortes (2007-2008)  $\widehat{\beta}_{11} = -0.21894$ , podemos esperar que a taxa de mortalidade diminua na segunda coorte (diagnóstico em 2007-2008) em relação à primeira (diagnóstico em 2000-2001) sendo a razão das taxas de mortalidade de 0.80 (de 0.72 a a 0.90 vezes com um intervalo de confiança de 95%).

Tendo em conta os coeficientes calculados, iremos representar graficamente a função de sobrevivência modelada por Cox, em função da variação das covariáveis selecionadas, fixando as restantes nos valores mais frequentes. Dado estarmos na presença de covariáveis categóricas foram utilizados os valores modais e não os valores medianos, para ambas as coortes.

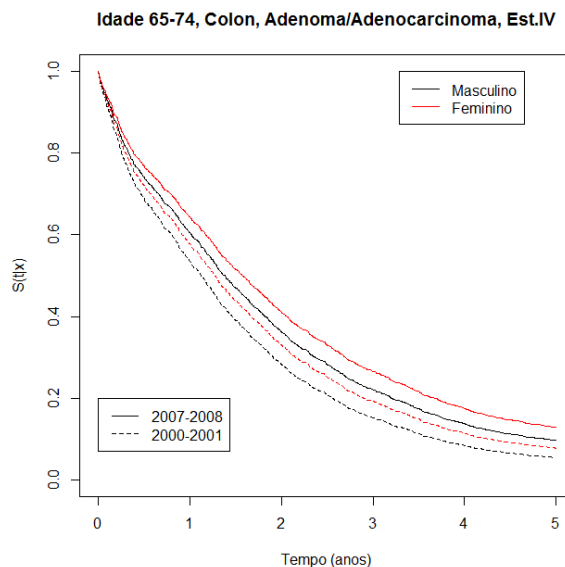
Figura 17 Cox – Sobrevivência em função da idade



Conseguimos verificar, para todas as idades consideradas, que a sobrevivência aparenta ser superior na coorte 2 relativamente à coorte 1 e que em ambas as coortes a mesma diminui, naturalmente, de forma proporcional ao aumento da idade do paciente aquando do diagnóstico.

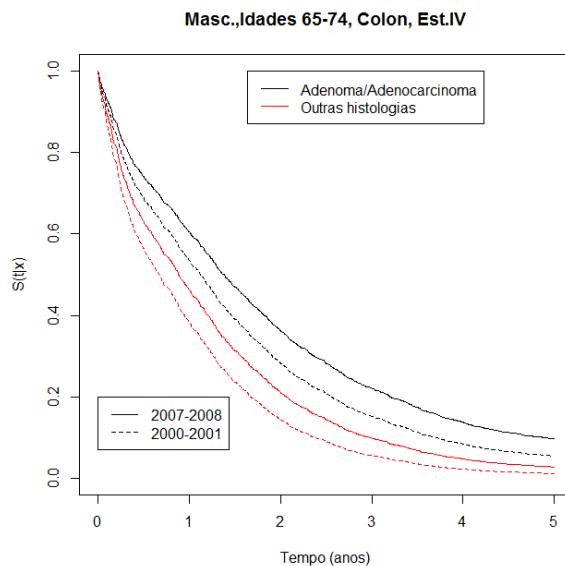


Figura 18 Cox – Sobrevivência em função do sexo



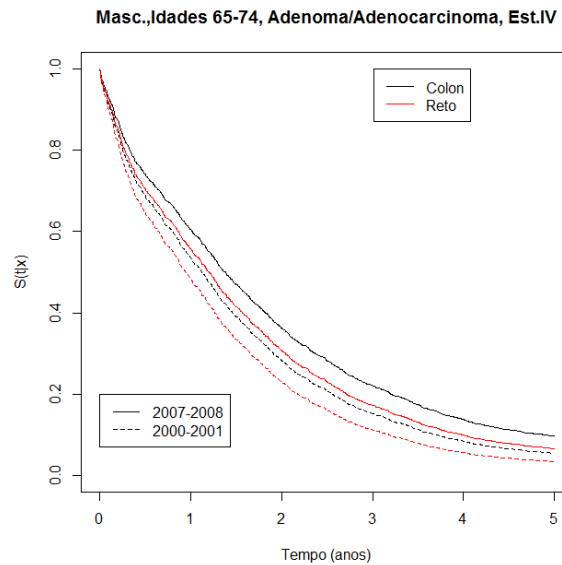
Conseguimos verificar que a sobrevivência aparenta ser superior na coorte 2 relativamente à coorte 1 e que em ambas as coortes a sobrevivência de pacientes do sexo feminino é superior à do sexo masculino.

Figura 19 Cox – Sobrevivência em função da histologia



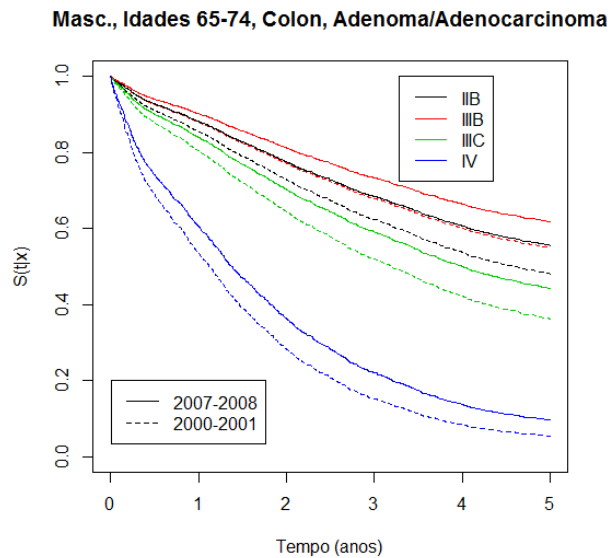
Conseguimos verificar que a sobrevivência aparenta ser superior na coorte 2 relativamente à coorte 1. Em ambas as coortes a sobrevivência de pacientes a cujo CCR tenha sido atribuído um código histológico do tipo Adenoma/Adenocarcinoma aparenta ser bastante superior à sobrevivência dos pacientes com outras histologias.

Figura 20 Cox – Sobrevivência em função da topografia



A sobrevivência aparenta ser superior na coorte 2 relativamente à coorte 1 e que em ambas as coortes a sobrevivência de pacientes cujo CCR tenha sido localizado no cólon parece ser superior relativamente à localização do mesmo no reto.

Figura 21 Cox – Sobrevivência em função do estadio



A sobrevivência aparenta ser superior na coorte 2 relativamente à coorte 1 e que em ambas as coortes a sobrevivência de pacientes cujo CCR é diagnosticado no estadio IV parece ser bastante inferior à dos outros estadios. Confirmando, também, as razões de taxa de mortalidade calculadas, verifica-se que a sobrevivência no estadio IIIB aparenta ser superior à sobrevivência no estadio IIB.

### 4.3. ASR – Estimação

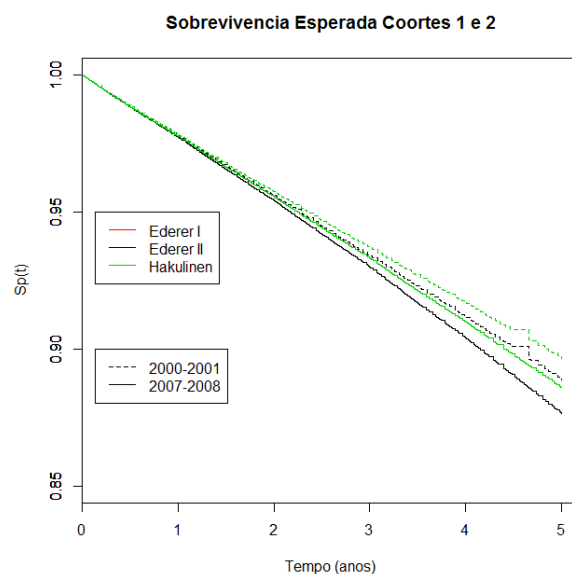
Os resultados abaixo apresentados têm em conta os estimadores para a sobrevivência esperada *Ederer I*, *Ederer II* e *Hakulinen* utilizados para o cálculo da taxa de sobrevivência relativa, bem como o estimador Pohar-Perme utilizado para o cálculo da *Net Survival*.

#### 4.3.1. Sobrevivência Relativa

##### Sobrevivência Esperada

A sobrevivência esperada para as 2 coortes, ao longo dos respectivos períodos de observação pode ser visualizada na seguinte figura, tendo em conta os estimadores *Ederer I*, *Ederer II* e *Hakulinen*:

Figura 22 Sobrevivência Esperada para as duas coortes



De notar que as curvas geradas pelos estimadores *Ederer I* e *Hakulinen* se sobrepõem, tendo o estimador *Ederer II* resultados inferiores a nível de sobrevivência esperada.

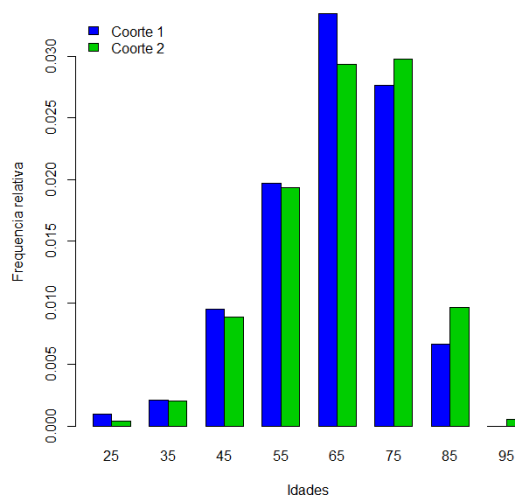
Estas curvas de sobrevivência têm como base as taxas de mortalidade esperadas para as coortes consideradas, tendo em conta as suas características demográficas, tais como o sexo e a idade.

Estas curvas são aproximadamente lineares e verifica-se que, apesar de existir tipicamente, uma diminuição das taxas de mortalidade ao longo dos anos considerados (2000 a 2013), a sobrevivência esperada para a coorte 2 é inferior à sobrevivência esperada para a coorte 1, qualquer que seja o estimador considerado.

Estas expectativas de menor sobrevivência para a coorte 2 que poderão, à partida não parecer fazer sentido (dada a diminuição da taxa de mortalidade ao longo dos anos), poderão ser facilmente explicadas pela diferença de características entre as duas coortes, nomeadamente, no que se refere à idade de diagnóstico.

Na coorte 2 cerca de 26.3% dos pacientes possuíam mais de 75 anos, enquanto que na coorte 1, essa percentagem era de apenas 17.9%. Na coorte 2 a média de idades é, também, mais elevada cerca de 1 ano e meio (de 64.54 para 66.05), e a mediana é mais alta 2 anos (de 66 para 68), relativamente à coorte 1. Esta diferença na distribuição das idades entre as 2 coortes pode ser facilmente visualizável no histograma abaixo, o que evidencia a existência de pacientes com idades mais elevadas na coorte 2, em termos de frequência relativa. De notar que na coorte 1 cerca de 56.8% dos pacientes são do sexo masculino e na coorte 2 essa percentagem é de cerca de 59.2%.

Figura 23 Histograma – distribuição das idades nas coortes



Assim, a sobrevivência esperada para os 2 coortes ao longo dos 5 anos de *follow up* é-nos dada pela seguinte tabela.

Tabela 23 ASR –Sobrevivência esperada da população para as 2 coortes aos 5 anos

Estimador	Coorte 1	Coorte 2	Global
Ederer I	0.897	0.886	0.889
Hakulinen	0.897	0.886	0.889
Ederer II	0.889	0.877	0.880

Conseguimos perceber, assim, a quase linearidade das curvas esperadas de sobrevivência sendo que a sobrevivência diminui por ano, para as coortes 1 e 2 respectivamente, aproximadamente 2.0% e 2.3% pelos estimadores *Ederer I* e *Hakulinen* e aproximadamente 2.2% e 2.5% pelo estimador *Ederer II*.

Verifica-se, também, que o estimador *Ederer II* apresenta valores inferiores de sobrevivência esperada da população de cerca de 0.8 e 0.9% em ambas as coortes, o que indicia uma taxa de sobrevivência relativa superior para esse estimador.

### Sobrevivência Relativa

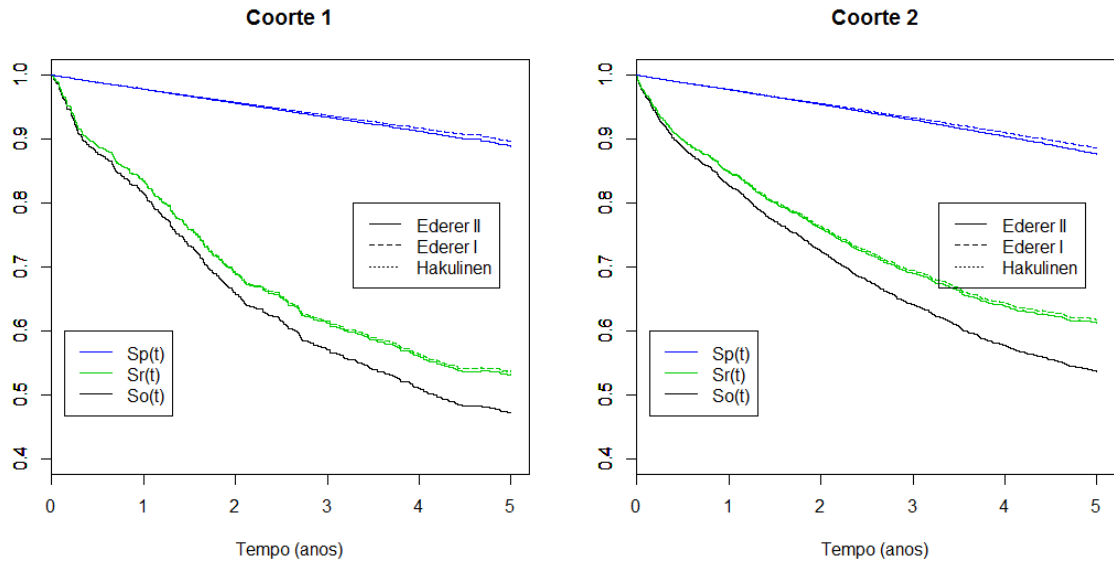
Tendo em conta que:

- A sobrevivência observada nas coortes em estudo  $S_o(t)$  é estimada por *Kaplan-Meier*
- A sobrevivência esperada para a população  $S_p(t)$  é estimada por *Ederer I*, *Hakulinen* ou *Ederer II*

A taxa de sobrevivência relativa é-nos dada pela equação (7).

Assim, abaixo visualizamos as curvas de sobrevivência relativa tendo em conta os estimadores para a sobrevivência esperada acima calculados, *Ederer I*, *Hakulinen* e *Ederer II* e o estimador de *Kaplan-Meier* para a sobrevivência observada.

Figura 24 Sobrevivência observada, esperada e relativa para as duas coortes



Verificamos, acima, que graficamente o resultado dos estimadores é muito semelhante em ambas as coortes, embora com valores ligeiramente diferentes para o estimador *Ederer II*, havendo uma sobreposição entre as curvas estimadas por *Ederer I* e *Hakulinen*.

Tabela 24 ASR – Sobrevivência Esperada, Observada e Relativa para as duas coortes – 5 anos

Coorte	$S_o(t)$	Ederer I / Hakulinen		Ederer II	
		$S_p(t)$	$S_r(t)$	$S_p(t)$	$S_r(t)$
1	0.472	0.897	0.536	0.889	0.531
2	0.536	0.886	0.617	0.877	0.612

Verificamos que a sobrevivência aos 5 anos é igual para as estimativas *Ederer I* e *Hakulinen*. Apesar das curvas de sobrevivência relativa serem muito semelhantes independentemente do método utilizado, adotamos o estimador *Ederer II*, de acordo com as razões avançadas no ponto 2.3.3. Verificamos, também, que para este conjunto de dados a estimativa *Ederer II* fornece os resultados para a sobrevivência a 5 anos mais aproximados da taxa de sobrevivência relativa. A estimativa para a sobrevivência do conjunto global das 2 coortes é de 0.593 pelos estimadores *Ederer I* e *Hakulinen*, e de 0.588 pelo estimador *Ederer II*.

Utilizando doravante, neste capítulo o estimador *Ederer II*, podemos verificar que a sobrevivência observada, pelo estimador de *Kaplan-Meier* já era superior na coorte 2 cerca de 6.4%, ao fim de 5 anos. Tendo em conta a mortalidade esperada para os 2 coortes nos respectivos períodos, a diferença ainda se torna mais evidente, alcançando os 8.1%.

Na coorte 1 a sobrevivência relativa é superior 5.9% relativamente à sobrevivência observada e na coorte 2 a diferença é de 7.6%.

Se estratificarmos os dados em função do sexo, obtemos as seguintes curvas de sobrevivência relativa:

Figura 25 Sobrevivência Relativa por Sexo

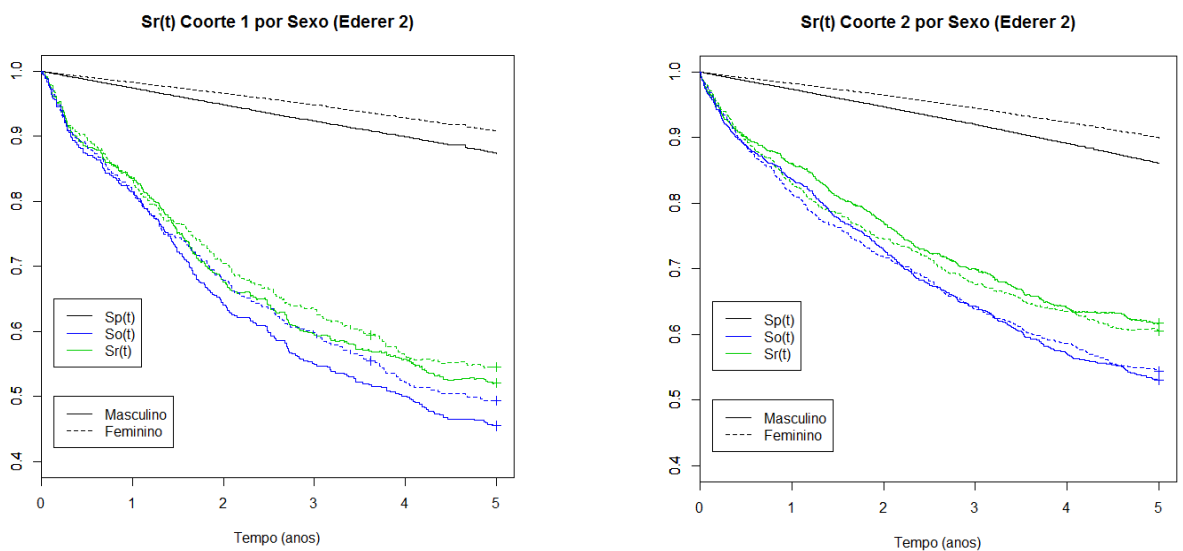


Tabela 25 ASR –Sobrevivência Relativa por sexo

Sexo	Coorte 1			Coorte 2		
	$S_p(t)$	$S_o(t)$	$S_r(t)$	$S_p(t)$	$S_o(t)$	$S_r(t)$
Masculino	0.874	0.455	0.520	0.861	0.531	0.617
Feminino	0.908	0.494	0.544	0.900	0.544	0.606

Em termos de sobrevivência observada, a mesma era superior para o sexo feminino em ambos as coortes, no entanto estas diferenças foram atenuadas na coorte 1 e na coorte 2 verifica-se, mesmo, uma ligeira melhoria da sobrevivência para o sexo masculino.

A estimativa para a sobrevivência do conjunto global das 2 coortes é de 0.589 para o sexo masculino e de 0.587 para o sexo feminino.

Estratificando os dados em função dos grupos de idade, conseguimos perceber que os valores de sobrevivência que decresciam com o aumento da idade, se atenuam, em termos de sobrevivência relativa deixando de haver essa tendência.

Tabela 26 ASR –Sobrevivência Relativa por grupo de idade

Idade	Coorte 1			Coorte 2		
	$S_p(t)$	$S_o(t)$	$S_r(t)$	$S_p(t)$	$S_o(t)$	$S_r(t)$
15-44	0.991	0.679	0.685	0.991	0.615	0.620
45-54	0.980	0.528	0.537	0.980	0.665	0.679
55-64	0.951	0.511	0.537	0.956	0.566	0.592
65-74	0.882	0.442	0.502	0.889	0.555	0.623
75+	0.693	0.364	0.521	0.692	0.410	0.592

A estimativa para a sobrevivência do conjunto global das 2 coortes é de 0.644 para o grupo (15-44), de 0.638 (45-54), 0.575 (55-64), 0.584 (65-74) e 0.577 (75+).

