

Universidade do Minho
Escola de Ciências

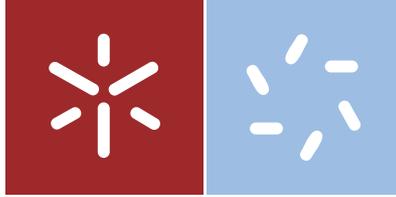
Paula Rosa Silva Lopes

Estudo de Sobrevivência das
Doentes com Cancro do Colo de Útero

Paula Rosa Silva Lopes | Estudo de Sobrevivência das
Doentes com Cancro do Colo de Útero

UMinho | 2016

dezembro de 2016



Universidade do Minho
Escola de Ciências

Paula Rosa Silva Lopes

Estudo de Sobrevivência das
Doentes com Cancro do Colo de Útero

Dissertação de Mestrado
2º Ciclo de Estudos de Mestrado em Estatística

Trabalho efetuado sob a orientação do
Professor Doutor Luís Filipe Meira Machado (UM)
Dr. Eugénio Ferreira Neves Cordeiro (ARSC)

AGRADECIMENTO

A realização deste estágio não seria possível sem o contributo de várias pessoas a quem dirijo os meus sinceros agradecimentos.

Em primeiro lugar aos coordenadores do Programa de Estágios Profissionais na Administração Central (PEPAC) por me terem proporcionado um estágio profissional em estatística na Administração Regional do Saúde do Centro, I.P., (ARSC).

À Prof.^a Doutora Arminda Manuela Gonçalves pelo seu contributo para a integração do meu projeto de tese, no âmbito do PEPAC, na ARSC.

Ao Diretor do Departamento de Saúde Pública (DSP) da ARSC, Dr. João Pedro Pimentel pela aceitação da realização do meu estágio, incluindo o meu projeto de tese, no departamento, bem como pela colaboração prestada.

Ao Dr. Eugénio Cordeiro por ter aceitado a realização do meu estágio e projeto de tese na Unidade de Investigação e Planeamento em Saúde (UIPS) do DSP sob sua orientação.

Ao prof. Doutor Luís Filipe Meira Machado, pela orientação, ajuda, paciência, e sobretudo pela partilha de conhecimentos.

À Dr.^a Joana Bastos, do Registo Oncológico Regional do Centro (RORC), pela partilha de conhecimentos e pela sua capacidade de ajuda na resolução prática de problemas.

Ao Dr. Pedro Valente do RORC pela ajuda na recolha de informações nos processos médicos do RORC do Instituto Português de Oncologia (IPO) de Coimbra.

À coordenadora do programa de rastreio do cancro do colo do útero da ARSC, Dr.^a Fernanda Loureiro pela sua ajuda na recolha de informação.

Às administrações do Centro Hospitalar e Universitário de Coimbra, das Unidades locais de Saúde da Guarda e de Castelo Branco, dos Centros Hospitalar da Cova da Beira, Tondela Viseu e Hospitalar do Baixo Vouga e dos Hospitais Distrital da Figueira da Foz e de Santo André, pela colaboração no preenchimento de informação em falta que não se encontrava disponível no RORC.

À colega e companheira de trabalho, Enfermeira Clara Rodrigues pela ajuda na recolha de dados nos processos médicos e interpretação dos registos clínicos.

À minha família que mesmo distante, me incentivou a prosseguir e me deu apoio moral.

Ao meu irmão Luís Roque Lopes, pelo incentivo e apoio moral.

Ao Nilton Gomes, pelo companheirismo, ajuda, apoio e incentivo durante a realização deste projeto.

A todos os meus amigos e pessoas que me acompanharam ao longo destes anos.

Muito Obrigado a todos!

RESUMO

Este trabalho demonstra de que forma a sobrevivência após o diagnóstico do CCU é influenciado pelo rastreio do CCU. O principal objetivo deste projeto foi comparar a sobrevivência das mulheres submetidas ao rastreio do CCU com a das mulheres não submetidas ao rastreio e verificar quais as covariáveis que influenciam a sobrevivência por CCU.

Da análise de sobrevivência, verificou-se que o tempo médio de sobrevivência por CCU é de 2381,3 dias e a sobrevivência global a 1, 3 e 5 anos após o diagnóstico do CCU são de 95,7%, 89,8% e 87,3%, respetivamente. Em média as mulheres submetidas ao rastreio do CCU sobrevivem 478 dias mais do que as mulheres não submetidas ao rastreio do CCU. As covariáveis que por si só influenciaram a sobrevivência por CCU foram, rastreio do CCU, faixa etária, comportamento do cancro, estadiamento do cancro à data diagnóstico, distrito de residência das pacientes, sintomas do cancro, nível de escolaridade, número de partos, número de gestações e o número de abortos. Através do teste de log-rank e o modelo de Cox simples, verificou-se que existem diferenças estatisticamente significativas na sobrevivência das mulheres submetidas ao rastreio e das não submetidas ao rastreio e que as mulheres não submetidas ao rastreio do CCU têm 7,831 vezes mais risco de morte por CCU do que as mulheres submetidas ao rastreio. Deste estudo conclui-se que existem diferenças estatisticamente significativas na sobrevivência após o diagnóstico do CCU para as mulheres rastreadas e não rastreadas, só a partir de 45 anos.

Do modelo de regressão de Cox múltiplo, conclui-se que para um nível de significância de 5% as mulheres não submetidas ao rastreio, com comportamento do cancro maligno, com localização primária na faixa etária dos [45,54], têm um risco acrescido estatisticamente significativa de ter morte por CCU.

PALAVRAS-CHAVE: Análise de sobrevivência; Cancro do colo do útero; Rastreio do cancro do colo do útero; Modelo de regressão de Cox.

ABSTRACT

This work shows how survival after CC diagnosis is influenced by CC screening. The main objective of this project was to compare the survival time of women undergoing CC screening with the survival time of women not screened and explore which socioeconomic, demographic and clinic covariates influence CC survival.

From the survival analysis it was concluded that the mean survival time for CC is 2381.3 days and overall survival at 1, 3 and 5 years after CC diagnosis was 95.7%, 89.8% and 87, 3%, respectively. On average, women undergoing CC screening survive 478 days longer than women not screened for CC. The covariates that single influenced CC survival were CC screening, age group, cancer behavior, cancer staging at diagnosis date, patients residence district, cancer symptoms, level of education, number of deliveries, number of pregnancies and the number of abortions. Through the use of the log-rank test and the simple Cox model, it was found that there were statistically significant differences in the survival of women undergoing screening and those not undergoing screening, and that women not undergoing CC screening had a 7,831-fold increased risk of death by CC than women undergoing screening. Statistically significant differences were found in survival after the diagnosis of CC among the screened and unscreened women from the age of 45 years.

From the fitted multiple Cox regression model, it is concluded that at a significance level of 5%, non-screened women with malignant cancers with a primary location in the age range of [45,54] have an increased risk statistically significant from CC death.

KEYWORDS: Survival analysis; Cervical cancer; Screening; Cox regression model.

ÍNDICE

CAPÍTULO 1	1
INTRODUÇÃO	1
1.1. OBJETIVOS DO ESTÁGIO	3
1.2. ESTRUTURA DA DISSERTAÇÃO	3
1.3. ENQUADRAMENTO TEÓRICO DO TEMA.....	4
1.3.1. MAGNITUDE DO CANCRO NO MUNDO	4
1.3.2. MAGNITUDE DO CANCRO EM PORTUGAL	5
1.3.3. MAGNITUDE DO CANCRO DE COLO DE ÚTERO DO MUNDO	6
1.3.4. CANCRO DO COLO DO ÚTERO	6
1.3.4.1. FATORES DE RISCO ASSOCIADO AO CANCRO DO COLO DO ÚTERO	7
1.3.4.2 SINTOMATOLOGIA DO CANCRO DO COLO DO ÚTERO.....	11
1.3.4.3 IMPORTÂNCIA DO RASTREIO NA DETEÇÃO PRECOCE DO CANCRO DO COLO DO ÚTERO	13
1.3.4.4 TÉCNICA DO RASTREIO.....	14
1.3.4.4.1 CITOLOGIA CERVICAL	14
1.3.4.5. INTERPRETAÇÃO DOS RESULTADOS CITOLÓGICO- CLASSIFICAÇÃO DE BETHESDA	16
1.3.4.6 MÉTODOS DE TRATAMENTO DE CANCRO DO COLO DO ÚTERO	19
1.3.4.7 VACINAÇÃO PARA O VÍRUS DE HPV E O SEU PAPEL NO COMBATE AO CANCRO DO COLO DO ÚTERO.....	21
CAPÍTULO 2	27
NOÇÕES BÁSICAS DE ANÁLISE DE SOBREVIVÊNCIA	27
2.1 INTRODUÇÃO	29
2.2 CENSURA	29
2.2.1 MECANISMO DE CENSURA À DIREITA.....	31

2.3	TRUNCATURA	32
2.3.1	TRUNCATURA À DIREITA	32
2.3.2	TRUNCATURA À ESQUERDA	32
2.4	FUNÇÕES DE INTERESSE NO CONTEXTO DE DADOS DE SOBREVIVÊNCIA.....	33
2.4.1	VARIÁVEL ALEATÓRIA CONTÍNUA	33
2.4.1.1	FUNÇÃO DE DISTRIBUIÇÃO	33
2.4.1.2	FUNÇÃO DENSIDADE DE PROBABILIDADE	33
2.4.1.3	FUNÇÃO DE SOBREVIVÊNCIA	34
2.4.1.4	FUNÇÃO DE RISCO.....	34
2.4.1.5	FUNÇÃO DE RISCO CUMULATIVA	35
2.4.2	VARIÁVEL ALEATÓRIA DISCRETA	35
2.5	RELAÇÕES ENTRE AS FUNÇÕES.....	36
2.6	MODELO DE SOBREVIVÊNCIA PARAMÉTRICO	37
2.6.1	DISTRIBUIÇÃO EXPONENCIAL	38
2.6.2	DISTRIBUIÇÃO WEIBULL	38
2.6.3	DISTRIBUIÇÃO GAMA	39
2.6.4	DISTRIBUIÇÃO LOG-NORMAL	40
2.6.5	DISTRIBUIÇÃO DE PARETO	41
2.6.6	DISTRIBUIÇÃO LOG-LOGÍSTICA	42
2.7	ESTIMAÇÃO PARAMÉTRICA DA FUNÇÃO DE SOBREVIVÊNCIA	42
2.8	ESTIMAÇÃO NÃO PARAMÉTRICA DA FUNÇÃO DE SOBREVIVÊNCIA	44
2.8.1	ESTIMADOR DE KAPLAN-MEIER	44
2.8.2	O ESTIMADOR DE GREENWOOD.....	48
2.8.3	INTERVALO DE CONFIANÇA PARA A FUNÇÃO DE SOBREVIVÊNCIA.....	49
2.8.4	ESTIMAÇÃO DA FUNÇÃO DE SOBREVIVÊNCIA E FUNÇÃO DE RISCO CUMULATIVA UTILIZANDO O ESTIMADOR DE NELSON-AALEN.....	51

2.8.5	TESTES DE HIPÓTESES PARA COMPARAR CURVAS DE SOBREVIVÊNCIA	51
2.9	MODELO DE REGRESSÃO DE COX	57
2.9.1	FORMULAÇÃO DO MODELO DE REGRESSÃO DE COX	57
2.9.2	ESTIMAÇÃO NO MODELO DE COX.....	58
2.9.3	SELEÇÃO DE COVARIÁVEIS NO MODELO DE COX	60
2.9.3.1	TESTES E HIPÓTESES	61
2.9.3.1.1	O TESTE DE WALD	61
2.9.3.1.2	TESTE DA RAZÃO DE VEROSIMILHANÇAS.....	62
2.9.4	SELEÇÃO DE UM MODELO ADEQUADO.....	62
2.9.4.1	TESTE DA RAZÃO DE VEROSIMILHANÇAS.....	63
2.9.4.2	CRITÉRIO DE INFORMAÇÃO DE AKAIKE.....	64
2.9.4.3	SELEÇÃO DO MODELO USANDO O TESTE DE WALD	64
2.9.5	VALIDAÇÃO DO PROSSUPOSTO DE RISCOS PROPORCIONAIS.....	65
2.9.5.1	RESÍDUOS DE SCHOENFELD.....	65
2.10	EXTENSÕES DE MODELO DE COX	66
2.10.1	MODELO DE REGRESSÃO DE COX COM COVARIÁVEIS DEPENDENTES NO TEMPO 67	
2.10.2	MODELO DE COX ESTRATIFICADO	69
CAPÍTULO 3	71
METODOLOGIA DO ESTUDO	71
3.1.	DESENHO DE ESTUDO	73
3.2.	DESCRIÇÃO DA BASE DE DADOS UTILIZADA.....	73
3.3.	TIPO DE VARIÁVEL ENVOLVIDO EM ESTUDO	75
3.3.1	VARIÁVEL DEPENDENTE OU VARIÁVEL RESPOSTA	75
3.3.2	VARIÁVEL INDEPENDENTE	75
3.4	DESCRIÇÃO DAS VARIÁVEIS ENVOLVIDAS NO ESTUDO	75

3.5 SOFTWARE UTILIZADO	78
3.6 METODOLOGIA UTILIZADA	79
CAPÍTULO 4	81
APRESENTAÇÃO DOS RESULTADOS SOBRE O ESTUDO DO CCU.....	81
4.1 ANÁLISE EXPLORATÓRIA.....	83
4.1.1 ANÁLISE EXPLORATÓRIA PARA VARIÁVEIS QUALITATIVAS	83
4.1.2 ANÁLISE EXPLORATÓRIA PARA VARIÁVEIS QUANTITATIVAS.....	86
4.1.3 ANÁLISE EXPLORATÓRIA PARA VARIÁVEIS LOCALIZAÇÃO TOPOGRÁFICA E MORFOLOGIA DO TUMOR.....	88
4.1.4 CRUZAMENTO DAS VARIÁVEIS	92
4.1.4.1 CRUZAMENTO DAS VARIÁVEIS RASTREIO, ESTADO DA PACIENTE E ESTADIAMENTO.....	93
4.1.4.2 CRUZAMENTO DAS VARIÁVEIS RASTREIO E COMPORTAMENTO DO CANCRO .	94
4.1.5. ESTATÍSTICA DESCRITIVA PARA O RESULTADO DE RASTREIO	95
4.2. ASSOCIAÇÃO ENTRE RASTREIO E O COMPORTAMENTO DO CANCRO DO COLO DO ÚTERO.....	96
4.2 1. TESTE DE ASSOCIAÇÃO DE QUI-QUADRADO	96
4.3. ANÁLISE DE SOBREVIVÊNCIA.....	99
4.3.1 CATEGORIZAÇÃO DA VARIÁVEL IDADE	99
4.3.2. CRUZAMENTO DA VARIÁVEL RASTREIO, FAIXA ETÁRIA E ÓBITO.....	101
4.3.3 SOBREVIVÊNCIA APÓS O DIAGNÓSTICO DO CCU.....	102
4.3.4 TESTE LOG-RANK.....	116
4.3.5 AJUSTAMENTO DO MODELO DE REGRESSÃO DE COX	118
4.3.5.1. MODELO DE REGRESSÃO DE COX SIMPLES	118
4.3.5.2 MODELO DE REGRESSÃO DE COX MÚLTIPLO	122
4.3.5.3. VALIDAÇÃO DO PRESSUPOSTO DE RISCOS PROPORCIONAIS	123
4.5.3.1. RESÍDUOS DE SCHOENFELD	123

4.3.5.4. INTERPRETAÇÃO DOS PARÂMETROS DO MODELO FINAL	125
CAPÍTULO 5	127
DISCUSSÃO.....	127
CAPÍTULO 6	133
CONCLUSÕES, LIMITAÇÕES E RECOMENDAÇÕES	133
6.1. CONCLUSÕES.....	135
6.2. LIMITAÇÕES.....	136
6.3. RECOMENDAÇÕES.....	137
BIBLIOGRAFIA.....	139
ANEXOS	147
Anexo I. REPRESENTAÇÃO GRÁFICA PARA COMPARAR CURVA DE SOBREVIVÊNCIA DE KAPLAN-MEIER.....	149
Anexo II. COMPARAÇÃO DE CURVA DE SOBREVIVÊNCIA USANDO O TESTE LOG-RANK	149
Anexo III. LEITURA DA BASE DE DADOS A UTILIZAR	149
Anexo IV. CATEGORIZAÇÃO DAS VARIÁVEIS NO R.....	149
Anexo V. VERIFICAÇÃO DOS PRESSUPOSTOS DE TESTE DE ASSOCIAÇÃO DE QUI- QUADRADO	150
Anexo VI. CÁLCULO DO RISCO RELATIVO	150
Anexo VII. BOXPLOT E ESTATÍSTICA DESCRITIVA PARA VARIÁVEL IDADE.....	150
Anexo VIII. CRIAÇÃO DA VARIÁVEL FAIXA ETÁRIA.....	151
Anexo IX. FREQUÊNCIAS ABSOLUTAS E RELATIVAS PARA FAIXA ETÁRIA	151
Anexo X. ESTIMAÇÃO DE FUNÇÃO DE SOBREVIVÊNCIA GLOBAL DE KAPLAN-MEIER E A SUA REPRESENTAÇÃO GRÁFICA.....	151
Anexo XI. TEMPO MÉDIO DE SOBREVIVÊNCIA POR CCU.....	151
Anexo XII. ESTIMAÇÃO DE FUNÇÃO DE SOBREVIVÊNCIA DE KALAN-MEIER POR RASTREIO E A SUA REPRESENTAÇÃO GRÁFICA.....	151
Anexo XIII. TEMPO MÉDIO DE SOBREVIVÊNCIA DO CCU POR RASTREIO DO CCU.....	152

Anexo XIV.	TEMPO MÉDIO DE SOBREVIVÊNCIA DO CCU POR COMPORTAMENTO DO CANCRO	152
Anexo XV.	ESTIMAÇÃO DE FUNÇÃO DE SOBREVIVENCIA GLOBAL USANDO O ESTIMADO DE NELSON-AALEN	152
Anexo XVI.	MODELO FINAL DE COX MULTIVARIADO	153
Anexo XVII.	RESÍDUOS DE SCHOENFELD.....	153

ÍNDICE DE FIGURAS

Figura 1 - Útero (adaptado de A.D.A.M. -Enciclopédia Multimédia).	6
Figura 2 - Progressão do cancro do colo de útero (adaptado de Sanofi Pasteur MSD).....	7
Figura 3 - HPV infetando as células epiteliais (Fonte: Pencubanexile.org-Health Medical Information).....	8
Figura 4 - Fases de pré-malignidade do cancro cervical (fonte: Emory University School of Medicine)	12
Figura 5 - Cervex -Brush (fonte: CervexBrush.com).	15
Figura 6 – Representação gráfica da curva de sobrevivência pelo método Kaplan-Meier.....	47
Figura 7 - Curva de sobrevivência estimada de Kaplan-Meier para os dois tipos de tratamento. 56	
Figura 8 - Boxplot da variável idade	100
Figura 9 – Gráfico de barras (barplot) da variável faixa etária	101
Figura 10 - Estimativa da curva de sobrevivência global do CCU de Kaplan-Meier	104
Figura 11 - Curva de sobrevivência estimada de Kaplan-Meier por rastreio.....	105
Figura 12 - Curva de sobrevivência estimada de Kaplan-Meier por faixa etária	105
Figura 13 - Curva de sobrevivência estimada de Kaplan-Meier para o cruzamento das faixas etárias e rastreio.....	106
Figura 14 - Curva de sobrevivência estimada de Kaplan-Meier por estadiamento	107
Figura 15 - Curva de sobrevivência estimada de Kaplan-Meier por distrito de residência das pacientes.....	108
Figura 16 - Curva de sobrevivência estimada de Kaplan-Meier por comportamento do cancro	109
Figura 17 - Curva de sobrevivência estimada de Kaplan-Meier por estado civil	110
Figura 18 - Curva de sobrevivência estimada de Kaplan-Meier por variável fumador	110
Figura 19 - Curva de sobrevivência estimada de Kaplan-Meier por infeção por HPV	111
Figura 20 - Curva de sobrevivência estimada de Kaplan-Meier por antecedentes familiares	111
Figura 21 - Curva de sobrevivência estimada de Kaplan-Meier por diabetes	112
Figura 22 - Curva de sobrevivência estimada de Kaplan-Meier por obesidade	113
Figura 23 - Curva de sobrevivência estimada de Kaplan-Meier por contraceção.....	113
Figura 24 - Curva de sobrevivência estimada de Kaplan-Meier por nível de escolaridade	114
Figura 25 - Curva de sobrevivência estimada de Kaplan-Meier por sintomas	115
Figura 26 - Curva de sobrevivência estimada de Kaplan-Meier por ano de diagnóstico	115

Figura 27 - Estimativa da curva de sobrevivência global do CCU de Nelson-Aalen	116
Figura 28 - Resíduos de Schoenfeld para o modelo final (apresentado na Tabela 53)	124

ÍNDICE DE TABELAS

Tabela 1 - Classificação das estirpes de HPV segundo o seu potencial oncogénico (adaptado de Munoz et al., 2003)	9
Tabela 2 - Percentagem de lesões atribuídas aos tipos específicos de HPV (adaptado de Munoz et al., 2003)	9
Tabela 3 - Resultados possíveis de uma citologia e procedimento a seguir	17
Tabela 4 - Outras classificações para alterações citológicas cervicais.	18
Tabela 5 - Estadiamento da FIGO (2009) para o CCU	20
Tabela 6 - Tratamento realizado em Portugal nos diferentes estádios do cancro do colo do útero (adaptado SPG,2007).	21
Tabela 7 - Alguns parâmetros de sobrevivência	37
Tabela 8 - Dados para ilustrar estimativa para a função de sobrevivência.....	46
Tabela 9 - Dados ordenados onde + indica presença de censura pela direita	46
Tabela 10 - Estimativa da função de sobrevivência pelo método de Kaplan-Meier.....	46
Tabela 11 - Estimativa de função de sobrevivência usando os pesos Kaplan-Meier.....	48
Tabela 12 - Dados de sobreviventes e mortes em dois grupos para cada tempo <i>t_i</i>	53
Tabela 13 - Teste log-rank.....	56
Tabela 14 - Descrição das variáveis envolvidas no estudo	76
Tabela 15 - Estatística descritiva para as variáveis qualitativas.....	84
Tabela 16 - Estatística descritiva para as variáveis quantitativas.....	87
Tabela 17 - Estatística descritiva para variável localização topográfica.....	88
Tabela 18 - Estrutura do código morfológico.....	89
Tabela 19 - 5º dígito - Código de comportamento biológico das neoplasias.	90
Tabela 20 - Estatística descritiva para a variável Comportamento do cancro	90
Tabela 21 - 6º dígito do Código para graduação e diferenciação histológicas.....	91
Tabela 22 - Estatística descritiva para variável Grau do cancro.....	91
Tabela 23 - Frequências absolutas para o cruzamento das três variáveis	93
Tabela 24 - Frequências absolutas para o cruzamento das duas variáveis	94
Tabela 25 - Frequência absoluta para resultado de rastreio.	95
Tabela 26 - Frequência absoluta e relativa para os resultados de rastreio e número de óbitos..	96
Tabela 27 - Resultado de teste de associação de Qui-quadrado	98

Tabela 28 - Risco relativo para associação das duas variáveis.....	98
Tabela 29 - Frequências absoluta e relativa.....	101
Tabela 30 - Frequência absoluta para o cruzamento da variável rastreio, faixa etária e óbito ..	101
Tabela 31 - Estimativa da função de sobrevivência global de Kaplan-Meier para o CCU.....	102
Tabela 32 - Estimativa de sobrevivência global a 1, 3 e 5 anos após o diagnóstico do CCU (n=324).....	103
Tabela 33 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por rastreio	104
Tabela 34 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por faixa etária ...	105
Tabela 35- Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por faixa etária e rastreio.....	106
Tabela 36 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por estadiamento do cancro.....	107
Tabela 37 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por distrito de residência.....	107
Tabela 38 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por comportamento do cancro	108
Tabela 39 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por estado civil...	109
Tabela 40 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico variável Fumador	110
Tabela 41 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por infeção por HPV	111
Tabela 42 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por antecedentes familiares	111
Tabela 43 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por diabetes.....	112
Tabela 44 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por obesidade	112
Tabela 45 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por contraceção .	113
Tabela 46 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por nível de escolaridade.....	114
Tabela 47 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por sintomas.....	114
Tabela 48 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por ano de diagnóstico.....	115
Tabela 49 - Teste de log rank para a variável rastreio do CCU	117
Tabela 50 - Teste log-rank.....	118

Tabela 51 - Modelo de regressão de Cox simples	119
Tabela 52 - Análise de modelo de regressão de Cox simples	120
Tabela 53 - Estimativa dos parâmetros do modelo final	123
Tabela 54 - Proporcionalidade global do risco.....	125

LISTA DE SIGLAS E ACRÓNIMOS

ADN - Ácido Desoxirribonucleico

AGC - Células Glandulares Atípicas

AIC - Critério de Informação de Akaike

AIS - Adenocarcinoma in situ

ARSC - Administração Regional do Saúde do Centro, IP

ASC - Células Escamosas Atípicas

ASC-H - Células Escamosas Atípicas não podendo ser excluída HSIL

ASC-US - Células Escamosas Atípicas de Significado Indeterminado

CCU - Cancro do Colo do Útero

CID-0 - Classificação Internacional de Doenças para Oncologia

CIN I, II, III - Neoplasia Intraepitelial Cervical I, II, III

DGS - Direção Geral de Saúde

DSP - Departamento de Saúde Pública

DST - Doença Sexualmente Transmissíveis

$f(t)$ - Função Densidade/Massa de Probabilidade

$F(t)$ - Função Distribuição

FIGO - Federação Internacional de Genecologia e Obstetrícia

FUTURE - Females United to Unilaterally Reduce Endo/ectocervical disease

$h(t)$ - Função de Risco

$H(t)$ - Função de Risco Acumulativa

HPV - Vírus da Papiloma Humano

HSIL - Lesão Pavimentosa Intraepitelial de Alto Grau

INE - Instituto Nacional de Estatística

IPO - Instituto Português de Oncologia

LSIL - Lesão Pavimentosa Intraepitelial de Baixo Grau

NILM - Negativo para Lesão Intraepitelial ou Neoplasia Maligna

PEPAC - Programa de Estágio Profissional na Administração Central

PNV - Programa Nacional de Vacinação

RORC - Registo Oncológico Regional do Centro

RR - Risco Relativo

S(t) - Função de Sobrevivência

SPG - Sociedade Portuguesa Ginecologia

SPGO - Sociedade Portuguesa Ginecologia Oncológica

UIPS - Unidade de Investigação e Planeamento em Saúde

VIH - Vírus da Imunodeficiência Humana

CAPÍTULO 1

INTRODUÇÃO

Este projeto realiza-se no âmbito curricular do 2º ciclo de estudos de Mestrado em Estatística e relata o estágio efetuado na ARSC em Coimbra, na UIPS, em parceria com o Departamento de Matemática e Aplicações da Universidade do Minho.

1.1. OBJETIVOS DO ESTÁGIO

A ARSC propôs como tema da tese “Análise de sobrevivência das Doentes com Cancro do Colo do Útero”, com o objetivo de chamar a atenção das mulheres para a necessidade de se submeterem ao rastreio do cancro do colo do útero como forma de prevenção do CCU.

Os objetivos definidos foram os seguintes:

- ✓ Estimar a sobrevivência por cancro do colo do útero na Região Centro;
- ✓ Verificar se as mulheres submetidas ao rastreio do cancro do colo do útero têm uma probabilidade de sobrevivência mais elevada do que as mulheres não submetidas ao rastreio;
- ✓ Após o diagnóstico do CCU, identificar os fatores que podem influenciar a sobrevivência por cancro do colo do útero.

1.2. ESTRUTURA DA DISSERTAÇÃO

O presente projeto está dividido em seis capítulos. No primeiro capítulo, é feita a introdução, onde se enunciam os objetivos da dissertação, apresenta-se a organização da mesma e o enquadramento do tema do projeto.

No capítulo seguinte, apresentam-se os fundamentos teóricos da técnica de Análise de Sobrevivência.

No Capítulo 3, é apresentada a metodologia do estudo, onde é definido o tipo de estudo e as descrições das variáveis envolvidas no referido estudo.