Figura 26 Sobrevivência Relativa por grupos de idade na coorte 1

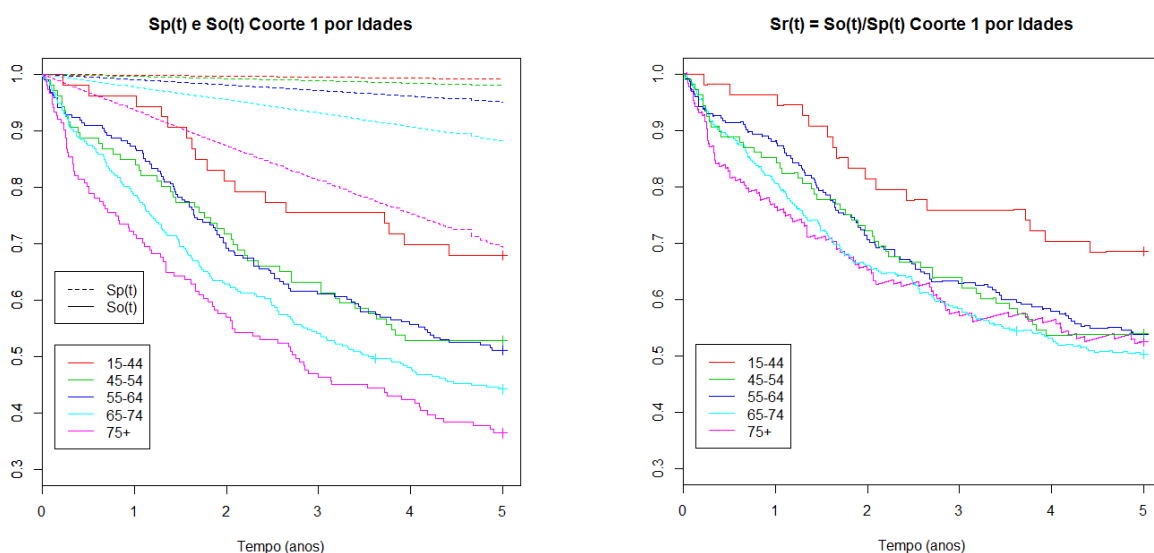
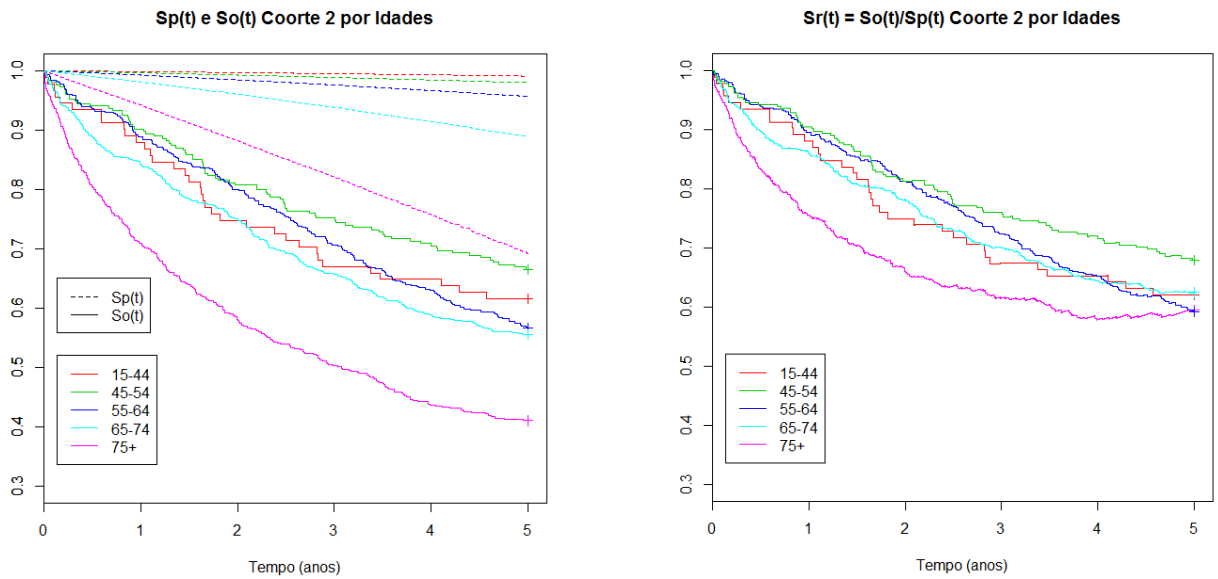




Figura 27 Sobrevivência Relativa por grupos de idade na coorte 2



Estratificando os dados por estadios, verificou-se um ligeiro reforço das diferenças entre estadios registadas na sobrevivência observada, dado que a sobrevivência esperada é tendencialmente crescente à medida que aumentam os estadios.

Tabela 27 ASR –Sobrevivência Relativa por estadio

Estadio	Coorte 1			Coorte 2		
	$S_p(t)$	$S_o(t)$	$S_r(t)$	$S_p(t)$	$S_o(t)$	$S_r(t)$
I	0.864	0.699	0.795	0.865	0.801	0.925
IIA	0.884	0.738	0.833	0.858	0.734	0.856
IIB	0.903	0.542	0.590	0.867	0.550	0.625
IIIA	0.887	0.633	0.681	0.876	0.830	0.910
IIIB	0.904	0.527	0.582	0.898	0.611	0.681
IIIC	0.932	0.310	0.333	0.897	0.449	0.500
IV	0.916	0.094	0.103	0.905	0.100	0.110

A estimativa para a sobrevivência do conjunto global das 2 coortes é de 0.884 para o estadio I, de 0.850 (IIA), 0.616 (IIB), 0.830 (IIIA), 0.657 (IIIB), 0.456 (IIIC) e 0.108 (IV).

Figura 28 Sobrevivência Relativa por estadio na coorte 1

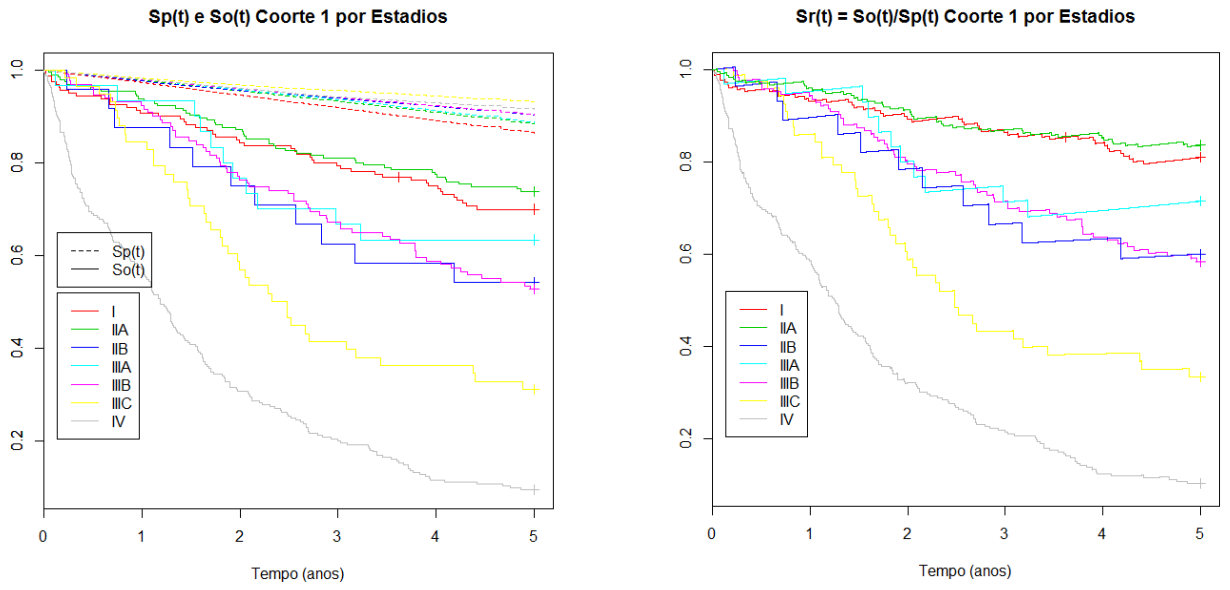
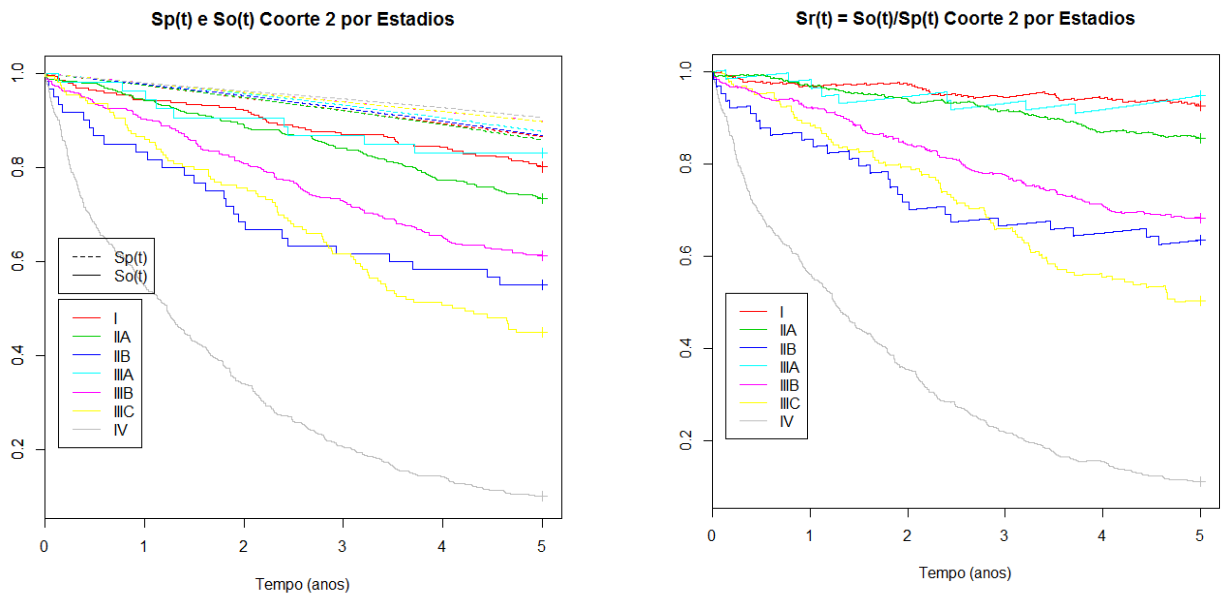


Figura 29 Sobrevivência Relativa por estadio na coorte 2



Estratificando os dados por distritos foram obtidos os seguintes resultados.

*Tabela 28 ASR –Sobrevivência Relativa por distrito*

Distrito	Coorte 1			Coorte 2		
	$S_p(t)$	$S_o(t)$	$S_r(t)$	$S_p(t)$	$S_o(t)$	$S_r(t)$
Braga	0.897	0.542	0.603	0.877	0.558	0.630
Bragança	0.914	0.571	0.611	0.855	0.525	0.605
Porto	0.882	0.445	0.505	0.881	0.542	0.616
V. Castelo	0.906	0.577	0.605	0.874	0.481	0.549
V. Real	0.884	0.292	0.329	0.854	0.508	0.591

A estimativa para a sobrevivência do conjunto global das 2 coortes é de 0.622 para o distrito de Braga, de 0.610 (Bragança), 0.584 (Porto), 0.562 (V. Castelo) e 0.520 (Vila Real).

Fazendo a mesma análise por grupo topográfico temos a tabela seguinte.

*Tabela 29 ASR –Sobrevivência Relativa por grupo topográfico*

Topgrp	Coorte 1			Coorte 2		
	$S_p(t)$	$S_o(t)$	$S_r(t)$	$S_p(t)$	$S_o(t)$	$S_r(t)$
Cólon	0.876	0.470	0.537	0.870	0.545	0.627
Reto	0.908	0.476	0.523	0.886	0.525	0.593

A estimativa para a sobrevivência do conjunto global das 2 coortes é de 0.599 para cólon e de 0.574 para o reto.

Da análise aos grupos histológicos verifica-se que após a aplicação da ASR, as diferenças entre os 2 grupos saem reforçadas devido ao facto da sobrevivência esperada ser inferior para os pacientes com adenomas/adenocarcinomas relativamente aos pacientes com outras histologias.

*Tabela 30 ASR –Sobrevivência Relativa por grupo histológico*

Histgrp	Coorte 1			Coorte 2		
	$S_p(t)$	$S_o(t)$	$S_r(t)$	$S_p(t)$	$S_o(t)$	$S_r(t)$
Aden\Adenoc	0.888	0.476	0.536	0.876	0.542	0.619
Outras hist.	0.927	0.346	0.371	0.889	0.350	0.370

A estimativa para a sobrevivência do conjunto global das 2 coortes é de 0.595 para adenomas e adenocarcinomas e de 0.384 para outras histologias.

### 4.3.2. Net Survival

Os resultados obtidos pelo estimador Pohar-Perme (*Net Survival*) são muito semelhantes aos obtidos pelo estimado *Ederer II*, anteriormente determinados, para a obtenção da taxa de sobrevivência relativa, tal como podemos verificar estratificando os dados por sexo, em ambas as coortes.

Figura 30 Comparação estimativas Ederer II vs Pohar-Perme por sexos

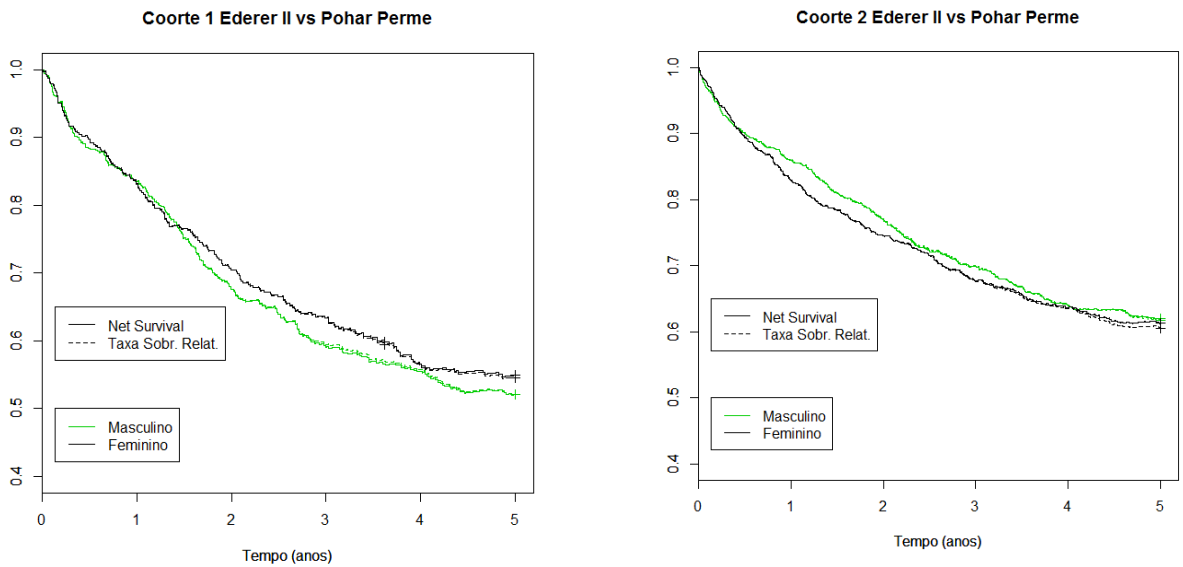


Tabela 31 ASR - Sobrevivência por sexo

Sexo	Observada Kaplan-Meier	Coorte 1		Coorte 2		
		Taxa Sobr. Relat. Ederer II	Net Survival Pohar-Perme	Observada Kaplan-Meier	Taxa Sobr. Relat. Ederer II	Net Survival Pohar-Perme
Masculino	0.455	0.520	0.520	0.531	0.617	0.619
Feminino	0.494	0.544	0.547	0.544	0.606	0.613

A estimativa para a sobrevivência do conjunto global das 2 coortes é de 0.591 para o sexo masculino e de 0.593 para o sexo feminino.

Pelo teste log-rank conseguimos perceber não existirem diferenças significativas para as curvas de sobrevivência relativa em função do sexo ( $p = 0.478$  e  $p = 0.676$  nas coorte 1 e 2, e  $p = 0.960$  no conjunto global).

Dado que graficamente do estimador de Pohar-Perme resultam curvas de sobrevivência muito semelhantes às já calculadas no capítulo anterior para o

estimador *Ederer II*, para as restantes covariáveis apresentamos apenas as tabelas com os valores de sobrevivência, a 5 anos.

Estes resultados de sobrevivência relativa já refletem, por inerência, o impacto de variáveis tais como idade, sexo e ano de calendário de ocorrência do evento de interesse ou da censura. No entanto e tal como já constatado em capítulos anteriores, a variável estadio apresenta bastante influência na sobrevivência, motivo pelo qual é feita a análise detalhada da sobrevivência apresentada na tabela acima tendo em conta os diferentes estadios. Pelo mesmo motivo esta análise será efetuada até ao final do presente subcapítulo, para as covariáveis relativas à idade, distrito, histologia e topografia, dotando o estudo de informação que permitirá atingir conclusões mais efetivas.

Tabela 32 ASR – Net Survival sexo por estadio

Sexo	Coorte	Net Survival (n) Sexo por Estadio						
		I	IIA	IIB	IIIA	IIIB	IIIC	IV
Masculino	1	0.747 (100)	0.855 (110)	0.611 (14)	0.527 (18)	0.556 (67)	0.181 (30)	0.106 (140)
	2	0.918 (188)	0.880 (326)	0.627 (28)	0.903 (24)	0.665 (248)	0.515 (81)	0.085 (296)
Feminino	1	0.866 (60)	0.788 (85)	0.525 (10)	0.885 (12)	0.595 (64)	0.481 (28)	0.092 (105)
	2	0.919 (103)	0.837 (222)	0.664 (32)	0.814 (29)	0.690 (156)	0.475 (75)	0.147 (204)

Tabela 33 ASR - Sobrevivência por grupo de idade

Idade	Coorte 1		Coorte 2		Global	
	Observada Kaplan-Meier	Net Survival Pohar-Perme	Observada Kaplan-Meier	Net Survival Pohar-Perme	Observada Kaplan-Meier	Net Survival Pohar-Perme
15-44	0.679	0.685	0.615	0.621	0.639	0.644
45-54	0.528	0.537	0.665	0.679	0.625	0.638
55-64	0.511	0.536	0.566	0.592	0.549	0.575
65-74	0.442	0.501	0.555	0.623	0.518	0.584
75+	0.364	0.527	0.410	0.599	0.400	0.583

Quanto à covariável idade, não existem diferenças significativas para as curvas de sobrevivência relativa ( $p = 0.053$  e  $p = 0.125$  nas coortes 1 e 2, respetivamente). O mesmo se verifica para o conjunto global com  $p = 0.109$ .

Tabela 34 ASR – Net Survival Idade por estadio

Net Survival (n) Idade por Estadio								
Idade	Coorte	I	IIA	IIB	IIIA	IIIB	IIIC	IV
15-44	1	- (8)	- (9)	- (2)	- (1)	0.690 (16)	- (3)	0.144 (14)
	2	- (10)	0.828 (23)	- (4)	- (4)	0.854 (20)	- (7)	0.131 (23)
45-54	1	0.851 (19)	0.865 (21)	- (2)	- (8)	0.579 (14)	- (7)	0.145 (35)
	2	0.934 (35)	0.889 (62)	- (8)	0.824 (11)	0.843 (58)	0.580 (21)	0.121 (59)
55-64	1	0.752 (39)	0.838 (57)	- (5)	- (6)	0.553 (30)	0.403 (18)	0.160 (66)
	2	0.903 (66)	0.781 (130)	0.702 (13)	- (13)	0.754 (108)	0.523 (42)	0.073 (128)
65-74	1	0.752 (59)	0.775 (70)	0.543 (10)	0.625 (12)	0.682 (43)	0.282 (23)	0.070 (95)
	2	0.949 (102)	0.854 (177)	0.573 (17)	- (12)	0.604 (124)	0.539 (49)	0.162 (157)
75+	1	0.807 (35)	0.865 (38)	- (5)	- (3)	0.365 (28)	- (7)	≈0 (35)
	2	0.863 (78)	0.935 (156)	0.549 (18)	0.957 (13)	0.514 (94)	0.361 (37)	0.073 (133)

Tabela 35 ASR - Sobrevivência por estadio

Estadio	Coorte 1		Coorte 2		Global	
	Observada Kaplan-Meier	Net Survival Pohar-Perme	Observada Kaplan-Meier	Net Survival Pohar-Perme	Observada Kaplan-Meier	Net Survival Pohar-Perme
I	0.699	0.793	0.801	0.924	0.765	0.884
IIA	0.738	0.835	0.734	0.864	0.735	0.857
IIB	0.542	0.584	0.550	0.651	0.548	0.633
IIIA	0.633	0.685	0.830	0.921	0.759	0.840
IIIB	0.527	0.576	0.611	0.677	0.591	0.652
IIIC	0.310	0.327	0.449	0.496	0.411	0.451
IV	0.094	0.100	0.100	0.111	0.098	0.107

Conseguimos, também, demonstrar pelo teste log-rank, que existem curvas com diferenças estatísticas significativas por estadio em ambas as coortes, com  $p < 0.001$ . Para o número total de pacientes verifica-se a mesma tendência, com  $p < 0.001$ , também.