O Capítulo 4, destina-se à apresentação dos resultados do estudo desenvolvido, procurando-se destacar as informações mais relevante retirados na realização deste projeto.

No Capítulo 5, é feita a discussão dos resultados obtidos, comparando os resultados obtidos com outros estudos existentes. Por fim, no Capítulo 6, são feitas as conclusões, apresentação das limitações do estudo e algumas recomendações para o referido estudo.

1.3. ENQUADRAMENTO TEÓRICO DO TEMA

Nesta seção é feita uma revisão bibliográfica do CCU no mundo e em Portugal, apresentando-se de acordo com bibliografia existentes os fatores de risco associado ao CCU, os principais sintomas, a forma de prevenção, a importância do rastreio do CCU na deteção precoce do CCU e os tipos de tratamento utilizados em Portugal para o CCU.

1.3.1. MAGNITUDE DO CANCRO NO MUNDO

Segundo a revista JAMA Oncology, entre 1990 e 2013, a proporção de mortes causadas por cancro em todo o mundo aumentou de 12% para 15%. Durante esse período, os anos de vida saudável perdidos para o cancro aumentaram 29%, de acordo com dados sobre 28 tipos de cancro em 188 países. No total, ocorreram 14,9 milhões de novos casos do cancro, 8,2 milhões de mortes e 196300 mil anos de uma vida saudável perdidos em 2013, de acordo com os pesquisadores. Entre o nascimento e 79 anos, um em cada três homens e uma em cada cinco mulheres desenvolveram o cancro (JAMA Oncology, 2015).

A principal causa de morte por cancro em 2013 foi o cancro da traqueia, o cancro dos brônquios e o cancro de pulmão, que causaram 1,6 milhões de mortes. O cancro da mama foi a principal causa de perda de anos de vida saudável entre as mulheres e, para os homens, foi o cancro de pulmão, de acordo com a revista JAMA Oncology.

Segundo os autores da mesma revista, o cancro foi a segunda causa de morte em 2013 perdendo apenas para as doenças cardiovasculares e a “epidemia” do cancro como parte das doenças não transmissíveis, deverá aumentar em todos os países, devido ao crescimento da população, o envelhecimento e uma crescente prevalência de certos fatores de risco. Cerca de um terço das mortes por cancro são provocadas pelos fatores comportamentais e alimentares nomeadamente: alto índice de massa corporal; baixa ingestão de frutas e legumes; falta de atividade física, tabagismo e o uso de álcool. O consumo do tabaco é um fator de risco mais

importante para o desenvolvimento do cancro, o tabaco é responsável por cerca de 20% das mortes por cancro globais e é responsável por cerca de 70% das mortes no mundo de cancro do pulmão.

Mais de 60% do total de novos casos de cancro anuais do mundo ocorrem na África, Ásia e América Central e do Sul. Estas regiões são responsáveis por 70% das mortes por cancro no mundo. Espera-se que os casos de cancro anuais aumentarão de 14 milhões em 2012 para 22 milhões nas próximas 2 décadas (Stewart & Wild, 2014).

O Cancro do Colo do Útero é o 7º Cancro mais frequente a nível mundial e o 2º Cancro mais frequente na mulher, com uma taxa de mortalidade em Portugal de cerca de 4 óbitos por 100000 habitantes. De todos os tumores maligno o CCU é o que pode ser controlada com maior efetividade, com possibilidade de redução de cerca de 80% de incidência, através de programa de rastreio citológico cervical organizado, de base populacional com periodicidade de 3 ou 5 anos (“ARS Algarve, IP,” 2016).

1.3.2. MAGNITUDE DO CANCRO EM PORTUGAL

Em 2014, os tumores malignos foram a segunda causa básica de morte em Portugal, com registo de 26 220 óbitos, o que correspondeu a 24,9% da mortalidade (29,3% para os homens). A idade média ao óbito devido a estas causas, situou-se nos 72,7 anos, mais elevada para as mulheres (73,7 anos) do que para os homens (71,9). Em 2014, das mortes provocadas por tumores malignos, destacaram-se as ocorridas por tumor maligno da traqueia, brônquios e pulmão, com 3937 óbitos (3084 homens e 853 mulheres); tumor maligno do cólon, reto e ânus, 3808 óbitos (2200 homens e 1608 mulheres); tumor maligno do estômago, 2293 óbitos (1382 homens e 911 mulheres); tumor maligno da próstata, 1791 óbitos; tumor maligno da mama, 1686 óbitos, dos quais 1664 óbitos foram da mama feminina; tumor maligno do pâncreas, 1362 óbitos (742 homens e 620 mulheres) e tumor maligno da bexiga com 940 óbitos (690 homens e 250 mulheres) (Instituto Nacional de Estatística, 2016).

1.3.3. MAGNITUDE DO CANCRO DE COLO DE ÚTERO DO MUNDO

De acordo com o *jornal o país* publicado em 22 de julho de 2013, por ano são diagnosticados, cerca de 500000 mulheres com CCU em todo o mundo e 250000 mulheres morreram por causa desta doença neste período de tempo (o que equivalente a quase 685 mulheres por dia ou 30 por hora) (O País, 2013).

Na Europa existem cerca de 60 mil novos casos/ano de CCU, com uma mortalidade de 30 mil/ano. Em Portugal surgem, anualmente, cerca de 750 novos casos. Ao IPO de Lisboa é referenciada uma média de 250 doentes/ano com este tipo de cancro (Atlas da Saúde, 2015).

Em Portugal, o CCU é o segundo mais frequente no sexo feminino e o sexto mais mortal, e representa cerca de 10% de todos os tipos de cancro na mulher (Sociedade Portuguesa de Ginecologia Oncológica, 2007).

1.3.4. CANCRO DO COLO DO ÚTERO

O útero tem a dimensão e a forma de uma pêra média e mede cerca de 7,5 cm de comprimento e 5 cm de largura (Seeley, Stephens, & Tate, 2001). Este órgão é ligeiramente achatado no sentido antero-posterior e localiza-se na cavidade pélvica com a proporção mais arredondada e de maior diâmetro transversal, designado de fundo, orientado para cima e a proporção mais estreita, denominado de colo uterino, dirigido para baixo. A proporção de útero entre fundo e o colo uterino é designada corpo. Uma ligeira constrição ou istmo marca a junção do colo com o corpo. No interior do útero situa-se a cavidade uterina que se estende através do canal cervical e abre para vagina através do orifício externo do canal cervical.

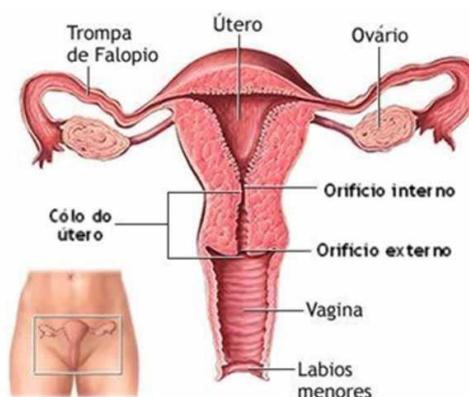


Figura 1 - Útero (adaptado de A.D.A.M. -Enciclopédia Multimédia).

O CCU desenvolve-se no colo de útero, que é a extremidade inferior do útero que liga o corpo do útero à vagina. Este tipo de cancro resulta do desenvolvimento anormal das células no colo de útero, que se multiplicam de forma descontrolada, originando assim lesões que por vezes evoluem para lesões cancerígenas (Figura 2). O CCU desenvolve-se de forma lenta, na sua fase inicial pode não causar sintomas, e que pode evoluir para quadros de sangramento vaginal de forma alternado ou após relação sexual, secreção vaginal anormal e dor abdominal associada com queixas urinárias ou intestinais, nos casos mais avançadas.

As células anormais evoluem de forma lenta, demorando muitos anos desde as fases iniciais até ao CCU (ECCA, 2016).

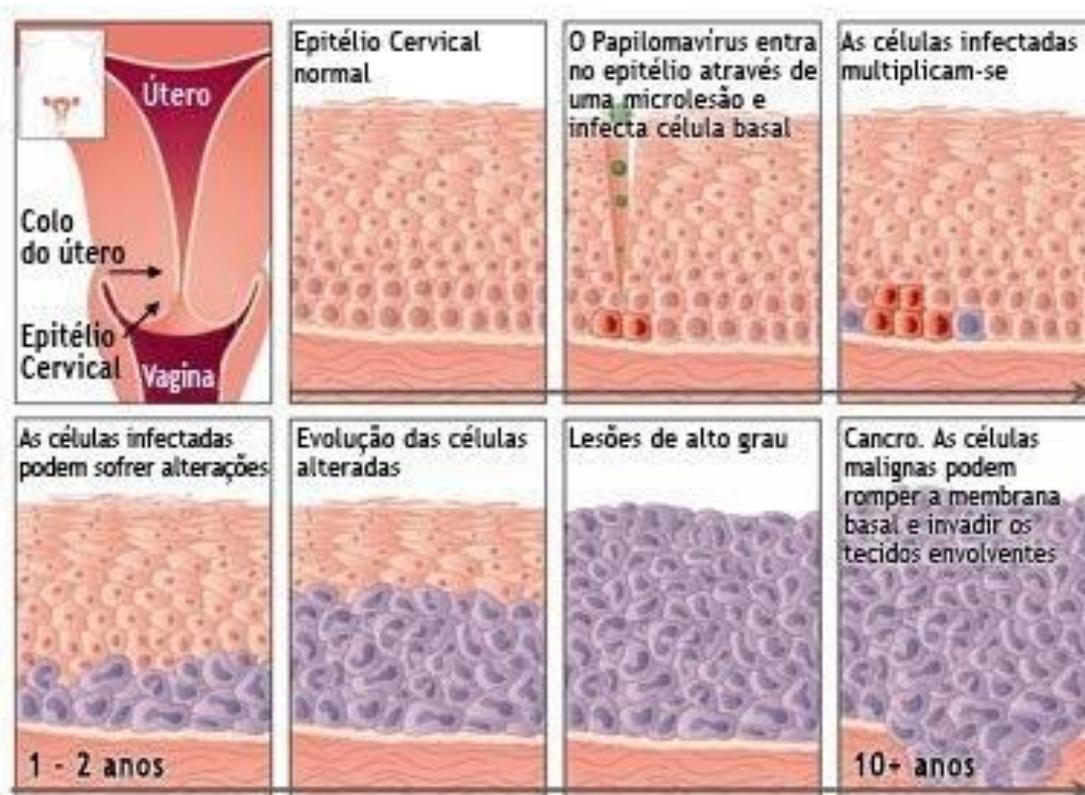


Figura 2 - Progressão do cancro do colo de útero (adaptado de Sanofi Pasteur MSD).

1.3.4.1. FATORES DE RISCO ASSOCIADO AO CANCRO DO COLO DO ÚTERO

Os médicos nem sempre conseguem explicar porque é que algumas mulheres desenvolvem o CCU e outras não. Porém, sabe-se que uma mulher com determinados fatores de risco está mais predisposta do que outras a desenvolver o CCU. Um fator de risco é algo que aumenta a possibilidade de se vir a desenvolver a doença (Info Cancro, 2016)

A etiologia do CCU não se encontra especificamente determinada. A descoberta mais importante nesta área é que esta patologia está geralmente associada a uma infeção por algumas estirpes do HPV (Alves, 2003; Castellsagué et al., 2006).

De acordo com a literatura existente destaca-se os seguintes fatores de risco:

- ✓ **Infeção por Vírus da Papiloma Humano**, o HPV pertence à família Papovaviridae, género Papiloma e espécie papilomavírus humanos.

O HPV (Figura 3) é um vírus de ADN de dupla cadeia capsulado. Apresenta estabilidade ambiental, resistindo à ação dos detergentes. Porém este vírus pode ser inativado por aquecimento a uma temperatura superior a 57°C e por ação de alguns desinfetantes.

O vírus do HPV é transmitido por contacto pessoal, podendo a transmissão ocorrer por via sexual, pelas mucosas e pelo contacto com a pele, infetando quer células internas, quer células externas do organismo. As manifestações clínicas de uma infeção por HPV podem ocorrer à superfície da pele, da boca, da língua, da garganta, das amígdalas, da vagina, do pénis, do colo do útero e do ânus (Sousa, 2011).



Figura 3 - HPV infetando as células epiteliais (Fonte: Pencubanexile.org-Health Medical Information)

O HPV é um vírus que infecta o epitélio. Em 2006, Trottier e Franco referiram a existência de mais de 120 estirpes de HPV diferentes, e que destas mais de 40 afetam o revestimento o epitelial do tracto anogenital (Trottiera & Franco, 2006).

As estirpes de HPV classificam-se de um modo geral, como tipo de baixo ou alto risco, de acordo com o seu potencial oncogénico e a sua associação ao CCU (Muñoz et al., 2003) (Tabela 1). Certas estirpes de HPV geralmente afetam o epitélio mucoso ou o cutâneo. As estirpes cutâneas provocam mais frequentemente verrugas na pele. As estirpes de HPV que afetam o epitélio

mucoso podem causar papilomas benignos da conjuntiva, orais nasais, da mucosa genital e da mucosa aerodigestiva, estando associados a malignidade (alto risco) (Widdice & Moscicki, 2008).

Tabela 1 - Classificação das estirpes de HPV segundo o seu potencial oncogénico (adaptado de Munoz et al., 2003)

Classificação de Risco	Tipo de Vírus
Alto Risco	Tipos 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, 82.
Provável Alto Risco	Tipos 26, 53 e 66.
Baixo Risco	Tipos 6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81 e CP108.
Risco Indeterminado	Tipos 34, 57 e 83

Apesar das estirpes de alto risco carcinogénico variarem entre as 10 e as 15, a sua contribuição na disseminação do CCU não é, de todo, uniforme sendo que as estirpes 16 e 18 são as que se encontram presentes em cerca de 70% dos casos e, mais concretamente, a estirpe 16 surge em cerca de 50% dos casos (Kahn, 2009).

Tabela 2 - Percentagem de lesões atribuídas aos tipos específicos de HPV (adaptado de Munoz et al., 2003)

Tipo de HPV	Tipo de Lesão	Percentagem
HPV 16, 18	Cancro Cervical	70%
	Lesão de Alto Grau Cervical	50%
	Lesão de Baixo Grau Cervical	30%
	Cancro Anal	~70%
	Cancro Vulva/Vagina/Pénis	~40%
	Cancro de Cabeça e Pescoço	~3-12%
HPV 6, 11	Lesão de Baixo Grau Cervical	5-25%
	Condilomas Genitais	90%
	Papilomatose Respiratória Recorrente	90%

No geral, uma infeção por HPV não conduz ao desenvolvimento do cancro, porém os estudos comprovam que 99% das mulheres com CCU estão infetadas por estirpes de alto risco de HPV (Sousa, 2011).

Importante salientar que a infeção por HPV por si só não leva ao desenvolvimento do CCU, a infeção por HPV é um fator necessário, mas não suficiente para o desenvolvimento do CCU. Assim, a persistência da infeção por genótipos oncogénicos de HPV, associados a cofatores, pode, em algumas mulheres, levar ao desenvolvimento do CCU. O CCU desenvolve-se lenta e progressivamente, podendo levar 20 anos ou mais desde a infeção persistente até ao desenvolvimento de lesões de alto grau e do cancro invasivo (Direcção Geral de Saúde, 2008). A idade de maior incidência da doença é entre os 45 e os 55 anos.

A par de correlação entre a infeção por HPV e o desenvolvimento do carcinoma do colo uterino, é possível estabelecer uma relação com outros cofatores de risco, nomeadamente (CDC, 2009):

- ✓ Início de vida sexual ativa antes dos 20 anos;
- ✓ Múltiplos parceiros sexuais;
- ✓ Parceiro sexual com historial de DST;
- ✓ Parceiro sexual com multiplicidade de contactos;
- ✓ Historial clínico com DST, especialmente virais;
- ✓ Uso de anticoncecionais hormonais associados a hábito tabágicos e excesso de peso;
- ✓ Imunossupressão

Dos principais fatores de risco ligado a infeção por HPV destaca-se os fatores, de natureza sexual, ou seja, o comportamento sexual tem um forte impacto na infeção causada pelo HPV (CDC, 2009).

A idade, por si só, não é considerada um fator de risco para CCU. Porém, a idade em que se ocorre primeira relação sexual é considerado como um dos principais fatores de risco, principalmente se essa relação tiver ocorrido numa idade precoce (Wee, Huang, Huskey, & McCarthy, 2008).

O risco das mulheres contraírem a infeção por HPV aumenta face ao número de parceiros sexuais, história sexual dos parceiros, frequência dos relações sexuais, anteriores infeções vaginais pelo vírus de herpes simples e pela presença de verrugas genitais (Moscicki et al., 2001).

No desenvolvimento do CCU, existem fatores de risco adicionais, tais como, os hábitos tabágicos, o uso de anticoncecionais orais, a exposição anterior a outras DST's, na qual se inclui a Chlamydia, a obesidade, entre outros fatores correlativos (Castellsagué et al., 2006; J. T. Cox, 1995).

Todos os indivíduos imunodeprimidos, quer pelo vírus da imunodeficiência humana (VIH), quer por imunossupressão pós transplante de órgão tem um risco potencial acrescido de infeção por HPV e conseqüentemente terá um maior risco de desenvolver doenças associadas a este vírus (Palefsky & Holly, 2003).

Outros fatores do risco no desenvolvimento de CCU que importa salientar são: a paridade, a idade da primeira gravidez e os fatores nutricionais.

Quanto a idade da primeira gravidez e paridade, vários autores referem que estes serão fatores de risco se a mulher adotar na sua vida sexual um comportamento de promiscuidade, de outra forma não parece ser possível estabelecer qualquer relação (Alves, 2003).

Os fatores nutricionais, segundo o Instituto Nacional do Cancro Americano, apresentam-se como fatores correlativos do risco de desenvolvimento do CCU. Associa-se uma diminuição da persistência do HPV a níveis mais elevados de consumo de vegetais e frutas, sendo vários os estudos a demonstrar que níveis mais baixos de antioxidantes, aliados a baixos níveis de ácido fólico aumentam o risco de neoplasia intraepitelial cervical (CIN). Baixos níveis de alfa-tocoferol foram encontrados em doentes com teste de HPV positivos com neoplasia intraepitelial cervical. Os estudos demonstraram também existir uma redução de 56% do risco de persistência de infeção por HPV nas mulheres com maiores concentrações de licopeno ao nível do plasma, pelo que se sugere o consumo de vegetais com licopeno (Sousa, 2011).

1.3.4.2 SINTOMATOLOGIA DO CANCRO DO COLO DO ÚTERO

Geralmente, os sintomas e os sinais associados à neoplasia do colo do útero dependem do estado da doença. Segundo Andrade JM et al., 2001, as lesões pré-neoplásicas e as micro-invasoras são assintomáticas. A maioria das pessoas infetadas por HPV não apresentam quaisquer sintomas e esta infeção será ultrapassada sem recursos a tratamentos terapêuticos. Outro dado relevante é que mais de 90% da sintomatologia detetada desaparece no prazo de 2 anos (Cutts et al., 2007).

A hemorragia vaginal anormal é o sintoma mais comum na apresentação do cancro invasivo do colo de útero. Nas mulheres sexualmente ativas, as coitorragias são os sintomas mais habituais, sendo também frequentes as hemorragias entre as menstruações e no período pós-menopáusico. Nos casos mais avançados podem surgir dores abdominais, sintomas urinários, digestivos, respiratórios ou outros (Sociedade Portuguesa de Ginecologia Oncológica, 2007).

É importante salientar que apesar da maioria das infeções por HPV serem transitórias, não apresentando geralmente sintomas, a infeção persistente é a principal causa do CCU, bem como

de outros cancros anogenitais. Desta forma, justifica-se a necessidade de um rastreamento populacional cada vez mais completo e abrangente, centrando-se as preocupações na melhoria do diagnóstico (Wolschick, Consolaro, Suzuki, & Boer, 2007).

Contudo, não devem ser menosprezados os sintomas, quer físicos quer psicológicos, que estas lesões provocam nos seus portadores, nomeadamente, quanto à sintomatologia física, o prurido, a sensação de ardência, o leve corrimento sanguíneo nas relações sexuais. Ao nível psicológico, os principais sintomas estão relacionados com a disfunção sexual, o sentimento de vergonha e culpa, a rejeição por parte dos parceiros, a perda do desejo e do prazer sexual (Sousa, 2011).

Deve ser referido, que a partir do exame citológico das células cervicais é possível detetar a fase de pré-malignidade do CCU. Estas alterações de pré-malignidade (Figura 4) podem ser representadas num espectro de anormalidades histológicas que vão desde a displasia mínima (CIN 1), a displasia moderada (CIN 2), a displasia severa/carcinoma in situ (CIN 3) (Schiffman & Castle, 2003).

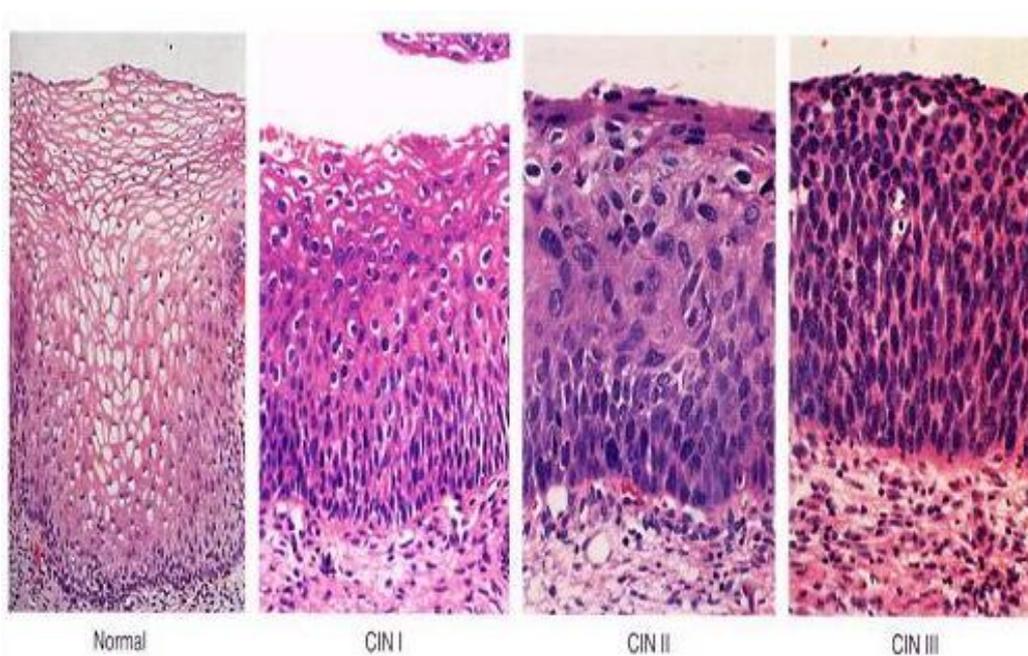


Figura 4 - Fases de pré-malignidade do cancro cervical (fonte: Emory University School of Medicine)

1.3.4.3 IMPORTÂNCIA DO RASTREIO NA DETEÇÃO PRECOCE DO CANCRO DO COLO DO ÚTERO

Segundo a publicação do dia 03-06-2015 no site de Direção Geral de Saúde (DGS,) a DGS destacou a importância do rastreio atempado do CCU.

A DGS reforçou, a propósito do dia de saúde feminina, assinalado a 28 de maio, a importância de realizar rastreios precoces ao CCU.

Segundo a DGS, este tipo de cancro, provocado por uma infeção persistente do vírus de HPV, é o segundo tumor ginecológico maligno mais frequente nas mulheres portuguesas com menos de 50 anos. A DGS lembra que o desenvolvimento do CCU é silencioso, pelo que não se deve esperar pelos sinais de alarme (corrimento vaginal anormal, dor e perdas de sangue durante ou após a relação sexual).

O rastreio é de extrema importância, uma vez que a deteção precoce permite, muitas vezes, evitar ou retardar o avanço do CCU, ainda que esta patologia seja das mais comuns entre as mulheres (Baseman & Koutsky, 2005).

O teste de Papanicolau, desde da sua implementação em 1940, tem sido o método de eleição para o rastreio citológico do colo uterino (Alves, 2003).

Existem vários estudos que apontam existir uma associação negativa entre o rastreio e a incidência da infeção invasora por HPV, ou seja, quantas mais pessoas forem rastreadas, menor será a incidência de infeção invasora por HPV. Este facto permite acreditar no efeito protetor do rastreamento da população (Andrade JM et al., 2001).

Nos países desenvolvidos, a percentagem de deteção de todos os cancros cervicais situa-se nos 80% (Baseman & Koutsky, 2005), o que demonstra a eficácia dos programas do rastreio nestes países. Os países que conseguiram implementar um programa de rastreio eficaz, reduziram a incidência de CCU em menos de 10 casos por cada 100000 mulheres/ano (Zeferino, 2008).

O rastreio citológico da população feminina a cada 3/5 anos pode reduzir a incidência do CCU em mais de 80%. Porém, para que tal aconteça é necessário que todas as etapas do programa do rastreio sejam implementadas, nomeadamente ao nível de informações prestadas, da convocatória das faixas da população feminina a serem rastreadas, no seguimento clínico dos

casos que apresentam resultados com anormalidade, no tratamento dos casos instalados (Arbyn et al., 2010). Em Inglaterra verificou-se que o aumento da abrangência do rastreio levou a redução da mortalidade por CCU. Este aumento de abrangência foi conseguido devido à introdução de uma convocatória periódica das mulheres para realizar o exame citológico, ou seja, uma melhor organização do sistema de saúde potencia ganhos ao nível da mortalidade em doenças passivas a serem rastreadas (Zeferino, 2008).

Em Portugal, apesar das recomendações oficiais e da comunidade científicas, o rastreio citológico da mulher ainda continua a ser oportunístico, o que poderá ser apontado como a causa dos elevados números de novos casos do CCU que surgem a cada ano (Direcção Geral de Saúde, 2008).

De acordo com vários estudos existentes, existe maior incidência do CCU nas mulheres nas faixas etária dos 40 anos (Naucler et al., 2007). Na opinião de vários investigadores, as mulheres com idade compreendidas entre os 30 e 40 anos deveriam ser submetidas à citologia convencional, vulgarmente designado por teste de Papanicolau e ao teste de deteção de infeção por HPV, de acordo com Naucler et al., (2007).

1.3.4.4 TÉCNICA DO RASTREIO

As técnicas de rastreio são atualmente, um dos componentes chave nos programas de prevenção do CCU. O diagnóstico precoce constitui ainda a ferramenta mais eficaz no combate a essa patologia.

1.3.4.4.1 CITOLOGIA CERVICAL

O exame citológico ou citologia cervical é recomendando como o método mais correto no rastreio de grande número de pessoas, apresentando uma sensibilidade na ordem dos 75% (Sousa, 2011).

Esta técnica de rastreio pode realizar-se segundo dois métodos:

- ✓ Em lâmina -Citologia ou teste convencional de Papanicolau;
- ✓ Em meio líquido.

Segundo o *Programa Harvard Medical School Portugal*, o teste convencional de Papanicolau é realizado pelo médico durante o exame ginecológico. A mulher deverá deitar-se de costas na marquesa com as pernas afastadas, com os pés apoiados em suportes próprios e com os joelhos fletidos. O médico irá inserir um instrumento denominado de espéculo na vagina para manter as paredes desta afastada e permitir a observação do colo do útero e a realização do teste Papanicolau.

O teste convencional de Papanicolau consiste na recolha de células esfoliadas do colo do útero e posterior preparação de esfregaços dessas células numa lâmina por parte do médico. A preparação da lâmina pode ser limitada pela sobreposição das células, colheitas insuficientes das células ou artefactos de fundo de células inflamatórias ou sangue (Widdice & Moscicki, 2008).

Depois de observar o colo do útero e a vagina para verificar se existem alguns problemas visíveis, o médico irá raspar suavemente a superfície do colo do útero com uma pequena espátula para obter uma amostra de células da parte de fora do colo do útero e usar uma pequena escova para colher células do interior do colo do útero (em determinados locais, são utilizados os escovilhões que permitem fazer ambas as colheitas em simultâneo). Tal como foi referido anteriormente, a citologia em lâmina é o teste de referência para o rastreamento de grandes quantidades populacionais (Wolschick et al., 2007). Recomenda-se a colheita com a escova Cervex-Brush (Figura 5) em lâmina única, não se recomenda colheitas ao fundo do saco posterior (Brenna, Hardy, Zeferino, & Namura, 2001).

A recolha de células com esta escova resulta numa amostra pura e precisa o que se traduz numa elevada fiabilidade dos esfregaços realizados sendo deste modo, o uso desta bastante eficaz na deteção precoce do CCU (CervexBrush, 2016).

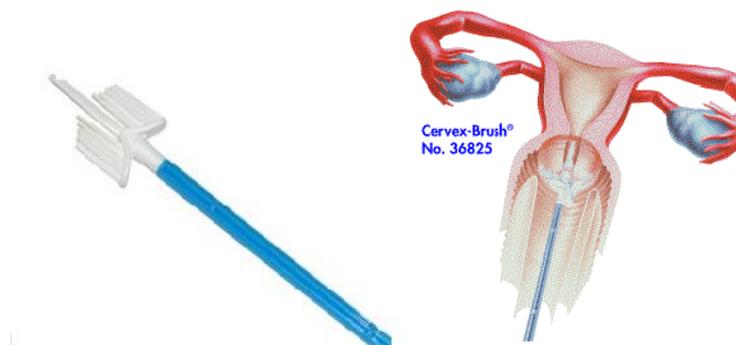


Figura 5 - Cervex -Brush (fonte: CervexBrush.com).

No teste em meio líquido, o médico coloca a amostra das células do colo do útero diretamente num líquido fixador e a preparação da lâmina ocorre no laboratório. Esta técnica facilita a leitura, permite a repetição de esfregaços da mesma amostra e a realização de outras técnicas de diagnóstico (por exemplo, o teste de HPV).

De acordo com vários estudos existentes, o teste em meio líquido apresenta uma taxa inferior de falsos negativos quando comparada com o teste convencional de Papanicolau (Widdice & Moscicki, 2008), no entanto apresenta um custo mais elevado do que a citologia convencional e é mais difícil de se realizar (Direcção Geral de Saúde, 2008).

É importante salientar que, os erros da citologia cervical ocorrem frequentemente no momento da colheita, levando a apresentação de falsos resultados e à consequente falência do rastreio. A colheita pode ser inadequada por várias razões (Brenna et al., 2001):

- ✓ Má execução do esfregaço e/ou fixação;
- ✓ Não ter sido feito diretamente no exo endocolo;
- ✓ Colheita insuficiente por característica própria da lesão nomeadamente:
 - Lesões pequenas que descamam poucas células;
 - Lesões que apresentam queratinização à superfície;
 - Lesões localizadas na parte alta do canal ou periferia do colo;
 - Lesões afastadas do orifício cervical externo.

A má interpretação dos resultados dos exames citológicos também contribuiu para a percentagem de falsos negativos associados a esta técnica.

1.3.4.5. INTERPRETAÇÃO DOS RESULTADOS CITOLÓGICO- CLASSIFICAÇÃO DE BETHESDA

A citologia cervical ou teste convencional de Papanicolau possui uma nomenclatura própria para a redação de relatórios citológicos aceite em todo o mundo, designada de “sistema de Bethesda” (Tabela 3). Esta classificação resultou da necessidade em uniformizar os resultados obtidos nas avaliações citológicas convencionais, a par da procura da otimização e da utilização de dados recolhidos num plano estatístico. Estas necessidades levaram a que diferentes peritos se reunissem, em Bethesda (EUA) em 1998, e criassem esta classificação, que sofreu uma revisão em 2001 (Solomon & Nayar, 2005).

De acordo com a fonte do Protocolo do rastreio do CCU, os resultados citológicos são classificados de acordo com a classificação de Bethesda. Os procedimentos recomendados no âmbito do rastreio de acordo com o resultado citológico, encontra-se apresentado na (Tabela 3).

Tabela 3 - Resultados possíveis de uma citologia e procedimento a seguir

Resultado citológico	Procedimento
Não satisfatória para avaliação	Repetir a citologia logo que possível
Negativo para lesão intraepitelial ou neoplasia maligna (NILM)	Convidar o utente a repetir a citologia depois de 3 anos (registo disponível no início do ano civil)
Negativa para lesão intraepitelial ou neoplasia maligna, sem representação da zona de transformação	Repetir a citologia após 6 meses
Atipia de significado indeterminado do epitélio escamoso (ASC-US)	1º ASC-US: repetir a citologia após 6 meses 2º ASC-US consecutivo: referência à unidade de patologia cervical (UPC)
Atipia de significado indeterminado do epitélio escamoso (ASC-US) (não inclui lesão de alto grau) (ASC-H)	Referência à UPC
Atipia do epitélio glandular	
Lesão de baixo grau (AGC)	
Lesão de baixo grau (LSIL)	
Lesão de alto grau n (HSIL)	
Carcinoma Espinho-celular	
Adenocarcinoma in situ (AIS)	
Adenocarcinoma ou outra neoplasia maligna	
Sem a presença de lesão intraepitelial, mas com presença de células endometriais pós-menopausa.	Referência para consulta ginecologia
Exceção: sempre que esteja registado no relatório do anatomo-patologista: <ul style="list-style-type: none"> • “Repetir após terapêutica, no rastreio” ou • “Repetir após terapêutica com estrogénios, no rastreio”, 	Repetir após terapêutica; os registos das utentes ficam imediatamente disponível no SiiMA Rastreios para registo de nova citologia.

A classificação de Bethesda permite que, de acordo com os resultados obtidos da citologia fosse adaptado um protocolo de procedimento de realização de exames complementares e/ou tratamentos (Tabela 4).

De seguida apresenta a correspondência entre diferentes sistemas de classificação para as alterações citológicas cervicais.

Tabela 4 - Outras classificações para alterações citológicas cervicais.

Displasia	Neoplasia Cervical intraepitelial (CIN)	Sistema de classificação de Bethesda, 2001
Atipia	Atipia	Alterações celulares de significado indeterminado do epitélio pavimentoso (ASC-US e ASC-H) e glandular (AGUS)
Alterações relacionados com HPV Displasia ligeira	Alterações relacionados com HPV CIN I	Lesão intraepitelial de baixo grau (LSIL) do epitélio pavimentoso
Displasia Moderada	CINII	Lesão intraepitelial de alto grau (HSIL) do epitélio pavimentoso
Displasia Grave/carcinoma in situ	CINIII	

É importante salientar que, sempre que exista suspeita de neoplasia micro-invasora, ou face a resultados anormais da citologia, a Sociedade Portuguesa de Genecologia (SPG) recomenda a colposcopia.

Esta irá permitir, antes de mais direccionar a biópsia e, sempre que essa não consegue definir com precisão a existência ou não de invasão, a colposcopia é o procedimento protocolar recomendado (Sociedade Portuguesa de Ginecologia Oncológica, 2007).

A colposcopia é o estudo de imagens da camada epitelial superficial do colo do útero, utilizando para tal um colposcópio, o qual permite a visualização desta área com um aumento de 40 e 60 vezes. Esta técnica é utilizada especialmente em mulheres que obtiveram um exame citológico anormal. Desta forma toda a área da zona de transformação deverá ser visualizada, observando-se possíveis alterações da vagina, colo do útero e pélvis (Poli & Silveira, 1994).

Convém salientar que, entre as várias organizações mundiais, não há consenso universal a cerca da periodicidade da implementação de programas de rastreios, sendo a regularidade do mesmo variável, de acordo com os fatores de risco para o desenvolvimento do CCU. Porém, há uma certa concordância em relação ao início do rastreio, que deve coincidir com a idade/início da atividade sexual (Alves, 2003).

De uma forma geral, recomenda-se a realização anual dos esfregaços cervicais em mulheres com multiplicidades de parceiros, portadoras de VIH e imunocomprometidas, enquanto que, para as mulheres sexualmente não ativas ou com relações monogâmicas estáveis consideram suficiente a despistagem de 3 em 3 anos (Arbyn et al., 2010).

Segundo Andrade JM et al., (2001) o intervalo entre citologia deverá variar entre um e três anos dependendo da presença ou não de fatores de risco, como sendo o início precoce da atividade sexual, a multiplicidade de parceiros e o baixo nível socioeconómico. Caso algum destes fatores esteja presentes o rastreio deve ser realizado anualmente.

1.3.4.6 MÉTODOS DE TRATAMENTO DE CANCRO DO COLO DO ÚTERO

Os métodos de tratamento recomendado para o CCU envolvem terapia multidisciplinares, como a radioterapia, quimioterapia e a cirurgia. A implementação destes métodos de tratamento deve ter sempre em consideração diversos fatores, nomeadamente a idade da paciente, a vontade de procriar, o estado geral de saúde, a resposta imunológica, as características do tumor, e sobretudo as preferências do doente (Burd, 2003; Wolschick et al., 2007).

A histerectomia, tratamento cirúrgico foi a técnica mais utilizada até um passado recente, porém com o desenvolvimento da investigação acerca da doença, bem como o aparecimento de métodos e técnicas mais conservadoras, têm restringindo esta técnica cirúrgica a relativamente a poucas circunstâncias, nomeadamente quando existe a patologia uterina associada, como mioma ou prolapso (Wolschick et al., 2007).

Segundo a SPG, o tratamento primário do carcinoma do colo do útero nos estados iniciais deve ser feito quer por cirurgia quer por radioterapia. A quimioterapia surge como coadjuvante em estado mais avançado do carcinoma do colo do útero, em que há extensões para outras paredes. No intuito do uso correto do procedimento de tratamento é necessário proceder ao correto estadiamento da doença, que no caso específico do CCU assenta fundamentalmente na avaliação clínica (Sociedade Portuguesa de Ginecologia Oncológica, 2007).

Em Portugal, utilizam-se os seguintes critérios de classificação de estadiamento do CCU, apresentado na Tabela 1 (adaptada pela Federação Internacional de Ginecologia e Obstetrícia (FIGO, 2016)).

Tabela 5 - Estadiamento da FIGO (2009) para o CCU

Estádio	Característica clínica
I	Carcinoma limitado ao colo
IA	Carcinoma invasivo que apenas é diagnosticado por microscópia, com invasão profunda ≤ 5 mm e extensão ≥ 7 mm
IA1	Invasão do estroma ≤ 3 mm em profundidade e ≤ 7 mm em extensão
IA2	Invasão do estroma > 3 mm e ≤ 5 mm em profundidade e ≤ 7 mm em extensão
IB	Lesões invasivas, clinicamente visíveis, mas limitadas ao colo do útero ou doença microscópica com dimensões superiores ao estado IA
IB1	Carcinoma clinicamente visível com ≤ 4 cm
IB2	Carcinoma clinicamente visível com > 4 cm
II	Carcinoma com invasão para além do útero, mas não atingindo a parede pélvica ou o 1/3 inferior da vagina
IIA	O tumor envolve até os 2/3 superiores da vagina, sem infiltração óbvia do paramétrio
IIA1	Carcinoma clinicamente visível com ≤ 4 cm, sem envolvimento do paramétrio
IIA2	Carcinoma clinicamente visível com > 4 cm, sem envolvimento do paramétrio
IIB	Carcinoma infiltra o paramétrio, sem atingir a parede pélvica
III	Carcinoma estende-se à parede pélvica, e/ou invade 1/3 inferior da vagina, e/ou hidronefrose ou rim não funcionante, mesmo por outra causa
IIIA	Invasão de 1/3 inferior da vagina, sem atingir a parede pélvica
IIIB	Extensão a parede pélvica, e/ou hidronefrose, e/ou rim não funcionante
IV	Carcinoma estende-se para além da pelve, e/ou invade a mucosa da bexiga e/ou do reto
IVA	O tumor invade a mucosa da bexiga e/ou do reto
IVB	Metástases à distancia, incluindo metástases ganglionares inguinais e lomboaórticos

O tratamento a adaptar depende do estadiamento da doença, para as lesões de alto grau pré-neoplásica e micro-invasoras, a conização é a técnica mais usualmente utilizada.

Nos estádios de carcinoma invasor não localmente avançado, o tratamento cirúrgico exclusivo é o tratamento mais adequado, existem vários estudos a demonstrar que a taxa de sobrevivência é maior para os pacientes com tratamento cirúrgico do que para os pacientes com tratamento de radioterapia (Andrade JM et al., 2001).

Em Portugal, de acordo com o estadiamento da doença recorre-se aos tratamentos representados na Tabela 6.

Tabela 6 - Tratamento realizado em Portugal nos diferentes estádios do cancro do colo do útero (adaptado SPG,2007).

Estádio	Tratamento
Estádio IA	Conização
Estádio IB1 e IIA≤4cm	Histerectomia radical
	Linfadenectomia pélvica
	Exérese gânglios
Estádio IB2 e IIA> 4cm	Histerectomia radical
	Linfadenectomia pélvica
Estádio IIB, III e IVA	Quimiorradioterapia
	Exéreses gânglios volumosos
	Tratamento paliativo individualizado
Estádio IVB	Quimiorradioterapia
	Braquiterapia vaginal
	Quimiorradioterapia pélvica e lomboaórtica

1.3.4.7 VACINAÇÃO PARA O VÍRUS DE HPV E O SEU PAPEL NO COMBATE AO CANCRO DO COLO DO ÚTERO

O CCU, ao ver estabelecida na sua génese uma relação causal entre a infeção persistente por HPV de alto risco e o aparecimento do cancro propriamente dito, tornou-se no primeiro cancro a ter como causa necessária, mas não suficiente, um agente infeccioso. Este fator contribuiu para que a investigação científica ao nível médico e farmacêutica se fosse na procura de vacinas com efeito profiláticos, dirigidas contra as infeções por HPV 's de alto risco 16 e 18 responsáveis por cerca de 70-75% dos carcinomas cervicais (Sociedade Portuguesa de Ginecologia Oncológica, 2007).

A partir do momento em que foi possível estabelecer correlação entre a infeção persistente do HPV de alto risco e o CCU, tornou-se óbvio que o campo de ação no combate a este tipo de cancro passava pela proteção relativamente aos agentes infecciosos causadores da sua origem.

O objetivo principal da vacinação estende-se ao curto, médio e longo prazo.

Desta forma, como o objetivo a curto prazo pretende-se que a vacinação contra o HPV produz uma redução significativa dos resultados citológicos cervicais anormais. De entre estes resultados inclui-se as atipias incertas escamosas e glandulares (ASG-US e AGC), as lesões intraepiteliais de baixo grau (LSIL), que na maioria das vezes apresenta uma resposta citohistológica aguda à presença do vírus, mas que é transitória na maioria dos casos (Wolschick et al., 2007). Associa-se a esse objetivo, no caso de vacinação quadrivalente, um outro que é a redução, no curto prazo, de verrugas genitais e da papilomatose respiratória recorrente causadas pelos tipos 6 e 11 do HPV (SPG, 2010).

Para o objetivo a médio prazo estabeleceu-se a prevenção das neoplasias precursora do cancro cervical (CIN) em especial a lesão CIN III e Adenocarcinoma in situ (AIS) (Stanley, 2008).

A longo prazo o principal objetivo da vacinação consiste em prevenir o CCU, tendo associados os objetivos que pretendem a prevenção de outros cancros correlacionados com infeção persistente por HPV (Kahn, 2009).

Segundo a publicação da DGS do dia 29-09-2014, as vacinas contra infeções por vírus do HPV foram comercializadas em Portugal, 2006 e 2007. A vacina bivalente inclui os genótipos 16 e 18 e a quadrivalente inclui os genótipos 6, 11, 16, 18.

Estas vacinas estão indicadas para a prevenção de lesões pré-malignas do colo do útero, da vulva e da vagina e do CCU. A vacina quadrivalente está ainda indicada para a prevenção de verrugas genitais/condilomas.

Em outubro de 2008 a vacina quadrivalente foi introduzida no Programa Nacional de Vacinação (PNV) para jovens com 13 anos de idades, nascidas a partir de 1995. De 2009 a 2011 decorreu uma campanha de vacinação das raparigas com 17 anos de idade (nascidas entre 1992 e 1994).

Existem duas vacinas para combater o vírus de HPV, nomeadamente: (Resumo das Características do Medicamento - Gardasil. Aprovado pelo INFARMED em 20 setembro de 2006).

- ✓ Vacina Bivalente- Cervarix
- ✓ Vacina Quadrivalente- Gardasil

Vacina Bivalente –Cervarix este tipo de vacina protege as pessoas contra as infeções provocadas pelos vírus 16 e 18 do HPV.

A eficácia da vacina bivalente foi avaliada em 2 ensaios clínicos, randomizados e duplamente cegos, sequenciados em três fases que incluíram 19757 mulheres com idades compreendidas entre os 15 e 25 anos (SPG, 2010).

No ensaio clínico, realizado em maio de 2004 e junho de 2005 participaram mulheres de 14 países distribuídas pela Ásia (Pacífica), Europa, América Latina e América do Norte.

O protocolo deste estudo prévio tem critérios de inclusão e exclusão. Foram motivos de inclusão do ensaio, as mulheres que não tiveram até à data do estudo mais de 6 parceiros sexuais, que demonstrassem concordar com os métodos de concepção proposta pelo estudo e não possuísem anormalidades citológicas. Excluíram dos ensaios as mulheres que já tivessem realizado colposcopia, estivessem grávidas, possuísem cancro da mama ou sofressem de qualquer tipo de doença crónica ou autoimune (Paavonen & Patricia, 2007).

Os resultados finais do estudo da fase III (PATRICIA- Papilloma Trial Against Cancer In Young Adults) foram publicados, após 18644 mulheres envolvidas nesta fase do estudo terem recebido as 3 doses de Cervarix preconizadas e passados 34,9 meses após a administração da 3ª e última dose (SPG, 2010).

A vacina bivalente Cevaxix protege contra os tipos 16 e 18 do HPV e, após 18 meses de “follow-up”, demonstrou possuir uma eficácia de 100% da prevenção da infeção persistente provocada por estes dois tipos de vírus nas mulheres incluídas no grupo “protocolo”. Relativamente à amostra “intensão de tratar”, que é mais representativo do que se passa na realidade clínica, a vacina bivalente demonstrou eficácia persistente por HPV 16 e 18, e eficácia de 93% dos casos associados a anormalidades citológicas (ASC-US) e outras de maior gravidade (Ault, 2006).

Em estudo prévio de proteção cruzada associadas a outros tipos de HPV, verificou-se que a vacina bivalente demonstrou eficácia relativamente a 5 tipos de HPV não presente na composição da vacina, designadamente 31, 33, 45, 52, 58, sendo estes os mais vulgarmente associados ao cancro invasivo (Paavonen & Patricia, 2007).

Vacina Quadrivalente-Gardasil este tipo de vacina protege as pessoas contra as infeções provocadas pelos vírus 6, 11, 16 e 18 do HPV.

Os primeiros protótipos de vacina contra HPV 16 remontam a meados da década de 90, apresentada pela empresa farmacêutica Merck Sharp & Dohme. Este laboratório realizou em 2000, os primeiros estudos visando a produção de uma única vacina contra os tipos 6, 11, 16 e 18 do HPV (Bragueto & Suzuki, 2008). Durante 5 anos realizaram múltiplos ensaios clínicos sequenciados em diferentes fases, e no final de 2005 foram dados a conhecer os resultados de dois ensaios clínicos denominados de FUTURE (Females United to Unilaterally Reduce Endo/ectocervical disease) I e II da fase III (Bragueto & Suzuki, 2008). Foram mais de 20 mil mulheres a nível mundial (participaram 33 países de todos os continentes) envolvidas neste processo de investigação, distribuídas por diferentes ensaios clínicos randomizados e duplamente cegos, controlados por placebos (Stanley, 2008).

Nos ensaios clínicos FUTURE I e II participaram 17622 mulheres com idades compreendidas entre os 16 e 26 anos, em que 94% já tinha iniciado a vida sexual.

A vacina quadrivalente do HPV demonstrou 100% de prevenção efetiva contra HPV 6, 11, 16 e 18 relacionados com anormalidades cervicais (CIN I-III), com neoplasia vulvares/vaginas e verrugas genitais, contra o HPV 16 e 18 relacionados com CIN II-III, com os AIS e cancro cervical (Ault, 2007).

De ressaltar que a eficiência da vacina obteve percentagem relativamente mais baixas (97%) relativamente às anormalidades cervicais no grupo não pertencentes ao protocolo.

Este grupo de mulheres que não cumpriu todos os parâmetros exigidos pelo protocolo (o protocolo incluiu grupos de mulheres com idades compreendidas entre os 16 e os 26 que concordaram em receber 3 doses de vacinação e aceitaram também serem submetidas ao teste de Papanicolau no espaço de 2 anos), recebeu a denominação de MITT-Modified Intention To

Treat (intenção de tratar) e serviu de estimativa para representar, o mais realisticamente, a prática clínica (Ault, 2007).

Mulheres com as idades compreendidas entre os 24 e os 45 anos foram também envolvidas nos estudos randomizados e duplamente cegos (FUTURE III). Ao fim de 2 anos e 2 meses, a eficácia de vacina da prevenção de lesões/infeções relacionadas com os 4 tipos de vírus persistente na vacina foi de 90,5% e de 83,1% para as lesões/infeções relacionadas com HPV 16 /18 e de 100% para as lesões associadas a HPV 6 e 11 (Muñoz et al., 2009).

De uma forma resumida, a vacina quadrivalente é uma vacina profilática recomendada para prevenção de doenças clínicas casualmente relacionadas com os tipos 6,11, 16 e 18 do HPV, incluindo as lesões genitais pré-cancerosas (colo do útero, vulva, vagina), CCU e verrugas genitais externas (condiloma acuminado) (Bragueto & Suzuki, 2008).

É importante salientar, que a vacinação contra o HPV reduz o risco de ter o CCU, mas não o elimina, uma vez que nem todos os tipos de HPV estão na composição das vacinas, por isso, a necessidade de se realizar rastreios periódicos, recomenda-se a citologia cervical.

CAPÍTULO 2

NOÇÕES BÁSICAS DE ANÁLISE DE SOBREVIVÊNCIA

2.1 INTRODUÇÃO

A Análise de Sobrevivência é um conjunto de métodos e técnicas estatísticas que permitem analisar uma variável aleatória não negativa. Normalmente essa variável aleatória é o tempo até ocorrência de uma determinada condição particular (evento de interesse). Esse evento de interesse pode ser a “morte”, a “recidiva” ou outra.

As suas origens estão associadas aos campos da Biologia, Medicina e ao uso de tabelas de mortalidade (o acontecimento usual era a ‘morte’ do indivíduo), mas estas técnicas não se restringem apenas a estes campos e encontram aplicações em áreas tão diversas como as Ciências Sociais, Económicas e Engenharias.

Com a análise de sobrevivência podemos estudar a distribuição do tempo de vida num grupo de indivíduos ou comparar as distribuições em dois ou mais grupos. Permite-nos também modelar e determinar a relação entre a distribuição do tempo de vida e variáveis explicativas (covariáveis) associadas a cada indivíduo e que representam fatores que se supõe ajudarem a explicar o tempo de vida.

Neste tipo de estudo é essencial definir corretamente:

Evento de interesse: Condição particular que se deseja estudar;

Janela temporal: Data de início e fim do estudo;

Escala do tempo: Unidade de medida do tempo (dias, meses, semanas, anos, etc.);

Variável de interesse: Tempo até o evento de interesse.

2.2 CENSURA

Em muitos estudos, nomeadamente aqueles em que exigem o acompanhamento do indivíduo por muito tempo, muitos indivíduos não observam o evento de interesse durante o tempo total de acompanhamento. Na análise estatística clássica, uma vez que esses indivíduos não foram acompanhados ao longo do tempo previsto, esses indivíduos têm de ser excluídos da análise, já que se desconhece o tempo até o evento de interesse. Estas observações podem ocorrer porque

os indivíduos abandonaram o estudo, foram perdidos no acompanhamento ou porque o estudo chegou ao fim sem que se tivesse observado o evento de interesse.

Nos estudos de Análise de Sobrevivência, os dados destes indivíduos são incluídos na análise, considerando-se que esses indivíduos constituem observações censuradas. Assim, censura significa que para determinado indivíduo, possuímos observações incompletas não se tendo observado o evento de interesse durante o período de observação do estudo, independentemente do motivo pelo qual tal aconteceu ou do que lhe possa ter acontecido posteriormente (Botelho, Silva, & Cruz, 2009).

O que distingue a análise de sobrevivência dos usuais métodos estatísticos é a presença de censura, causada pela observação incompleta da variável de interesse. Em geral, a presença de censura é causada pelo abandono dos indivíduos do estudo ou pelo término do experimento, sem que o evento de interesse tenha ocorrido.

A grande vantagem da análise de sobrevivência é que permite utilizar a informação de todos os participantes, até ao momento em que desenvolvem o evento ou até à censura. Assim, a análise de sobrevivência é a técnica ideal para analisar uma variável resposta (em geral contínua) não negativa para a qual se conhece informação de uma variável binárias (ter ou não ter o evento).

Consideramos uma amostra de n indivíduos. Seja T_i , $1 \leq i \leq n$, uma variável aleatória não negativa que representa o tempo potencial ou real de observação para cada indivíduo.

Seja Y_i ($i = 1, \dots, n$) a variável aleatória que representa o tempo de vida do i -ésimo indivíduo que é totalmente observado. Cada indivíduo tem um potencial tempo de censura (aleatório) que denotamos por C_i . Para cada indivíduo as observações são da forma (Y_i, δ_i) , onde, $Y_i = \min(T_i, C_i)$ e $\delta_i = I(T_i \leq C_i)$ indica a presença ($\delta_i = 1$) ou ausência ($\delta_i = 0$) de um evento.

Note-se que á priori uma observação censurada é diferente de uma omissa, uma vez que a ordem de uma observação censurada relativamente a outra não censurada, pode ser conhecida e como tal, transmite informação sobre a distribuição amostral.

Dado que existem diversos motivos para o aparecimento da censura, podem ser considerados vários tipos de censuras:

Censura à direita - é a mais comum em Análise de Sobrevida e ocorre quando no momento em que termina o estudo, existem indivíduos para os quais não se conhece o tempo exato em que ocorreu o evento de interesse. Só se sabe que terá sido posterior ao final do estudo. Neste tipo de censura, o valor exato do tempo de vida será superior ao valor observado. O mesmo ocorre quando não se pode observar o momento exato do evento de interesse devido à perda do acompanhamento dos indivíduos em estudo.

Censura à esquerda - é o tipo de censura em que, o tempo de sobrevida é menor do que o tempo observado.

Censura intervalar - este tipo de censura ocorre quando não conhecemos o momento exato em que ocorreu o evento de interesse, mas sabemos que este ocorreu dentro de um determinado intervalo de tempo. Este tipo de censura é muito frequente em situações de ensaios clínicos em que os doentes efetuam visitas regulares ao médico.

Convém realçar que, quer a censura à direita, quer a censura à esquerda são casos particulares de censura intervalar.

O tipo de censura mais comum em Análise de Sobrevida é a censura à direita. A partir de agora, e sempre que utilizarmos o termo censura ou dados censurados, estaremos a referir-nos a este tipo de censura.

2.2.1 MECANISMO DE CENSURA À DIREITA

- **Censura do tipo I:** O estudo termina após um período pré-estabelecido de tempo, ou seja, existe um tempo t_c a partir do qual, todos os dados são censurados. Se o acontecimento de interesse acontecer antes de t_c , podemos observar o tempo real de vida do indivíduo e caso isso não aconteça, temos então as chamadas observações censuradas.
- **Censura do tipo II:** O estudo termina após ter ocorrido o evento de interesse, num número pré-determinado de indivíduos.
- **Censura do tipo III (aleatória):** Este tipo de censura não está relacionado com o tempo de sobrevida, ocorre de forma aleatória. O indivíduo é retirado do estudo por uma causa alheia ao próprio estudo. E essa causa pode ser o fato do estudo ter chegado ao

fim ou por qualquer outro motivo que não esteja relacionado com o evento de interesse (morte por outras causas, por exemplo).

2.3 TRUNCATURA

Um outro fenómeno muitas vezes presente na Análise de sobrevivência, mas muitas vezes confundido com a censura, é a existência de truncatura. A truncatura ocorre, devido a um processo inerente ao planeamento do estudo. Esta é caracterizada por excluir indivíduos que não são importantes para o estudo que se pretende efetuar.