Na coorte 1 as diferenças são estatisticamente significativas entre o estadio IIIC e os restantes e entre o estadio IV e os restantes, não o sendo entre os estadios (I, IIA, IIB, IIIA) e (IIB, IIIA, IIIB).

Na coorte 2 as diferenças são estatisticamente significativas entre o estadio IV e os restantes e entre os estadios IIIB e IIIC e os restantes excetuando o estadio IIB.

Tabela 36 ASR - Sobrevivência por distrito

Distrito	Coorte 1		Coorte 2		Global	
	Observada Kaplan-Meier	Net Survival Pohar-Perme	Observada Kaplan-Meier	Net Survival Pohar-Perme	Observada Kaplan-Meier	Net Survival Pohar-Perme
Braga	0.542	0.604	0.558	0.629	0.552	0.622
Bragança	0.571	0.607	0.525	0.609	0.538	0.611
Porto	0.445	0.508	0.542	0.621	0.514	0.589
V. Castelo	0.577	0.608	0.481	0.546	0.495	0.559
V. Real	0.292	0.317	0.508	0.603	0.449	0.526

O mesmo teste permitiu verificar não existirem diferenças para as curvas de sobrevivência relativa em função do distrito para a coorte 2 ( $p = 0.594$ ), embora na coorte 1 pareçam existir curvas com essas diferenças ( $p = 0.0146$ ), sendo possível, por exemplo, verificar que no distrito de Vila Real (31.7%), nesta coorte a sobrevivência relativa aos 5 anos é muito inferior à dos outros distritos. No conjunto global das duas coortes a diferença também não é significativa com  $p = 0.273$ . Verificou-se que na coorte 1 existem diferenças estatisticamente significativas entre o distrito de Vila Real e os restantes excetuando o Porto, e entre o distrito do Porto e o de Braga, não existindo essas diferenças entre os distritos de Braga, Bragança e Viana do Castelo.

Tabela 37 ASR – Net Survival distrito por estadio

Distrito	Coorte	Net Survival (n) Distrito por Estadio						
		I	IIA	IIB	IIIA	IIIB	IIIC	IV
Braga	1	0.828 (19)	0.849 (64)	- (8)	- (7)	0.731 (40)	0.381 (19)	0.116 (55)
	2	0.988 (44)	0.746 (98)	- (6)	- (10)	0.712 (71)	0.606 (32)	0.099 (69)
Bragança	1	- (4)	0.917 (14)	- (-)	- (1)	0.604 (14)	- (6)	0.205 (10)

	2	0.851 (22)	0.800 (37)	- (3)	- (3)	0.657 (16)	0.472 (12)	0.071 (29)
Porto	1	0.771 (129)	0.815 (106)	0.449 (14)	0.601 (22)	0.482 (62)	0.278 (27)	0.072 (148)
	2	0.917 (184)	0.895 (334)	0.637 (38)	0.944 (31)	0.683 (264)	0.516 (88)	0.124 (335)
V. Castelo	1	- (4)	- (3)	- (2)	- (-)	- (5)	- (4)	- (8)
	2	0.848 (23)	0.860 (46)	- (9)	- (4)	0.563 (32)	- (7)	≈0 (37)
Vila Real	1	- (4)	- (8)	0.525 (-)	0.885 (-)	0.325 (10)	- (2)	0.131 (24)
	2	0.871 (18)	0.856 (33)	- (4)	- (5)	0.595 (21)	0.369 (17)	0.161 (30)

Tabela 38 ASR - Sobrevivência por grupo topográfico

Topgrp	Coorte 1		Coorte 2		Global	
	Observada Kaplan-Meier	Net Survival Pohar-Perme	Observada Kaplan-Meier	Net Survival Pohar-Perme	Observada Kaplan-Meier	Net Survival Pohar-Perme
Cólon	0.470	0.541	0.545	0.634	0.522	0.605
Reto	0.476	0.518	0.525	0.595	0.511	0.574

Verificou-se também não existirem diferenças significativas para as curvas de sobrevivência relativa em função dos grupos topográficos sendo  $p = 0.965$  para a coorte 1,  $p = 0.308$  para a coorte 2 e  $p = 0.459$  para o conjunto global.

Tabela 39 ASR – Net Survival grupo topográfico por estadio

Topografia	Coorte	Net Survival (n) Topografia por Estadio						
		I	IIA	IIB	IIIA	IIIB	IIIC	IV
Cólon	1	0.806 (92)	0.863 (118)	0.618 (21)	0.748 (19)	0.580 (67)	0.403 (29)	0.106 (163)
	2	0.951 (182)	0.932 (307)	0.645 (40)	0.928 (20)	0.693 (174)	0.504 (101)	0.122 (313)
Reto	1	0.772 (68)	0.779 (77)	0.338 (3)	0.573 (11)	0.566 (64)	0.248 (29)	0.089 (82)
	2	0.875 (109)	0.775 (241)	0.650 (20)	0.853 (33)	0.659 (220)	0.480 (55)	0.092 (187)



Tabela 40 ASR - Sobrevivência por grupo histológico

Histgrp	Coorte 1		Coorte 2		Global	
	Observada Kaplan-Meier	Net Survival Pohar-Perme	Observada Kaplan-Meier	Net Survival Pohar-Perme	Observada Kaplan-Meier	Net Survival Pohar-Perme
Aden\Adenoc	0.476	0.537	0.542	0.623	0.523	0.598
Outras hist.	0.346	0.370	0.350	0.372	0.349	0.390

Não existem, também, diferenças significativas para as curvas de sobrevivência relativa em função dos grupos histológicos na coorte 1 sendo  $p = 0.096$ . Na coorte 2,  $p < 0.001$  e no conjunto global,  $p < 0.001$  as diferenças são significativas.

Tabela 41 ASR – Net Survival grupo histológico por estadio

Histologia	Coorte	Net Survival (n) Topografia por Estadio						
		I	IIA	IIB	IIIA	IIIB	IIIC	IV
Adenomas Adenocarcinomas	1	0.795 (154)	0.833 (193)	0.563 (23)	0.685 (30)	0.579 (127)	0.333 (57)	0.105 (233)
	2	0.916 (283)	0.862 (547)	0.655 (58)	0.936 (51)	0.677 (399)	0.481 (147)	0.117 (467)
Outras histologias	1	- (6)	- (2)	- (1)	- -	- (4)	- (1)	≈0 (12)
	2	- (8)	- (1)	- (2)	- (2)	- (5)	- (9)	0.031 (33)

A análise efetuada em termos de mortalidade permite, também, calcular qual a componente da taxa de mortalidade acumulada que é devida à causa específica (CCR) e qual a componente que é devida à mortalidade geral da população em questão.

Figura 31 Mortalidade acumulada – Causas específica e populacional na coorte 1

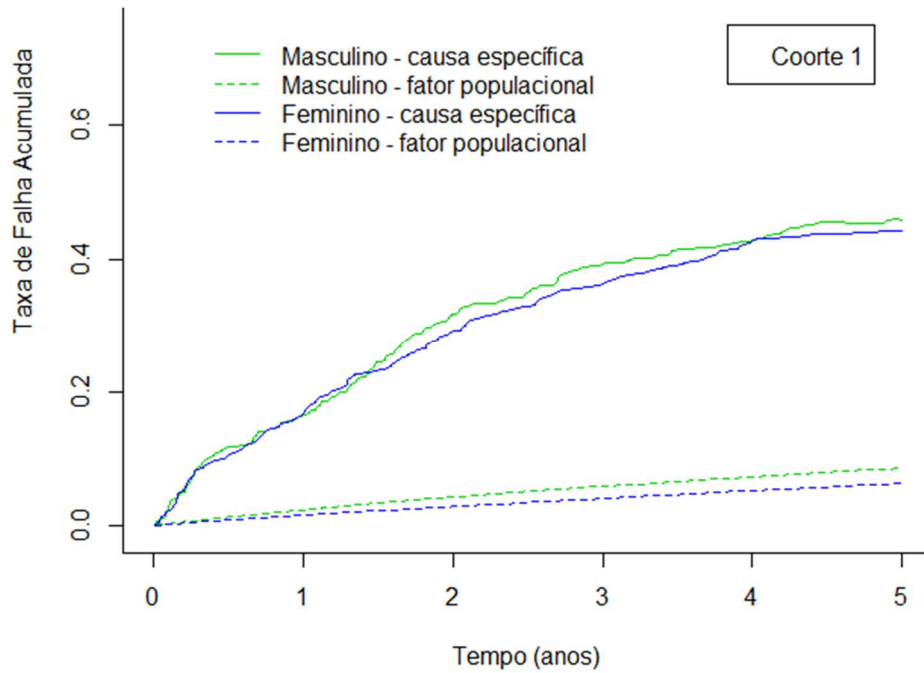


Figura 32 Mortalidade acumulada – Causas específica e populacional na coorte 2

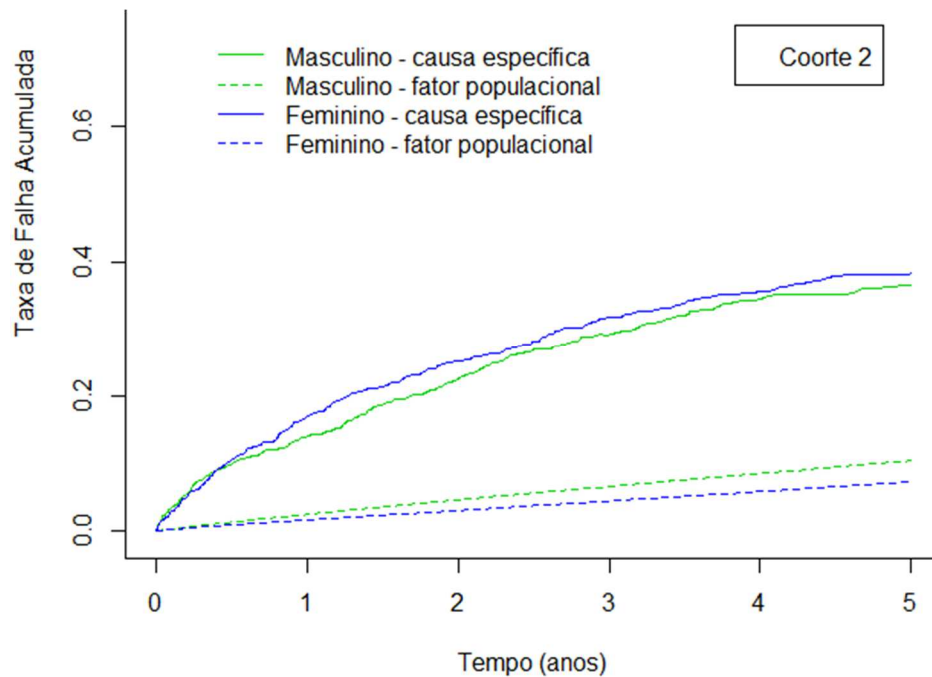


Tabela 42 ASR – Taxa de mortalidade acumulada – Causas específica e populacional

Sexo	Coorte	Causa	Ano 1	Ano 2	Ano 3	Ano 4	Ano 5
Masculino	1	Causa CCR	0.163	0.316	0.392	0.427	0.459
		Geral	0.023	0.042	0.059	0.072	0.086
		<b>Total</b>	<b>0.186</b>	<b>0.358</b>	<b>0.451</b>	<b>0.499</b>	<b>0.545</b>
	2	Causa CCR	0.139	0.224	0.291	0.344	0.366

		Geral	0.024	0.046	0.065	0.085	0.103
		<b>Total</b>	<b>0.163</b>	<b>0.270</b>	<b>0.356</b>	<b>0.429</b>	<b>0.469</b>
Feminino	1	Causa CCR	0.166	0.291	0.359	0.425	0.442
		Geral	0.015	0.028	0.040	0.052	0.062
		<b>Total</b>	<b>0.181</b>	<b>0.319</b>	<b>0.399</b>	<b>0.477</b>	<b>0.504</b>
	2	Causa CCR	0.170	0.251	0.317	0.356	0.382
		Geral	0.016	0.030	0.044	0.058	0.072
		<b>Total</b>	<b>0.186</b>	<b>0.281</b>	<b>0.361</b>	<b>0.414</b>	<b>0.454</b>

A tabela seguinte permite comparar a sobrevivência observada e a sobrevivência relativa recorrendo ao estimador *Net Survival* ao longo de todo o período de observação. Na tabela está representado, também, o número absoluto de mortes acumuladas no final de cada ano. De notar que o número de mortes acumuladas estimadas pelo estimador *Net Survival* foi arredondado à unidade. Aqui podemos verificar que do total de 2855 pacientes considerados no estudo, verificaram-se 1378 mortes ao final dos 5 anos, sendo que, considerando a sobrevivência relativa *Net Survival*, dessas 1378 mortes, se estimou que 1165 das mesmas se ficaram a dever ao CCR (84.5%), podendo ser atribuída as restantes 213 (15.5%) à mortalidade da população.

Tabela 43 ASR – Resultados globais sobrevivência e nº de mortes

		Sobrevivência e nº de mortes acumuladas											
		Sobrevivência		1		2		3		4		5	
Coorte 1	Observada	0.814	157	0.657	289	0.571	362	0.510	413	0.472	445		
	Relativa	0.833	141	0.688	263	0.610	329	0.559	372	0.532	395		
Coorte 2	Observada	0.828	346	0.725	553	0.641	722	0.577	851	0.536	933		
	Relativa	0.847	308	0.760	483	0.690	624	0.639	726	0.617	771		
Total	Observada	0.824	502	0.705	842	0.620	1085	0.557	1265	0.517	1378		
	Relativa	0.843	448	0.738	748	0.666	954	0.615	1099	0.592	1165		

Da tabela acima podemos constatar, em termos de mortalidade relativa, dessas 1165 mortes, que na coorte 1, mortalidade foi alta durante os dois primeiros anos de follow-up (35.7% e 30.9% respetivamente), enquanto que na coorte 2 a mortalidade foi alta, apenas, no primeiro ano (40%), baixando para 22.7% no segundo ano. Em ambas as coortes a mortalidade foi decrescendo ao longo dos 5 anos.

## 4.4. ASR – Modelação

### 4.4.1. Modelos Aditivos

Para a determinação destes modelos assumimos que durante o período de *follow-up* a mortalidade é constante em períodos anuais, sendo então considerados 5 intervalos (0-1), (1-2), (2-3), (3-4) e (4-5), aos quais será atribuído um coeficiente pelo modelo.

A esses intervalos de tempo, poderemos associar ou não mais covariáveis. Iremos considerar no modelo as covariáveis grupo de idade, sexo e coortes, as quais estão relacionadas com os dados das tábuas de mortalidade e tentar conseguir o melhor ajuste, considerando inicialmente estadio, distrito, grupo histológico e grupo topográfico.

De acordo com as considerações do capítulo 2.3.5, tendo sido considerado o modelo aditivo de *Estève* o mais adequado, vamos começar por considerá-lo.

Considerando, então, todas as covariáveis, para o modelo de *Estève* obtemos:

Tabela 44 ASR – Modelo Aditivo *Estève* com todas as covariáveis

	Coef	p
sexo feminino	-0.08334	0.203932
Idade 45-54	0.16469	0.325004
Idade 55-64	0.31362	0.042213
Idade 65-74	0.43992	0.003993
Idade 75+	0.88047	2.41e-08
Estadio IIA	0.44126	0.053512
Estadio IIB	1.51737	4.93e-08
Estadio IIIA	0.59876	0.093925
Estadio IIIB	1.27204	4.09e-09
Estadio IIIC	1.91773	<2e-16
Estadio IV	3.05844	<2e-16
Bragança	0.01466	0.925324
Porto	0.09908	0.261497
V. Castelo	0.27644	0.055692
V. Real	0.09276	0.510917
outras histologias	0.53997	0.000214

Reto	0.13264	0.044212
2007-2008	-0.22302	0.001291
<i>follow-up</i> [0,1)	-4.08325	<2e-16
<i>follow-up</i> [1,2)	-4.07928	<2e-16
<i>follow-up</i> [2,3)	-4.12977	<2e-16
<i>follow-up</i> [3,4)	-4.20517	<2e-16
<i>follow-up</i> [4,5)	-4.57388	<2e-16

Tal como no capítulo 4.2 foram consideradas como valores referência para a taxa base o sexo masculino, o grupo de idades 15-44, o estadio I, o distrito de Braga, as histologias referentes a adenomas e adenocarcinomas, a localização topográfica cólon e a coorte 2000-2001.

Tendo em conta o teste de ajuste das covariáveis ao modelo pela formação de uma BB, obtemos os seguintes resultados:

*Tabela 45 ASR – Brownian Bridge para Modelo Aditivo Estêve com todas as covariáveis*

	<b>max</b>	<b>p</b>
sexo feminino	1.250	8.78e-02
Idade 45-54	0.624	8.31e-01
Idade 55-64	2.491	8.12e-06
Idade 65-74	0.548	9.25e-01
Idade 75+	2.477	9.41e-06
Estadio IIA	0.745	6.36e-01
Estadio IIB	0.937	3.44e-01
Estadio IIIA	0.952	3.25e-01
Estadio IIIB	1.090	1.85e-01
Estadio IIIC	1.635	9.55e-03
Estadio IV	1.296	6.97e-02
Bragança	0.665	7.68e-01
Porto	1.289	7.20e-02
V. Castelo	0.607	8.55e-01
V. Real	0.779	5.79e-01
outras histologias	1.724	5.24e-03
Reto	2.417	1.69e-05
2007-2008	1.016	2.53e-01
GLOBAL	1.641	9.14e-03

Para obtermos um ajuste mais adequado o valor máximo absoluto da BB não deverá ultrapassar o valor 1.36 (95% da BB), sendo, no entanto, aceitável um valor

até 1.63 (99% da BB). Os valores mais próximos de zero são os mais adequados, pois indicam uma menor variabilidade das covariáveis ao longo do tempo.

Da Tabela 44 concluímos que a inclusão da covariável distrito não é significativa. Da Tabela 45 constatamos que a inclusão das covariáveis grupo histológico (valor máximo da BB: 1.724) e grupo topográfico (valor máximo da BB: 2.417) prejudicam o ajuste do modelo, o qual apresenta um valor máximo da BB, global, de 1.641. Apesar de nos grupos de idade termos os grupos 55-64 e 75+ com valores não ajustados, dada a importância da covariável no contexto, decidimos mantê-la no modelo.

Excluimos, assim, do modelo as covariáveis distrito, grupo histológico e grupo topográfico, de forma a conseguirmos um modelo globalmente mais ajustado.

Tabela 46 ASR – Modelo Aditivo Estêve ajustado

	Coef (SE)	Exp(coef)	p
sexo feminino	-0.06957 (0.0646)	0.9328	0.28145
Idade 45-54	0.15426 (0.1668)	1.1668	0.35513
Idade 55-64	0.27521 (0.1535)	1.3168	0.07300
Idade 65-74	0.39229 (0.1515)	1.4804	0.00963
Idade 75+	0.84512 (0.1569)	2.3283	7.22e-08
Estadio IIA	0.41531 (0.2288)	1.5148	0.06949
Estadio IIB	1.50963 (0.2782)	4.5251	5.72e-08
Estadio IIIA	0.58787 (0.3588)	1.8002	0.10134
Estadio IIIB	1.27045 (0.2164)	3.5625	4.32e-09
Estadio IIIC	1.90581 (0.2229)	6.7249	<2e-16
Estadio IV	3.05193 (0.2030)	21.1561	<2e-16
2007-2008	-0.20595 (0.0683)	0.8139	0.00257
follow-up [0,1)	-3.89092 (0.2521)	0.0204	<2e-16
follow-up [1,2)	-3.89122 (0.2513)	0.0204	<2e-16
follow-up [2,3)	-3.94686 (0.2552)	0.0193	<2e-16
follow-up [3,4)	-4.02529 (0.2575)	0.0179	<2e-16
follow-up [4,5)	-4.39449 (0.2770)	0.0123	<2e-16

Da tabela acima podemos constatar que a influência da covariável sexo no modelo não é significativa a 5%, bem como dos grupos de idade 45-54 e 55-64. O mesmo se verifica para os estádios IIA e IIIA.