2.3.1 TRUNCATURA À DIREITA

Este tipo de truncatura ocorre quando são considerados na análise estatística, apenas os indivíduos para os quais se observou um determinado acontecimento de interesse, durante o período de estudo estabelecido. Este tipo de truncatura ocorre quando os indivíduos não são acompanhados a partir do tempo inicial, mas só são acompanhados após experimentarem um certo evento de interesse durante o período em estudo.

2.3.2 TRUNCATURA À ESQUERDA

Este tipo de truncatura é o mais comum, ocorre quando são incluídos na amostra indivíduos que sobrevivem tempo suficiente, para que ocorra um determinado acontecimento antes do evento do interesse. Um exemplo deste tipo de truncatura pode ser encontrado num estudo de sobrevivência de doentes com leucemia. Por exemplo, pacientes que foram submetidos a transplante de medula e apenas aqueles que tiveram a recaída da doença, é que são analisados pelo médico.

É de realçar que, o facto de existir censura não impede a existência em simultâneo de truncatura, sendo muito usual nos depararmos com dados que são censurados à direita e truncados à esquerda.

2.4 FUNÇÕES DE INTERESSE NO CONTEXTO DE DADOS DE SOBREVIVÊNCIA

Seja T uma variável aleatória não negativa, que representa o tempo decorrido até um determinado evento de interesse. A distribuição desta variável pode ser definida à das seguintes funções: (i) função densidade de probabilidade $f(t)$, (ii) função de sobrevivência $S(t)$, (iii) função de distribuição $F(t)$, (iv) função de risco $h(t)$ e (v) função de risco cumulativa $H(t)$. Cada uma destas funções (definidas em baixo) ilustra diferentes aspetos dos dados.

2.4.1 VARIÁVEL ALEATÓRIA CONTÍNUA

Nesta secção vamos assumir que T denota uma variável absolutamente contínua.

2.4.1.1 FUNÇÃO DE DISTRIBUIÇÃO

Seja $F(t)$ a função de distribuição de T , definida como a probabilidade de observar o evento de interesse no intervalo de $[0, t]$ e é representada da seguinte forma:

$$F(t) = P(T \leq t) = \int_0^t f(u) du, t \geq 0 \quad (2.1)$$

$F(t)$ é uma função:

- Monótona não decrescente;
- $F(t) = 0$ e $F(t) = 1$ quando $t \rightarrow \infty$.

2.4.1.2 FUNÇÃO DENSIDADE DE PROBABILIDADE

A função densidade de probabilidade (fdp) é, por definição:

$$f(t) = P(T = t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}, t \geq 0 \quad (2.2)$$

Esta função representa a probabilidade de um individuo sofrer o evento no intervalo instantâneo de tempo t .

2.4.1.3 FUNÇÃO DE SOBREVIVÊNCIA

A função de sobrevivência representa a probabilidade de um indivíduo não ter observado o evento de interesse, pelo menos, até ao instante t . Define-se a função de sobrevivência $S(t)$ da variável aleatória T por:

$$S(t) = P(T > t) = 1 - F(t), t \geq 0 \quad (2.3)$$

Onde o $F(t)$ é a função de distribuição correspondente.

A função de sobrevivência goza das seguintes propriedades:

- Monótona não crescente
- $S(t) = 1$ para $t = 0$ e $S(t) = 0$ para $t \rightarrow +\infty$. O gráfico de $S(t)$ é chamado de curva de sobrevivência e é decrescente com o tempo.

2.4.1.4 FUNÇÃO DE RISCO

A função *hazard* ou função de risco corresponde à probabilidade de falha no instante t dado que o indivíduo sobreviveu até esse mesmo instante. Permite quantificar o risco instantâneo de um acontecimento ocorrer por unidade de tempo. A função de risco é dada por:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, t \geq 0 \quad (2.4)$$

A função de risco satisfaz as seguintes propriedades:

- $h(t) \geq 0$
- $\int_0^{+\infty} h(t) dt = +\infty$.

As condições apresentadas anteriormente são necessárias e suficientes para que qualquer função real seja uma função de risco. De facto, a função de risco pode ser estendida a qualquer variável aleatória não necessariamente positiva.

2.4.1.5 FUNÇÃO DE RISCO CUMULATIVA

Define-se ainda a função de risco cumulativa como sendo:

$$H(t) = \int_0^t h(u)du, t \geq 0 \quad (2.5)$$

Esta função mede o risco de ocorrência de acontecimento de interesse até ao instante t , sendo ele uma função monótona não decrescente.

2.4.2 VARIÁVEL ALEATÓRIA DISCRETA

Se os tempos de sobrevivência estão agrupados ou se referem um número inteiro de ciclos a variável T , deve ser considerada como uma variável aleatória discreta. As cinco funções introduzidas na secção anterior são definidas do seguinte modo:

- **Função Massa de Probabilidade**

$$P(T = t_i) = P(t_i) = p_i \quad (2.6)$$

- **Função de Distribuição**

$$F(t) = P(T \leq t) = \sum_{t_i \leq t} P(t_i) \quad (2.7)$$

- **Função de Sobrevivência**

$$S(t) = P(T > t) = \sum_{t_i > t} P(t_i) \quad (2.8)$$

- **Função de Risco**

$$h(t_i) = \frac{P(t_i)}{S(t_i)} \quad (2.9)$$

- Função de Risco Cumulativa

$$H(t) = \sum_{t_i \leq t} h(t_i) \quad (2.10)$$

2.5 RELAÇÕES ENTRE AS FUNÇÕES

Uma vez introduzidos os conceitos anteriores, podemos deduzir várias relações entre as funções $f(t)$, $F(t)$, $S(t)$, $h(t)$ e $H(t)$. As relações apresentadas em baixo são validas para distribuições discretas e contínuas.

$$S(t) = \exp(-H(t)) \quad (2.11)$$

$$h(t) = \frac{f(t)}{S(t)} \quad (2.12)$$

$$f(t) = h(t)S(t) \quad (2.13)$$

$$F(t) = 1 - S(t) \quad (2.14)$$

$$H(t) = -\ln S(t) \quad (2.15)$$

Das relações apresentadas podemos concluir que a distribuição da variável aleatória T fica univocamente definida a partir de qualquer uma das seguintes funções: função distribuição, função densidade de probabilidade, função de sobrevivência, função de risco e função de risco cumulativa.

Tanto para variáveis aleatórias discretas como contínuas existem parâmetros de interesse associados às curvas de sobrevivência, que podem ser facilmente calculadas. (Veja-se Tabela 7).

Tabela 7 - Alguns parâmetros de sobrevivência

Variável Aleatória T	Sobrevivência Média	Sobrevivência Mediana
Discreta	$\mu = E(T) = \sum_i i \times P(i)$	$S(t) = 0.5$
Continua	$\mu = E(T) = \int_0^{\infty} tf(t)dt$	$S(t) = 0.5$

2.6 MODELO DE SOBREVIVÊNCIA PARAMÉTRICO

As técnicas tradicionais utilizadas em análise de sobrevivência podem ser agrupadas em três grupos: paramétricos, não paramétricos e semiparamétricos. Os semiparamétricos e os paramétricos são os mais utilizados e as estimativas que se obtém com estes métodos servem de auxílio a posterior análise estatística mais detalhadas. O processo de estimação da função de sobrevivência varia consoante o modelo e o método que se pretender ajustar aos dados.

Nesta secção vamos dar maior ênfase ao modelo de sobrevivência paramétrico.

Verifica-se que qualquer distribuição com suporte não negativo pode ser usada como modelo para o tempo de vida (Collett, 2015; DR Cox & Oakes, 1984; Lawless, 1982)

Os modelos paramétricos têm-se mostrado bastante adequados na análise estatística de dados de sobrevivência. Estes modelos também são designados por modelos probabilísticos ou paramétricos para o tempo de falha (Colosimo, 1995).

Existem vários modelos probabilísticos ou paramétricos em análise de sobrevivência e alguns destes modelos ocupam a posição de destaque, devido à sua comprovada adequação em várias situações práticas, sendo elas usadas com bastante frequência ((Bolfarine, Rodrigues, & Achcar, 1991). Entre esses modelos paramétricos destacam-se as seguintes distribuições: exponencial, weibull, log-normal, gama e gama generalizada. A distribuição gama generalizada inclui como casos especiais, a distribuição exponencial e a weibull.

Em seguida estão representadas algumas distribuições paramétricas muito utilizadas em análise de sobrevivência. Para cada distribuição é apresentada a respetiva função distribuição, a função densidade de probabilidade, a função de sobrevivência e a função de risco.

2.6.1 DISTRIBUIÇÃO EXPONENCIAL

Nesta secção introduzimos a distribuição exponencial com um único parâmetro, $\lambda > 0$, onde $1/\lambda$ representa o tempo médio de vida e tem a mesma unidade de tempo de falha t .

A função densidade de probabilidade para a variável aleatória tempo de falha T com distribuição exponencial é dado por:

$$f(t) = \lambda \exp(-\lambda t), t \geq 0 \quad (2.16)$$

Então a função distribuição é dado por:

$$F(t) = 1 - \exp(-\lambda t), t \geq 0 \quad (2.17)$$

A função de sobrevivência referente a esta distribuição, $S(t)$ é representada pela seguinte expressão:

$$S(t) = \exp(-\lambda t), t \geq 0 \quad (2.18)$$

A taxa de falha $h(t)$ também é chamada de função de risco e é constante. O risco constante é a principal característica desta distribuição. Isso significa que tanto uma unidade antiga como uma nova, que ainda não falharam, têm o mesmo risco de falha num intervalo futuro. Esta propriedade é chamada de falta de memória da distribuição exponencial. A referida função pode ser apresentada por:

$$h(t) = \lambda, t \geq 0 \quad (2.19)$$

2.6.2 DISTRIBUIÇÃO WEIBULL

A distribuição weibull foi originalmente proposta por Weibull em 1939 e a sua aplicabilidade também foi discutida por este mesmo autor (Weibull, 1939, 1949, 1951). Esta distribuição é muito útil na prática devido à sua grande variedade de formas, todas com uma propriedade básica: a função de risco é monótona, podendo ser crescente, decrescente ou constante.

Seja T uma variável aleatória com distribuição weibull de parâmetros λ e α , isto é, $T \sim W(\lambda, \alpha)$, $\lambda > 0$ é o parâmetro de escala e $\alpha > 0$ é o parâmetro da forma.

A distribuição weibull pode ser definida pela sua função densidade de probabilidade da seguinte forma:

$$f(t) = \alpha \lambda t^{\alpha-1} \exp(-\lambda t^\alpha), t \geq 0 \quad (2.20)$$

A sua função de distribuição é dada por:

$$F(t) = 1 - \exp(-\lambda t^\alpha), t \geq 0 \quad (2.21)$$

A função de sobrevivência é dada por:

$$S(t) = \exp(-\lambda t^\alpha), t \geq 0 \quad (2.22)$$

A sua função de risco é definida por:

$$h(t) = \alpha \lambda (\lambda t)^\alpha t^{\alpha-1} \quad (2.23)$$

A função de risco da distribuição weibull é monótona crescente para $\alpha > 1$, decrescente para $0 < \alpha < 1$ ou constante para $\alpha = 1$. De um modo geral pode dizer-se que a distribuição weibull é uma generalização da distribuição exponencial.

Esta distribuição é provavelmente a mais utilizada em análise de sobrevivência, nomeadamente em aplicações biomédicas e em experiência com animais em laboratório.

2.6.3 DISTRIBUIÇÃO GAMA

Seja T uma variável com distribuição gama, é caracterizada por dois parâmetros, β e λ , em que $\beta > 0$ é chamado de parâmetro de forma e $\lambda > 0$ de escala, a sua função densidade de probabilidade é definida por:

$$f(t) = \frac{\lambda^\beta}{\Gamma(\beta)} t^{\beta-1} \exp(-\lambda t) \quad (2.24)$$

A função de sobrevivência é dada por:

$$S(t) = 1 - I(\lambda t, \beta) \quad (2.25)$$

Onde $\Gamma(\beta)$ é a função gama e $I(\lambda t, \beta) = \frac{1}{\Gamma(\beta)} \int_0^t u^{\beta-1} e^{-u} du$ que é conhecida como função gama incompleta.

A sua função de risco é dada por:

$$h(t) = \frac{f(t)}{S(t)} \text{ onde } t \geq 0, \lambda > 0, \beta > 0 \quad (2.26)$$

não tendo uma expressão simples. A sua função de risco é monótona crescente quando ($\beta > 1$), monótona decrescente quando $0 < \beta < 1$ e é contante quando $\beta = 1$.

A distribuição weibull e a distribuição gama têm propriedades semelhantes. Contudo, a distribuição gama é menos utilizada pelo facto de depender da função gama incompleta, tornando assim a especificação da função de sobrevivência e da função de risco mais difícil.

2.6.4 DISTRIBUIÇÃO LOG-NORMAL

A distribuição log-normal é muito utilizada para caracterizar o tempo de vida de itens e indivíduos, como por exemplo descrever situações clínicas, como o tempo de vida dos doentes. Existem também diversas aplicações deste modelo em testes do tempo de falha de produtos ou itens (Nelson, 1990). Informação mais detalhada sobre a distribuição log-normal pode ser encontrada em (Crow & Shimizu, 1988).

Seja T uma variável aleatória com distribuição log-normal de parâmetros μ e σ . Nestas situações a variável aleatória $Y = \ln T$ tem distribuição normal com valor esperado μ e variância σ^2 .

A função densidade de probabilidade de T é dada por:

$$f(t) = \frac{1}{2\pi\sigma t} \exp\left[-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2\right] \text{ onde } \mu \in \mathbb{R} \text{ e } \sigma > 0 \quad (2.27)$$

A sua função de sobrevivência é da forma:

$$S(t) = 1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \quad (2.28)$$

Onde $\Phi(\cdot)$ denota a distribuição de uma variável com distribuição normal.

A sua função de risco é dada por:

$$h(t) = \frac{f(t)}{S(t)} \quad (2.29)$$

não tendo uma expressão simplificada.

A sua função de risco é crescente até atingir um valor máximo e depois decresce, com

$$\lim_{t \rightarrow 0^+} h(t) = \lim_{t \rightarrow \infty} h(t) = 0.$$

A distribuição log-normal tem sido frequentemente utilizada para modelar tempos de sobrevivência. A principal desvantagem desta distribuição é que em situações em que o risco de falha é crescente e deve-se ao facto da função de risco ser decrescente a partir de um dado instante. Esta distribuição é adequada quando os valores elevados de t não têm interesse.

2.6.5 DISTRIBUIÇÃO DE PARETO

Seja T uma variável aleatória com distribuição de Pareto com parâmetros θ e λ , onde $\theta > 0, \lambda > 0$ e $t \geq \lambda$.

De seguida apresentamos a representação da sua função densidade de probabilidade, função de sobrevivência e a sua função de risco, respetivamente.

$$f(t) = \frac{\theta \lambda^\theta}{t^{\theta+1}} \quad (2.30)$$

$$S(t) = \frac{\lambda^\theta}{t^\theta} \quad (2.31)$$

$$h(t) = \frac{\theta}{t} \quad (2.32)$$

2.6.6 DISTRIBUIÇÃO LOG-LOGÍSTICA

A distribuição Log-Logística de parâmetros α e λ , com função densidade de probabilidade, função de sobrevivência e função de risco dada por, respetivamente:

$$f(t) = \frac{\alpha t^{\alpha-1} \lambda}{[1 + \lambda t^\alpha]^2}, \alpha \text{ e } \lambda > 0, t \geq 0 \quad (2.33)$$

$$S(t) = \frac{1}{1 + \lambda t^\alpha} \quad (2.34)$$

$$h(t) = \frac{\alpha t^{\alpha-1} \lambda}{1 + \lambda t^\alpha} \quad (2.35)$$

A sua função de risco é monótona decrescente se $\alpha \leq 1$. Se $\alpha > 1$ cresce até $\frac{(\alpha-1)^{\frac{1}{\alpha}}}{\lambda}$, para decrescer depois.

2.7 ESTIMAÇÃO PARAMÉTRICA DA FUNÇÃO DE SOBREVIVÊNCIA

Os modelos paramétricos apresentados na seção anterior são caracterizados por quantidades geralmente desconhecidas, denominadas de parâmetros, esses parâmetros variam de um modelo para outro e em cada caso representam características como por exemplo, a forma, a escala, a média ou o desvio-padrão. Os modelos weibull, log-normal, gama, pareto e log-logística são caracterizados por dois parâmetros enquanto o modelo exponencial é caracterizado por apenas um parâmetro. Essas quantidades conferem uma forma geral aos modelos paramétricos. Para cada estudo que envolve os tempos de falha, os parâmetros devem ser estimados a partir das observações amostrais para que o modelo fique determinado e, assim seja possível responder às questões de interesse.

Entre os métodos de estimação mais conhecidos na literatura, está o método dos mínimos quadrados e o da verossimilhança. O primeiro método referido não é apropriado para o estudo de tempos de vida, devido à sua incapacidade de incorporar as censuras e truncaturas no seu processo de estimação. O método de verossimilhança constitui uma opção válida para este tipo

de dados. Este método permite incorporar a censura e truncatura, é relativamente simples de ser estendido e possui propriedades ótimas para amostras de maior dimensão.

Resumindo, pode-se dizer que nos modelos paramétricos a função de sobrevivência depende de um vetor de parâmetros θ , que pode ser estimado a partir do método de verossimilhança, ou seja esses estimadores de θ são obtidos maximizando o logaritmo natural da função de verossimilhança.

A função de verossimilhança para um parâmetro genérico θ desta população é então expressa por:

$$L(\theta) = \prod_{i=1}^n f(t_i; \theta) \quad (2.36)$$

Nesta expressão θ pode estar representado um único parâmetro ou um vetor de parâmetros. Por exemplo, no modelo log-normal, $\theta = (\mu, \sigma)$.

A expressão da função de verossimilhança depende do mecanismo de censura. Consideremos o caso de mecanismo de censura do tipo II. Assumindo que T_1, \dots, T_n , são independentes e identicamente distribuídas, e que segue uma distribuição contínua com função densidade de probabilidade $f(t)$ e função distribuição $F(t)$, tem-se que a função densidade de probabilidade conjunta de $(t_{(1)}, \dots, t_{(r)})$ (estatística ordinais) é dada, por:

$$f_{(t_{(1)}, \dots, t_{(r)})}((t_1, \dots, t_r); \theta) = \frac{n!}{(n-r)!} \left[\prod_{i=1}^r f(t_i) \right] [1 - F(t_r)]^{n-r} \quad (2.37)$$

com $0 \leq t_1 \leq \dots \leq t_r$.

No caso de a experiência terminar num instante específico t_0 e no caso de mecanismo de censura do tipo I, a função densidade de probabilidade conjunta é dada por:

$$f_{(t_{(1)}, \dots, t_{(r)})}((t_1, \dots, t_r); \theta) = \frac{n!}{(n-r)!} \left[\prod_{i=1}^r f(t_0) \right] [1 - F(t_i)]^{n-r} \quad (2.38)$$

Onde $0 \leq t_1 \leq \dots \leq t_r \leq t_0$.

2.8 ESTIMAÇÃO NÃO PARAMÉTRICA DA FUNÇÃO DE SOBREVIVÊNCIA

Em análise de sobrevivência, usam-se os modelos não paramétricos quando o processo de estimação é realizado sem que se faça nenhuma suposição sobre a distribuição de probabilidade do tempo de sobrevivência, ou seja, não é imposto nenhum modelo teórico para as falhas, censuras ou truncaturas observadas. Entre esses modelos destaca-se o estimador de Kaplan-Meier (Kaplan & Meier, 1958).

2.8.1 ESTIMADOR DE KAPLAN-MEIER

O estimador de Kaplan-Meier é o mais utilizado em análise de sobrevivência e também é o mais utilizado em estudos clínicos e vem ganhando cada vez mais espaço nos processos biológicos (Colosimo & Giolo, 2006).

Através do estimador de Kaplan-Meier podemos obter estimativas para a função de sobrevivência de um determinado grupo de indivíduos. Além disso a sua representação gráfica evidencia informação sobre os percentis, dispersão dos dados, e em geral evidencia todos os aspetos importantes da distribuição amostral do tempo de sobrevivência, construindo assim uma ferramenta básica para a seleção de um ou mais modelos probabilísticos que possam ser um bom ajuste (Andersen & Keiding, 2002).

Na ausência de observações censuradas, para uma amostra de dimensão n , a função de sobrevivência num instante t , pode ser estimada a partir dos tempos de vida observados. Um estimador natural da função de sobrevivência, é o estimado empírico ou amostral, dado por:

$$\hat{S}(t) = \frac{\text{número de observações} > t}{n}, t \geq 0 \quad (2.39)$$

Este estimador da função de sobrevivência é uma função em escada com 'saltos' nos tempos observados de falha/evento. O estimador de Kaplan-Meier na sua construção, considera tantos intervalos de tempo quantos forem os números de tempos distintos de falha. Os limites dos intervalos de tempo são os tempos distintos de falha.

Na presença de dados censurados são necessários métodos alternativos. No caso de haver observações censuradas, o estimador empírico terá que ser modificado porque o número de tempos de sobrevivência maiores ou iguais a t não é, em geral conhecido.

Kaplan e Meier em 1958 propuseram um estimador não paramétrico, para a função de sobrevivência que incorpora a censura denominado por Kaplan-Meier ou estimador produto-limite, sendo esta uma generalização do estimador empírico para dados censurados.

Sejam $y_1 < y_2 < \dots < y_k$ os k limites distintos de evento de uma amostra de dimensão n com ($k \leq n$).

Então o estimador de Kaplan-Meier da função de sobrevivência, é dada por:

$$\hat{S}(t) = \prod_{i:y_i < t} \left(\frac{n_i - d_i}{n_i} \right) = \prod_{i:y_i < t} \left(1 - \frac{d_i}{n_i} \right) \quad (2.40)$$

Onde,

d_i - “número de eventos ocorridos em y_i ”;

n_i - “número de indivíduos em risco no instante y_i ou seja, número de indivíduos vivos e não censurados imediatamente antes do instante y_i , $i = 1, 2, \dots, k$ ”;

$\hat{S}(t)$ - “estimativa da função de sobrevivência”.

O estimador de Kaplan-Meier apresentado em cima, é uma função em escada que decresce logo após a observação de eventos y_i . Os saltos da função dependem não só do número de eventos, mas também do número de indivíduos em risco no momento, sendo expressos por um fator $1 - \frac{d_i}{n_i}$. A função de sobrevivência não decresce nos tempos censurados.

Refira-se que, se a maior observação registada for um evento, então a função de sobrevivência termina em zero, caso contrário nunca termina em zero.

O estimador de Kaplan-Meier também pode ser escrito recorrendo aos pesos de Kaplan-Meier.

$$\hat{S}(t) = 1 - \sum_{i=1}^n w_i I(Y_i \leq t) \quad (2.41)$$

Onde $I(\cdot)$ é denominado da função indicatriz e W_i os pesos de Kaplan-Meier associados a Y_i .

$$W_i = \frac{\delta_{[i]}}{n - i + 1} \prod_{j=1}^{i-1} \left[1 - \frac{\delta_{[j]}}{n - j + 1} \right] \quad (2.42)$$

$$\delta_i = \begin{cases} 1 & \text{se o observado é evento de interesse} \\ 0 & \text{se ocorrer observação censurada} \end{cases}$$

Convém referir que no caso de ausência de censura, os pesos Kaplan-Meier são todos iguais com peso $W_i = \frac{1}{n}$; na presença de censura, as observações censuradas têm peso nulo ($W_i = 0$).

Por convenção, em caso de empates, as ordens das observações Y_i com censura são maiores do que a ordem dos eventos (não censurados).

Vamos considerar um exemplo ilustrativo apenas para uma melhor compreensão da estimação da função de sobrevivência, usando as duas formas alternativas do estimador Kaplan-Meier.

Exemplo: consideremos os seguintes tempos de vida (em dias) de 12 doentes, onde as observações censuradas estão representadas por "+".

Tabela 8 - Dados para ilustrar estimativa para a função de sobrevivência

55	61	74	55+	93+	122+
138	151	168	202+	220+	238+

Primeiramente, deve-se começar por ordenar os valores $y: y_1 \leq \dots \leq y_n$, da Tabela 8, colocando em caso de empate primeiro os eventos e só depois as observações censuradas.

Tabela 9 - Dados ordenados onde + indica presença de censura pela direita

i	1	2	3	4	5	6	7	8	9	10	11	12
Y_i	55	55+	61	74	93+	122+	138	151	168	202+	220+	238+
Status (δ_i)	1	0	1	1	0	0	1	1	1	0	0	0

De seguida, são calculadas as estimativas para a função de sobrevivência através do método de Kaplan-Meier na Tabela 10.

Tabela 10 - Estimação da função de sobrevivência pelo método de Kaplan-Meier

y_j	d_j	n_j	Estatutos (δ_i)	$\hat{S}(t)$
55	1	12	1	0.917
61	1	10	1	0.825
74	1	9	1	0.733
93	0	8	0	0.733
122	0	7	0	0.733
138	1	6	1	0.611
151	1	5	1	0.489
168	1	4	1	0.367
202	0	3	0	0.367
220	0	2	0	0.367
238	0	1	0	0.367

Onde,

y_j - “representa os tempos de sobrevivência distintos”;

d_j - “representa em cada tempo o número de evento de interesse”;

n_j - “representa o número de indivíduos em risco imediatamente antes do tempo y_j ”;

δ_i - “representa o evento ou censura”;

$\hat{S}(t)$ - “representa estimativa da função de sobrevivência pelo estimador Kaplan-Meier”.

Podemos verificar através da Tabela 10, que a função de sobrevivência é decrescente, mas não termina no valor zero, porque o tempo mais elevado corresponde a uma observação censurada.

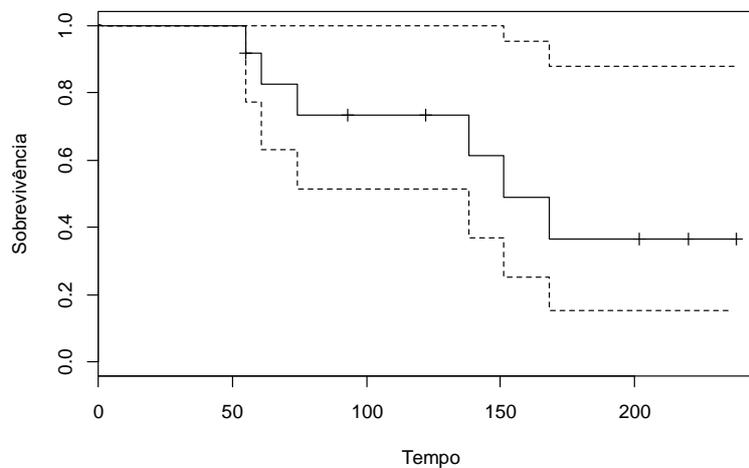


Figura 6 – Representação gráfica da curva de sobrevivência pelo método Kaplan-Meier

Da Figura 6, verifica-se que o sinal “+” apresentado no gráfico representa as observações censuradas e podemos verificar que a função de sobrevivência se apresenta em forma de escada e é decrescente à medida que o tempo passa.

Estimação de função de sobrevivência usando os pesos Kaplan-Meier

Considerando os mesmos dados da Tabela 8.

Tabela 11 - Estimação de função de sobrevivência usando os pesos Kaplan-Meier

i	y_j	d_j	n_j	Estatutos	w_i	$\hat{S}(t)$
1	55	1	12	1	$w_1 = 0.0833$	$\hat{S}(1) = 0.917$
2	55	0	11	0	$w_2 = 0$	$\hat{S}(2) = 0.825$
3	61	1	10	1	$w_3 = 0.0917$	$\hat{S}(3) = 0.825$
4	74	1	9	1	$w_4 = 0.0917$	$\hat{S}(4) = 0.733$
5	93	0	8	0	$w_5 = 0$	$\hat{S}(5) = 0.733$
6	122	0	7	0	$w_6 = 0$	$\hat{S}(6) = 0.733$
7	138	1	6	1	$w_7 = 0.122$	$\hat{S}(7) = 0.611$
8	151	1	5	1	$w_8 = 0.122$	$\hat{S}(8) = 0.489$
9	168	1	4	1	$w_9 = 0.122$	$\hat{S}(9) = 0.367$
10	202	0	3	0	$w_{10} = 0$	$\hat{S}(10) = 0.367$
11	220	0	2	0	$w_{11} = 0$	$\hat{S}(11) = 0.367$
12	238	0	1	0	$w_{12} = 0$	$\hat{S}(12) = 0.367$

Neste caso a estimativa para função de sobrevivência foi obtida da seguinte forma:

$$\hat{S}(1) = 1 - w_1 \quad (2.43)$$

$$\hat{S}(2) = 1 - (w_1 + w_2) \quad (2.44)$$

$$\hat{S}(3) = 1 - (w_1 + w_2 + w_3) \quad (2.45)$$

Assim, foi calculada a função de sobrevivência para os 12 tempos.

2.8.2 O ESTIMADOR DE GREENWOOD

Breslow & Crowley (1974) e Meier (1975) provaram que sobre condições muito gerais, o estimador $\hat{S}(t)$ pode ser considerado como um estimador da máxima verosimilhança não

paramétrico de $S(t)$. A partir do método de verosimilhança, chega-se à seguinte expressão para a estimativa da variância do estimador $\hat{S}(t)$ dado por:

$$\hat{\sigma}^2[\hat{S}(t)] = \widehat{\text{var}}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i - d_i)} \quad (2.46)$$

O estimador em cima apresentado é conhecido como o estimador de Greenwood (Greenwood, 1926). Nota-se também no caso de não existir observação censurada o estimador da $\widehat{\text{var}}(\hat{S}(t))$ é dado por:

$$\widehat{\text{var}}[\hat{S}(t)] = \frac{\hat{S}(t)[1 - \hat{S}(t)]}{n} \quad (2.47)$$

2.8.3 INTERVALO DE CONFIANÇA PARA A FUNÇÃO DE SOBREVIVÊNCIA

Um intervalo de confiança calculado no determinado instante, pertence ao domínio da curva de sobrevivência, indicando a credibilidade da estimativa nesse instante.

O estimador de Kaplan-Meier, para o t fixo, tem distribuição assintótica normal. Portanto, o intervalo aproximado de $100(1 - \alpha)\%$ de confiança para a função de sobrevivência no instante específico é dado por:

$$\hat{S}(t) \pm Z_{\frac{\alpha}{2}} \sqrt{\widehat{\text{var}}[\hat{S}(t)]} \quad (2.48)$$

Onde $Z_{\frac{\alpha}{2}}$ é o quantil de probabilidade $1 - \frac{\alpha}{2}$ de uma distribuição normal reduzida, $N(0,1)$. Para a estimativa de variância de $\hat{S}(t)$ podemos recorrer à proposta por Greenwood (1926).

Os intervalos de confiança obtidos por este processo são simétricos, sendo possível que os limites do intervalo de confiança sejam maiores do que um ou menores do que zero, nos instantes em que está próximo desses valores. Uma forma de resolver esse problema é a substituição de qualquer limite maior do que um pelo valor 1 e qualquer limite menor que zero pelo valor 0.

Uma alternativa é transformar $\hat{S}(t)$ num valor pertencente ao intervalo $(-\infty, +\infty)$ e obter um intervalo de confiança para o valor transformado. Existem várias transformações possíveis sendo duas delas a transformação logística $\log\left[\frac{S(t)}{1-S(t)}\right]$ e a transformação complementar log-log, ou seja, $\log[-\log(S(t))]$.

Quando se aplica transformação log-log obtém-se um intervalo de confiança assimétrico $S(t)$ para a partir de um intervalo simétrico para $\log[-\log(S(t))]$. Este método para além de assegurar que os limites do intervalo de confiança são positivos e menores ou iguais a um, é bastante preciso, porque $\log[-\log(\hat{S}(t))]$ tem uma distribuição mais próxima da distribuição normal do que $\hat{S}(t)$. Assim o intervalo com um grau de confiança $100(1-\alpha)\%$ para $\log[-\log(S(t))]$ é dado por:

$$\hat{W} \pm Z_{\frac{\alpha}{2}} \sqrt{\widehat{\text{var}}(\hat{W})} \quad (2.49)$$

$$\hat{W} = \log[-\log(\hat{S}(t))] \quad (2.50)$$

Onde $\frac{Z}{\alpha}$ representa o quantil de probabilidade $1 - \frac{\alpha}{2}$ da distribuição $N(0,1)$.

A sua variância é definida da seguinte forma:

$$\widehat{\text{var}}(\hat{W}) = \frac{1}{[\log(\hat{S}(t))]^2} \sum_{i:t_i < t} \frac{d_i}{n_i(n_i - d_i)} \quad (2.51)$$

O estimador proposto por Kaplan e Meier (1958) é o estimador mais utilizado para obter uma estimativa não paramétrica da função de sobrevivência. Frequentemente interessa-nos também estimar a função de risco cumulativa. Um estimador natural desta função, $H(t)$, poderá ser obtido recorrendo ao estimador produto-limite de Kaplan e Meier. Este estimador é dado por $\hat{H}(t) = -\log(\hat{S}(t)) = -\sum_{i:y_i \leq t} \left[\log\left(1 - \frac{d_i}{n_i}\right)\right]$, onde $\hat{S}(t)$ é o estimador de Kaplan-Meier. No entanto um estimador alternativo, sugerido por Nelson (1972) e estudado por Aalen (1978) é uma outra opção que se torna cada vez mais usual. Este estimador é designado por estimador de Nelson-Aalen.

2.8.4 ESTIMAÇÃO DA FUNÇÃO DE SOBREVIVÊNCIA E FUNÇÃO DE RISCO CUMULATIVA UTILIZANDO O ESTIMADOR DE NELSON-AALEN

Sejam $y_1 < \dots < y_k$ os instantes de interesse distintos e ordenados numa amostra de dimensão n ($k \leq n$), d_i o número de eventos ocorridos em y_i e n_i o número de indivíduos em risco justamente antes do tempo y_i .

O estimador de Nelson-Aalen para a função de risco cumulativa define-se por:

$$\hat{H}(t) = \sum_{i: y_i \leq t} \frac{d_i}{n_i} \quad (2.52)$$

O estimador de Nelson-Aalen estima diretamente a função de risco cumulativa, embora como é óbvio possa ser utilizado para estimar a função de sobrevivência. Assim sendo, o estimador de Nelson-Aalen para a função de sobrevivência é definido da seguinte forma:

$$\hat{S}(t) = \exp\left(-\sum_{i: y_i \leq t} \frac{d_i}{n_i}\right) \quad (2.53)$$

Quando comparado com o estimador de Kaplan-Meier, o estimador de Nelson-Aalen é reconhecido como um estimador competitivo obtendo um melhor comportamento, para amostras de pequena dimensão. Em muitas circunstâncias as estimativas serão muito semelhantes, principalmente para tempos para os quais ainda existem muitos indivíduos em risco.

2.8.5 TESTES DE HIPÓTESES PARA COMPARAR CURVAS DE SOBREVIVÊNCIA

Em análise de sobrevivência interessa muitas vezes analisar os fatores endógenos ou exógenos aos indivíduos que contribuem para a ocorrência de um evento de interesse, ou seja, características como o género, a idade, tratamento, entre outras, podem ter um papel primordial no tempo de sobrevivência e irão originar curvas de sobrevivência distintas. O estimador de Kaplan-Meier pode ser utilizado para comparar as diferentes curvas correspondentes aos vários grupos de uma variável qualitativa (por exemplo, género, tratamento, etc.). Para variáveis quantitativas (por exemplo, idade, peso, tamanho, etc.) poder-se-á proceder a uma categorização

da variável e posteriormente efetuar uma comparação das diferentes curvas que se obtêm para cada categoria. Este processo consiste na estimação das funções de sobrevivências, separadamente para cada um dos grupos, o que possibilita de modo informal, avaliar se a variável tem influência no tempo de sobrevivência. O estimador Kaplan-Meier é vantajoso, na medida em que possibilita interpretações e representações gráficas intuitivas.

Depois da estimação das funções de sobrevivência e representação gráfica das mesmas, será importante testar se existem diferenças estatisticamente significativas entre as curvas de sobrevivência para os vários grupos. Para comparar as curvas de função de sobrevivência são geralmente utilizados testes não paramétricos, denominados por testes de ordem (rank test), visto que a sua estatística de teste depende unicamente das ordens das observações. Mais adiante vamos falar sobre o teste não paramétrico que permite fazer comparação entre tempo de sobrevivência para dois ou mais grupos.

Um objetivo crucial da Análise de Sobrevivência é a comparação das curvas de sobrevivência de dois ou mais grupos de indivíduos que diferem numa ou mais características. Existem vários métodos não paramétricos para avaliar se existem diferenças entre as distribuições dos tempos de sobrevivência dos G amostras.

Os testes não paramétricos mais utilizados para avaliar se existem diferenças significativas nas curvas de sobrevivência de Kaplan-Meier, são o teste de log-rank, teste de Wilcoxon (Breslow) e teste de Tarone-Ware. No entanto, destes testes destaca-se o teste de log-rank e por isso vamos utilizar esse teste para o nosso estudo.

O teste log-rank compara a distribuição da ocorrência dos acontecimentos observados em cada grupo, com a distribuição que seria esperada, se a incidência fosse igual em todos os grupos. Se a distribuição observada for equivalente à distribuição esperada, dizemos que a função de sobrevivência dos indivíduos pertencentes ao grupo coincide com a função de sobrevivência dos indivíduos em geral, ou seja, que a covariável introduzida não influencia a sobrevivência.

O teste considera a seguinte hipóteses:

$$H_0: S_1(t) = S_2(t) = \dots = S_r(t) \text{ (as curvas de sobrevivências são todos iguais)}$$

Versus

H_1 : existem pelo menos um par em que as curvas de sobrevivência se diferenciam

Considerando inicialmente apenas 2 grupos, suponha que no conjunto de dados (amostra conjunta) se registarem r tempos de mortes distintos $t_i, i = 1, 2, \dots, r, t_1 < t_2 < \dots < t_r$, a informação relevante pode ser resumida numa tabela de contingência 2×2 .

Tabela 12 - Dados de sobreviventes e mortes em dois grupos para cada tempo t_i

Grupo	Nº de mortes no tempo t_i	Nº de sobreviventes no tempo t_i	Nº de indivíduos em risco em t_i
1	d_{1i}	$n_{1i} - d_{1i}$	n_{1i}
2	d_{2i}	$n_{2i} - d_{2i}$	n_{2i}
Total	d_i	$n_i - d_i$	n_i

Seja:

- d_{ji} é o número de mortes no tempo t_i para amostra $j, 1 \leq j \leq G$.
- n_{ji} é o número de indivíduos em risco no tempo t_i para amostra $j, 1 \leq j \leq G$.
- d_i é o número de morte no tempo t_i (amostra conjunta) e é definida por:

$$d_i = \sum_{j=1}^G d_{ji} \quad (2.54)$$

- n_i o número de indivíduos em risco no tempo t_i , é definida por:

$$n_i = \sum_{j=1}^G n_{ji} \quad (2.55)$$

- Sob a hipótese nula espera-se que $\frac{d_{ji}}{n_{ji}} \cong \frac{d_i}{n_i}, 1 \leq j \leq G e 1 \leq i \leq r$.

Mantel & Haenszel (1959) consideraram a distribuição das frequências observadas de cada célula, dados os totais marginais, sob a validade de hipótese nula. Então, supondo que H_0 é verdadeira, a distribuição de d_{1i} , condicional aos valores marginais é hipergeométrica e é dada por:

$$p(d_{1i}|d_i, n_i) = \frac{\binom{d_i}{d_{1i}} \binom{n_i - d_i}{n_{1i} - d_{1i}}}{\binom{n_i}{n_{1i}}} \quad (2.56)$$

O valor médio e a variância condicionada a d_{1i} são respetivamente:

$$e_{1i} = \frac{n_{1i}d_i}{n_i} \quad (2.57)$$

$$v_{1i} = \frac{n_{1i}n_{2i}d_i(n_i - d_i)}{n_i^2(n_i - 1)} \quad (2.58)$$

Sendo e_{1i} o número esperado de mortes no instante t_i no grupo 1. Sob a validade da hipótese nula, a probabilidade de morte no instante t_i não depende do grupo no qual o indivíduo se encontra, logo a probabilidade de morte no instante t_i é $\frac{d_i}{n_i}$, se multiplicamos esta probabilidade pelo número de indivíduos no grupo 1, obtemos o valor esperado para o grupo 1.

O passo final é, analisar as diferenças entre o número observados de indivíduos que morreram nos dois grupos em cada instante e o número esperado de mortes sob a validade de hipótese nula.

A estatística de teste de log-rank para comparar o tempo de sobrevivência em dois grupos é dada por:

$$W_L = \frac{U_L^2}{V_L}, \text{ em que } U_L = \sum_{i=1}^r (d_{1i} - e_{1i}) \text{ e } V_L = \text{Var}(U_L) = \sum_{i=1}^r v_{1i} \quad (2.59)$$

Onde,

$$\frac{U_L}{\sqrt{V_L}} \sim N(0,1) \text{ e } W = \frac{U_L^2}{V_L} \sim X_1^2$$

É de realçar que W_L , estatística de teste log-rank tem distribuição de Qui-quadrado com 1 grau de liberdade.

Rejeita-se a hipótese nula (igualdade de curva de sobrevivência) se $W_L > X_{1-\alpha}^2$, em que $X_{1-\alpha}^2$ representa o quantil de probabilidade $1 - \alpha$ da distribuição Qui-quadrado com um grau de liberdade.

Também deve-se realçar que a forma de colocar as informações na tabela de contingência 2×2 para cada instante de morte, foi proposta por Mantel & Haenszel (1959) e é conhecido pelo teste de Mantel-Haenszel ou teste log-rank.

Para o caso de $G > 2$ (mais de que 2 grupo) Teste log-rank

$$U_L = \sum_{i=1}^r \left[d_{ji} - n_{ji} \frac{d_i}{n_i} \right], 1 \leq j \leq G - 1 \quad (2.60)$$

E a matriz de variância e covariância $(G - 1) \times (G - 1)$ é dada por:

$$V_L = \sum_{i=1}^r \frac{n_{ji} d_i (n_i - d_i)}{n_i (n_i - 1)} \left(\delta_{jk} - \frac{n_{ki}}{n_i} \right) \quad (2.61)$$

Onde,

$$\delta_{jk} = \begin{cases} 1 & \text{se } j = k \\ 0 & \text{se } j \neq k \end{cases}$$

Deve ser enfatizado que o teste de log-rank tem maior potência quando se verifica o pressuposto de riscos proporcionais, ou seja, quando existe evidência que as funções de risco não se cruzam. Este pressuposto será abordado mais à frente quando introduzirmos o modelo de regressão de Cox.

Consideremos a base de dado **aml** disponível na biblioteca **survival** do *software* estatístico **R**. Vamos recorrer a esta base de dados para ilustrar o processo de comparação de curvas de sobrevivência em dois grupos, recorrendo ao teste log-rank.

A base de dados **aml** é uma base de dados referente a doentes com leucemia para os quais foram registadas três variáveis: time - “tempo de sobrevivência ou censura”, status - “variável indicador do evento ou censura” e x - “variável binária que representa se o paciente fez ou não o tratamento”.

Primeiramente, vamos efetuar uma representação da curva de sobrevivência de Kaplan-Meier para verificar de uma forma intuitiva, se as curvas de sobrevivência diferem para os dois tipos de tratamento (ver o output do software R no Anexo I).

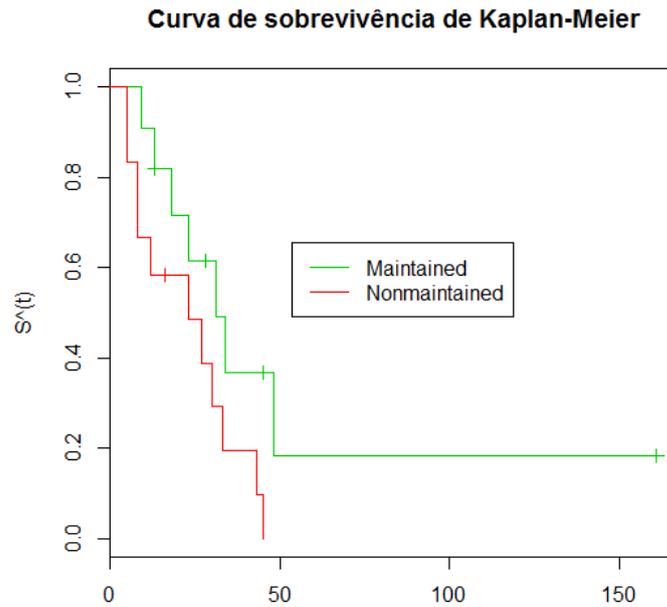


Figura 7 - Curva de sobrevivência estimada de Kaplan-Meier para os dois tipos de tratamento

A Figura 7, indica diferenças entre as duas curvas de sobrevivência, mas sugere que esta diferença poderá não ser estatisticamente significativa.

A representação gráfica é uma forma intuitiva, mas pouco adequada para avaliar se existem ou não diferenças significativas entre as curvas de sobrevivência para os dois tipos de tratamento.

As formas mais apropriadas de avaliar se existem diferenças significativas entre as curvas de sobrevivência consiste na utilização do teste de hipóteses anteriormente introduzido, neste caso para comparar se as duas curvas de sobrevivência diferem.

Hipótese a testar:

H_0 : As curvas de sobrevivência são iguais para os dois tratamentos

Versus

H_1 As curvas de sobrevivência não são iguais para os dois tratamentos

A função do R utilizada para comparar se as curvas de sobrevivência diferem é o `survdiff`.

Aplicando o teste de log-rank no *software* R, obtemos o seguinte resultado.

Tabela 13 - Teste log-rank

Estadística de teste Qui-quadrado X^2	P-valor	Grau de liberdade gf
$X^2 = 3.4$	0.0653	1

Para um nível de significância de $\alpha = 5\%$, verifica-se que a covariável tratamento, quando considerada isoladamente, não tem influência significativa na sobrevivência dos doentes com leucemia, ou seja, como $p - value = 0.0653 > 0.05$, não se rejeita a hipótese nula e conclui-se que não há evidência estatística de que as curvas de sobrevivência diferem para os dois tipos de tratamento (ver o output do teste em Anexo II).

Pela análise da Tabela 13, verifica-se que os graus de liberdade representam o número de grupo da covariável menos um (no caso em análise, como a variável tratamento tem dois níveis, ficamos com um grau de liberdade). A estatística de teste de Qui-quadrado foi calculada com base na expressão apresentada anteriormente. O valor de prova foi calculado com base na estatística de Qui-quadrado com 1 grau de liberdade, ou seja, pode ser calculada no *software R* da seguinte forma, $1 - pchisq(3.4, 1)$.

Convém realçar que ao usarmos a função **survdiff** para comparar as curvas de sobrevivência quando especificamos o argumento $rho = 0$, o *software R* efetua o teste de log-rank ou Mantel-Haenszel, que é o teste que o *software R* faz por defeito. No caso $rho = 1$ faz o teste de Gehan-Wilcoxon.

2.9 MODELO DE REGRESSÃO DE COX

Em análise de sobrevivência, interessa muitas vezes estudar a relação entre as covariáveis ou variáveis explicativas e o tempo de vida. Uma forma adequada de o fazer é através de modelos de regressão. Se no caso ajustarmos os dados a uma distribuição conhecida, estaremos perante um modelo de regressão paramétrico. Porém, em várias situações, existem dúvidas sobre a distribuição adequada para os nossos dados. Neste contexto, o modelo mais utilizado na literatura é o chamado modelo de regressão semiparamétrico de riscos proporcionais de Cox, sendo abreviadamente designado por modelo de regressão de Cox (David Cox, 1972).

2.9.1 FORMULAÇÃO DO MODELO DE REGRESSÃO DE COX

Os modelos de regressão de Cox geralmente são utilizados para analisar o tempo até ocorrência de um determinado evento de interesse. Estes modelos estimam a função de risco, onde a função de risco depende de um conjunto de covariáveis.

David Cox (1972) propôs um modelo em que a função de risco no instante t é definida por:

$$h(t; Z) = h_0(t) \exp(\beta_1 Z_1 + \dots + \beta_k Z_k), \quad (2.62)$$

onde $h_0(t)$ representa a função de risco suporte, não negativa, ou seja, representa a função de risco associado a um indivíduo, com vetor de covariáveis nulo, $Z = 0$, Z_i , representa o vetor das covariáveis associadas a esses indivíduos e β_i representa os coeficientes de regressão para as k covariáveis.

É de salientar que o modelo de regressão de Cox é conhecido como modelo de riscos proporcionais, pelo fato da razão da função de risco para dois grupos com vetores de covariáveis fixos ser constante no tempo, ou seja, não depende do tempo. O modelo de regressão de Cox requer o pressuposto de risco proporcionais, ou seja, este modelo só é válido se for verificado o pressuposto de riscos proporcionais, isto é, o risco relativo para dois vetores de covariáveis Z_1 e Z_2 é constante no tempo,

$$\frac{h_0(t; Z_1)}{h_0(t; Z_2)} = \frac{h_0(t) \exp(\beta Z_1)}{h_0(t) \exp(\beta Z_2)} = \exp\{\beta(Z_1 - Z_2)\} \quad (2.63)$$

Como se pode constatar pela equação em cima, a função de risco para dois quaisquer indivíduos é constante ao longo do tempo (não depende do tempo t). Notemos ainda, que embora o efeito das covariáveis seja especificado parametricamente, a função $h_0(t)$ não é especificada, por esse motivo o modelo de Cox é denominado de modelo semiparamétrico, de risco proporcional.

As covariáveis têm um efeito multiplicativo na função de risco, pelo fato da função de risco depender das covariáveis através do fator $\exp(\beta Z)$, que é designado de risco relativo.

2.9.2 ESTIMAÇÃO NO MODELO DE COX

Como no modelo de regressão de Cox, a função de risco suporte $h_0(t)$, não é especificada, nesta situação não é possível utilizar a função de verosimilhança usual para estimar o vetor dos parâmetros β . David Cox (Cox, 1972) construiu uma função de verosimilhança que não depende da especificação de $h_0(t)$, permitindo assim a realização de inferência sobre β , sem que seja necessário fazer qualquer restrição à forma de $h_0(t)$.

Sejam $t_{(1)} < \dots < t_{(r)}$ os r tempos de vida observados e R_j o conjunto de risco no instante $t_{(j)}$, ou seja, o conjunto de indivíduos em risco imediatamente antes do instante $t_{(j)}$.

A função de verosimilhança parcial proposta por David Cox (1972) é definida por:

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta Z_j)}{\sum_{l \in R_j} \exp(\beta Z_l)}, \quad (2.64)$$

Em que Z_j é o vetor de covariáveis associado ao indivíduo que observou o evento no tempo $t_{(j)}$.

Como se consta, as observações censuradas não interferem diretamente na estimação do vetor β , no entanto a ocorrência de censura é importante para a identificação correta do conjunto de indivíduos em risco. Se o indivíduo j tem tempo de vida censurado c_j , com $c_j = t_{(j)}$ para algum j , então é considerado um indivíduo em risco em $t_{(j)}$ e é desprezado daí em diante. Verifica-se que a probabilidade condicionada de 'morte' em $t_{(j)}$ depende de toda a informação anterior a t_j , através da identificação do conjunto de risco, mas não depende dos valores e da distribuição das características ordinais $t_j, j = 1, \dots, r$. De fato, a função não se altera se forem aplicadas transformações monótonas no tempo. David Cox (1975) denominou esta função por função de verosimilhança parcial. Embora esta função não seja uma função de verosimilhança no sentido usual, ela permite obtenção de estimadores para β que verifica as propriedades usuais dos estimadores de máxima verosimilhança. Assim, $\hat{\beta}$ é um estimador consistente, assintoticamente normal com valor médio β e com matriz de covariância $I(\beta)^{-1}$, em que $I(\beta)$ é a matriz de informação de Fisher, definida por:

$$I(\beta) = - \left[E \left(\frac{\partial^2 \log L}{\partial \beta_j \partial \beta_k} \right) \right]_{p \times p} \quad (2.65)$$

Os coeficientes de regressão β_j são estimados pelos valores de $\hat{\beta}_j$ que maximizam a função verosimilhança parcial $L(\beta)$ ou, de forma equivalente dessa função verosimilhança. O sistema de equação verosimilhança pode ser resolvido através do método iterativo de Newton-Raphson.

Apesar de T ser uma variável contínua, é possível que existam observações repetidas provavelmente por causa de arredondamento, uma vez que a probabilidade de ocorrência de ligações, num modelo contínuo é nula. Desta forma, a função de verosimilhança deve ser

modificada. Então, no caso em que tenha ocorrido evento de interesse de mais do que um indivíduo em simultâneo originando valores iguais, neste caso a função de verossimilhança parcial não é adequada. Nesta situação temos de arranjar outras alternativas mais adequadas.

Breslow (1974) propôs a seguinte aproximação da função da verossimilhança parcial, definida por:

$$L(\beta) = \prod_{j=1}^r \frac{\exp(S_j \beta)}{\left[\sum_{l \in R_j} \exp(Z_l \beta) \right]^{d_j}} \quad (2.66)$$

Em que, d_j representa o número de indivíduos que observou o evento de interesse no instante $t_{(j)}$ e em que $S_j = \sum_{l=1}^{d_j} Z_{jl}$ representa a soma dos vetores de covariáveis dos d_j indivíduos.

Deve-se destacar que, esta aproximação proposta por Breslow (1974) só é válida se o número d_j de indivíduos que se observaram no evento de interesse no instante $t_{(j)}$ for pequeno, quando comparado com o número de indivíduos que pertencem ao conjunto de risco R_j .

2.9.3 SELEÇÃO DE COVARIÁVEIS NO MODELO DE COX

Sempre que, pretendemos desenvolver o modelo de regressão para análise dos dados procuramos construir modelos que sejam úteis e que se ajustem bem aos nossos dados e que sigam o princípio da parcimónia, ou seja, que utilizem o menor número possível de covariáveis. É neste sentido que se levantam inúmeras questões sobre quais as covariáveis a incluir no nosso modelo. No entanto, quando se realiza um estudo de sobrevivência em parceria com uma instituição de saúde, deve-se seguir as suas orientações e ainda deve-se consultar as informações que se encontram disponíveis na literatura, deve-se utilizar as ferramentas estatísticas adequadas para essas escolhas.

Nesta secção serão abordadas técnicas estatísticas que nos vão auxiliar na seleção das covariáveis a incluir no modelo.

2.9.3.1 TESTES E HIPÓTESES

Estamos interessados em identificar quais as covariáveis que têm influências significativas na sobrevivência dos indivíduos, entre todas as que foram registadas.

Seja T uma variável aleatória contínua que representa o tempo de vida e $Z = Z_1, \dots, Z_k$ o vetor das covariáveis associado a cada indivíduo.

O risco de 'morte' de um indivíduo, no instante t é dado por $h(t; Z) = h_0(t) \times g(\beta Z)$, onde $\beta = \beta_1, \dots, \beta_k$ é o vetor dos parâmetros de regressão desconhecidos, h_0 é a função de risco subjacente e g é uma função arbitrária de Z e β .

Pretendemos testar a seguinte hipótese:

$$H_0: \beta = 0 \text{ vs } H_1: \beta \neq 0, j = 1, \dots, k$$

ou seja, na hipótese nula está a testar-se a covariável Z_j não influencia a sobrevivência na presença de outras covariáveis, contra a hipótese alternativa, que essa covariável tem uma influencia estatisticamente significativa para a sobrevivência. Neste caso se rejeitamos a hipótese nula, significa que a covariável em causa é importante ou influencia a sobrevivência. Os métodos de inferência baseiam-se, fundamentalmente, no método da máxima verosimilhança. De acordo com o método da máxima verosimilhança, existem dois testes e hipóteses para avaliar se a covariável em causa influencia ou não a sobrevivência. Sendo elas: o teste de Wald e o teste da razão verosimilhança.

2.9.3.1.1 O TESTE DE WALD

O teste de Wald é baseado na distribuição assintótica de $\hat{\beta}_j$ e é uma generalização do teste T de Student e a sua estatística de teste é dada por:

$$Z = \frac{\hat{\beta}_j - 0}{\sqrt{\text{var}(\hat{\beta}_j)}} \sim N(0,1). \quad (2.67)$$

A estatística Z têm distribuição assintótica normal centrada e reduzida. Neste caso rejeitamos a hipótese nula se o valor da prova for inferior a α (nível de significância). O valor de prova no

software estatístico **R** é obtido da seguinte forma: $2 \times (1 - pnorm(Z))$ (duas vezes por dizer respeito a um teste bilateral). É de realçar que ao rejeitamos a hipótese nula, significa que a covariável influencia a sobrevivência.