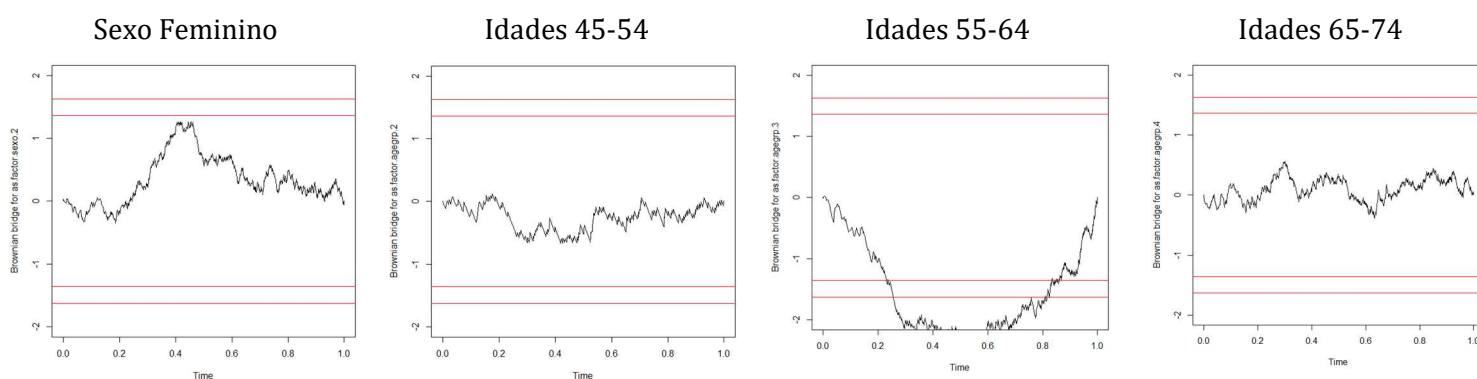
Podemos constatar também que a a mortalidade decresce da coorte 1 para a coorte 2 e que é superior nos primeiros anos de *follow-up*.

Tabela 47 ASR – Brownian Bridge para Modelo Aditivo Estêve ajustado

	Max	p
sexo feminino	1.269	8.00e-02
Idade 45-54	0.666	7.66e-01
Idade 55-64	2.490	8.22e-06
Idade 65-74	0.565	9.07e-01
Idade 75+	2.489	8.35e-06
Estadio IIA	0.759	6.13e-01
Estadio IIB	0.922	3.63e-01
Estadio IIIA	0.951	3.26e-01
Estadio IIIB	1.103	1.75e-01
Estadio IIIC	1.633	9.64e-03
Estadio IV	1.321	6.10e-02
2007-2008	1.058	2.13e-01
GLOBAL	1.591	1.27e-03

O valor máximo da BB ficou agora ligeiramente melhor ajustado (de 1.641 do modelo inicial para 1.591). A visualização gráfica da BB representada na tabela acima para este modelo é apresentada nas seguintes figuras.

Figura 33 Modelo Aditivo Estêve – Brownian Bridge para as covariáveis



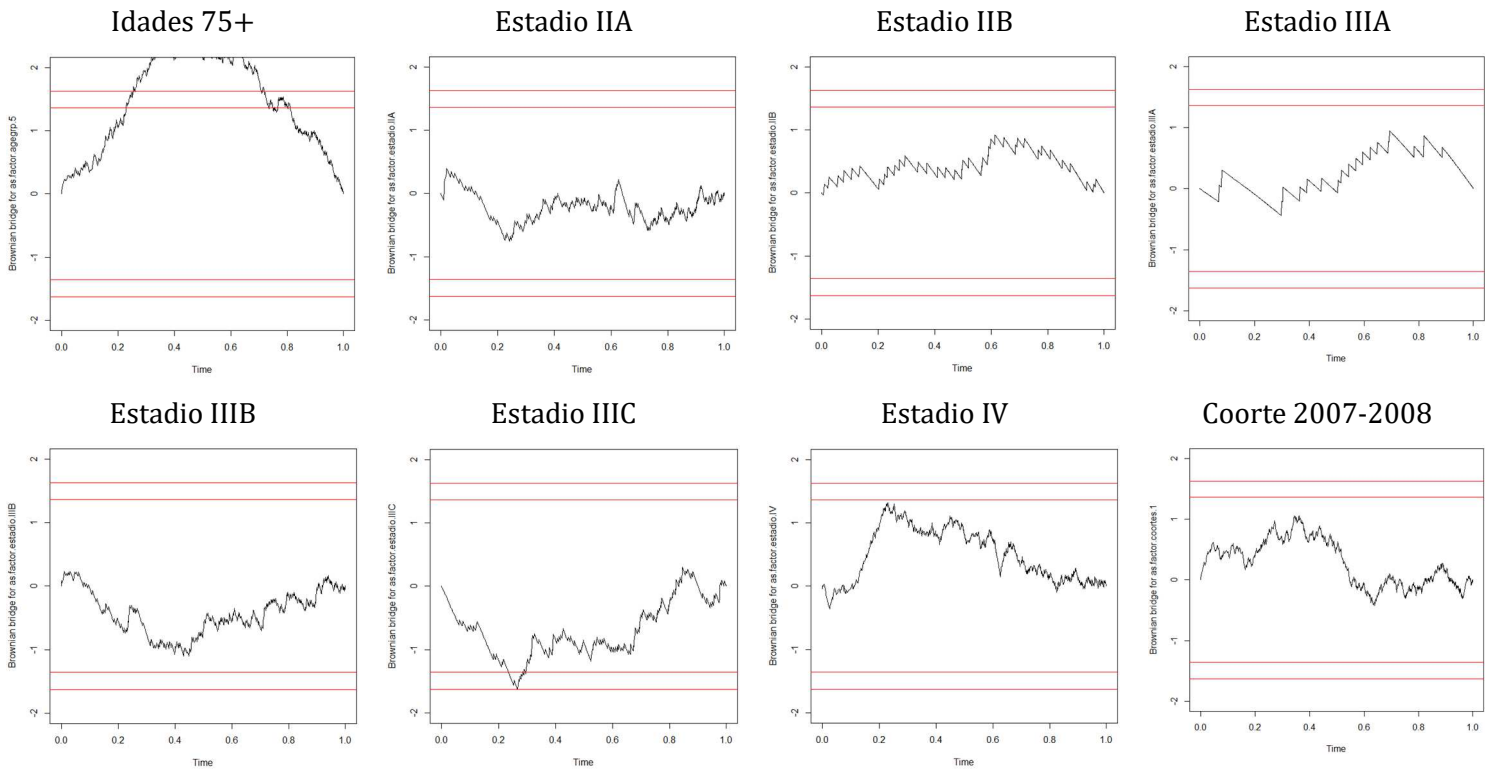
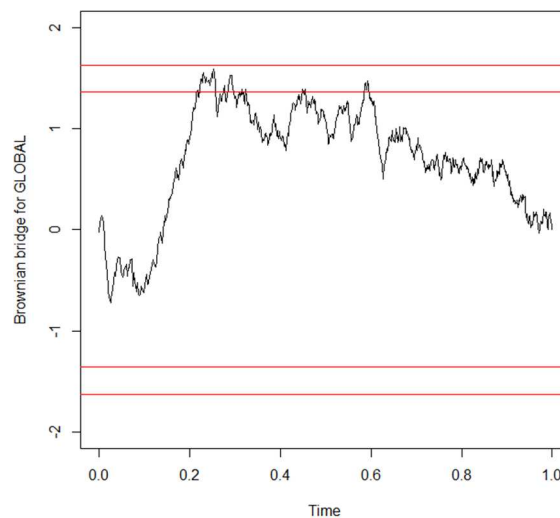


Figura 34 Modelo Aditivo Estève – Brownian Bridge para o modelo global



Com a inclusão do mesmo conjunto de covariáveis passaremos a determinar os modelos aditivos de *Hakulinen-Tenkanen* e de *Poisson*.

Para ambos os modelos incluindo todas as covariáveis o número observado de mortes é inferior ao esperado em 2429 dos 3270 (74.3%) dos  $k$  estratos referidos no capítulo 3.4.2, o que se compreende dado o número de estratos calculados ter dimensão superior ao número total de pacientes, 2855. Utilizando apenas as covariáveis sexo, grupo de idade, estadio e coorte, tal como fizemos para o modelo



de *Estève*, a estratificação é menor (665), sendo que o número de mortes é inferior ao esperado em 306 dos mesmos (46%). Este facto indicia, também para estes modelos um melhor ajuste retirando as covariáveis distrito, grupo histológico e grupo topográfico.

Tabela 48 ASR – Modelos Aditivos – comparação entre coeficientes e seus erros padrão

	Coeficientes (SE)		
	Estève	Poisson	Hakulinen
sexo feminino	-0.07 (0.06)	-0.07(0.06)	-0.07(0.07)
Idade 45-54	0.15 (0.17)	0.16(0.17)	0.14(0.17)
Idade 55-64	0.28 (0.15)	0.28(0.15)	0.27(0.15)
Idade 65-74	0.39(0.15)	0.40(0.15)	0.38(0.15)
Idade 75+	0.85 (0.16)	0.86(0.16)	0.83(0.16)
Estadio IIA	0.42 (0.23)	0.36(0.23)	0.37(0.23)
Estadio IIB	1.51 (0.28)	1.44(0.28)	1.45(0.28)
Estadio IIIA	0.59(0.36)	0.52(0.37)	0.52(0.38)
Estadio IIIB	1.27 (0.22)	1.25(0.21)	1.25(0.21)
Estadio IIIC	1.91 (0.22)	1.87(0.22)	1.88(0.22)
Estadio IV	3.05 (0.20)	3.00(0.20)	2.99(0.20)
2007-2008	-0.21 (0.07)	-0.21(0.07)	-0.22(0.07)
<i>follow-up</i> [0,1)	-3.89 (0.25)	-3.84(0.25)	-3.86(0.25)
<i>follow-up</i> [1,2)	-3.89 (0.25)	-3.85(0.25)	-3.82(0.25)
<i>follow-up</i> [2,3)	-3.95 (0.26)	-3.90(0.25)	-3.87(0.25)
<i>follow-up</i> [3,4)	-4.03 (0.26)	-3.99(0.25)	-3.97(0.26)
<i>follow-up</i> [4,5)	-4.39 (0.28)	-4.37(0.27)	-4.35(0.27)

Nos modelos de *Poisson* e *Hakulinen-Tenkanen*, tal como no modelo de *Estève*, a influência da covariável sexo não é significativa a 5%, bem como dos grupos de idade 45-54 e 55-64, e estadios IIA e IIIA.

O valor máximo da BB do modelo de *Estève* (1.591), indica que este é ligeiramente mais ajustado que o de *Poisson* (1.600) e o de *Hakulinen-Tenkanen* (1.632).

Outra forma de verificar o comportamento das covariáveis no tempo para os modelos de regressão de sobrevivência relativa poderá ser efetuada, tal como para o modelo de *Cox*, analisando os gráficos de resíduos de Schoenfeld para as covariáveis incluídas no modelo, conforme ilustrado nas próximas figuras.

Figura 35 Modelo Aditivo Estève – Resíduos de Schoenfeld por sexo

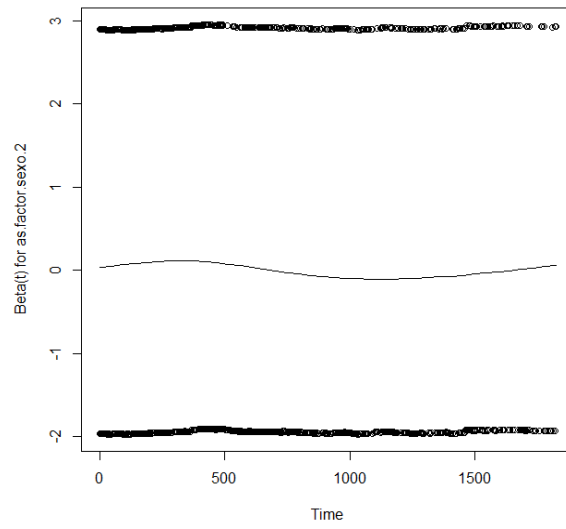


Figura 36 Modelo Aditivo Estève – Resíduos de Schoenfeld por grupo de idade

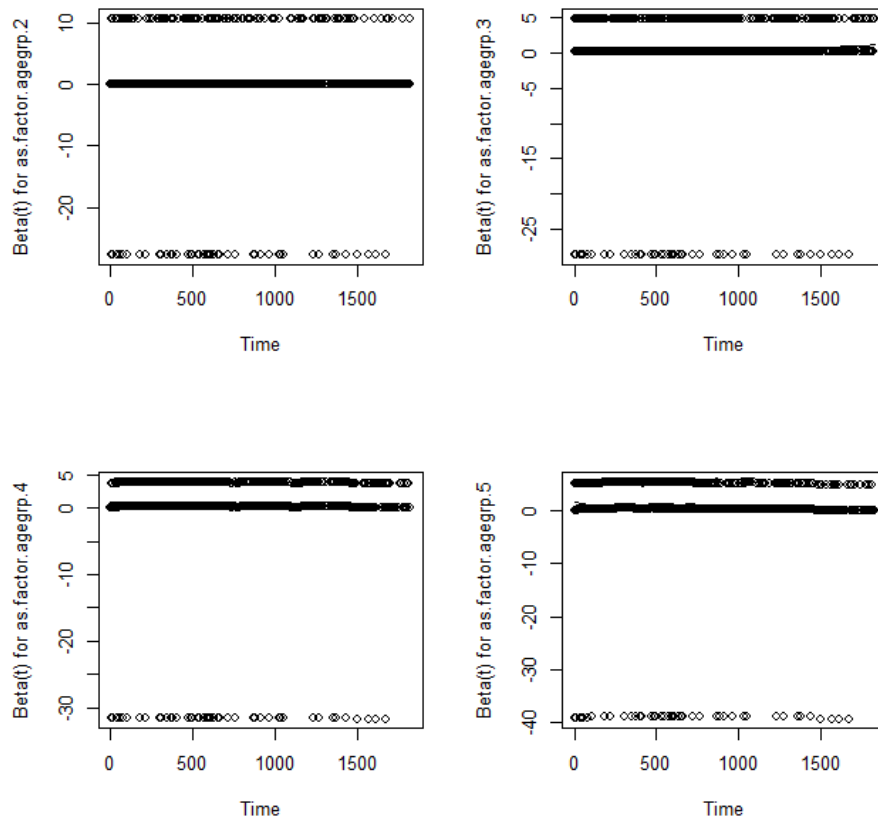


Figura 37 Modelo Aditivo Estève – Resíduos de Schoenfeld por estadio

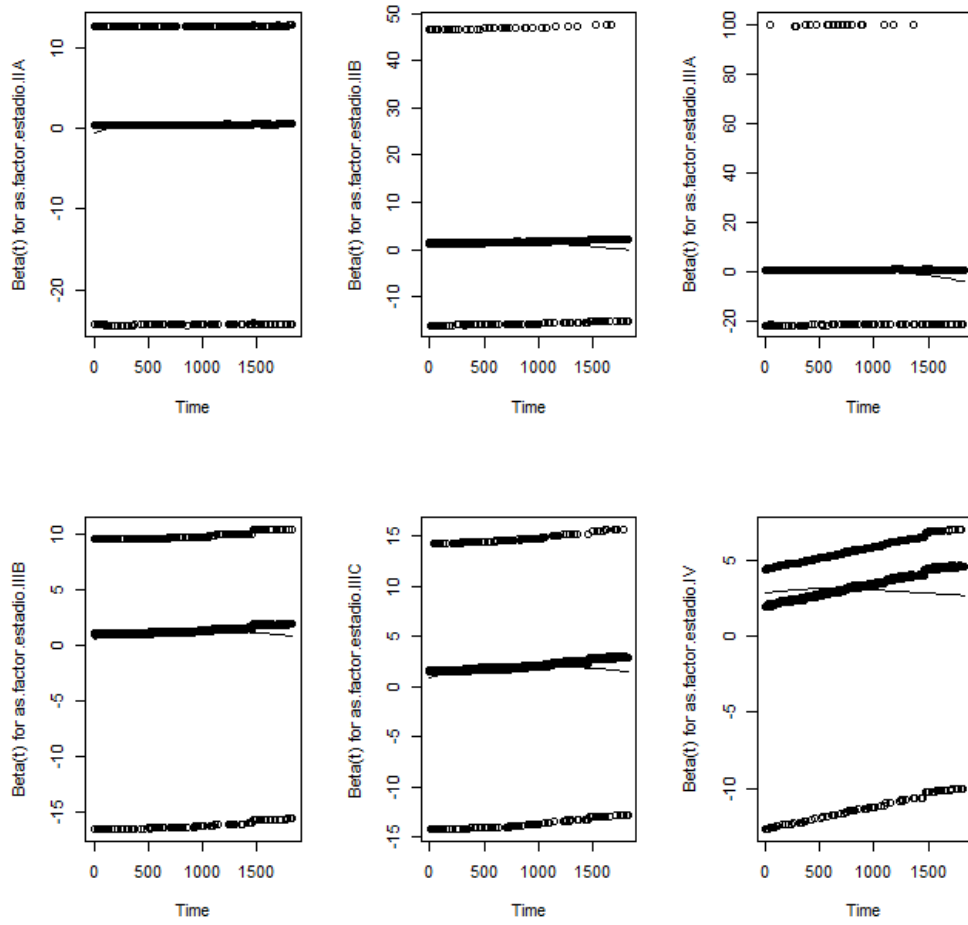
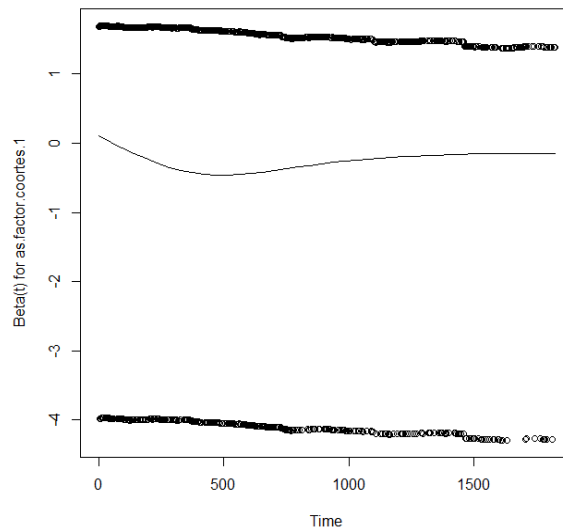


Figura 38 Modelo Aditivo Estève – Resíduos de Schoenfeld por coortes



Pelos gráficos, uma vez que as curvas adaptadas não se distanciam muito de uma reta horizontal, podemos esperar que o modelo de aditivo de *Estève* seja adequado, pois aparenta haver proporcionalidade entre as taxas de mortalidade ao longo do tempo entre os valores representados das covariáveis e os respectivos valores de referência.

Os modelos aditivos pressupõem a existência da taxa de mortalidade em excesso, a qual é representada no gráfico abaixo, para ambas as coortes.

Figura 39 Mortalidade em Excesso para ambas as coortes

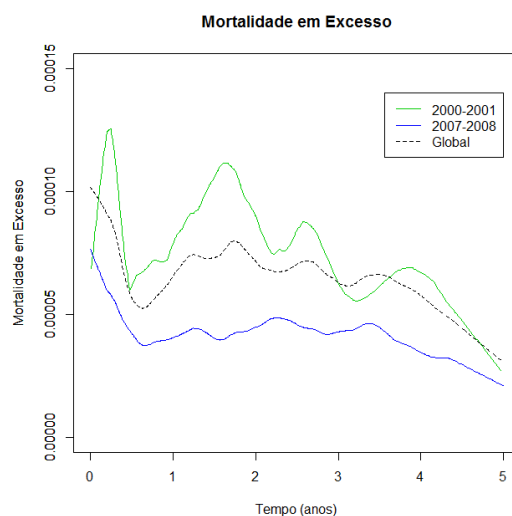
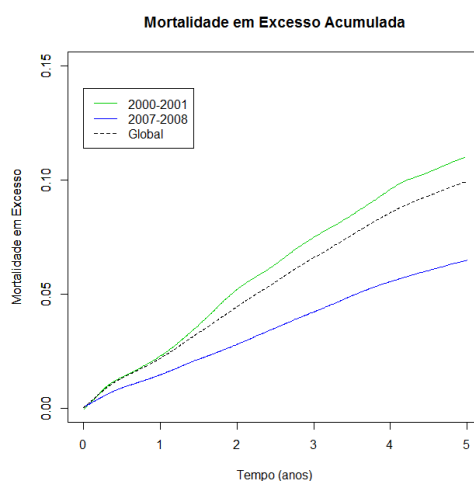


Figura 40 Mortalidade em Excesso acumulada para ambas as coortes



Da análise das 2 figuras acima confirmamos uma maior mortalidade em excesso na coorte 1, com uma elevada mortalidade no segundo ano de *follow-up*.

## 4.4.2. Modelo Multiplicativo

Vamos considerar todas as covariáveis, para o modelo multiplicativo de *Andersen*, conforme a seguinte tabela.

Tabela 49 ASR – Modelo Multiplicativo de Andersen com todas as covariáveis

	<b>Coef</b>	<b>p</b>
sexo feminino	0.5116	< 2e-16
Idade 45-54	-0.6765	3.61e-05
Idade 55-64	-1.2531	< 2e-16
Idade 65-74	-2.0220	< 2e-16
Idade 75+	-2.5879	< 2e-16
Estadio IIA	0.1513	0.211461
Estadio IIB	0.8563	6.47e-06
Estadio IIIA	0.1244	0.610083
Estadio IIIB	0.8272	4.17e-12
Estadio IIIC	1.3627	< 2e-16
Estadio IV	2.3774	< 2e-16
Bragança	0.0296	0.8203
Porto	0.1077	0.1439
V. Castelo	0.2704	0.0286
V. Real	-0.0055	0.9634
outras histologias	0.5354	0.0001
Reto	0.1834	0.0010
2007-2008	-0.1065	0.0729

Na tabela acima foram consideradas como valores referência o sexo masculino, o grupo de idades 15-44, o estadio I, o distrito de Braga, as histologias referentes a adenomas e adenocarcinomas, a localização topográfica cólon e a coorte 2000-2001, tal como nos modelos aditivos.

De forma a determinarmos um modelo comparável com os aditivos, excluimos do mesmo as covariáveis distrito, grupo histológico e grupo topográfico.

Tabela 50 ASR – Modelo Multiplicativo de Andersen ajustado

	Coef (SE)	Exp(coef)	p
sexo feminino	0.51641 (0.0559)	1.6760	<2e-16
Idade 45-54	-0.67519 (0.1635)	0.5091	3.63e-05
Idade 55-64	-1.28001 (0.1497)	0.2780	<2e-16
Idade 65-74	-2.05659 (0.1279)	0.1279	<2e-16
Idade 75+	-2.63199 (0.1488)	0.0719	<2e-16
Estadio IIA	0.12806 (0.1208)	1.1366	0.289
Estadio IIB	0.83600 (0.1895)	2.3071	<2e-16
Estadio IIIA	0.11470 (0.2439)	1.1215	0.638
Estadio IIIB	0.82977 (0.1189)	2.2928	<2e-16
Estadio IIIC	1.33922 (0.1330)	3.8161	<2e-16
Estadio IV	2.35495 (0.1065)	10.5376	<2e-16
2007-2008	-0.07937 (0.0587)	0.9237	0.176

Da tabela acima podemos constatar que a influência da covariável sexo é significativa, ao contrário do que acontecia nos modelos aditivos. Aqui todos os grupos de idade são significativos. Os estadios IIA e IIIA não são, tal como nos modelos aditivos significativos a 5%, e neste modelo em específico também não o é a covariável coortes.

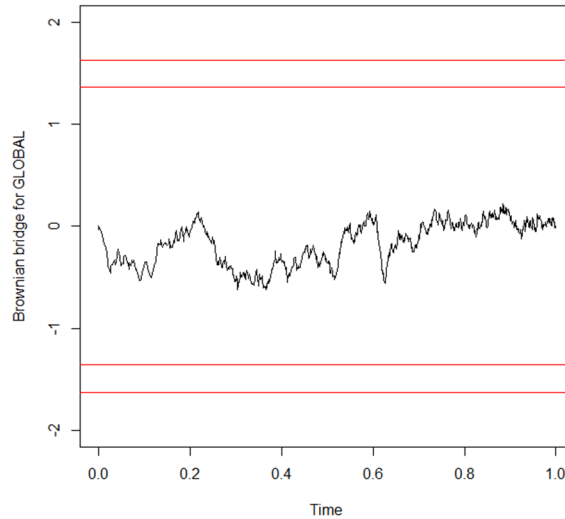
Podemos constatar também que a mortalidade decresce da coorte 1 para a coorte 2.

Tabela 51 ASR – Brownian Bridge para Modelo Multiplicativo Anderson ajustado

	Max	p
sexo feminino	1.303	0.0672
Idade 45-54	0.472	0.9791
Idade 55-64	2.174	0.0002
Idade 65-74	0.780	0.5764
Idade 75+	1.847	0.0022
Estadio IIA	0.869	0.4372
Estadio IIB	0.997	0.2729
Estadio IIIA	0.944	0.3353
Estadio IIIB	0.944	0.3344
Estadio IIIC	1.582	0.0134
Estadio IV	1.270	0.0793
2007-2008	1.246	0.0894
GLOBAL	0.622	0.8341

O valor máximo da BB para o modelo global é muito mais adequado que o dos modelos aditivos, sendo de 0.622. A visualização gráfica da BB global representada na tabela acima para este modelo é apresentada na seguinte figura:

Figura 41 Modelo Multiplicativo de Andersen – Brownian Bridge para o modelo global



Outra forma de verificar o comportamento das covariáveis no tempo para os modelos de regressão de sobrevivência relativa poderá ser efetuada, tal como para o modelo aditivo, analisando os gráficos de resíduos de Schoenfeld para as covariáveis incluídas no modelo.

Figura 42 Modelo Multiplicativo Andersen – Resíduos de Schoenfeld por sexo

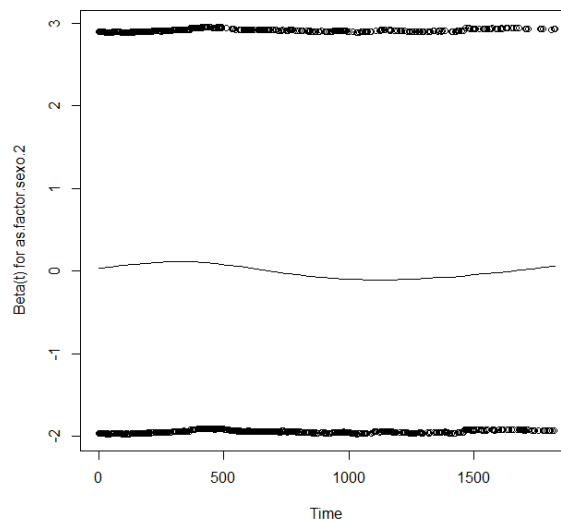


Figura 43 Modelo Multiplicativo Andersen – Resíduos de Schoenfeld por grupo de idade

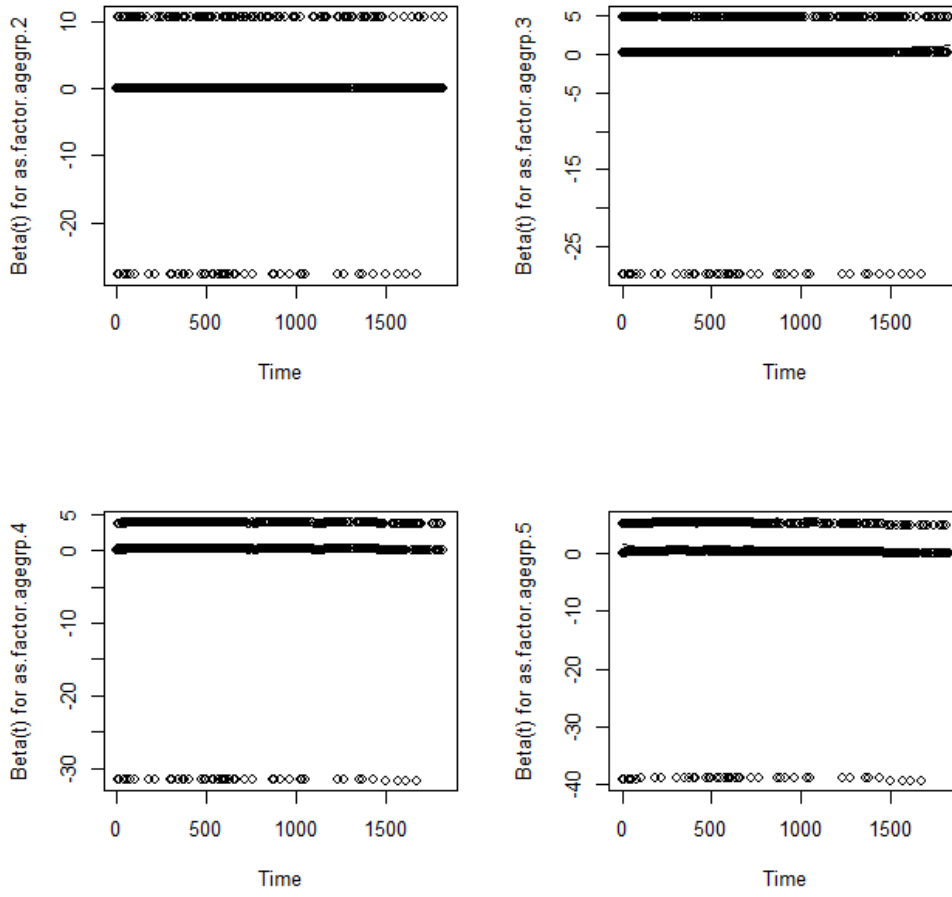


Figura 44 Modelo Multiplicativo Andersen – Resíduos de Schoenfeld por coortes

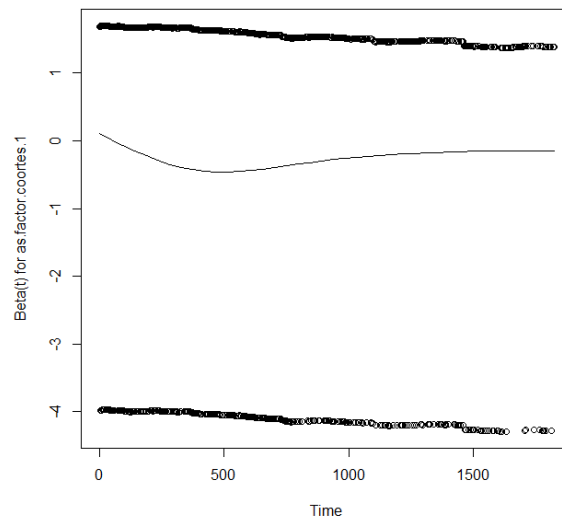
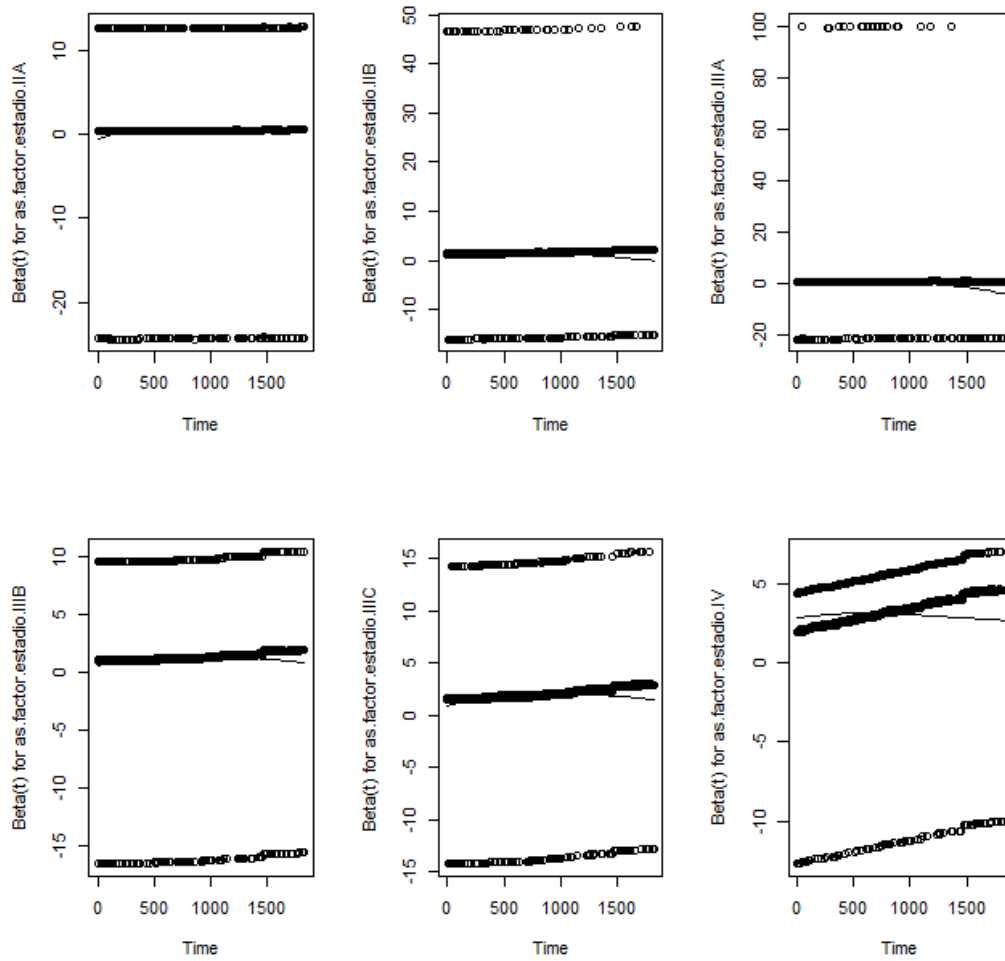




Figura 45 Modelo Multiplicativo Andersen – Resíduos de Schoenfeld por estadio



Tal como no modelo aditivo também podemos constatar que o efeito das covariáveis parece ser constante no tempo, pelo que os pressupostos do modelo são satisfeitos.

## 5. Considerações Finais

Este último capítulo da dissertação é apresentado com foco em 4 componentes principais. Na primeira, e mais extensa, são apresentados os principais resultados obtidos. Na segunda são apresentadas as aprendizagens e competências adquiridas no decorrer do trabalho e na prossecução dos objetivos que se pretendeu alcançar. Na terceira são apresentadas as limitações do estudo e as perspectivas futuras de continuidade do mesmo. Na quarta e última é apresentada uma síntese do trabalho desenvolvido e respetivos objetivos.

### 5.1. Principais resultados obtidos

Depois do tratamento de dados efetuado, visando, por um lado, a completude dos mesmos, e por outro, a formatação necessária para que pudessem servir como *inputs* da ferramenta computacional utilizada, começámos, no presente trabalho, por fazer uma abordagem aos métodos da análise de sobrevivência tradicional, tendo obtido resultados para o estimador de *Kaplan-Meier* e para o modelo de *Cox* de riscos proporcionais.

Em análise de sobrevivência relativa apresentámos resultados dos estimadores *Ederer I*, *Ederer II* e *Hakulinen*, os quais têm por base a estimação da sobrevivência esperada, sendo a sobrevivência relativa uma taxa resultante da razão entre a sobrevivência observada e a sobrevivência esperada representada pela equação (7). Para além dessas estimativas, foram apresentados os resultados do estimador de Pohar-Perme, o qual pelas suas características e unânime adoção pela comunidade científica foi considerado como o mais adequado para a interpretação dos resultados finais deste trabalho.

Quanto aos modelos de regressão foram considerados e calculados modelos semiparamétricos para análise de sobrevivência e para análise de sobrevivência relativa. No primeiro caso foi considerado o modelo de *Cox* de riscos proporcionais e no segundo foram considerados e calculados os modelos aditivos de *Hakulinen-Tenkanen*, *Poisson* e *Estève*, e o modelo multiplicativo de *Andersen*.

## 5.1.1 Estimativas

Neste trabalho consideramos os resultados do estimador *Net Survival* como os mais pertinentes para o mesmo, dadas as conclusões retiradas no capítulo 3.4.1. A tabela seguinte apresenta uma súmula desses mesmos resultados estratificados pelas diversas covariáveis e aqui com os valores de sobrevivência NS (*Net Survival*) representados em forma de percentagem, de forma a proporcionar uma mais fácil leitura dos resultados.

Tabela 52 Resultados gerais Net Survival

Sexo	2000-2001				2007-2008				Total			
	n	NS	IC95%	p	n	NS	IC95%	p	n	NS	IC95%	p
Masculino	479	52.0	47.0-57.5	0.478	821	61.9	58.7-65.4	0.676	1300	59.1	56.3-62.0	0.690
Feminino	364	54.7	49.2-60.8		1191	61.3	57.5-65.3		1555	59.3	56.2-62.6	
<b>Idade</b>												
15-44	53	68.5	57.0-82.2	0.053	91	62.1	52.8-72.9	0.125	144	64.4	57.0-72.8	0.109
45-54	106	53.7	44.9-64.1		254	67.9	62.2-74.1		360	63.8	58.9-69.1	
55-64	221	53.6	47.2-61.0		500	59.2	54.8-63.9		721	57.5	53.8-61.4	
65-74	312	50.1	44.3-56.8		638	62.3	58.1-66.8		950	58.4	54.9-62.1	
75+	151	52.7	42.5-65.3		529	59.9	53.9-66.4		680	58.3	53.1-64.1	
<b>Estadio</b>												
I	160	79.3	71.2-88.3	< 0.001	291	92.4	86.8-98.3	< 0.001	451	88.4	83.6-93.4	< 0.001
IIA	195	83.5	76.7-91.0		548	86.4	82.0-91.0		743	85.7	82.0-89.6	
IIB	24	58.4	40.6-84.0		60	65.1	52.0-81.5		84	63.3	52.2-76.8	
IIIA	30	68.5	52.4-89.5		53	92.1	81.6-100		83	84.0	74.4-94.7	
IIIB	131	57.6	48.8-68.1		404	67.7	62.4-73.4		535	65.2	60.6-70.2	
IIIC	58	32.7	22.3-48.0		156	49.6	41.6-59.1		214	45.1	38.4-53.1	
IV	245	10.0	6.8-14.8		500	11.1	8.6-14.4		745	10.7	8.7-13.3	
<b>Distrito</b>												
Braga	212	60.4	53.3-68.4	0.0146	330	62.9	57.0-69.5	0.594	542	62.2	57.5-67.2	0.273
Bragança	49	60.7	47.7-77.3		122	60.9	51.4-72.1		171	61.1	53.2-70.3	
Porto	508	50.8	46.0-56.1		1274	62.1	59.0-65.4		1782	58.9	56.3-61.7	
V. Castelo	26	60.8	44.3-83.5		158	54.6	46.2-64.5		184	55.9	48.1-65.0	
V. Real	48	31.7	20.1-50.1		128	60.3	51.0-71.3		176	52.6	44.6-62.0	
<b>Topografia</b>												
Cólon	509	54.1	49.3-59.4	0.965	1137	63.4	60.0-66.9	0.308	1646	60.5	57.8-63.4	0.459
Reto	334	51.8	46.1-58.1		875	59.5	55.8-63.4		1209	57.4	54.2-60.7	
<b>Histologia</b>												
Aden\Adenoc.	817	53.7	49.9-57.8	0.096	1952	62.3	59.8-65.0	< 0.001	2769	59.8	57.7-62.0	< 0.001
Outras hist.	26	37.0	22.2-61.8		60	37.2	26.5-52.2		86	39.0	29.3-51.8	

Coortes	n	NS	IC95%	$\rho$
2000-2001	843	53.2	49.5-57.2	< 0.001
2007-2008	2012	61.7	59.2-64.3	
Global	2855	59.2	57.1-61.4	

NS – Net Survival

### Sobrevivência por Sexo

Comparando a sobrevivência observada para os dois sexos (masculino e feminino) pelo estimador de *Kaplan-Meier* obtivemos os valores de 50.9% e 52.9% respectivamente, para o conjunto global. Verificámos, também, que as diferenças entre essas curvas de sobrevivência não apresentam diferenças significativas ( $p = 0.394$ ). Tendo em conta as coortes esses valores foram de 45.5% e 49.4% ( $p = 0.276$ ) para a coorte 2000-2001 e de 53.1% e 54.4% ( $p = 0.69$ ) para 2007-2008. Assim, a diferença entre a sobrevivência observada por sexos não é significativa, embora apresente valores superiores para o sexo feminino.