2.9.3.1.2 TESTE DA RAZÃO DE VEROSIMILHANÇAS

O teste de razão de verosimilhanças é baseado na comparação dos valores do logaritmo da função de verosimilhança maximizada sob a validade da hipótese nula.

A sua estatística de teste é dada por:

$$LRT = -2\log\left[\frac{L(\text{modelo reduzido})}{L(\text{modelo saturado})}\right] = -2\log(Lr) + 2\log(Ls) \quad (2.68)$$

Onde o Lr é a verosimilhança do modelo reduzido e o Ls é a verosimilhança do modelo saturado. É de salientar que se o modelo reduzido tem k parâmetro o modelo saturado terá $k + r$ parâmetros. A estatística de teste, LRT, segue uma distribuição Qui-quadrado, $LRT \sim X_r^2$, onde o r é o número de parâmetro estimado no modelo ($k + r - k$). De acordo com o teste de razão de verosimilhança rejeita-se a hipótese nula, se o valor da prova for inferior ao nível de significância. O valor de prova no *software* estatístico **R** pode ser calculado da seguinte forma: $1 - pchisq(LRT, r)$.

2.9.4 SELEÇÃO DE UM MODELO ADEQUADO

Os testes anteriormente apresentados avaliam a influência de uma covariável na presença das outras, mas não nos informa qual o melhor modelo. Por essa razão é necessário recorrer a métodos alternativos que permitam fazer essa comparação.

A escolha de um modelo apropriado, do ponto de vista estatístico, é um tópico fundamental na análise de dados, deve-se buscar o modelo mais parcimonioso, ou seja, o modelo que envolve o mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento de uma variável resposta. Dentro dos critérios para a seleção de modelos, os critérios baseados no máximo da função verosimilhança são os mais utilizados, vamos dar mais ênfase ao teste da razão verosimilhança e ao critério de informação de Akaike (AIC).

2.9.4.1 TESTE DA RAZÃO DE VEROSIMILHANÇAS

Para efetuar comparações entre vários modelos, precisamos de alguma medida de qualidade de ajustamento de um determinado modelo. Uma forma de efetuar comparação de modelos é através da função de verosimilhança uma vez que a função de verosimilhança resume informações sobre os parâmetros desconhecidos contida nos dados, uma boa estatística resumo é o valor da função de verosimilhança, quando os parâmetros são substituídos pelas respetivas estimativas da máxima verosimilhança. Assim, para um determinado conjunto de dados, quanto maior for o valor da função de verosimilhança maximizada, melhor será o ajustamento do modelo aos dados. Porém, por uma questão prática a medida que se irá utilizar é $-2\log\hat{L}$, em que \hat{L} representa a verosimilhança maximizada. Esta escolha é devido ao facto de \hat{L} resultar do produto de probabilidade condicionadas, logo o seu valor ser inferior à unidade. Desta forma, $-2\log\hat{L}$ será sempre positivo, quanto menor for essa quantidade melhor será o modelo.

Consideremos um modelo com p covariáveis (modelo 1) e um outro modelo com q covariáveis (modelo 2) com $p < q$, ou seja,

$$\text{modelo 1: } h_0(t)g(\beta_1Z_1 + \dots + \beta_pZ_p)$$

$$\text{modelo 2: } h_0(t)g(\beta_1Z_1 + \dots + \beta_pZ_p + \beta_{p+1}Z_{p+1} + \dots + \beta_{p+q}Z_{p+q}).$$

Pretende-se averiguar se a inclusão das q covariáveis melhora significativamente o ajustamento do modelo aos dados. Por outras palavras, queremos testar a hipótese $H_0: \beta_{p+1} = \dots = \beta_{p+q} = 0$. Para averiguar essa questão vamos utilizar o teste de razão de verosimilhança. A sua estatística de teste de dado por:

$$LRT = -2\log\left(\frac{\hat{L}_p}{\hat{L}_{p+q}}\right) \quad (2.69)$$

Onde \hat{L}_p e \hat{L}_{p+q} são designados de valor da verosimilhança maximizada dos modelos que inclui p e $p + q$ parâmetros, respetivamente.

A sua estatística de teste sob a validade de H_0 , tem distribuição assintótica de Qui-quadrado, onde o seu grau de liberdade é igual à diferença entre o grau de liberdade de cada modelo, ou seja q graus de liberdade.

Devemos salientar que o teste de razão de verossimilhança é útil para comparar dois modelos aninhados, ou seja quando o conjunto de covariáveis de um modelo é subconjunto do outro.

2.9.4.2 CRITÉRIO DE INFORMAÇÃO DE AKAIKE

O Critério de Informação de Akaike (AIC) admite a existência de um modelo real que descreve os dados que é desconhecido, e tenta escolher de entre um grupo de modelos avaliados, o que minimiza a divergência de Kullback-Leibler, ou seja, escolhe o melhor modelo como sendo aquele que apresenta o menor valor de AIC.

A estimativa de AIC para um determinado modelo é dada por: $AIC = -2L + 2r$ em que, o L é o logaritmo da função de verossimilhanças com os parâmetros θ e r é o número do parâmetro do modelo.

2.9.4.3 SELEÇÃO DO MODELO USANDO O TESTE DE WALD

Para selecionar covariáveis de forma sequencial podemos utilizar três métodos distintos: Backward (seleção regressiva) este método incorpora inicialmente todas as covariáveis disponíveis, por etapas cada uma pode ou não ser eliminada. Se por exemplo alguma das covariáveis não for estatisticamente significativa, remove primeiro a covariável que apresente maior valor da prova, assim sucessivamente até que todas as covariáveis sejam estatisticamente significativas. O método de Forward (progressiva) este método começa por considerar o modelo mais simples, com apenas uma covariável. De seguida passa-se a considerar o modelo com duas covariáveis, depois três assim sucessivamente, parando-se quando as covariáveis que se acrescentam não são estatisticamente significativas. Esta metodologia não é o mais adequado, uma vez que, tem o problema da determinação do melhor modelo em cada uma das fases, para além de ser muito dispendioso em termos de cálculos pois envolve a estimação de um número elevado de modelos durante o processo de seleção. Stepwise este método combina os dois métodos anteriores, basicamente é um procedimento forward pois vai adicionar covariáveis uma a uma. Porém, em cada passo é feita uma análise das covariáveis já introduzidas até aí, como forma de garantir que permanecem relevantes após a introdução de novas covariáveis. Este é o método mais completo e adequado dos três apresentados.

2.9.5 VALIDAÇÃO DO PROSSUPOSTO DE RISCOS PROPORCIONAIS

Como foi realçada anteriormente o modelo de Cox só é válido se for verificado o prossuposto de riscos proporcionais, que pode ser verificado através de um gráfico $\log(-\log(S(t)))$ contra t . Assim o prossuposto de riscos proporcionais é verificado se as curvas obtidas forem paralelas, no caso das curvas cruzarem o prossuposto de riscos proporcionais não é válido. Uma outra forma de avaliar o prossuposto de risco proporcional foi proposto por Grambsch & Therneau (1994) baseado em resíduos ponderados. O teste utiliza o seguinte teste de hipóteses:

H_0 : o prossuposto de risco proporcionais é valido

vs

H_1 : o prossuposto de risco proporcionais não é valido

Para validar o prossuposto de riscos proporcionais, tiramos a conclusão com base no valor da prova global, rejeita-se a hipótese nula se valor da prova global for inferior ao nível de significância estabelecido.

Ainda uma outra forma de validar o prossuposto de risco proporcional é através do gráfico dos resíduos de Schoenfeld, proposto por Schoenfeld (1982) são muitos uteis na validação do prossuposto de riscos proporcionais, após ajustamento de um modelo de regressão de Cox.

2.9.5.1 RESÍDUOS DE SCHOENFELD

Schoenfeld (1982) apresenta uma outra definição para resíduos, que permite avaliar a hipótese de riscos proporcionais. Baseia-se a sua construção na estatística “Score”. Segundo Schoenfeld não existe apenas um resíduo para cada indivíduo, mas sim vários, tantos quantas as covariáveis incluídas no modelo. Além disso, não é necessário a obtenção de uma estimativa da função de risco acumulativa.

Para o i -ésimo indivíduo em estudo, o resíduo correspondente à covariável $Z_j, j = 1, \dots, p$ é dado por:

$$r_{ji} = \delta_i(Z_{ji} - a_{ji}) \quad (2.70)$$

Onde $\alpha = \frac{\sum_{l \in R_i} Z_{jl} \exp(Z_l \hat{\beta})}{\sum_{l \in R_i} \exp(Z_l \beta)}$, sendo R_i o conjunto dos indivíduos em risco no instante t_i .

Estes resíduos podem ser interpretados como a diferença entre os valores observados das covariáveis de um determinado indivíduo, cuja morte foi observada em t_i , e é uma média ponderada dos valores dessas covariáveis, para todos os indivíduos em risco em t_i . Os pesos correspondentes a esses indivíduos são $\exp(Z_l \beta)$.

Os resíduos de Schoenfeld têm a particularidade de, para grandes amostras, serem não correlacionados e terem valor esperado nulo. Estes resíduos são necessariamente nulos para um indivíduo cujo tempo de vida seja censurado.

De acordo com Schoenfeld, se o pressuposto de riscos proporcionais for satisfeito, não pode existir nenhuma tendência no gráfico de resíduos de Schoenfeld contra os tempos de vida, ou contra as ordens dos tempos de vida, ou seja, a representação gráfica deverá apresentar o aspeto de uma nuvem de pontos em torno do zero.

É de salientar que, caso o pressuposto de risco proporcional não for verificada não devemos usar modelo de regressão de Cox, temos de arranjar um modelo alternativo, como por exemplo o modelo de tempo de falha acelerado ou extensões do modelo de regressão de Cox, por exemplo, o modelo de regressão de Cox com covariáveis dependentes no tempo, na próxima seção vamos falar sobre esse modelo.

2.10 EXTENSÕES DE MODELO DE COX

Algumas situações práticas envolvendo medidas longitudinais não são ajustadas adequadamente usando-se o modelo de Cox na sua forma original, como apresentado na secção 2.9.1. Existem covariáveis que são monitoradas durante o estudo e seus valores podem mudar ao longo desse tempo. Por exemplo, pacientes podem mudar de grupo durante o tratamento ou, a dose de quimioterapia aplicada em pacientes com cancro pode sofrer alterações durante o tratamento. Se esses valores forem incorporados na análise estatística, resultados mais precisos podem ser obtidos comparados a aqueles que fazem uso somente das mesmas medidas registadas no início de estudo. Em outros exemplos, as exclusões desses valores podem acarretar sérios problemas. Este tipo de covariável é chamado de dependente de tempo e o modelo de Cox pode

ser estendido para incorporar as informações longitudinais registradas para estas covariáveis (COLOSIMO & GIOLO, 2005).

Em outras situações, quando o pressuposto de riscos proporcionais é violado e o modelo de Cox não é adequado, existem modelos alternativos para resolver esta situação. Um deles é a extensão do próprio modelo de Cox, denominado de modelo de Cox estratificado. Este modelo supõe-se que os riscos são proporcionais em cada estrato, mas não entre estratos (COLOSIMO & GIOLO, 2005).

O objetivo deste capítulo é essencialmente apresentar duas generalizações do modelo de Cox bastante úteis em situações práticas, a modelagem envolvendo covariáveis dependentes do tempo e o modelo estratificado. Tais generalizações são apresentadas nas seções 2.10.1 e 2.10.2 respetivamente.

2.10.1 MODELO DE REGRESSÃO DE COX COM COVARIÁVEIS DEPENDENTES NO TEMPO

Neste capítulo veremos como incluir covariáveis que mudam ao longo do tempo, situação relativamente comum em estudos de sobrevivência. Durante o período de acompanhamento de uma coorte, algumas variáveis dos indivíduos não mudam ao longo do tempo, por exemplo o gênero e a raça. Outras, como idade, tratamento, residências, hábitos podem mudar, sendo necessário adaptar os modelos a tal situação. Assim sendo, uma forma de resolver este problema consiste em fazer a modelação de covariáveis dependentes no tempo utilizando a abordagem de processo de contagem, ainda que a construção da base de dados exija cuidado (COLOSIMO & GIOLO, 2005).

O fenómeno em estudo deve ser registado de forma mais próxima possível da realidade, com o acompanhamento da coorte e registro de todas as ocorrências sofridas pelos indivíduos durante o período de observação. O modelo de Cox que incorpora covariáveis dependentes no tempo é denominado de **Modelo de Regressão de Cox com Covariáveis Dependentes no Tempo**. Este modelo tem as mesmas premissas do modelo de Cox e tem a forma genérica dada por:

$$h(t; Z) = h_0(t) \exp(Z(t)\beta), \quad (2.71)$$

onde $Z(t)$ é o valor da covariável no tempo t .

Definido desta forma, a equação anterior deixa de verificar o pressuposto de riscos proporcionais, pois os valores dos vetores de covariáveis Z_1 e Z_2 dependem do tempo t e a razão das funções de risco no tempo t para dois indivíduos diferentes fica sendo:

$$\frac{h(t; Z_1)}{h(t; Z_2)} = \exp(\beta(Z_1(t) - Z_2(t))) \quad (2.72)$$

que é dependente no tempo. A interpretação dos coeficientes β do modelo deve considerar o tempo t . Cada coeficiente β_l , para $l = 1, \dots, p$, pode ser interpretado como o logaritmo da razão de riscos cujo o valor da l – ésima covariável no tempo t difere de uma unidade, quando os valores das outras covariáveis são mantidos fixos neste tempo (COLOSIMO & GIOLO, 2005).

As estimativas dos parâmetros do modelo de regressão Cox com covariáveis dependentes no tempo, pode ser obtido estendendo-se o logaritmo da função verossimilhança parcial para:

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta Z_j(t_j))}{\sum_{l \in R_j} \exp(\beta Z_l(t_j))} \quad (2.73)$$

A expressão acima apresentada é uma extensão do estimador da máxima verossimilhança parcial apresentada anteriormente. Este estimador tem propriedades assintóticas dos estimadores da máxima verossimilhança parcial, para que se possa construir intervalo de confiança e testar hipóteses sobre os coeficientes do modelo, foram obtidos por Andersen e Gill em 1982. Eles apresentaram provas bastantes gerais das propriedades para o modelo de Cox incluindo covariáveis dependentes no tempo, também conseguirem provar que esses estimadores são consistentes e assintoticamente normais sob certas condições e regularidade. Desta forma, pode-se usar as conhecidas estatísticas de Wald e da razão de verossimilhanças para a realização de inferências sobre os parâmetros do modelo de regressão de Cox com covariáveis dependentes do tempo (COLOSIMO & GIOLO, 2005).

Para finalizar o nosso estudo de modelo de Cox com covariáveis dependentes no tempo, vamos citar um exemplo bastante analisado na literatura é o do programa de transplante de coração de Stanford (Crowley, Hu, 1977). Neste estudo, os pacientes eram aceites no programa quando se tornavam candidatos a um transplante de coração. Quando surgia um doador, os médicos

escolhiam, de acordo com alguns critérios, o candidato que iria receber o coração. Alguns pacientes morreram sem receber o transplante. A forma de locação estava fortemente viciada na direção daqueles pacientes com maior tempo de sobrevivência, pois somente estes pacientes viveram suficiente para receber o coração. O uso de uma covariável, assumindo o valor zero para aqueles esperando o transplante e um para aqueles com coração novo, serve para minimizar este problema. Esta covariável muda de valor assim que o transplante é realizado e é, portanto, dependente do tempo (Maria Sá Carvalho, Andreozzi, Codeço, Barbosa, & Shimakura, 2005).

2.10.2 MODELO DE COX ESTRATIFICADO

Nesta secção, devemos salientar que se o pressuposto de riscos proporcionais for violado, não devemos usar o modelo de Cox na sua forma original. Nesses casos, caso o problema esteja numa variável qualitativa, uma solução para o problema consiste na estratificação da covariável de modo que o pressuposto de riscos proporcionais seja válido em cada estrato. Por exemplo, os riscos podem não ser proporcionais entre homens e mulheres, mas o tal pressuposto pode ser verificado no estrato formado somente por homens e naquele formado somente por mulheres (COLOSIMO & GIOLO, 2005).

A análise estratificada consiste em dividir os dados da sobrevivência em m estratos, de acordo com uma indicação da violação do pressuposto. Devemos salientar que a variável que estratifica o risco basal não poderá ter o seu efeito estimado.

Numa situação genérica em que existem s estratos, o modelo estratificado para estrato j é definido por:

$$h_j(t) = h_{0j}(t)\exp(Z\beta), j = 1, \dots, s. \quad (2.74)$$

A estratificação não cria nenhuma complicação na estimação do vetor de parâmetros β . Uma função de verossimilhança parcial, como apresentada na equação (2.64) é construída para cada estrato e a estimação do vetor β é baseada na soma dos logaritmos das funções de verossimilhança parciais, isto é, em:

$$l(\beta) = [l_1(\beta) + \dots + l_s(\beta)] \quad (2.75)$$

com $l_j(\beta) = \log(L_j(\beta))$ obtida usando-se somente os dados nos indivíduos do j-ésimo estrato (Kalbfleisch & Prentice, 1980). As derivadas para a equação acima apresentada são encontradas por meio da soma das derivadas obtidas para cada estrato e então, $l(\beta)$ é maximizada com respeito a β , de modo análogo, apresentada na seção 2.9.2. As propriedades assintóticas dos estimadores são obtidas a partir dos estimadores do modelo não estratificado (Maria Sá Carvalho et al., 2005)

Note-se que o modelo de Cox estratificado, assume que as covariáveis atuam de modo similar na função de risco de base de cada estrato, ou seja, assume-se que os coeficientes de regressão são os mesmos em todos os estratos, embora o risco da base varie (COLOSIMO & GIOLO, 2005).

Devemos salientar, que o modelo de Cox estratificado deve ser usado somente em caso realmente necessário, ou seja, na presença de violação do pressuposto de riscos proporcionais. O uso desnecessário de estratificação acarreta em uma perda de eficiência das estimativas obtidas (Maria Sá Carvalho et al., 2005).

Deve-se realçar que a violação do pressuposto de riscos proporcionais no modelo de regressão de Cox, muitas vezes deve-se ao facto de introduzir o efeito das covariáveis de forma linear (ou log-linear) quando na verdade a covariável tem um efeito não linear. Por esse motivo, sempre que é violado o pressuposto de riscos proporcionais, deve-se verificar primeiramente se a covariável que levou a rejeição do pressuposto de riscos proporcionais tem efeito não linear. A função no *software* R que permite verificar se uma variável quantitativa tem um efeito não linear é o *pspline*. Em caso de uma variável qualitativa ou categórica quando é violado o pressuposto de riscos proporcionais podemos estratificar essa variável de modo que esse pressuposto seja válido para cada estrato.

CAPÍTULO 3

METODOLOGIA DO ESTUDO

3.1. DESENHO DE ESTUDO

Há dois tipos de estudos que podem utilizar o tempo como variável de interesse. Um deles é o **estudo experimental (ensaios clínicos controlados aleatorizados)**, indicado para avaliar formas de tratamento. Outro tipo são os **estudos de coorte observacionais**, cujos dados podem ser obtidos pela coleta direta em prontuários médicos ou em bases de dados já existentes (dados secundários). Essas fontes de dados secundários podem ser de base hospitalar (registros hospitalares de cancro) ou populacional (registros de cancro de base populacional) (Bustamante-Teixeira, Faerstein, & Latorre, 2002).

O presente estudo é um estudo de **coorte prospetivo**, visto que os dados foram obtidos a partir de consulta do processo dos doentes já existente em processos médicos e são todas as mulheres com diagnóstico de cancro do CCU.

Compreende-se que os estudos de coorte são estudos observacionais em que a situação dos participantes quanto à exposição de interesse determina a sua seleção para o estudo, ou a sua classificação após a inclusão no estudo. Para isso, os indivíduos são monitorados ao longo do tempo para avaliar a incidência de doença ou de outro desfecho de interesse (ROBERTO A. MEDRONHO, 2009).

O estudo de coorte mostra-se interessante porque pode ser conduzido com diversas finalidades, incluindo a avaliação da etiologia de doenças ou da sua história natural e o estudo do impacto de fatores prognósticos, intervenções diagnósticas e terapêuticas (ROBERTO A. MEDRONHO, 2009).

A análise de sobrevivência é utilizada quando o tempo é o objeto de interesse, seja esse interpretado como o tempo até a ocorrência de um evento ou o risco de ocorrência de um evento por unidade de tempo. Desta forma, estes estudos procuram enfatizar e detalhar ao máximo possível a informação relacionada com o tempo (Marília Sá Carvalho et al., 2011).

3.2. DESCRIÇÃO DA BASE DE DADOS UTILIZADA

Foram analisados dados referente ao tempo de vida de 324 doentes do RORC do IPO de Coimbra e que foram diagnosticadas com CCU durante o período de 1 de janeiro de 2009 a 31

de dezembro de 2010. As doentes foram seguidas, pelo menos por um período de 5 anos e 3 meses.

Foram considerados como critérios de inclusão todas as mulheres com idades compreendidas entre os 25 e 64 anos (idade de rastreio do CCU), com registo de CCU confirmado pelo RORC por qualquer uma das fontes, com exceção do Registo Oncológico Regional do Norte (RORENO) e do Registo Oncológico Regional do Sul (RORSUL), com diagnóstico a partir de 1 de janeiro de 2009 a 31 de dezembro de 2010.

Para poder consultar os arquivos das doentes no RORC, foi preciso submeter um pedido formal ao diretor do RORC. A coleta dos dados foi realizada nos meses de março a setembro de 2016, após aprovação da comissão de ética de ARSC.

Para a coleta dos dados utilizou-se um ficheiro de Excel disponibilizado pelo RORC. O ficheiro disponibilizado continha algumas informações necessárias para o presente estudo nomeadamente: a data do nascimento da mulher na primeira consulta, estado da doente (vivo ou óbito por CCU), distrito de residência da doente, data do diagnóstico, o comportamento do tumor, localização topográfica, histologia do cancro, grau do tumor e a variável fonte que indica em que hospital a doente tem o processo. Para complementar as informações do RORC, foi necessário consultar os arquivos das doentes uma a uma para a recolha das informações necessárias, para estudar quais as variáveis explicativas que influenciam o tempo de sobrevivência. No entanto nem todas as doentes com registo do CCU no RORC de IPO de Coimbra, tinham o processo no IPO, em alguns casos foi necessário contactar com outros hospitais para complementar a nossa base de dados e esse processo revelou-se bastante demorado.

O passo seguinte consistiu em comparar a data de follow-up do indivíduo em estudo, com a sua data de diagnóstico (ou seja, a sua data de entrada em estudo). Foi estipulado pelo RORC que a data de saída do estudo seria a 31/03/2016. Por exemplo para um indivíduo diagnosticado com CCU em 2009 ou 2010 e que no final de estudo se encontra vivo, a sua data de follow-up foi inserida como 31/03/2016, mas no caso de ocorrer óbito por CCU antes da data do término do estudo, o tempo de follow-up será inserido como a data do óbito. Mas no caso de ocorrer morte por outras causas antes do final do estudo, o seu tempo de follow-up será inserido como data de óbito e a respetiva observação é considerado uma censura pela direita.

De seguida, através da função *DATADIF(Data_diag; Data_saida; "d")* do Excel, criou-se a variável “tempo de follow-up” que se encontra em dias, que é a diferença entre a data de entrada no estudo e a data de saída ou data de óbito.

3.3. TIPO DE VARIÁVEL ENVOLVIDO EM ESTUDO

Nesta seção vão ser definidos os tipos de variáveis presente no referente estudo.

3.3.1 VARIÁVEL DEPENDENTE OU VARIÁVEL RESPOSTA

É importante saber que qualquer que seja o tipo de estudo epidemiológico, geralmente há uma variável de interesse também chamada de variável dependente ou variável resposta. Essa variável pode ser o número de casos de determinada doença, a sua incidência, a sua probabilidade de ocorrência ou outra medida que viesse descrever a frequência com que a doença ocorre (Bustamante-Teixeira et al., 2002).

No nosso caso em particular a variável dependente é o tempo até à morte, que como já foi referido anteriormente foi calculada a partir da data de entrada no estudo (data diagnóstico) e a data de óbito e /ou data da última consulta. Tempo de vida de mulheres que continuaram vivas após o término do estudo (31/03/2016) foram consideradas censuras.

3.3.2 VARIÁVEL INDEPENDENTE

As variáveis independentes ou também chamadas de covariáveis são variáveis que podem ou não influenciar a variável de interesse.

As variáveis independentes surgiram de acordo com carácter socioeconómicos, demográfico, da mulher e também de acordo com a caracterização do tumor.

3.4 DESCRIÇÃO DAS VARIÁVEIS ENVOLVIDAS NO ESTUDO

Nesta seção serão apresentadas e descritas as informações de todas as variáveis recolhidas para o presente estudo.

Tabela 14 - Descrição das variáveis envolvidas no estudo

Variável	Tipo	Definição	Categorização
Rastreio do CCU (Rastreio_uterio)	Independente	Variável que indica se a mulher fez ou não fez o rastreio antes do diagnóstico de CCU	0-"Não Rastreio" 1- "Rastreio"
Estado (status)	Dependente	Variável que indica se o paciente está vivo ou ocorreu óbito por CCU	0- "Censura" 1-"Óbito"
Estadiamento (Estadio)	Independente	Variável que indica o estadiamento da doença à data do diagnóstico do CCU	1-"Estádio I" 2-"Estádio II" 3-"Estádio III"
Infeção por HPV (Infecao HPV)	Independente	Variável que indica se o doente teve ou não teve infeção por HPV	0-"Sem Infeção" 1-"Com infeção"
Diabetes	Independente	Variável que indica se o paciente teve ou não teve diabetes	0-" Não diabetes" 1-" Diabetes"
Obesidade	Independente	Variável que indica se o paciente teve ou não teve obesidade	0- "Não obesidade" 1-"Obesidade"
Antecedentes de familiares com cancro (Antecedentes familiares)	Independente	Variável que indica se o paciente tem ou não familiares com alguns tipos de cancro	0-"Sem antecedentes" 1-" Com antecedentes"
Ano do diagnóstico do CCU (Ano_diag)	Independente	Variável que indica o ano do diagnóstico do CCU do paciente	1-"Ano 2009" 2-"Ano 2010"
Fumador	Independente	Variável que indica se o paciente fuma ou não fuma	0-"Não fuma" 1-" Fuma"
Sintomas do CCU (Sintomas)	Independente	Variável que indica se o paciente teve ou não sintomas	0- "Assintomática" 1-"Sintomática"
Comportamento do cancro (Comp_cancro)	Independentes	Variável que indica a maneira como o tumor se comporta no corpo	2- "In situ ou não invasivo" 3-"Maligno, localização primária"

Continuação da Tabela 14

Variável	Tipo	Definição	Categorização
Localização Topográfica (loc_top)	Independentes	Variável que indica o local de origem do tumor	1-"C53.0 - Neoplasia maligna do endocérvix"
			2-C53.1 - "Neoplasia maligna do exocérvix"
			3-C53.8 - "Neoplasia maligna do colo do útero com lesão invasiva"
			4-C53.9 -" Neoplasia maligna do colo do útero, não especificado"
Grau do cancro (Grau)	Independente	Variável que indica a diferenciação ou grau histológico do cancro	1-Grau 1 - "Bem diferenciado"
			2-Grau 2 - "Moderadamente diferenciado"
			3-Grau 3 - "Pouco diferenciado"
			9-Grau 9 - "Não determinado"
Nível de escolaridade (Nível_escol")	Independente	Variável que indica o nível de escolaridade do paciente	1- "Se o paciente tem o 1º ciclo (1º a 4º ano)"
			2- "Se o paciente tem o 2º ciclo (5º e 6º ano)"
			3- "Se o paciente tem o 3º ciclo (7º,8º e 9º ano)"
			4- "Secundário (foram incluído os com 10º a 12º ano)"
			5- "Licenciatura (foram incluídos os pacientes com licenciatura e bacharelato)"
Estado civil (Estado_civil)	Independente	Variável que indica o estado civil do paciente	0- "Solteira"
			1-"Casada"
			2-"Viúva"
			3-"Divorciada"
Resultado do Rastreio(res_rastreio)	Independente	Variável que indica se o resultado de rastreio foi normal ou anormal	0-"Normal"
			1-"Anormal"
Contraceção (contracecao)	Independente	Variável que indica se a mulher fez ou não fez contraceção	0-"Não faz contraceção"
			1-"Faz contraceção"
Distrito de residência (Distrito)	Independente	Variável que indica o distrito de residência dos pacientes	1-"Aveiro"
			2-"Castelo Branco"
			3- "Coimbra"
			4-" Guarda"
			5-"Leiria"
			6- "Viseu"

Continuação da Tabela 14

Variável	Tipo	Definição
Tempo de follow-up (em dias) (temp_followp)	Dependente	Variável que resulta da diferença entre a primeira data registada e a última, isto é, trata-se do tempo de follow-up, ou período em que o indivíduo esteve em estudo
Data do diagnóstico (Data_diag)	Data	Variável que indica a data do diagnóstico do CCU do paciente
Data de saída(Data_saida)	Data	Variável que indica a data do termino do estudo ou data de óbito por CCU
Data do Rastreio (Data_rastreio)	Data	Variável que indica a data em que a paciente foi submetida ao rastreio de CCU
Idade da 1º relação sexual (1º_relacao)	Independente	Variável que indica a idade da 1º relação sexual
Idade da Menarca (Menarca)	Independente	Variável que indica a idade da paciente na 1ª menstruação
Idade na 1º gravidez (Idade_1º gravidez)	Independente	Variável que indica a idade da paciente na 1º gravidez
Nº de parceiros (Parceiros)	Independente	Variável que indica o número de parceiros sexuais da paciente
Nº de partos (Partos)	Independente	Variável que indica o número de partos da paciente
Nº de gestações (Gestacao)	Independente	Variável que indica o número de gestações da paciente
Nº de abortos (Abortos)	Independente	Variável que indica o número de abortos da paciente

3.5 SOFTWARE UTILIZADO

Durante esse trabalho foram utilizados os seguintes *software*:

- **Microsoft Excel 2013**
Utilizou-se este programa uma vez que a base de dados se encontra no formato do Excel e porque é um software de fácil manipulação, ideal para cruzamento de dados, bem como para a construção de tabelas e gráficos.
- **R Versão 3.3.1**

Neste trabalho foi dado mais ênfase ao *software* R na análise de dados de sobrevivência, pelo facto do *software* R ser de livre acesso e de permitir a utilização de mais recente técnica estatística.