Em termos de sobrevivência relativa, pelo estimador *Net Survival* de Pohar-Perme, obtivemos os valores de 59.1% e 59.3% ( $p = 0.69$ ), sendo para 2000-2001 os valores 52% e 54.7% ( $p = 0.478$ ) e para 2007-2008 os valores 61.9% e 61.3% ( $p = 0.676$ ). Entre as duas coortes verificou-se que a sobrevivência no sexo feminino é superior na primeira coorte e inferior na segunda coorte. As diferenças não são, também, consideradas significativas.

### Sobrevivência por Grupos de Idade

Comparando a sobrevivência observada para os grupos etários considerados, pelo estimador de *Kaplan-Meier* obtivemos valores decrescentes desde o grupo (15-44) com 63.9% até ao grupo (75+) com 40%, para o conjunto global. Verificámos, também, que existem curvas que apresentam diferenças significativas, tanto para esse conjunto global, como para as coortes. Tendo em conta as coortes esses valores variaram entre 67.9% e 36.4% para a coorte 2000-2001 e entre 61.5% e 41% para 2007-2008. Na coorte 1 a diferença entre a sobrevivência observada entre o grupo etário mais avançado e os restantes grupos

é estatisticamente significativa, bem como entre os grupos 15-44 e 65-74. Na coorte 2 verifica-se, também, uma diferença estatisticamente significativa entre o grupo etário mais avançado e os restantes bem como entre o grupo etário 45-54 e os grupos etários superiores.

Em termos de sobrevivência relativa, pelo estimador *Net Survival* de Pohar-Perme, obtivemos uma atenuação acentuada no decréscimo em função da idade, variando os valores entre 64.4% (15-44) e 58.3% (75+), para o conjunto global ( $p = 0.109$ ). Para 2000-2001 essa variação foi entre 68.5% e 52.7% ( $p = 0.053$ ) e para 2007-2008 essa variação foi entre 62.1% e 59.9% ( $p = 0.125$ ). Assim, tendo a sobrevivência observada apresentado diferenças significativas entre os diferentes grupos etários, em termos de sobrevivência relativa essas diferenças deixaram de ser significativas, sendo este facto mais visível na coorte 2007-2008. Estes resultados fazem com que o efeito da idade na sobrevivência não tenha o impacto que seria de esperar tendo em conta a sobrevivência observada, pois uma maior mortalidade da população em geral para os grupos etários que englobam idades mais avançadas, justifica essa atenuação.

### **Sobrevivência por Estadio**

Comparando a sobrevivência observada, pelo estimador de *Kaplan-Meier* verifica-se existirem diferenças estatisticamente significativas entre as curvas de sobrevivência por estadio entre, tipicamente, os estadios IIIC e IV e os restantes estadios, em ambas as coortes. Obtivemos valores que variam entre 76.5% e 9.8% (estadios I e IV) para o conjunto global, sendo que para a coorte 2000-2001 variam entre 69.9% e 9.4%, e para a coorte 2007-2008 variam entre 80.1% e 10%.

Em termos de sobrevivência relativa, pelo estimador *Net Survival* de Pohar-Perme, as diferenças entre as curvas de sobrevivência no conjunto global e nas 2 coortes continuam a ser estatisticamente significativas entre os estadios IIIC e IV e os restantes, tal como acontece com a sobrevivência observada. Obtivemos valores que variam entre 88.4% e 10.7% (estadios I e IV) para o conjunto global, sendo que para a coorte 2000-2001 variam entre 79.3% e 10%, e para a coorte 2007-2008 variam entre 92.4% e 11.1%. Relativamente à sobrevivência observada verifica-se

que as diferenças entre os estadios se reforçam, tendo a diferença de sobrevivência observada entre os estadios I e IV de 66.7% e a diferença de sobrevivência relativa entre esses mesmos estadios de 77.7%, para o conjunto global. Verificou-se, também que entre as duas coortes se verificou uma melhoria substancial na sobrevivência relativa nos estadios I (13.1%), IIIA (23.6%) e IIIC (16.9%). Facto interessante é o do estadio IIB apresentar uma sobrevivência bastante inferior à do estadio IIIA e ligeiramente inferior à do estadio IIIB. Comparando os estadios IIB e IIIA, essa uma diferença de 20.7% parece indicar que a sobrevivência é superior quando existe invasão de 1 a 3 nódulos (N1), com invasão da submucosa ou da muscularis própria (T1-T2), relativamente a uma doença onde não exista invasão dos nódulos linfáticos próximos (N0), mas cujo tumor tenha invadido diretamente ou aderido a outros órgãos e/ou tenha penetrado na superfície do peritoneu visceral (T4).

### **Sobrevivência por Distrito**

Comparando a sobrevivência observada, pelo estimador de *Kaplan-Meier* verifica-se que existem diferenças estatisticamente significativas entre as curvas de sobrevivência do distrito de Vila Real e dos restantes distritos excetuando o Porto, e entre o distrito do Porto e Viana ou Braga para a coorte 2000-2001. Essas diferenças não se verificam para a coorte 2007-2008 ( $p = 0.455$ ) nem para o conjunto global ( $p = 0.13$ ), tendo na primeira coorte os valores oscilado entre 29.2% para Vila Real e 57.1% para Bragança, na segunda coorte entre 50.8% para Vila Real e 55.8% para Braga.

Em termos de sobrevivência relativa, pelo estimador *Net Survival* de Pohar-Perme, verificou-se existirem, também, as diferenças entre as curvas de sobrevivência do distrito de Vila Real e dos restantes distritos excetuando o Porto, e entre o distrito do Porto e o de Braga para a coorte 2000-2001. Essas diferenças não se verificam a coorte 2007-2008 ( $p = 0.594$ ) nem para o conjunto global ( $p = 0.273$ ). Em termos de sobrevivência relativa os valores oscilam na primeira coorte entre 31.7% para Vila Real e 60.8% para Viana do Castelo e na segunda coorte entre 54.6% para Viana do Castelo e 62.1% para o Porto. O valor de sobrevivência

no distrito de Vila Real para a coorte 2000-2001 destaca-se pela negativa, o que se poderá ficar a dever ao facto de nesse pequeno estrato, 50% dos pacientes terem sido diagnosticados no estadio IV (variando esse valor, nos restantes distritos, entre 20.4% e 30.8%). Entre os dois períodos, 2000-2001 e 2007-2008 os níveis sobrevivência em Braga e Bragança foram semelhantes, tendo-se registado uma ligeira diminuição no distrito de V. do Castelo (-6.2%) e melhorias significativas no Porto (11.3%) e principalmente em Vila Real (28.6%).

### **Sobrevivência por Grupo Topográfico**

Comparando a sobrevivência observada para os dois grupos topográficos considerados, cólon e reto, pelo estimador de *Kaplan-Meier* obtivemos os valores de 52.2% e 51.1% respetivamente, para o conjunto global. Verificámos, também, que as diferenças entre essas curvas de sobrevivência não apresentam diferenças significativas ( $p = 0.835$ ). Tendo em conta as coortes esses valores foram de 47% e 47.6% ( $p = 0.425$ ) para a coorte 2000-2001 e de 54.5% e 52.5% ( $p = 0.709$ ) para 2007-2008. Assim, a diferença entre a sobrevivência observada por grupos topográficos não é significativa sendo ligeiramente superior para o cólon.

Em termos de sobrevivência relativa, pelo estimador *Net Survival* de Pohar-Perme, obtivemos os valores de 60.5% e 57.4% ( $p = 0.459$ ), sendo para 2000-2001 os valores 54.1% e 51.8% ( $p = 0.965$ ) e para 2007-2008 os valores 63.4% e 59.5% ( $p = 0.308$ ). Verificou-se que a sobrevivência apresenta valores superiores para a localização cólon em ambas as coortes. As diferenças não são, também, consideradas significativas entre ambas as curvas de sobrevivência relativa. As diferenças entre a sobrevivência observada e a sobrevivência *Net Survival* indiciam que o diagnóstico do cancro no cólon é efetuado em idades ligeiramente superiores relativamente ao diagnóstico no reto, tal como se pode verificar nos valores de sobrevivência esperada apresentados na Tabela 27.

### **Sobrevivência por Grupo Histológico**

Comparando a sobrevivência observada para os dois grupos histológicos considerados, o primeiro englobando os adenomas e adenocarcinomas e o

segundo englobando todas as outras histologias, pelo estimador de *Kaplan-Meier* obtivemos os valores sobrevivência ao fim de 5 anos de 52.3% e 34.9% respectivamente, para o conjunto global. Verificámos, também, que as diferenças entre essas curvas de sobrevivência apresentam diferenças significativas ( $p < 0.001$ ). Tendo em conta as coortes, esses valores foram de 47.6% e 34.6% ( $p = 0.057$ ) para a coorte 2000-2001 e de 54.2% e 34.9% ( $p < 0.001$ ) para 2007-2008. Assim, a diferença entre a sobrevivência observada por grupos histológicos é significativa no conjunto global sendo os seus valores bastante superiores para os adenomas e adenocarcinomas.

Em termos de sobrevivência relativa, pelo estimador *Net Survival* de Pohar-Perme, obtivemos os valores de 59.8% e 39% ( $p < 0.001$ ), sendo para 2000-2001 os valores 53.7% e 37.0% ( $p = 0.096$ ) e para 2007-2008 os valores 62.3% e 37.2% ( $p < 0.001$ ). As diferenças são consideradas significativas, também, entre as curvas de sobrevivência relativa para a coorte 2007-2008 e para o conjunto global. Verificou-se que a sobrevivência é superior para o grupo histológico dos adenomas e carcinomas em ambas as coortes, mas no entanto, dado o número de casos relativos a outras histologias ser bastante pequeno, apenas 26 na primeira coorte e 60 na segunda, (constituindo apenas 3% do total), os respetivos valores de sobrevivência poderão não ser muito relevantes. Verificou-se, ainda, que entre as coortes houve uma evidente evolução na sobrevivência dos pacientes diagnosticados com histologias adenomas e adenocarcinomas (8.6%), sendo essa diferença muito pequena (0.2%) relativamente às outras histologias.

### **Sobrevivência por Coortes e Global**

Comparando as coortes 2000-2001 e 2007-2008, em análise de sobrevivência, pelo estimador de *Kaplan-Meier* obtivemos os valores de 47.2% e 53.6% respectivamente. Verificámos, também, que as diferenças entre as curvas de sobrevivência entre essas coortes apresentam diferenças significativas ( $p = 0.001$ ). A sobrevivência global estimada nas duas cortes é de 51.7%, obviamente mais próxima da sobrevivência de 2007-2008 dado que a dimensão desta coorte é cerca de 2.4 vezes superior à da coorte 2000-2001.



Em termos de sobrevivência relativa, pelo estimador *Net Survival* de Pohar-Perme, obtivemos os valores de 53.2% e 61.7% para as duas coortes, com diferenças ainda mais significativas ( $p < 0.001$ ). A sobrevivência relativa nas duas coortes é de 59.2%. A diferença da sobrevivência entre as duas coortes aumentou de 6.4% para 8.5%, o que se deve ao facto da sobrevivência esperada da população ser menor em 2007-2008, dado a média de idades nessa corte ser superior. A nível de mortalidade conseguimos perceber que em 2000-2001 a mortalidade foi muito elevada nos 2 primeiros anos (16.7% e 14.5% respetivamente), enquanto que em 2007-2008 a mesma foi elevada no primeiro ano, com 15.3%, tendo decaído para 8.7% no segundo ano. Verificaram-se 1378 mortes no conjunto global sendo que 84.5% das mesmas, ou seja, 1165 se estimou serem causadas pelo CCR, sendo as restantes 213 atribuídas à mortalidade da população. Para as coortes 2000-2001 e 2007-2008 das 445 e 933 mortes observadas, estimámos que 395 (88.8%) e 771 (82.6%), respetivamente, se devam ao CCR. A sobrevivência superior na coorte 2007-2008, relativamente a 2000-2001 poderá, eventualmente, ser justificada pela evolução científica na área clínica, embora este fator não seja objeto deste trabalho.

### **5.1.2. Modelos de regressão**

Quanto aos modelos de regressão aplicados à análise de sobrevivência relativa, pelo teste BB podemos constatar que o efeito do conjunto das covariáveis é mais constante no tempo no modelo multiplicativo de *Andersen*, relativamente ao modelo aditivo de *Estève*. Pelo teste dos resíduos ambos os modelos apresentam valores muito similares indicando que a razão entre as taxas de mortalidade das várias respostas às covariáveis envolvidas e os seus valores de referência são constantes no tempo. No entanto as diferenças entre os coeficientes desses dois modelos são bastante evidentes.

Do modelo aditivo de *Estève* constatamos que, sendo o coeficiente da covariável sexo de sinal negativo (-0.07), a sobrevivência nos pacientes do sexo feminino é relativamente superior à dos pacientes do sexo masculino.

A idade não parece ser um fator muito importante, pelo menos até aos 65 anos, o que significa que para as idades mais baixas, as diferenças de sobrevivência entre os grupos de idade podem ser quase totalmente atribuídas à diferença de mortalidade das populações gerais entre esses grupos de idades.

Os coeficientes de *follow-up* por ano são bastante significativos no modelo e indicam valores de mortalidade semelhantes nos 2 primeiros anos (mortalidade de 0.0204), decrescendo gradualmente no terceiro e quarto ano, sendo que no quinto a descida é mais acentuada para um valor de 0.0123.

As covariáveis que não estão incluídas nas tábuas de mortalidade, como o estadio apresentam o mesmo sinal de coeficiente. No modelo aditivo os coeficientes são superiores aos do modelo multiplicativo, devido ao facto da mortalidade em excesso ser adicionada à mortalidade da população (Perme M.P. 2007), enquanto que no modelo multiplicativo a mortalidade observada é multiplicada por este valor.

Constatamos, também, que no modelo multiplicativo as covariáveis sexo e idade são mais significativas que no modelo aditivo e os seus coeficientes apresentam sinal oposto. Isto significa que essas covariáveis têm o mesmo efeito na mortalidade em excesso no modelo aditivo e na população (Perme M.P. 2007) pois, por exemplo, grupos de idade mais avançados resultam num incremento de mortalidade tanto na população como no excesso de mortalidade.

Tabela 53 Comparação entre razões de taxa de mortalidade

Covariável	Estimador Net Survival		Modelos Regressão	
	$\lambda_i/\lambda_1$	IC95%	Aditivo	Multip.
Sexo				
Masculino				
Feminino	1.00	(0.90-1.10)	0.93	1.68
Idade				
15-44				
45-54	1.01	(0.82-1.24)	1.17	0.51
55-64	1.12	(0.93-1.35)	1.32	0.28
65-74	1.10	(0.92-1.33)	1.48	0.13
75+	1.10	(0.89-1.25)	2.33	0.07
Estadio				
I				
IIA	1.03	(0.93-1.14)	1.51	1.14

IIB	1.40	(1.09-1.79)	4.53	2.31
IIIA	1.05	(0.88-1.26)	1.80	1.12
IIIB	1.36	(1.19-1.54)	3.56	2.29
IIIC	1.96	(1.57-2.43)	6.72	3.82
IV	8.26	(6.29-10.74)	21.16	10.54
Coortes				
2000-2001				
2007-2008	0.86	(0.77-1.03)	0.81	0.92

Na tabela acima foram consideradas para cada covariável incluída nos modelos a razão entre as taxas de mortalidade acumuladas aos 5 anos obtidas a partir do estimador *Net Survival* (com IC95%) e as razões entre as taxas obtidas nos modelos aditivo de Estève e multiplicativo de Andersen.

Pelo que se consegue constatar dessa mesma tabela, as covariáveis que estão incluídas nas tábuas de mortalidade (sexo e idade), parecem apresentar coeficientes mais adequados no modelo aditivo. Para outras covariáveis tais como estadio o modelo multiplicativo parece ser mais adequado. Estas conclusões mostram que nenhum dos modelos parece totalmente adequado, parecendo o modelo aditivo apresentar melhores resultados quando inclui na maioria covariáveis incluídas nas tábuas de mortalidade. Inversamente conjectura-se que o modelo multiplicativo deverá apresentar melhores resultados quando a maior parte das covariáveis incluídas no mesmo não estão diretamente relacionadas com as referidas tábuas.

Antes de utilizarmos a componente relativa da análise de sobrevivência foi calculado o modelo de Cox de riscos proporcionais, que, nesse contexto apresenta resultados mais adequados globalmente (tendo em conta todas as covariáveis envolvidas), relativamente aos modelos aditivo e multiplicativo acima referidos, cuja adequação varia em função das covariáveis consideradas.

## 5.2. Aprendizagem e competências adquiridas

Em termos teóricos a presente dissertação assentou em três pilares fundamentais: o cancro colo-retal como objeto de estudo, enunciado no capítulo 2.1; os conceitos e metodologias utilizados na análise de sobrevivência tradicional, introduzidos no capítulo 2.2; e os conceitos e metodologias utilizados na análise de sobrevivência relativa, introduzidos no capítulo 2.3.

Se relativamente à análise de sobrevivência tradicional, neste estudo se revisitaram os conceitos e metodologias apreendidas durante a fase curricular do mestrado, aplicando-as, no entanto a dados reais, quanto ao cancro colo-retal e à análise de sobrevivência relativa foi gratificante o conjunto de novas aprendizagens e competências adquiridas.

Sobre o cancro colo-retal foram adquiridos conhecimentos sobre a forma como o mesmo pode ser caracterizado em termos da sua localização topográfica (que indica o local onde inicialmente surgiu) e em termos da sua histologia (isto é quanto à caracterização do seu tecido biológico). Interessante foi também perceber que essas características topográficas e histológicas estão perfeitamente identificadas através de sistemas de codificação criados pela IARC, constituindo padrões também adotados pelo RORENO, e que se traduzem em variáveis utilizadas no estudo. Novos conhecimentos foram adquiridos relativamente ao estadiamento do mesmo, pois antes do início do trabalho era conhecido o estadiamento anatómico, vulgarmente identificado pelos estadios I a IV, no entanto foi interessante perceber que esse estadiamento tem como base o estadiamento TNM, criado pela UICC, o qual, através da análise das suas componentes T (tumor), N (nódulo) e M (metástase), permite uma correspondência direta para o estadiamento anatómico, a qual é específica para cada tipo de cancro. Na Tabela 1 é evidenciada essa mesma correspondência para o caso específico do cancro colo-retal.

Antes de iniciar o presente trabalho não tinha a perceção da vasta panóplia de entidades, agências e universidades, formando uma comunidade científica em permanente cooperação, e com uma constante preocupação na adoção e partilha de padrões, que têm ao longo dos anos investido em investigação do cancro, com

os diversos propósitos de prevenção, controlo e inovação na pesquisa e desenvolvimento de novos tratamentos.

De destacar também o papel fundamental dos registos oncológicos, no caso específico do Registo Oncológico Regional do Norte, na sua missão de contribuir para a diminuição da incidência do cancro e para a melhoria da prestação de cuidados de saúde a doentes oncológicos, através da manutenção de uma base de dados de todos os casos ocorridos na região norte de Portugal e respetivo seguimento, papel esse cujas características desconhecia antes do presente trabalho.

Havia já um conhecimento adquirido a nível da análise de sobrevivência tradicional, quanto à definição de conceitos, funções, estimadores e modelos, e da ferramenta computacional R muito focada no pacote mais adequado para esse tratamento, o pacote *Survival*. O mesmo não acontece com a análise de sobrevivência relativa, a qual levou à aquisição de novos conceitos e abordagens computacionais. Essa componente relativa da análise de sobrevivência levou a que para além do conjunto de dados das observações (capítulo 3.1) obtidos a partir do registo oncológico, fossem também considerados dados de mortalidade geral (capítulo 3.2). Estes dados de mortalidade geral foram obtidos a partir de tábuas de mortalidade disponibilizadas pelo estudo CONCORD-2, tendo como origem o Instituto Nacional de Estatística e sendo específicas para a região norte de Portugal. Foram adquiridos, ainda, conhecimentos teóricos sobre os estimadores normalmente adotados de sobrevivência relativa e de *Net Survival* introduzidos no capítulos 2.3.3 e 2.3.4, respetivamente, bem como dos modelos de regressão introduzidos no capítulo 2.3.5. A nível computacional foi utilizado o pacote do R *RelSurv*, o qual, apesar de baseado nos conceitos do pacote *Survival*, obrigou a uma nova aprendizagem, de forma a, numa primeira fase, conseguirmos tratar adequadamente os *inputs* do mesmo, tanto a nível do conjunto de dados observados com também dos dados oriundos das tábuas de mortalidade. Para além dessas operações necessárias para preparar adequadamente os *inputs* computacionais do pacote *RelSurv*, foi necessária toda uma aprendizagem relativamente à formalização matemática dos estimadores e modelos teóricos considerados para a análise de sobrevivência relativa. Ao contrário de outros

pacotes, os quais normalmente se dedicam a tópicos muito específicos, este pacote, criado por Maja Pohar, pela sua versatilidade permitiu abranger os diversos métodos utilizados em análise de sobrevivência relativa abordados neste trabalho (ver anexo 6.1.3), através da utilização de processos homogéneos. Essa característica do pacote *RelSurv* permitiu, assim, uma otimização no tempo de aprendizagem e na aplicação dos métodos.

### **5.3. Limitações do estudo e perspectivas futuras**

Este trabalho apresenta algumas limitações a nível de interpretação clínica. Este facto deve-se, primeiramente, ao desconhecimento específico do autor sobre essas mesmas áreas, mas também à ausência, nos dados obtidos do registo, de informação clínica, para além do estadio, localização topográfica e localização histológica. Isto é, não existe nos dados qualquer informação sobre os diferentes tratamentos: cirurgias ou terapias aplicadas aos pacientes em ambas as coortes tais como quimioterapia, radioterapia e imunoterapia, as quais têm evoluído ao longo do tempo tal como referido no capítulo 2.1.5.

Outra das limitações do estudo prende-se com a impossibilidade de utilizarmos a totalidade dos registos devido à existência de campos não preenchidos em parte significativa dos mesmos, o que levou à sua exclusão. Estas operações de tratamento dos dados referidas no capítulo 3.1, levaram a que dos inicialmente extraídos 7536 registos (2928 e 4608 correspondendo, às coortes cujo diagnóstico foi efetuado em 2000-2001 e 2007-2008, respetivamente), tenham sido incluídos apenas 2855 registos, número que constitui apenas cerca de 38% dos registos iniciais. Dado que a variável estadio, não preenchida em grande parte dos registos, está intrinsecamente ligada às variáveis T (tumor), N (nódulo) e M (metástase), foi possível completá-la a partir destas em muitos dos registos, utilizando a correspondência estabelecida pelo AJCC na Tabela 1. Nessa operação, e dado que, de acordo com a tabela referida, a existência de metástases torna inequívoca a associação ao estadio IV (mesmo que T e N não estejam preenchidas), levou a que a percentagem de pacientes com esse estadio tenha crescido significativamente no conjunto de registos (de 7% para 26%), o que leva a que, em termos globais, se espere um impacto negativo na sobrevivência geral.

Tendo em conta as conclusões retiradas no capítulo 5.1, relativamente aos modelos de regressão não foram incluídos no trabalho modelos, eventualmente mais adequados, que combinam os modelos aditivos baseados na mortalidade em excesso e nos modelos multiplicativos baseados na mortalidade relativa, tais como

sugerido nos artigos (Elie C. 2011) e (Elie C. 2014). Esta não inclusão ficou a dever-se à inexistência desse tipo de tratamento na ferramenta computacional utilizada.

Não foram, ainda, incluídos neste trabalho modelos flexíveis paramétricos para análise de sobrevivência relativa. Os dados de sobrevivência são frequentemente modelados através de modelos baseados no modelo de Cox de riscos proporcionais, que tem a vantagem de estimar o efeito das covariáveis como a transformação logarítmica  $\log$  (riscos) sem necessidade de estimar o risco de base. Todavia o comportamento da função de risco pode ter interesse clínico pois é diretamente relacionado com o curso do tempo de uma doença. A taxa de risco de base pode ajudar a entender a história da doença analisando a forma como a taxa de risco varia ao longo do tempo. Alguns modelos tais como o de Weibull são uma alternativa à abordagem de Cox, no entanto estes modelos não são populares devido às restrições impostas à forma do risco (Nelson 2007). O modelo flexível paramétrico utilizando *splines* (funções definidas em segmentos por polinómios) foi proposto por Royston e Parmar (Royston P. 2002). De acordo com o artigo (Lambert P. 2014), a utilização destes modelos é mais adequada nos casos em que a assunção dos riscos proporcionais não é apropriada e apresentam vantagens a nível de compreensão do modelo, predição, extrapolação, quantificação (medidas relativas e absolutas do risco) e modelação de efeitos dependentes do tempo.

Pelos motivos acima referenciados, o conjunto dos resultados obtidos, a nível de modelação, no presente trabalho poderá ser enriquecido, em futuras abordagens, alargando o âmbito do mesmo de forma a incluir modelos que combinam as características dos modelos aditivos e multiplicativos, bem com os modelos flexíveis paramétricos. Os resultados obtidos pelo estimador *Net Survival*, pelas suas características poderão ser utilizados de forma a poderem ser comparados com os resultados de outras regiões ou mesmo de outros países.

Relativamente aos dados do registo oncológico da região norte verificou-se existir uma evolução positiva no preenchimento dos dados dos pacientes, sendo que os registos não completos ou não completáveis eliminados do estudo relativamente ao período de 2000-2001 foram de cerca de 71% do total de registos, tendo diminuído esse número para cerca de 51% no período de 2007-2008. Ao



manter-se esta tendência de um melhor preenchimento dos registos, é de esperar que em períodos futuros a percentagem de registos analisados relativamente aos totais aumente, levando a que esse incremento de informação enriqueça a análise e melhore a fiabilidade e qualidade da mesma.

No futuro poder-se-á ainda estudar a alternativa de aplicar métodos de imputação de dados para variáveis selecionadas, e verificar o ganho efetivo na estimação, comparando com os ajustamentos que utilizam apenas registos observados (amostra incompleta).

## 5.4. Síntese e conclusões finais

Tendo como objetivo a análise de sobrevivência ao cancro colo-retal na região norte de Portugal para pacientes cujo diagnóstico foi efetuado nos períodos 2000-2001 e 2007-2008, foram obtidos dados oriundos do RORENO relativas a esses períodos. Foram considerados 7536 indivíduos de ambos os sexos e residentes nos distritos de Braga, Bragança, Porto, Viana do Castelo e Vila Real. Após tratamento dos dados, no presente estudo foram considerados 2855 registos de indivíduos com idades entre os 21 e os 98 anos.

Para a análise de sobrevivência foi considerado um período de observação de 5 anos e o evento de interesse a morte. Nos registos em que não ocorreu o evento de interesse durante o período de observação, foi considerado a data de *follow-up* no final do período de observação como censura (censura à direita). Considerando análise de sobrevivência tradicional foram introduzidos os métodos para o cálculo de estimativas e modelos de regressão considerados mais adequados tais como o estimador de *Kaplan-Meier* e o modelo de *Cox* de riscos proporcionais.

Tendo em conta que nos registos a causa do evento de interesse não é conhecida e que esse evento é a morte, interessou-nos comparar a sobrevivência observada com a sobrevivência da população geral, de forma a estimar qual a mortalidade que pode ser associada à doença em questão e qual a mortalidade esperada da população análoga à observada, qualquer que seja a causa, recorrendo assim a métodos de análise de sobrevivência relativa. De forma a utilizar este tipo de análise foi necessário recorrer a um segundo conjunto de dados, obtidos a partir de tábuas de mortalidade da população em geral específicas para a região norte de Portugal.

Foram de seguida introduzidos métodos para o cálculo de estimativas e modelos de regressão. A nível de estimadores foram abordados os de *Ederer I*, *Ederer II*, *Hakulinen* e *Net Survival*, tendo sido este último considerado o mais adequado, pois ao contrário dos três primeiros este método não é resultado da razão entre a sobrevivência observada e a sobrevivência da população. O estimador *Net Survival*, para além de ser o único considerado centrado, pela

comunidade científica, resulta apenas da taxa de mortalidade em excesso o que torna possível a comparação de sobrevivência entre países. Foram também abordados modelos aditivos de *Poisson*, *Hakulinen-Tenkanen* e *Estève*, e o modelo multiplicativo de *Andersen*.

Em ambas as coortes foram comparadas as estimativas da análise de sobrevivência tradicional com as de sobrevivência relativa, tendo-se verificado nesta última que as diferenças entre as curvas de sobrevivência apenas são significativas em função do estadio e das coortes. Pela análise de sobrevivência tradicional, existem curvas de sobrevivência com diferenças estatisticamente significativas em função dos grupos de idade, no entanto recorrendo à análise de sobrevivência relativa essas diferenças deixam de o ser. Isto significa que o efeito da idade na sobrevivência não tem o impacto esperado tendo em conta a sobrevivência observada, pois a maior mortalidade para os pacientes com idades mais avançadas é justificada, em parte, pela maior mortalidade da população em geral.

Comparando com outros estudos, os resultados obtidos para as coortes 2000-2001 e 2007-2008 são similares aos da RORENO para esses períodos, publicados respetivamente em (Bento M.J. 2010) e em (Bento M.J. 2014). Este resultados também estão alinhados com os do artigo *Global surveillance of cancer survival 1995–2009* (Allemani C. 2015).

Quanto aos modelos de regressão, a nível de análise de sobrevivência tradicional foram obtidos resultados satisfatórios com o modelo de *Cox*, e a nível de análise de sobrevivência relativa foram obtidos resultados parcialmente satisfatórios nos modelos aditivos e no modelo multiplicativo, podendo os resultados neste campo ser melhorados em futuras abordagens, nomeadamente ao nível de modelos paramétricos flexíveis.

Podemos concluir, também, e dado o número de diagnósticos registados nos dois diferentes períodos de tempo, que a frequência da doença (ou o seu diagnóstico) tem aumentado ao longo do tempo, mas o aumento da sobrevivência entre esses mesmos períodos indicia, também, que os tratamentos se têm tornado cada vez mais eficazes.

## 6. Anexos

### 6.1. Comandos R utilizados

Foi utilizado o software R versão 3.2.3 (2015-12-10) – Copyright © 2015 The Foundation for Statistical Computing, para todos os cálculos computacionais e elaboração de gráficos. Foi utilizado este software dada a enorme abrangência do mesmo relativamente às abordagens existentes no campo da análise de sobrevivência, contribuindo para isso os pacotes que vão sendo desenvolvidos e atualizados pela comunidade científica.

Abaixo estão identificados os comandos utilizados na produção dos resultados obtidos no capítulo 4, para além dos comandos de base, utilitários e de gráficos tais como o *read.table*, *table*, *summary*, *plot*, *multhist* e outros.

#### 6.1.1. Kaplan-Meier

Package MUHAZ <https://cran.r-project.org/web/packages/muhaz/muhaz.pdf>

Author/Maintainer: S original by Kenneth Hess, <khess@odin.mdacc.tmc.edu> R port by R. Gentleman

Version: 1.2.6 (20/02/2015)

- **Muhaz** - Estimate hazard function from right-censored data. Estimates the hazard function from right-censored data using kernel-based methods. Options include three types of bandwidth functions, three types of boundary correction, and four shapes for the kernel function. Uses the global and local bandwidth selection algorithms and the boundary kernel formulations described in Mueller and Wang (1994). The nearest neighbor bandwidth formulation is based on that described in Gefeller and Dette (1992). The statistical properties of many of these estimators are reported and compared in Hess et al (1999). Based on the HADES program developed by H.G. Mueller. Returns an object of class 'muhaz.'

Package SURVIVAL <https://cran.r-project.org/web/packages/survival/survival.pdf>

Author\Maintainer: Terry M Therneau <therneau.terry@mayo.edu>

Version: 2.38-3 (02/07/2015)

- **Survdiff** - Test Survival Curve Differences. Tests if there is a difference between two or more survival curves using the Gp family of tests, or for a single curve against a known alternative.
- **Survfit** - Create survival curves. This function creates survival curves from either a formula (e.g. the Kaplan-Meier), a previously fitted Cox model, or a previously fitted accelerated failure time model.

## 6.1.2. Cox

Package SURVIVAL <https://cran.r-project.org/web/packages/survival/survival.pdf>

Author\Maintainer: Terry M Therneau <therneau.terry@mayo.edu>

Version: 2.38-3 (02/07/2015)

- **Basehaz** - Alias for the survfit function. Compute the predicted survival curve for a Cox model.
- **coxph** - Fit Proportional Hazards Regression Model. Fits a Cox proportional hazards regression model. Time dependent variables, time dependent strata, multiple events per subject, and other extensions are incorporated using the counting process formulation of Andersen and Gill.
- **cox.zph** - Test the Proportional Hazards Assumption of a Cox Regression. Test the proportional hazards assumption for a Cox regression model fit (coxph).

## 6.1.3. ASR

Package SURVIVAL <https://cran.r-project.org/web/packages/survival/survival.pdf>

Author\Maintainer: Terry M Therneau <therneau.terry@mayo.edu>

Version: 2.38-3 (02/07/2015)

- **Survexp** - Compute Expected Survival. Returns either the expected survival of a cohort of subjects, or the individual expected survival for each subject.

Package RELSURV <https://cran.r-project.org/web/packages/relsurv/relsurv.pdf>

Author\Maintainer: Maja Pohar Perme <maja.pohar@mf.uni-lj.si> [aut,cre], Klemen Pavlic [ctb]

Version: 2.0-7 (15/01/2016)

- **cmp.rel** - Compute crude mortality curves. Estimates the probability of dying due to disease and due to population reasons.
- **rsadd** - Fit an Additive model for Relative Survival. The function fits an additive model to the data. The methods implemented are the maximum likelihood method, the semiparametric method, a glm model with a binomial error and a glm model with a poisson error.
- **rsmul** - Fit Andersen et al Multiplicative Regression Model for Relative Survival. Fits the Andersen et al multiplicative regression model in relative survival. An extension of the coxph function using relative survival.
- **rs.br** - Test the Proportional Hazards Assumption for Relative Survival Regression Models. Test the proportional hazards assumption for relative survival models (rsadd, rsmul or rstrans) by forming a Brownian Bridge.
- **rs.diff** – Test Net Survival Curve Differences. Tests if there is a difference between two or more net survival curves using a log-rank type test.
- **rs.surv** - Compute a Relative Survival Curve. Computes an estimate of the relative survival curve using the Ederer II method, Pohar-Perme method or the Hakulinen method.
- **rs.zph** - Behaviour of Covariates in Time for Relative Survival Regression Models. Calculates the scaled partial residuals of a relative survival model (rsadd, rsmul or rstrans).
- **Transrate** – Reorganize Data into a Ratetable Object. The function assists in reorganizing certain types of data into a ratetable object.