Depois de recolha dos dados procedeu-se ao tratamento de dados que foi realizado na maior parte com o *software* R.

3.6 METODOLOGIA UTILIZADA

As estimativas da função de sobrevivência foram obtidas aplicando o **método de Kaplan-Meier**. O teste de **Log-rank** foi utilizado para a comparação das funções de sobrevivência para as variáveis qualitativas. Foram analisados como fatores de risco as seguintes variáveis independentes: idade da mulher, rastreio do CCU, antecedentes familiares com cancro, estado civil, número de partos, número de gestações, número de abortos, idade da menarca, idade da 1^o relação sexual, nível de escolaridade, infeção por HPV, diabetes, obesidade, sintomas do cancro, fumador, contraceção, distrito de residência das mulheres, comportamento do cancro, estadiamento do cancro à data do diagnóstico, idade na 1^o gravidez e o ano de diagnóstico.

De seguida a partir da variável idade da mulher criou-se uma nova variável “Idade_grupo” com 4 categorias para as mulheres com idades entre os [25, 34,]; [35, 44,]; [45, 54] e [55, 64] anos, esta variável foi criada, para averiguar se existe diferença na sobrevivência por faixa etária.

Para avaliar quais as covariáveis que isoladamente influenciam a sobrevivência após o diagnóstico do CCU, utilizou-se o modelo de Cox simples, onde foram calculados a razão de riscos, doravante designada por hazard ratios (HR) e o seu respetivo intervalo com 95% de confiança e o seu p-valor.

Apenas as covariáveis que apresentaram valores de HR estatisticamente significativa (p-valor<0.05) no modelo de Cox simples foram incluídos no modelo de Cox múltiplo. Em todos os testes o nível de significância usada foi de 5%.

É de salientar que foram excluídas algumas covariáveis no modelo de regressão de Cox múltiplo nomeadamente: sintomas, estadiamento do cancro à data do diagnóstico, nível de escolaridade, número de gestação, número de partos e números de abortos, pelo facto dessas variáveis apresentarem vários valores em falta e por levar à não convergência do modelo de regressão de

Cox múltiplo. Como temos variáveis com valor ausente ("NA") não podemos utilizar o teste de razão de verossimilhança nem o critério de AIC para chegar o melhor modelo. A estratégia utilizada para chegar ao modelo final foi seguinte, primeiro foi construído um modelo de Cox simples para todas as covariáveis. De seguida foi construído um modelo de Cox múltiplo com apenas as covariáveis que foram estatisticamente significativas no modelo de Cox simples. No entanto as covariáveis estadiamento do cancro à data do diagnóstico e a variável sintomas do cancro apesar de serem estatisticamente significativas no modelo de Cox simples não foram incluídos no modelo de Cox múltiplo por apresentarem 248 e 257 valores em falta respetivamente. Para chegar a um modelo de Cox múltiplo válido começamos com o modelo com uma única variável e fomos adicionando outras variáveis que se revelaram estatisticamente significativas. Acabamos por excluir algumas variáveis que levaram a não convergência do modelo por apresentarem vários valores em falta, assim chegamos um modelo final válido onde foi verificado a premissa de proporcionalidade de risco. No entanto, não temos a garantia que esse modelo é o melhor modelo, uma outra pessoa pode chegar a um outro modelo, mas, no entanto, é um modelo válido.

CAPÍTULO 4

APRESENTAÇÃO DOS RESULTADOS SOBRE O
ESTUDO DO CCU

4.1 ANÁLISE EXPLORATÓRIA

Nesta seção será apresentada uma breve análise exploratória da base de dados envolvida no estudo, apresentam-se as medidas e os gráficos que são mais apropriados para variáveis quantitativas e qualitativas.

O grande objetivo de uma análise exploratória ou análise descritiva é conhecer a base de dados que pretendemos trabalhar, é o primeiro passo que se deve dar quando se pretende fazer um estudo mais complexo.

O primeiro passo a dar, quando se utiliza o *software* R, é proceder à leitura dos dados, o referido código do R encontra-se no Anexo III.

Após a leitura dos dados é necessário classificar corretamente as variáveis qualitativas e isso faz-se da seguinte maneira (ver Anexo IV).

4.1.1 ANÁLISE EXPLORATÓRIA PARA VARIÁVEIS QUALITATIVAS

As medidas mais adequadas para as variáveis qualitativas são as frequências absolutas e relativas. Os gráficos mais apropriados para esse tipo de variáveis são o gráfico circular e o gráfico de barras.

Para fazer uma análise descritiva rápida no R para uma variável qualitativa, usa-se a função `table`.

De seguida será apresentada a análise descritiva das variáveis qualitativas envolvidas no nosso estudo.

Tabela 15 - Estatística descritiva para as variáveis qualitativas

Variável Categórica	n =324	%	Nº de Óbito=44	% de Óbito
Rastreio do CCU				
0 - Não Rastreio	110	33,95	32	72,73
1 - Rastreio	161	49,69	7	15,91
Valores em falta ("NA")	53	16,36	5	11,36
Estadiamento				
Estádio 1	41	12,65	3	6,82
Estádio 2	11	3,4	2	4,55
Estádio 3	24	7,41	21	47,73
Valores em falta ("NA")	248	76,54	18	40,91
Fumador				
0 - Não fumador	158	48,77	22	50
1 - Fumador	74	22,84	6	13,64
Valores em falta ("NA")	92	28,4	16	36,36
Infeção por HPV				
0 - Não	7	2,16	0	0
1 - Sim	95	29,32	7	15,91
Valores em falta ("NA")	222	68,52	37	84,09
Diabetes				
0 - Não	58	17,9	7	15,91
1 - Sim	12	3,7	4	9,09
Valores em falta ("NA")	254	78,4	33	75
Obesidade				
0 - Não	42	12,96	5	11,36
1 - Sim	18	5,56	5	11,36
Valores em falta ("NA")	264	81,48	34	77,27
Antecedentes familiares				
0 - Não	88	27,16	14	31,82
1 - Sim	105	32,41	12	27,27
Valores em falta ("NA")	131	40,43	18	40,91
Contraceção				
0 - Não	25	7,72	5	11,36
1 - Sim	178	54,94	17	38,64
Valores em falta ("NA")	121	37,35	22	50
Sintomas				
0 - Assintomático	25	7,72	2	4,55
1 - Sintomático	42	12,96	14	31,82
Valores em falta ("NA")	257	79,32	28	63,64
Estado Civil				
0 - Solteira	34	10,49	5	11,36
1 - Casada	135	41,67	19	43,18
2 - Viúva	10	3,09	0	0
3 - Divorciada	20	6,17	2	4,55
Valores em falta ("NA")	125	38,58	18	40,91

Continuação da Tabela 15

Variável Categórica	n =324	%	Nº de Óbito=44	% de Óbito
Nível de escolaridade				
1 - 1º Ciclo	57	17,59	14	31,82
2 - 2º Ciclo	30	9,26%	1	2,27
3 - 3ºCiclo	47	14,51	1	2,27
4 - Ensino Secundário	44	13,58	4	9,09
5 - Licenciatura	25	7,72	5	11,36
Valores em falta ("NA")	121	37,35	19	43,18
Estado				
0 - Vivo	280	86,42		
1 - Óbito	44	13,58		
Distrito de residência				
1 - Aveiro	49	15,12	4	9,09
2 - Castelo Branco	13	4,01	1	2,27
3 - Coimbra	82	25,31	14	31,82
4 - Guarda	19	5,86	1	2,27
5 - Leiria	127	39,2	14	31,82
6 - Viseu	34	10,49	10	22,73
Fonte/instituição de diagnóstico				
C.H. COVA BEIRA - COVILHÃ	2	0,62	0	0
C.H. COIMBRA - MATERNIDADE BISSAYA BARRETO	45	13,89	2	4,55
C.R.O.C. - I.P.O.F.G.	179	55,25	24	54,55
HOSPITAIS UNIVERSIDADE COIMBRA	28	8,64	9	20,45
HOSPITAL AMATO LUSITANO - CASTELO BRANCO	1	0,31	0	0
HOSPITAL DISTRITAL DA FIGUEIRA FOZ	2	0,62	0	0
HOSPITAL INFANTE D. PEDRO - AVEIRO	1	0,31	0	0
HOSPITAL SÃO TEOTÓNIO - VISEU	4	1,23	2	4,55
HOSPITAL SOUSA MARTINS - GUARDA	15	4,63	1	2,27
HOSPITAL STº. ANDRÉ - LEIRIA	35	10,8	6	13,64
MICRODIAG	12	3,7	0	0
Ano Diagnóstico do cancro				
2009	150	46,3	25	56,82
2010	174	53,7	19	43,18

Da Tabela 15, observa-se que das 324 mulheres envolvidas no estudo, 280 foram censuras (continuaram vivas ou morreram por outra causa), que correspondem a 86,42% (280/324) e 44 morreram por CCU, que correspondem a 13,58% (44/324). Das 324 mulheres envolvidas no estudo, 161 que correspondem a 49,69% (161/324) foram submetidas ao rastreio do CCU, 110 que correspondem a 33,95% (110/324) não foram submetidas ao rastreio do CCU e 53 que correspondem a 16,36% (53/324) não têm indicação do rastreio do CCU. Dos 44 óbitos por

CCU, 32 não fizeram rastreio, que correspondem a 72,73% (32/44), 7 fizeram rastreio do CCU que correspondem a 15,91% (7/44) e 5 não tem indicação do rastreio, que correspondem a 11,36% (5/44). Observa-se ainda que a maioria das mulheres diagnosticadas com CCU (127) pertencem ao distrito de Leiria. Pela Tabela 15, verifica-se ainda que das 324 mulheres diagnosticadas com CCU a partir de 1 de janeiro de 2009 a 31 de dezembro de 2010, 150 (46,30%) tiveram diagnóstico em 2009 e 174 (53,70%) tiveram diagnóstico em 2010. A maior percentagem (56,82%) de óbitos foi registada em 2009. Dos 44 óbitos por CCU, 21 óbitos que correspondem a 47,73% foram registados nas mulheres com estadiamento do cancro à data diagnóstico com estágio 3.

4.1.2 ANÁLISE EXPLORATÓRIA PARA VARIÁVEIS QUANTITATIVAS

Para fazer uma análise exploratória das variáveis quantitativas, as medidas mais adequadas são medidas de localização nomeadamente: média, mediana, mínimo, máximo e os quartis e as medidas de dispersão são nomeadamente: variância; desvio padrão e amplitude interquartil.

Os gráficos mais apropriados para as variáveis quantitativas são o diagrama de extremos e quartis (uma representação em caixa com bigodes-boxplot) e o histograma.

O boxplot é útil para ver se existem outliers ou valores atípicos, assim como para identificar uma possível assimetria nos dados. O histograma é uma representação gráfica alternativa que também permite averiguar se existe assimetria nos dados.

Geralmente, quando se pretende fazer uma análise exploratória rápida das variáveis quantitativas no R, utiliza-se a função *summary*. Com este comando obtemos para as variáveis quantitativas os quartis, mínimo, máximo e a mediana.

De seguida, encontra-se apresentada a análise descritiva das variáveis quantitativas envolvidas no nosso estudo.

Tabela 16 - Estatística descritiva para as variáveis quantitativas.

Variável Quantitativa	Média	Mediana	1º Quartil	3º Quartil	Mínimo	Máximo	Desvio Padrão
Tempo de follow-up em dias (n=324)	2064	2184	2010	2435	2	2635	588,77
Idade em anos (n=324)	42,62	41	34	50	25	64	10,75
Nº de partos (n=246)	1,82	2	1	3	0	10	1,53
Nº de gestações (n=243)	2,16	2	1	3	0	10	1,71
Nº de abortos (n=228)	0,36	0	0	1	0	5	0,72
Nº de parceiros sexuais (n=193)	2,11	2	1	3	1	10	1,59
Idade na 1º gravidez (n=137)	22,99	22	19	26	15	42	4,67
Idade da 1º relação sexual (n=220)	18,64	18	17	20	9	31	2,77
Idade da menarca (n=241)	12,84	13	12	14	9	19	1,75

Na Tabela 16, encontra-se apresentada a estatística descritiva para as variáveis quantitativas.

O tempo médio de follow-up (seguimento) foi de 2064 dias e o tempo mediano foi de 2184 dias. O tempo mínimo de seguimento foi de 2 dias e o tempo máximo de seguimento foi de 2635 dias. A mulher mais nova tem 25 anos e a mais velha tem 64 anos, idade média das mulheres é de 42,62 anos. A idade mediana das mulheres é de 41 anos, isto significa que 50% das mulheres envolvidas no estudo têm idades inferiores ou iguais a 41 anos e 50 % têm idades iguais ou superiores a 41 anos. O 1º quartil da variável idade da mulher é de 34 anos, isto significa que 25% das mulheres envolvidas no estudo têm idades inferiores ou iguais a 34 anos e 75% das mulheres têm idades iguais ou superiores a 34 anos. O 3º quartil para a variável idade é de 50 anos, isto significa que 75% das mulheres envolvidas no estudo têm idades inferiores ou iguais a 50 anos e 25% destas mulheres têm idades iguais ou superiores a 50 anos.

O número mínimo de gestações e partos é igual a 0 e número máximo é igual a 10, para ambas.

O número mínimo de abortos é 0 e o máximo é 5.

Em média as mulheres envolvidas no respetivo estudo iniciam a atividade sexual a partir dos 18,64 anos.

4.1.3 ANÁLISE EXPLORATÓRIA PARA VARIÁVEIS LOCALIZAÇÃO TOPOGRÁFICA E MORFOLOGIA DO TUMOR

Nesta secção será apresentada a estatística descritiva para variável localização topográfica e morfologia do cancro, apresentando os seus conceitos e a respetiva frequência absoluta e relativa do número de óbitos para cada uma das variáveis.

4.1.3.1 CONCEITO DE TOPOGRAFIA DO CANCRO

O código topográfico indica a localização de origem de um tumor, onde ele nasce. Na 3ª edição da Classificação Internacional de Doenças para Oncologia (CID-O) -códigos C00-C80 são baseados no capítulo II-CID-10. Todas as neoplasias, maligna, benigna, in situ ou incerta quanto a benignidade e malignidade são registadas com o mesmo código topográfico na CID-O (Organização Panamericana da Saúde & Organização Mundial da Saúde, 2000).

De seguida, encontram-se apresentados na Tabela 17, os códigos para localização topográfica do CCU e a respetiva análise descritiva.

Tabela 17 - Estatística descritiva para variável localização topográfica.

	Localização Topográfica	n=324	%	Nº de óbito (44)	% de óbito
CID-10-C53 Neoplasia maligna do colo de útero	C53.0 - Neoplasia maligna do endocérvix	17	5,25%	3	6,82%
	C53.1 - Neoplasia maligna do exocérvix	44	13,58%	1	2,27%
	C53.8 - Neoplasia maligna do colo do útero com lesão invasiva	1	0,31%	1	2,27%
	C53.9 - Neoplasia maligna do colo do útero, não especificado	262	80,86%	39	88,64%

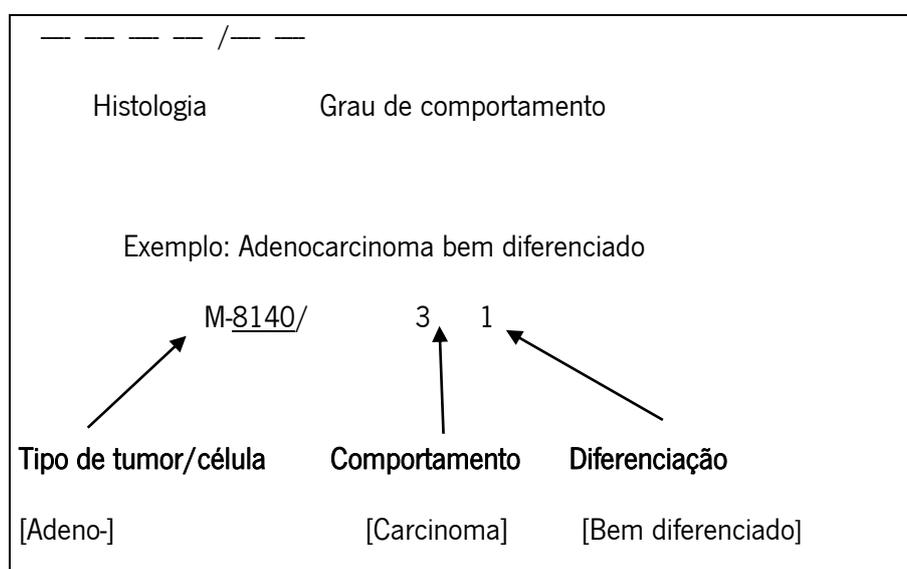
Pela Tabela 17, verifica-se que das 324 mulheres diagnosticadas com CCU, 262 mulheres que correspondem os 80,86 % tiveram diagnóstico de cancro, neoplasia maligna do colo do útero não especificado. Verifica-se ainda que 88,64% (39/44) dos óbitos são causados por neoplasia maligna do colo do útero, não especificado.

Antes de prosseguir com a análise estatística da variável morfologia do cancro, será apresentado o conceito de morfologia.

4.1.3.2 CONCEITO DE MORFOLOGIA DO CANCRO

O código morfológico regista o tipo de células que se tornou neoplásica e o seu comportamento biológico, em outras palavras, regista o tipo de tumor e o seu comportamento. O código morfológico completo é composto por 6 dígitos (ver Tabela 18) em que 4 dígitos correspondem ao tipo de celular (histologia), 1 dígito corresponde o comportamento biológico (quanto a malignidade e benignidade) e 1 dígito que corresponde o grau de diferenciação ou fenótipo (Organização Panamericana da Saúde & Organização Mundial da Saúde, 2000)

Tabela 18 - Estrutura do código morfológico



4.1.3.2 .1 CÓDIGO DE COMPORTAMENTO NA MORFOLOGIA (5º DÍGITO)

O comportamento biológico de um tumor é a maneira como ele se comporta no corpo. Os patologistas usam diversas formas para determinar o comportamento de um tumor. A Tabela 19 mostra um código de comportamento tumoral. Um tumor pode crescer num local e sem potencial de disseminação (0, benigno), ele pode ser maligno e estar limitado ao local de origem (2 não invasivo ou in situ), ele pode invadir os tecidos adjacentes (3 maligno, localização primária), ou também pode ser disseminado, estar longe do seu ponto de origem e começar a crescer em outra localização (6 metastático) (Organização Panamericana da Saúde & Organização Mundial da Saúde, 2000).

Tabela 19 - 5º dígito - Código de comportamento biológico das neoplasias.

Código	
0	Benigno
1	Incerto se benigno ou maligno Baixo potencial maligno Incerto sobre a malignidade
2	Carcinoma in situ Não infiltrativo Não invasivo
3	Maligno, localização primária
6	Maligno, localização metastática Maligno, localização secundária
9*	Maligno, incerto se é primário ou metastático * Não se usa em registo de cancro

É importante salientar que a maioria dos registos do cancro coleta dados sobre neoplasia maligna (3) ou in situ (2). Os tumores codificados como 6 (metastático) ou 9 (incerto se é primário ou metastáticos) não são geralmente usados nos registos do cancro (Organização Panamericana da Saúde & Organização Mundial da Saúde, 2000).

De seguida encontramos apresentada a estatística descritiva para a variável comportamento do cancro e o respetivo óbito para cada tipo de comportamento.

Tabela 20 - Estatística descritiva para a variável Comportamento do cancro

Comportamento cancro	n=324	%	Nº de óbito (44)	% de óbito
2 - In situ, não invasiva	197	60,80%	3	6,82%
3 - Maligna, localização primária	127	39,20%	41	93,18%

Pela Tabela 20, observa-se, que das 324 mulheres com CCU 197 que correspondem 60,80% tiveram CCU com comportamento in situ, não invasivo e 127 mulheres que correspondem os 39,20% tiveram CCU maligna, com localização primária. Dos totais 44 óbitos, 41 que correspondem os 93,18% foram registados nas mulheres com comportamento do cancro

maligno, com localização primária e apenas 3 óbito que correspondem os 6,82% foram registadas nas mulheres com cancro com comportamento in situ, não invasiva.

4.1.3.2 .2 CÓDIGO PARA GRADUAÇÃO E DIFERENCIAÇÃO HISTOLÓGICA (6º DÍGITO)

No sistema da CID-O, um único dígito é destinado à codificação da diferenciação ou grau histológico das neoplasias malignas. Somente os tumores malignos são graduados.

A prática de graduação dos tumores varia muito entre as patologias, nas diversas partes do mundo e muitos tumores malignos não são rotineiramente graduados (Organização Panamericana da Saúde & Organização Mundial da Saúde, 2000).

De seguida encontra-se apresentado na Tabela 21, os códigos para graduação do cancro apresentado pelo CID-O.

Tabela 21 - 6º dígito do Código para graduação e diferenciação histológicos

Código	Designação
Grau 1	Bem diferenciado
Grau 2	Moderadamente diferenciado Moderadamente bem diferenciado Diferenciação intermédia
Grau 3	Pouco diferenciado
Grau 4	Indiferenciado
Grau 9	Grau ou diferenciação não determinado, não estabelecido ou não se aplica.

Encontra-se de seguida na Tabela 22, a estatística descritiva para graduação e diferenciação histológicas tendo em consideração os casos de óbito para cada grau.

Tabela 22 - Estatística descritiva para variável Grau do cancro

Grau do Cancro	n=324	%	Nº de óbito (44)	% de óbito
Grau 1 - Bem diferenciado	20	6,17%	6	13,64%
Grau 2 - Moderadamente diferenciado	24	7,41%	5	11,36%
Grau 3 - Pouco diferenciado	4	1,23%	1	2,27%
Grau 9 - Não determinado	215	66,36%	23	52,27%
Valores em falta ("NA")	61	18,83%	9	20,45%

Pela Tabela 22, observa-se que 215 mulheres que correspondem os 66,36% tiveram grau do cancro, correspondente ao grau 9, ou seja tiveram grau não determinado, 24 mulheres que correspondem os 7,41% tiveram grau 2, ou seja, grau moderadamente diferenciado, 20 mulheres que correspondem os 6,17% tiveram grau 1, ou seja grau bem diferenciados, 4 mulheres que correspondem os 1,23% tiveram grau 3, ou seja, grau pouco diferenciado e 61 mulheres que correspondem os 18,83 não apresentavam informações em relação ao grau do cancro. A maior percentagem de óbito (52,27%) foi registada nas mulheres com grau 9, ou seja, grau do cancro não determinado.

Uma vez que o nosso grande objetivo de estudo é comparar a sobrevivência do CCU das mulheres submetidas ao rastreio e das não submetidas ao rastreio, seria interessante fazer o cruzamento de algumas variáveis.

4.1.4 CRUZAMENTO DAS VARIÁVEIS

Nesta seção, será apresentada a estatística descritiva para o cruzamento de algumas variáveis importantes para o respetivo estudo.

Por exemplo, seria interessante cruzar as informações do estadiamento do cancro à data diagnóstico, com a variável rastreio e estado da doente. Espera-se que as mulheres que não foram submetidas ao rastreio do CCU tenham um estadiamento mais avançado à data do diagnóstico do cancro, quando comparado com as mulheres submetidas ao rastreio. Espera-se ainda, que se registre um maior número de óbitos no grupo das mulheres não submetidas ao rastreio do que no grupo das mulheres submetidas ao rastreio.

4.1.4.1 CRUZAMENTO DAS VARIÁVEIS RASTREIO, ESTADO DA PACIENTE E ESTADIAMENTO

Tabela 23 - Frequências absolutas para o cruzamento das três variáveis

Cruzamento do estado da paciente e estadiamento para os rastreados					
Estado	Estadiamento				Total
	Estádio 1	Estádio 2	Estádio 3	Valores em falta ("NA")	
0 - Vivo	14	1	1	138	154
1 - Óbito	1	0	1	5	7
Total	15	1	2	143	161
Cruzamento do estado da paciente e estadiamento para os não rastreados					
Estado	Estadiamento				Total
	Estádio 1	Estádio 2	Estádio 3	Valores em falta ("NA")	
0 - Vivo	24	8	2	44	78
1 - Óbito	2	2	19	9	32
Total	26	10	21	53	110

Pela Tabela 23, observa-se que dos totais, 161 pacientes rastreadas apenas faleceram 7 por CCU. Destes 7 óbitos, 1 apresentava estadiamento do cancro à data diagnóstico com estágio 1, 0 apresentava estágio 2, 1 apresentava estadiamento do cancro à data diagnóstico com estágio 3 e 5 destas pacientes não tinham informação em relação ao estadiamento.

Em relação às 110 pacientes não submetidas ao rastreio, registou-se no total de 32 óbitos por CCU. Destes 32 óbitos, 19 apresentaram estágio 3 à data do diagnóstico, 2 doentes apresentaram estágio 1, 2 doentes apresentaram estágio 2 à data do diagnóstico e 9 doentes não tinha informação em relação ao estadiamento.

Verificou-se que no geral, as mulheres submetidas ao rastreio do CCU apresentaram um estadiamento mais baixo do que as mulheres não submetidas ao rastreio, o que está de acordo com o esperado.

4.1.4.2 CRUZAMENTO DAS VARIÁVEIS RASTREIO E COMPORTAMENTO DO CANCRO

Tabela 24 - Frequências absolutas para o cruzamento das duas variáveis

Rastreio	Comportamento do Cancro		Total
	2 - In situ, não invasivo	3 - Maligno, com localização primária	
0 - Não	29	81	110
1 - Sim	130	31	161
Valores em falta ("NA")	38	15	53
Total	197	127	324

Pela Tabela 24, verifica-se que das 324 mulheres envolvidas no estudo, 197 que correspondem os 60,80% (197/324) tiveram comportamento do cancro in situ, não invasivo e 127 que correspondem os 39,20% (127/324) tiveram comportamento de cancro maligno, com localização primária. Das 197 mulheres com comportamento do cancro in situ, 130 que correspondem os 65,99% (130/197) foram submetidas ao rastreio do CCU, 29 que correspondem os 14,72% (29/197) não foram submetidas ao rastreio do CCU e 38 que correspondem os 19,29% (38/197) não tem informações em relação ao rastreio do CCU. Das 127 mulheres com comportamento do cancro maligno, com localização primária, 81 que correspondem os 63,78% (81/127) não foram submetidas ao rastreio do CCU, 31 que correspondem os 24,41% (31/127) foram submetidas ao rastreio do CCU e 15 que correspondem os 11,81% (15/127) não tiveram informação em relação ao rastreio do CCU.