# 6.2. Ratetables

Rate table with dimension(s): age sex year ,, 2000		84 3.465304e-04 2.367659e-04 85 3.850641e-04 2.702098e-04 86 4.235340e-04 3.001142e-04 87 4.657798e-04 3.473913e-04 88 5.287934e-04 3.806746e-04 89 5.817274e-04 4.229524e-04 90 6.331955e-04 4.765017e-04 91 6.963237e-04 5.517929e-04 92 7.824055e-04 6.215926e-04 93 8.603999e-04 7.002174e-04 94 9.269650e-04 7.887856e-04 95 1.063149e-03 9.134240e-04 96 1.145366e-03 1.001244e-03 97 1.259614e-03 1.127887e-03 98 1.445961e-03 1.270553e-03 99 1.590117e-03 1.471322e-03		69 7.724565e-05 4.023494e-05 70 8.373638e-05 4.364066e-05 71 9.345325e-05 5.009673e-05 72 1.023349e-04 5.676652e-05 73 1.122978e-04 6.175504e-05 74 1.267842e-04 7.108778e-05 75 1.390884e-04 7.855834e-05 76 1.527563e-04 8.809283e-05 77 1.693345e-04 9.886960e-05 78 1.861740e-04 1.110344e-04 79 2.047699e-04 1.265376e-04 80 2.252998e-04 1.422848e-04 81 2.523963e-04 1.600673e-04 82 2.753640e-04 1.827122e-04 83 3.088688e-04 2.057496e-04 84 3.396687e-04 2.317791e-04 85 3.777563e-04 2.648602e-04 86 4.165119e-04 2.944345e-04 87 4.588752e-04 3.413492e-04 88 5.200923e-04 3.743201e-04 89 5.719119e-04 4.162856e-04 90 6.237373e-04 4.695506e-04 91 6.887993e-04 5.446516e-04 92 7.733444e-04 6.142980e-04 93 8.500501e-04 6.928586e-04 94 9.199984e-04 7.814766e-04 95 1.055695e-03 9.065604e-04 96 1.138996e-03 9.955195e-04 97 1.254129e-03 1.121772e-03 98 1.404289e-03 1.265333e-03 99 1.585848e-03 1.468001e-03		54 1.998565e-05 8.456787e-06 55 2.165473e-05 9.280827e-06 56 2.334140e-05 1.020405e-05 57 2.558178e-05 1.124304e-05 58 2.753912e-05 1.240894e-05 59 3.014832e-05 1.343502e-05 60 3.269905e-05 1.488263e-05 61 3.596224e-05 1.651280e-05 62 3.926544e-05 1.809433e-05 63 4.273113e-05 2.014010e-05 64 4.696149e-05 2.229299e-05 65 5.149691e-05 2.453733e-05 66 5.687554e-05 2.782123e-05 67 6.166823e-05 3.136385e-05 68 6.735872e-05 3.442027e-05 69 7.506391e-05 3.889884e-05 70 8.151844e-05 4.219167e-05 71 9.085663e-05 4.847432e-05 72 9.953391e-05 5.496695e-05 73 1.094268e-04 5.980947e-05 74 1.232965e-04 6.891936e-05 75 1.353462e-04 7.619820e-05 76 1.488451e-04 8.552159e-05 77 1.649948e-04 9.604949e-05 78 1.816471e-04 1.079568e-04 79 2.000914e-04 1.231647e-04 80 2.204971e-04 1.386218e-04 81 2.466319e-04 1.560963e-04 82 2.699590e-04 1.783985e-04 83 3.025501e-04 2.011078e-04 84 3.328056e-04 2.267955e-04 85 3.704461e-04 2.595090e-04 86 4.094919e-04 2.887514e-04 87 4.519687e-04 3.353080e-04 88 5.113924e-04 3.679646e-04 89 5.620945e-04 4.096191e-04 90 6.142774e-04 4.625976e-04 91 6.812780e-04 5.375089e-04 92 7.642842e-04 6.070022e-04 93 8.397021e-04 6.855020e-04 94 9.130342e-04 7.741652e-04 95 1.048237e-03 8.996949e-04 96 1.132628e-03 9.877915e-04 97 1.248646e-03 1.115658e-03 98 1.434619e-03 1.260113e-03 99 1.581586e-03 1.464680e-03		39 6.760045e-06 2.604999e-06 40 7.204720e-06 2.826991e-06 41 7.682415e-06 3.049000e-06 42 8.256308e-06 3.301181e-06 43 8.841308e-06 3.553385e-06 44 9.528088e-06 3.846738e-06 45 1.023977e-05 4.115445e-06 46 1.097638e-05 4.441767e-06 47 1.177369e-05 4.759899e-06 48 1.259322e-05 5.121957e-06 49 1.348730e-05 5.497780e-06 50 1.447799e-05 5.961458e-06 51 1.554339e-05 6.430702e-06 52 1.664502e-05 7.004333e-06 53 1.811362e-05 7.536901e-06 54 1.975750e-05 8.187654e-06 55 2.118839e-05 8.978650e-06 56 2.280299e-05 9.866474e-06 57 2.494898e-05 1.086038e-05 58 2.684505e-05 1.198542e-05 59 2.937056e-05 1.249743e-05 60 3.184578e-05 1.435964e-05 61 3.495820e-05 1.529291e-05 62 3.814912e-05 1.744673e-05 63 4.151611e-05 1.942032e-05 64 4.539148e-05 2.149743e-05 65 5.002144e-05 2.365898e-05 66 5.515133e-05 2.683675e-05 67 5.991581e-05 3.026736e-05 68 6.545684e-05 3.322002e-05 69 7.288392e-05 3.86638e-05 70 7.929947e-05 4.074345e-05 71 8.825965e-05 4.685287e-05 72 9.673014e-05 5.316857e-05 73 1.065588e-04 5.76429e-05 74 1.198103e-04 6.674986e-05 75 1.316033e-04 7.384009e-05 76 1.449307e-04 8.292731e-05 77 1.606363e-04 9.322662e-05 78 1.771247e-04 1.048798e-04 79 1.954120e-04 1.197932e-04 80 2.156696e-04 1.349759e-04 81 2.408677e-04 1.521253e-04 82 2.645556e-04 1.740886e-04 83 2.962308e-04 1.964620e-04 84 3.259412e-04 2.218091e-04 85 3.631366e-04 2.515192e-04 86 4.024770e-04 2.830710e-04 87 4.450634e-04 3.292647e-04 88 5.022905e-04 3.616083e-04 89 5.528786e-04 4.029497e-04 90 6.048159e-04 4.526641e-04 91 6.737528e-04 5.303648e-04 92 7.552214e-04 5.997088e-04 93 8.293524e-04 6.781406e-04 94 9.060686e-04 7.668589e-04 95 1.040783e-03 8.928314e-04 96 1.126263e-03 9.810642e-04 97 1.243160e-03 1.10946e-03 98 1.428497e-03 1.254895e-03 99 1.577321e-03 1.461358e-03	
,, 2001		male female 0 1.727039e-05 1.521299e-05 1 2.076129e-06 1.361084e-06 2 1.032390e-06 7.420764e-07 3 7.831566e-07 4.642250e-07 4 6.380093e-07 5.777617e-07 5 5.585923e-07 5.312077e-07 6 5.202539e-07 4.956081e-07 7 5.120386e-07 4.737008e-07 8 5.284692e-07 4.600089e-07 9 5.695462e-07 4.545322e-07 10 6.352707e-07 4.572705e-07 11 7.311218e-07 4.627473e-07 12 8.571026e-07 4.764392e-07 13 1.021434e-06 4.928696e-07 14 1.224123e-06 5.147770e-07 15 1.470657e-06 5.421615e-07 16 1.758308e-06 5.722847e-07 17 2.076129e-06 6.051468e-07 18 2.402208e-06 6.407478e-07 19 2.714622e-06 6.763493e-07 20 3.002404e-06 7.174286e-07 21 3.262804e-06 7.639858e-07 22 3.501297e-06 8.105438e-07 23 3.731586e-06 8.625801e-07 24 3.950927e-06 9.200951e-07 25 4.159318e-06 9.830891e-07 26 4.354012e-06 1.051563e-06 27 4.554206e-06 1.128255e-06 28 4.732472e-06 1.210427e-06 29 4.916237e-06 1.303559e-06 30 5.091784e-06 1.399434e-06 31 5.275572e-06 1.511748e-06 32 5.451142e-06 1.626807e-06 33 5.640442e-06 1.755568e-06 34 5.832498e-06 1.898035e-06 35 6.041031e-06 2.051469e-06 36 6.282511e-06 2.207652e-06 37 6.534990e-06 2.380285e-06 38 6.817684e-06 2.577594e-06 39 7.144328e-06 2.766695e-06 40 7.517683e-06 2.988699e-06 41 7.924040e-06 3.205240e-06 42 8.432070e-06 3.457435e-06 43 8.959422e-06 3.706912e-06 44 9.613261e-06 4.003023e-06 45 1.031398e-05 4.279972e-06 46 1.106985e-05 4.625511e-06 47 1.191667e-05 4.971094e-06 48 1.280227e-05 5.363356e-06 49 1.377346e-05 5.780367e-06 50 1.486887e-05 6.285255e-06 51 1.603356e-05 6.801216e-06 52 1.719875e-05 7.424340e-06 53 1.882754e-05 8.011908e-06 54 2.039662e-05 8.723200e-06 55 2.212391e-05 9.580290e-06 56 2.388268e-05 1.054209e-05 57 2.621472e-05 1.162246e-05 58 2.823338e-05 1.285328e-05 59 3.092631e-05 1.390279e-05 60 3.355260e-05 1.540572e-05 61 3.696666e-05 1.709681e-05 62 4.038221e-05 1.873932e-05 63 4.394669e-05 2.086006e-05 64 4.832942e-05 2.309013e-05 65 5.297316e-05 2.541597e-05 66 5.860084e-05 2.880606e-05 67 6.341897e-05 3.246077e-05 68 6.926193e-05 3.562105e-05		male female 0 1.621808e-05 1.443671e-05 1 1.977491e-06 1.265211e-06 2 9.830891e-07 6.873037e-07 3 7.420764e-07 6.024083e-07 4 6.051468e-07 5.446516e-07 5 5.284692e-07 4.956081e-07 6 4.901312e-07 4.654857e-07 7 4.819160e-07 4.63171e-07 8 4.983465e-07 4.536366e-07 9 5.366846e-07 4.298869e-07 10 5.996698e-07 4.326253e-07 11 6.900423e-07 4.435787e-07 12 8.105438e-07 4.572705e-07 13 9.639169e-07 4.764392e-07 14 1.158384e-06 4.983465e-07 15 1.391216e-06 5.257308e-07 16 1.624221e-06 5.585838e-07 17 1.963791e-06 5.887157e-07 18 2.270676e-06 6.243165e-07 19 2.566632e-06 6.626564e-07 20 2.832472e-06 7.009968e-07 21 3.070928e-06 7.420764e-07 22 3.287475e-06 7.858953e-07 23 3.490332e-06 8.351924e-07 24 3.684979e-06 8.872293e-07 25 3.871414e-06 9.420059e-07 26 4.048950e-06 1.007739e-06 27 4.235717e-06 1.076213e-06 28 4.395147e-06 1.155645e-06 29 4.570661e-06 1.243297e-06 30 4.740700e-06 1.364430e-06 31 4.924465e-06 1.443263e-06 32 5.105499e-06 1.555579e-06 33 5.305747e-06 1.681598e-06 34 5.511497e-06 1.818582e-06 35 5.736468e-06 1.969271e-06 36 5.999872e-06 2.125449e-06 37 6.277023e-06 2.298078e-06 38 6.589880e-06 2.495380e-06 39 6.952180e-06 2.698216e-06 40 7.361197e-06 2.907214e-06 41 7.803225e-06 3.128489e-06 42 8.344187e-06 3.380677e-06 43 8.898919e-06 3.630147e-06 44 9.527484e-06 3.926251e-06 45 1.020750e-05 4.197707e-06 46 1.102312e-05 4.535008e-06 47 1.184518e-05 4.864123e-06 48 1.269975e-05 5.242654e-06 49 1.363037e-05 5.640442e-06 50 1.467342e-05 6.123352e-06 51 1.578847e-05 6.614580e-06 52 1.692325e-05 7.215701e-06 53 1.846918e-05 7.757676e-06		,, 2002 male female 0 1.516619e-05 1.366064e-05 1 1.876117e-06 1.169341e-06 2 9.337893e-07 6.352707e-07 3 7.009968e-07 5.58538e-07 4 5.695462e-07 4.983465e-07 5 4.983465e-07 4.600089e-07 6 4.627473e-07 4.326253e-07 7 4.545322e-07 4.161952e-07 8 4.682241e-07 4.079802e-07 9 5.065617e-07 4.052419e-07 10 5.640693e-07 4.107185e-07 11 6.489635e-07 4.216719e-07 12 7.612471e-07 4.381020e-07 13 9.091398e-07 4.572705e-07 14 1.089908e-06 4.819160e-07 15 1.311777e-06 5.093002e-07 16 1.569277e-06 5.421615e-07 17 1.851459e-06 5.750232e-07 18 2.141890e-06 6.106239e-07 19 2.415909e-06 6.462250e-07 20 2.662551e-06 6.845651e-07 21 2.879065e-06 7.229058e-07 22 3.070928e-06 7.612471e-07 23 3.249098e-06 8.050663e-07 24 3.419056e-06 8.516250e-07 25 3.583541e-06 9.036621e-07 26 3.737070e-06 9.611780e-07 27 3.904316e-06 1.026912e-06 28 4.057863e-06 1.100864e-06 29 4.225128e-06 1.183036e-06 30 4.392404e-06 1.270689e-06 31 4.576146e-06 1.374780e-06 32 4.759899e-06 1.484354e-06 33 4.968351e-06 1.604890e-06 34 5.187791e-06 1.741870e-06 35 5.429195e-06 1.889816e-06 36 5.714519e-06 2.043249e-06 37 6.021824e-06 2.215872e-06 38 6.364839e-06 2.413169e-06			







38 4.781841e-06	1835020e-06	51 1.382024e-05	5.135672e-06	64 3.600940e-05	1.592340e-05	77 1.301243e-04	7.347445e-05	90 5.385953e-04	4.069811e-04
39 5.412735e-06	2.035029e-06	52 1.470370e-05	5.530701e-06	65 3.969598e-05	1.751561e-05	78 1.454475e-04	8.333272e-05	91 6.210876e-04	4.803689e-04
40 6.106887e-06	2.265195e-06	53 1.561774e-05	5.879142e-06	66 4.307321e-05	1.995255e-05	79 1.626665e-04	9.618277e-05	92 6.917904e-04	5.486498e-04
41 6.842387e-06	2.500861e-06	54 1.670837e-05	6.323675e-06	67 4.765790e-05	2.259318e-05	80 1.820859e-04	1.093128e-04	93 7.569103e-04	6.266306e-04
42 7.646722e-06	2.763954e-06	55 1.792346e-05	6.878069e-06	68 5.213029e-05	2.481360e-05	81 2.005097e-04	1.243392e-04	94 8.573163e-04	7.156949e-04
43 8.429323e-06	3.018850e-06	56 1.902604e-05	7.503956e-06	69 5.761642e-05	2.821401e-05	82 2.267241e-04	1.439061e-04	95 9.885888e-04	8.447855e-04
44 9.236872e-06	3.298440e-06	57 2.051799e-05	8.215115e-06	70 6.377205e-05	3.059958e-05	83 2.520000e-04	1.639511e-04	96 1.081689e-03	9.339840e-04
45 9.978709e-06	3.534195e-06	58 2.198591e-05	9.017107e-06	71 7.008481e-05	3.550178e-05	84 2.778992e-04	1.869067e-04	97 1.204775e-03	1.066757e-03
46 1.064928e-05	3.797387e-06	59 2.392963e-05	9.687445e-06	72 7.711329e-05	4.058227e-05	85 3.119722e-04	2.167072e-04	98 1.389247e-03	1.218353e-03
47 1.127878e-05	4.024959e-06	60 2.587750e-05	1.069601e-05	73 8.647318e-05	4.424440e-05	86 3.533159e-04	2.433024e-04	99 1.547454e-03	1.438103e-03
48 1.186168e-05	4.269003e-06	61 2.792633e-05	1.184243e-05	74 9.540304e-05	5.157223e-05	87 3.967269e-04	2.869691e-04		
49 1.248596e-05	4.513069e-06	62 3.033103e-05	1.293156e-05	75 1.054033e-04	5.732843e-05	88 4.417862e-04	3.171227e-04		
50 1.311588e-05	4.825724e-06	63 3.300940e-05	1.438992e-05	76 1.175372e-04	6.485681e-05	89 4.835685e-04	3.562831e-04		

# Referências

- ACS (2005). *"Colorectal Cancer Facts & Figures Special Edition 2005."* American Cancer Society.
- ACS (2014). *"What are the risk factors for colorectal cancer?"* acessado em 16/04/2015, disponível em <http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-risk-factors>.
- ACS (2014). *"What`s new in colorectal cancer research and treatment?"* acessado em 16/04/2015, disponível em <http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-new-research>.
- ACS (2014). *"What is colorectal cancer?"* acessado em 16/04/2015, disponível em <http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-what-is-colorectal-cancer>.
- ACS (2015). *"The TNM System."* acessado em 06/05/2015, disponível em <http://www.cancer.org/treatment/understandingyourdiagnosis/staging>.
- Allemani C. et al. (2015). *"Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2)"*. *www.thelancet.com*. 385.
- Allison, P. D. (2010). *"Survival analysis, The Reviewer's Guide to Quantitative Methods in the Social Sciences"*, edited by Gregory R. Hancock and Ralph O. Mueller. New York: Routledge.: 413-425.
- AJCC (2002). *AJCC Cancer Staging Manual 6ed*, Springer-Verlag New York.
- AJCC (2010). *AJCC Cancer Staging Handbook*, Springer-Verlag New York.
- AJCC (2010). *AJCC Cancer Staging Manual 7ed*, Springer-Verlag New York.
- AJCC (2012). *AJCC Cancer Staging Atlas, A Companion to the Seventh Editions of the AJCC Cancer Staging Manual and Handbook*, Springer-Verlag New York.
- AJCC (2015). *"What is Cancer Staging."* disponível em <https://cancerstaging.org/references-tools/Pages/What-is-Cancer-Staging.aspx>.
- Andersen, P. K., Gill R. (1982). *"Cox's Regression Model for Counting Processes: A Large Sample Study"*, *Annals of Statistics*. 10: 1100-1200.
- Andersen P.K., K. B.-J., T. Deckert, A. Green, P. Hougaard, N. Keiding, S. Kreiner (1985). *"A Cox Regression Model for Relative Survival Mortality and its Application to Diabetes Mellitus Survival Data"*, *Biometrics*. 41: 921-932.
- Bento M.J., Antunes L., Castro C. (2010). *"Análise de Sobrevivência, Principais Cancros da Região Norte 2000-2001"*. IPO Porto, ROENO.
- Bento M.J., Antunes L., Castro C., Sousa F. (2014). *"Sobrevivência Global, doentes diagnosticados em 2007-2008, região norte"*. IPO Porto, ROENO.

- Blakely T., Soeberg M., Carter K., Costilla R., Atkinson J., Sarfati D. (2012). "Bias in relative survival methods when using incorrect life-tables: Lung and bladder cancer by smoking status and ethnicity in New Zealand." *IJC - International Journal of Cancer*(Epidemiology).
- Buckley, J. D. (1984). "Additive and multiplicative models for relative survival rates", *Biometrics*. 40: 51-62.
- CDC (2012). "How to Collect High Quality Cancer Surveillance Data." disponível em <http://www.cdc.gov/cancer/npcr/pdf/abstracting/colon.pdf>.
- Colosimo E. A., G. S. R. (2006). *Análise de Sobrevivência Aplicada*, Editora Blucher.
- Corazziari I, Q. and C. R. M (2015). "Standard cancer patient population for age standardising survival ratios", *Eur J Cancer*. 2004 Oct;40(15):2307-16. Table 2, Cancer sites for which each of the three proposed cancer standard populations apply; p.2310. [PubMed]. Acedido em 02/11/2015, disponível em <http://seer.cancer.gov/stdpopulations/survival.html>.
- Cox, D. R. (1972). "Regression Models and Life Tables". *Journal Royal Statistical Society*. B,34: 187-202.
- Cox, D. R. (1975). "Partial Likelihood", *Biometrika*. 62.
- CSG (2016). "UK Life Tables 2016"., disponível em <http://csg.lshtm.ac.uk/tools-analysis/uk-life-tables/>.
- Dickman P.W., Lambert P.C. (2014). "Estimating net survival – past and present. Cancer survival: principles, methods and applications." LSHTM - London School of Hygiene & Tropical Medicine.
- Dickman P.W., Sloggett A., Hills M., Hakulinen T. (2003). "Regression Models for Relative Survival", *J. R. Stat. Soc. Ser. C* 23: 51-64.
- Ederer, F., Axtell, L. M., and Cutler, S. J. (1961). "The relative survival rate: a statistical methodology", *National Cancer Institute Monograph*. 6: 101-121.
- Ederer F., Heise H. (1959). "Instructions to IBM 650 programmers in processing survival computations", *National Cancer Institute, Bethesda, MD*. methodological note no. 10, end results evaluation section. Tech. rep.
- Elie C., Y. D. R., Lais J.P., Landais P. (2011). "Appraising relative and excess mortality in population-based studies of chronic diseases such as end-stage renal disease", *Clinical Epidemiology Journal*.
- Elie C., L. P., Rycke Y. (2014). "A model combining excess and relative mortality for population-based studies." *Statistics in Medicine*, John Wiley & Sons, Ltd. 33: 275-288.
- Estève J., E. B., Croasdale M., Raymond M. (1990). "Relative Survival and the Estimation of Net Survival: Elements for further discussion", *Stat. Med*. 9: 529-538.
- Grambsch, P. M., Therneau T.M. (1994). "Proportional Hazard Tests and Diagnostics based on Weighted Residuals", *Biometrika*. 81: 515-526.
- Hanahan D., W. R. A. (2011). "Hallmarks of Cancer: The Next Generation." ISREC Switzerland, MIT Department of Biology.

- Hakulinen, T. (1982). "Cancer survival corrected for heterogeneity in patient withdrawal", *Biometrics*: 933–942.
- Hakulinen T., Tenkanen L. (1987). "Regression Analysis of Relative Survival Rates", *J. R. Stat. Soc. Ser. C* 36: 309-317.
- Hyunsoon C. , N. H., Angela B. Mariotto and Kathleen A. Cronin (2011). "Estimating relative survival for cancer patients from the SEER Program using expected rates based on Ederer I versus Ederer II method", *Surveillance Research Program, NCI. Technical Report #2011-01*.
- Jooste V., G. P., Remontet L, Launoy G, Baldi I, Molini'e F., et al.. (2013). "Unbiased estimates of long-term net survival of solid cancers in France." , *International Journal of Cancer* 132: 2370-2377.
- Kaplan E. L., Meier P. (1958). "Nonparametric estimation from incomplete observations." *Journal of the American Statistical Association*. 53: 457-481.
- Kyngas H., Rissane M. (2001). "Support as a crucial predictor of good compliance of adolescents with a chronic disease." *Journal of Clinical Nursing, Blackwell Science Ltd* 10: 767-774.
- Lambert P.C. (2014). "The use of flexible parametric survival in epidemiology. *Research Seminar in Medicine*", *Epidemiology and Public Health Department of Public Health, Aarhus University*.
- Lambert P.C., Dickman P. W., Nelson C.P., Royston P. (2009). "Modelling Relative Survival: Flexible Parametric Models and the Estimation of Net and Crude Mortality." *Department of Health Sciences, University of Leicester, UK, Medical Epidemiology & Biostatistics, Karolinska Institutet, Stockholm, Sweden, MRC Clinical Trials Unit, London, UK*.
- Monnereau A, T. X., Belot A, Guizard AV, Woronoff AS, Bara S, et al. (2013). "Unbiased estimates of long-term net survival of hematological malignancy patients detailed by major subtypes in France". *International Journal of Cancer*: 2378-2387.
- Nelson C., Lambert P.C., Squire I.B., Jones D.R. (2007). "Flexible parametric models for relative survival, with application in coronary heart disease." *Wiley InterScience, Statistics in Medicine*. 26: 5486-5498.
- Perme M.P., Stare J., Estève J. (2012). "On Estimation in Relative Survival", *The International Biometric Society. Biometrics DOI:10.1111/j.1541-0420-2011.01640.x*.
- Perme, M. P. (2007). "Prileganje refresijskih modelov za relativno preživetje". *Phd, Ljubljana University*.
- Perme, M. P., Pavlic K. (2015). *Package 'relsurv', CRAN*.
- Pohar M., Stare J. (2006). "Relative Survival Analysis in R", *Elsevier. Computer Methods and Programs in Biomedicine*: 272-278.
- Pohar M., Stare J. (2007). "Making Relative Survival Analysis relatively easy", *Elsevier. Computers in Biology and Medicine*: 1741-1749.
- Rao C.R. (2007). "Epidemiology and Medical Statistics", *North Holland*.

- Roche L, Danieli C., Belot A, Grosclaude P, Bouvier A.M., Velten M., et al (2012). "Cancer net survival on registry data: Use of the new unbiased Pohar-Perme estimator and magnitude of the bias with the classical methods", *Int J Cancer*. 132: 2359–2369.
- Royston P, Parmar M. K. (2002). "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*. 21: 2175-2197.
- RORENO (2015). "Registo Oncológico Regional do Norte 2010". Instituto Português de Oncologia do Porto.
- Schoenfeld D. (1982). "Partial Residuals for the Proportional Hazard Regression Model", *Biometrika*. 69: 239-241.
- Silva L. S. (2011). "Qual é a diferença entre tumor e cancro?", acessado em 16/04/2015, disponível <http://www.portaldasauade.pt/portal/conteudos/enciclopedia+da+saude/harvard/Cancro/diferenca+tumor+cancro.htm>.
- Spika D, R. B., Bannon F., Woods L.M., Maringe C., and C. M. Bonaventure A., Allemani C. "Life tables for the CONCORD-2 study." Acessado em 14/10/2015, disponível em <http://csg.lshtm.ac.uk/life-tables>.
- Therneau, T. M., Grambsch P.M. (2000). "Modelling Survival Data: Extending the Cox Model." New York, Springer-Verlag.
- UICC (2015). "What is TNM?" acessado em 06/05/2015, disponível em <http://www.uicc.org/resources/tnm>.
- Wikipedia. (2016). "Bathtub curve", acessado em 25/07/2016, 2016, disponível em [https://en.wikipedia.org/wiki/Bathtub\\_curve](https://en.wikipedia.org/wiki/Bathtub_curve).