Do resultado da Tabela 24, podemos concluir que das 161 mulheres submetidas ao rastreio do CCU, 130 que correspondem os 80,75% (130/161) tiveram comportamento do cancro in situ, não invasivo. Das 110 mulheres não submetidas ao rastreio do CCU, 81 que correspondem os 73,64% (81/110) tiveram comportamento do cancro, maligno, com localização primária. Deste resultado podemos afirmar que o rastreio do CCU parece influenciar o comportamento do CCU.

Seria interessante ver se existe associação estatisticamente significativa entre a variável rastreio do CCU e comportamento do CCU, ou seja, verificar se o comportamento do CCU é influenciado pelo rastreio, mas mais adiante falaremos sobre isso.

Agora o passo seguinte será saber, das 161 pacientes submetidas ao rastreio de CCU, quantas é que tiveram resultados normais e anormais e verificar a eficiência do rastreio.

4.1.5. ESTATÍSTICA DESCRITIVA PARA O RESULTADO DE RASTREIO

De seguida, encontram-se apresentados os resultados de rastreio para as 161 doentes submetidas ao rastreio do CCU.

Tabela 25 - Frequência absoluta para resultado de rastreio.

Resultado Rastreio	Nº
Adenocarcinoma Endocervical	1
AGC	2
Anormal	3
ASC-H	11
ASC-US	4
Carcinoma	1
Carcinoma Espinho Celular	1
Carcinoma de Células Escamosas	1
Carcinoma Epidermoide	2
Carcinoma Espinho Celular	2
Carcinoma Espinho Celular Invasivo	1
Células Glandulares Endocervicais Atípicas Favorecendo Adenocarcinoma in situ do endocolo	1
CIN I	1
Fenómeno de Atipia Celular	2
HSIL	91
LSIL	18
NILM	7
Valores em falta ("NA")	11
Total Geral	161

A Tabela 25 descreve os resultados obtidos para as 161 mulheres submetidas ao rastreio do CCU. Para complementar a variável resultado de rastreio, essa variável foi categorizada em dois níveis, em que temos as pacientes com resultado normal e anormal, em que os resultados normais de rastreio são os NILM e os outros restantes resultados são as alterações, ou seja, os resultados anormais. Observa-se ainda que o resultado de rastreio mais frequente é o HSIL (lesão intraepitelial escamosa de alto grau).

A seguir apresenta-se a tabela para essa variável.

Tabela 26 - Frequência absoluta e relativa para os resultados de rastreio e número de óbitos

Resultado rastreio	n=161	%	Nº de óbito (7)	% de óbito
0 - Normal	7	4,35	0	0,00
1 - Anormal	144	89,44	6	85,71
Valores em falta ("NA")	10	6,21	1	14,29

Pela Tabela 26, verifica-se que das 161 pacientes submetidas ao rastreio do CCU, 144 (89,44%) tiveram resultados anormais, 7 (4,35%) tiveram resultados normais e 10 (6,21%) não tiveram a informação do resultado de rastreio. Dos totais 7 óbitos registados nas mulheres submetidas ao rastreio, 6 ocorreram nas mulheres que tiveram citologia anormal antes da data do diagnóstico e 1 óbito ocorreu numa mulher sem informação do resultado do rastreio.

4.2. ASSOCIAÇÃO ENTRE RASTREIO E O COMPORTAMENTO DO CANCRO DO COLO DO ÚTERO

Na Secção 4.1.4.2, os dados indicaram que existem uma associação entre o rastreio e o comportamento do CCU, ou seja, as mulheres não submetidas ao rastreio tendem a ter comportamento do cancro invasivo, com localização primária e as mulheres submetidas ao rastreio do CCU, tendem a ter comportamento do cancro in situ, não invasivo (ver Tabela 24). Vamos ver se de facto existe uma associação estatisticamente significativa entre as duas variáveis. Vai ser utilizado o teste de associação de Qui-quadrado.

4.2.1. TESTE DE ASSOCIAÇÃO DE QUI-QUADRADO

O teste de Qui-quadrado permite analisar a relação de independência entre variáveis qualitativas, aplica-se tanto para análise univariada, como bivariada ou multivariada.

Na análise univariada, as observações de uma única população são subdivididas em categorias, geralmente os dados resultam da contagem de objetos em cada uma das categorias, pretendendo-se comparar as frequências observadas com as frequências esperadas.

Na análise bivariada, analisa-se a relação entre duas variáveis nominais independentes, cada uma com duas categorias (por exemplo, sim e não ou feminino e masculino), tratando-se por isso de variáveis dicotómicas, sendo os dados apresentados em tabelas de contingência 2x2.

As tabelas de contingência são utilizadas para estudar a relação entre duas ou mais variáveis categóricas, descrevendo as frequências das categorias de uma das variáveis relativamente às categorias das outras.

O teste do Qui-quadrado mede a probabilidade de as diferenças encontradas nos dois grupos da amostra serem devidas ao acaso, partindo do pressuposto que, na verdade, não há diferenças entre esses grupos na população donde provêm. Se a probabilidade for alta pode-se concluir que não há diferenças estatisticamente significativas. Se a probabilidade for baixa (particularmente menor que 5%) pode-se concluir que um grupo é diferente do outro grupo, quanto à característica estudada e de forma estatisticamente significativa.

Geralmente utiliza-se o teste de associação de Qui-quadrado para ver se existe associação entre duas variáveis qualitativas, como no nosso caso, onde temos as seguintes hipóteses a testar:

H₀: As duas variáveis não estão associadas

VS

H₀: As duas variáveis estão associadas

Mas, no entanto, para aplicar o teste de associação de Qui-quadrado temos que verificar os pressupostos de aplicabilidade do teste que são os seguintes:

- ✓ Dimensão da amostra grande, maior de que 20;
- ✓ As observações devem ser independentes;
- ✓ Ter pelo menos 2 grupos;
- ✓ Nenhuma das frequências esperadas podem ser inferiores a 1;
- ✓ Não se aplica se 20% das observações ou valores esperados forem inferiores a 5;
- ✓ Cada observação pertence a uma e uma só categoria.

Caso falhe algum dos pressupostos acima referidos, temos de substituir o teste de associação de Qui-quadrado pelo teste exato de Fisher, embora este teste só possa ser utilizado para tabelas 2x2.

Como foram verificados os pressupostos de teste de associação de Qui-quadrado (ver Anexo V), estamos em condições de aplicar o teste. No Anexo V, também se encontra o código do R utilizado, para aplicar o teste de associação de Qui-quadrado.

Tabela 27 - Resultado de teste de associação de Qui-quadrado

X^2	DF	P-valor
77,477	1	2,2e-16

Pela Tabela 27, conclui-se como $p\text{-valor}=2,2e-16 < 0,05$, logo para um nível de significância de 5% conclui-se que existe associação estatisticamente significativa entre o rastreio e o comportamento do CCU. O teste de associação de Qui-quadrado apenas diz se existe ou não associação entre as variáveis, mas, no entanto, não quantifica esta associação. Para quantificar essa associação vamos utilizar uma outra medida de associação, o risco relativo (RR) que é a medida mais apropriada a utilizar em caso de estudo de coorte. Para calcular o RR vamos utilizar a biblioteca **epitools**, que é uma biblioteca específica de epidemiologia.

De seguida, encontra-se apresentado o código e o Output do R utilizado para calcular o RR (mais informações ver Anexo VI).

Tabela 28 - Risco relativo para associação das duas variáveis

```
> epitab(dados$Rastreio_uter, dados$Comp_cancro, method="riskratio", pvalue="chi2", rev="rows")
$tab
      Outcome
Predictor 2      p0 3      p1 riskratio  lower  upper  p.value
      1 130 0.8074534 31 0.1925466  1.00000  NA     NA     NA
      0  29 0.2636364 81 0.7363636  3.82434 2.734325 5.34888 4.349907e-19

$measure
[1] "wald"

$conf.level
[1] 0.95

$pvalue
[1] "chi2"
```

Recordar, que calculamos essa medida de associação (RR) para ver se o comportamento do cancro é influenciado por rastreio do CCU. Sendo assim a variável Predictor que aparece na linha da Tabela 28, representa a variável rastreio do CCU em que o 0 representa as mulheres não submetidas ao rastreio e 1 representa as mulheres submetidas ao rastreio e a variável Outcome que aparece na coluna, indica o comportamento do cancro, em que o 2 representa comportamento do cancro in situ, não invasivo e o código 3 representa o comportamento do cancro maligno, com localização primária. Da Tabela 28, verifica-se que a categoria de

referência é o 1 (sim), porque é o que apresenta $RR=1$. Através da Tabela 28, conclui-se que as mulheres não submetidas ao rastreio do CCU têm aproximadamente 4 vezes mais possibilidades de ter cancro com comportamento maligno, com localização primária do que as mulheres submetidas ao rastreio do CCU.

4.3. ANÁLISE DE SOBREVIVÊNCIA

A análise de sobrevivência é utilizada quando o tempo for o objeto de interesse, seja esse interpretado como o tempo até a ocorrência de um evento ou o risco de ocorrência de um evento por unidade de tempo. Desta forma, estes estudos buscam enfatizar e detalhar ao máximo possível a informação relacionada com o tempo (Marília Sá Carvalho et al., 2011).

É importante salientar que para realizar um estudo de sobrevivência precisamos de duas variáveis chave, a variável tempo e a variável “status” que indica se o indivíduo observou ou não o evento de interesse.

Neste estudo em particular a nossa variável de interesse “temp_followp” o tempo de seguimento, ou seja, o tempo desde a data do diagnóstico até ao término do estudo ou morte por CCU e a variável “status” que representa o estado da doente em que o 0 corresponde a uma censura pela direita e o 1 corresponde ao evento de interesse (morte ou óbito por CCU). A data final dos seguimentos das doentes foi estipulada para 31/03/2016 pelo RORC do IPO de Coimbra.

Primeiramente, antes de prosseguir com o estudo de análise de sobrevivência, por decisão clínica, a variável idade foi categorizada.

4.3.1 CATEGORIZAÇÃO DA VARIÁVEL IDADE

Nesta secção vamos estudar a distribuição da variável idade e prosseguir com a sua categorização.

De seguida, encontra-se apresentada a distribuição da variável idade das mulheres à data do diagnóstico.

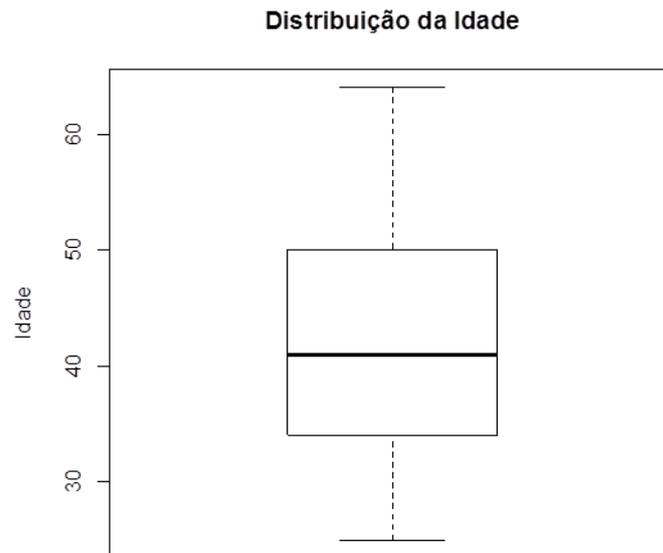


Figura 8 - Boxplot da variável idade

Pela Figura 8, observa-se que a variável idade apresenta uma ligeira assimetria positiva. A idade média das mulheres foi de 42,62 anos e a idade mediana foi de 41 anos. A mulher mais nova no estudo tem 25 anos e a mais velha tem 64 anos. O 1º quartil foi de 34 anos, o que significa que 25% das mulheres têm idades inferiores ou iguais a 34 anos e 75% têm idade iguais ou superiores a 34 anos. O 3º quartil foi de 50 anos o que significa que 75% das mulheres têm idades inferiores ou iguais a 50 anos e 25% têm idades iguais ou superiores a 50 anos. A variável idade não apresenta outliers (Ver Anexo VII para ver o código do R utilizado).

A partir da variável idade foi criada uma nova variável designado de faixa etária (*Idade_grupo*) com 4 categorias. Ver o Anexo VIII os códigos do R utilizado para criar a variável faixa etária.

A nova variável faixa etária encontra-se definida da seguinte forma:

$$faixa\ etária = \begin{cases} 1, se\ Idade\ [25; 34]\ anos \\ 2\ se\ Idade\ [35,44]\ anos \\ 3, se\ Idade\ [45; 54]\ anos \\ 4, se\ Idade\ [55; 64]\ anos \end{cases}$$

Através do comando `table` e `prop.table`, criaram-se as frequências absoluta e relativa para a variável faixa etária ver (Anexo IX).

De seguida, encontram-se apresentadas as frequências absolutas, relativas e o barplot para a variável faixa etária.

Tabela 29 - Frequências absoluta e relativa

para a variável faixa etária

Faixa Etária	n	%
Idade [25;34] anos	89	27,47%
Idade [35;44] anos	111	34,26%
Idade [45;54] anos	68	20,99%
Idade [55;64] anos	56	17,28%
Total	324	100,00%

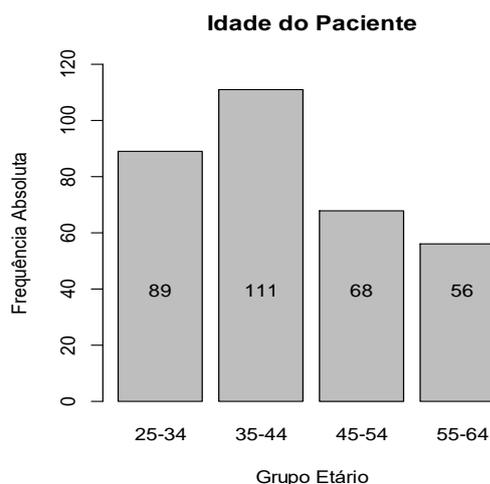


Figura 9 – Gráfico de barras (barplot) da variável faixa etária

Pela Tabela 29 e Figura 9, observa-se que a faixa etária predominante é a dos 35 a 44 anos, a que correspondem os 34,26% (111/324). A faixa etária com menor frequência é a faixa etária entre os 55 e 64 anos, a que correspondem os 17,28% (56/324).

4.3.2. CRUZAMENTO DA VARIÁVEL RASTREIO, FAIXA ETÁRIA E ÓBITO

É do nosso interesse perceber, em qual faixa etária foram registados maior número de óbitos e se foram nas mulheres submetidas ou não ao rastreio do CCU.

Tabela 30 - Frequência absoluta para o cruzamento da variável rastreio, faixa etária e óbito

Nº de Óbito por Rastreio e Faixa Etária				
Faixa Etária	Rastreio			Total
	0-Não	1-Sim	Valores em falta ("NA")	
Idade [25;34] anos	1	0	0	1
Idade [35;44] anos	4	3	0	7
Idade [45;54] anos	16	2	2	20
Idade [55;64] anos	11	2	3	16
Total	32	7	5	44

Pela Tabela 30, observa-se que do total dos 44 óbitos, 16 foram registados nas mulheres não submetidas ao rastreio, com idades compreendidas entre os 45 e 54 anos. Foram registados 11 óbitos em mulheres não submetidas ao rastreio com idades compreendidas entre os 55 e 64 anos e 4 óbitos foram registadas nas mulheres não submetidas ao rastreio com idade

compreendidas entre os 35 e 44 anos. Para as mulheres com idades compreendidas entre os 25 e 34, apenas se registou um único óbito para essa faixa etária.

Das mulheres submetidas ao rastreio, registaram-se 3 óbitos para as mulheres com idades compreendidas entre os 35 e 44 anos, 2 óbitos para as mulheres com idades compreendidas entre os 45 e 54 anos, 2 óbitos para as mulheres com idades compreendidas entre os 55 e 64 anos e não se registou nenhum óbito para as mulheres com idades compreendidas entre os 25 e 34 anos. Das mulheres sem indicação de rastreio, 3 óbitos ocorreram na faixa etária de 55 a 64 anos, 2 óbitos na faixa etária de 45 a 54 anos e não se registou nenhum óbito na faixa etária dos 25 a 34 anos e dos 35 a 44 anos.

4.3.3 SOBREVIVÊNCIA APÓS O DIAGNÓSTICO DO CCU

Para a estimação da função de sobrevivência, utilizou-se o estimador de Kaplan-Meier. Apresenta-se de seguida, a estimativa de Kaplan-Meier da função de sobrevivência após o diagnóstico para as mulheres com CCU e a sua representação gráfica. O código do R para obter esses resultados encontra apresentado no Anexo X.

Tabela 31 - Estimativa da função de sobrevivência global de Kaplan-Meier para o CCU

time	n.risk	n.event	survival	std.err	lower 95% CI	upe 95% CI
2	324	1	0.997	0.00308	0.978	1.000
18	323	1	0.994	0.00435	0.976	0.998
68	322	1	0.991	0.00532	0.972	0.997
82	321	1	0.988	0.00613	0.967	0.995
92	320	1	0.985	0.00685	0.963	0.994
133	319	1	0.981	0.00749	0.959	0.992
172	318	1	0.978	0.00808	0.955	0.990
190	317	1	0.975	0.00862	0.951	0.988
259	316	1	0.972	0.00913	0.947	0.985
263	315	1	0.969	0.00961	0.943	0.983
289	314	1	0.966	0.01006	0.940	0.981
328	313	1	0.963	0.01049	0.936	0.979
353	312	1	0.960	0.01090	0.932	0.977
360	311	1	0.957	0.01130	0.928	0.974
384	310	1	0.954	0.01167	0.924	0.972
.....
2463	72	1	0.855	0.02213	0.805	0.893

Na Tabela 31 são apresentadas as estimativas da função de sobrevivência de Kaplan-Meier, para os primeiros 15 tempos e o tempo máximo onde ocorreram os eventos de interesse e o seu respetivo intervalo de confiança. Foi utilizado o estimador de Greenwood com a transformação log-log, apresentado na equação 2.49 da secção 2, para obter o respetivo intervalo de confiança.

Da Tabela 31, verificamos que a coluna com a designação **time** que representa os tempos onde ocorreu o evento de interesse, ou seja, morte por CCU. A coluna com designação **n.risk** indica o número de indivíduos em risco justamente antes do tempo t . A coluna com designação **n.event** representa o número de óbitos por CCU em cada tempo. A coluna com designação **survival** representa a estimativa da função de sobrevivência de Kaplan-Meier em cada tempo. A coluna com designação **std.err** representa o desvio padrão da estimativa da função de sobrevivência de Kaplan-Meier, que foi calculado com base no estimador de Greenwood apresentado na Secção 2.8.2. As colunas com designação **lower 95% CI** e **upe 95% CI** representam o limite inferior e superior do intervalo de confiança para o estimador Kaplan-Meier com 95% de confiança, utilizando a transformação log-log apresentada na equação 2.49.

É importante salientar, que por defeito o R faz intervalo de confiança de Greenwood pela transformação log, mas, no entanto, na Tabela 31 utilizamos a transformação log-log. O tempo médio de sobrevivência após o diagnóstico do CCU é de 2381,3 dias (mais informação no Anexo XI).

De seguida encontra-se apresentado na Tabela 32 e Figura 10, a estimativa de Kaplan-Meier de sobrevivência global do CCU para 1, 3 e 5 anos após o diagnóstico do CCU e foi utilizado a transformação log-log, para obter os respetivos intervalo de Confiança (ver Anexo X).

Tabela 32 - Estimativa de sobrevivência global a 1, 3 e 5 anos após o diagnóstico do CCU (n=324)

1 ano		3 ano		5 ano	
$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
95,7	(92,8-97,4)	89,8	(86,0-92,6)	87,3	(83,2-90,5)

Curva de Sobrevivência global de Kaplan-Meier em 2635 dias do CCU, n=324

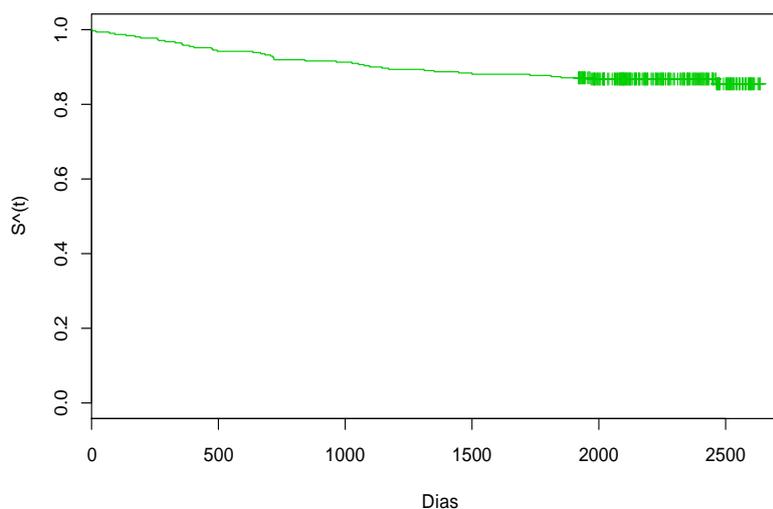


Figura 10 - Estimativa da curva de sobrevivência global do CCU de Kaplan-Meier

Da Figura 10, observa-se que a estimativa da função de sobrevivência por si só não nos traz grande informação. Observamos que à medida que o tempo passa, a estimativa da probabilidade de sobrevivência decresce lentamente, tendendo a estabilizar em valores próximos de 0,80. Os símbolos “+” que aparece no gráfico localiza as censuras.

Da Tabela 32, verifica-se que a probabilidade de uma mulher sobreviver a 1 ano (365 dias) após o diagnóstico do CCU é de 95,7%, a probabilidade de uma mulher sobreviver a 3 anos (1095) dias é de 89,8% e a probabilidade de sobrevivência a 5 anos (1825 dias) é de 87,3%.

De seguida, vamos proceder à estimação da função de sobrevivência e da curva de sobrevivência para os grupos distintos definidos pelos valores das variáveis qualitativas e é utilizado a transformação log-log para obter os respetivos intervalos de confiança. No Anexo XII encontra o código de R utilizado para obter estimativa e a curva de sobrevivência de Kaplan-Meier para a variável rastreio do CCU, foram utilizados os mesmos códigos para as outras variáveis, mudando-se apenas o nome das variáveis.

Tabela 33 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por rastreio

Rastreio	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
Não	110	90,0	(82,7-94,3)	78,2	(69,2-84,8)	71,8	(62,4-79,3)
Sim	161	100,0	(100,0-100,0)	96,9	(92,7-98,7)	96,9	(92,7-98,7)

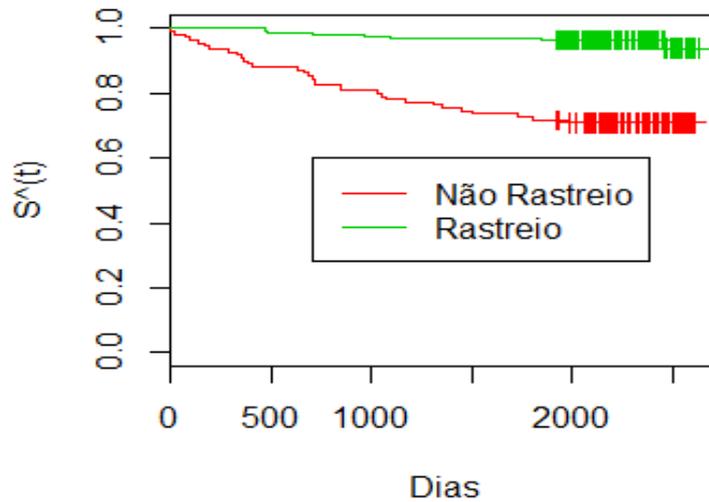


Figura 11 - Curva de sobrevivência estimada de Kaplan-Meier por rastreio

Pela observação da Tabela 33 e Figura 11, pode-se concluir que as mulheres submetidas ao rastreio do CCU têm uma maior sobrevivência do que as mulheres não submetidas ao rastreio.

O tempo médio de sobrevivência para as mulheres submetidas ao rastreio do CCU é de 2556 dias e o tempo médio das mulheres não submetidas ao rastreio é de 2078 dias. As mulheres submetidas ao rastreio do CCU em média sobrevivem 478 dias a mais do que as mulheres não submetidas ao rastreio do cancro do CCU (ver Anexo XIII).

Tabela 34 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por faixa etária

Faixa etária	n	1 ano		3 ano		5 ano	
		$\widehat{S}(t)\%$	95% IC	$\widehat{S}(t)\%$	95% IC	$\widehat{S}(t)\%$	95% IC
[25;34] anos	89	100,0	(100,0-100,0)	100,0	(100,0-100,0)	98,9	(92,3-99,8)
[35;44] anos	111	99,1	(93,8-99,9)	95,5	(89,5-98,1)	94,6	(88,4-97,5)
[45;54] anos	68	86,8	(76,1-92,9)	73,5	(61,3-82,4)	72,1	(59,8-81,2)
[55;64] anos	56	92,9	(82,1-97,3)	82,1	(69,4-90,0)	73,2	(59,6-82,9)

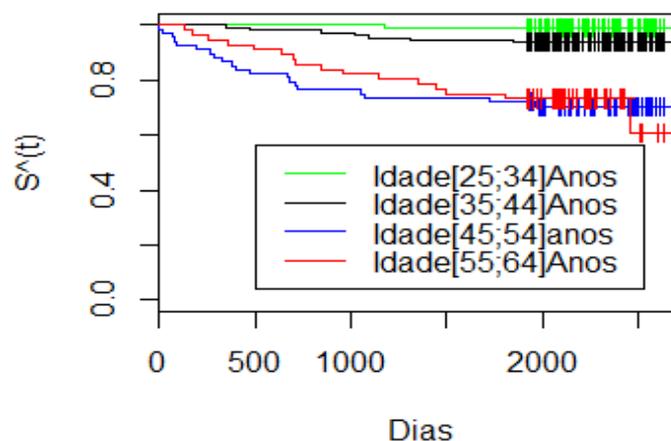


Figura 12 - Curva de sobrevivência estimada de Kaplan-Meier por faixa etária

Observa-se na Tabela 34 e Figura 12, que quanto à idade na altura do diagnóstico, as mulheres nas faixas etárias dos [25,34] anos apresentam, sempre, uma probabilidade de sobrevivência superior à das restantes faixas etárias.

Tabela 35- Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por faixa etária e rastreio

Rastreio			Faixa etária			
			[25,34] Anos	[25,34] Anos	[45,54] Anos	[55,64] Anos
1 ano	Não	$\hat{S}(t)\%$	100,0	96,7	79,4	89,3
		95% IC	(100,0-100,0)	(78,6-99,5)	(61,6-89,6)	(70,4-96,4)
	Sim	$\hat{S}(t)\%$	100,0	100,0	100,0	100,0
		95% IC	(100,0-100,0)	(100,0-100,0)	(100,0-100,0)	(100,0-100,0)
3 ano	Não	$\hat{S}(t)\%$	100,0	90,0	58,8	75,0
		95% IC	(100,0-100,0)	(72,1-96,7)	(40,6-73,2)	(54,6-87,2)
	Sim	$\hat{S}(t)\%$	100,0	96,6	92,6	94,4
		95% IC	(100,0-100,0)	(86,9-99,1)	(73,5-98,1)	(66,6-99,2)
5 ano	Não	$\hat{S}(t)\%$	94,4	86,7	55,9	60,7
		95% IC	(66,6-99,2)	(68,3-94,8)	(37,8-70,6)	(40,4-76,0)
	Sim	$\hat{S}(t)\%$	100,0	96,6	92,6	94,4
		95% IC	(100,0-100,0)	(86,9-99,1)	(73,5-98,1)	(66,6-99,2)

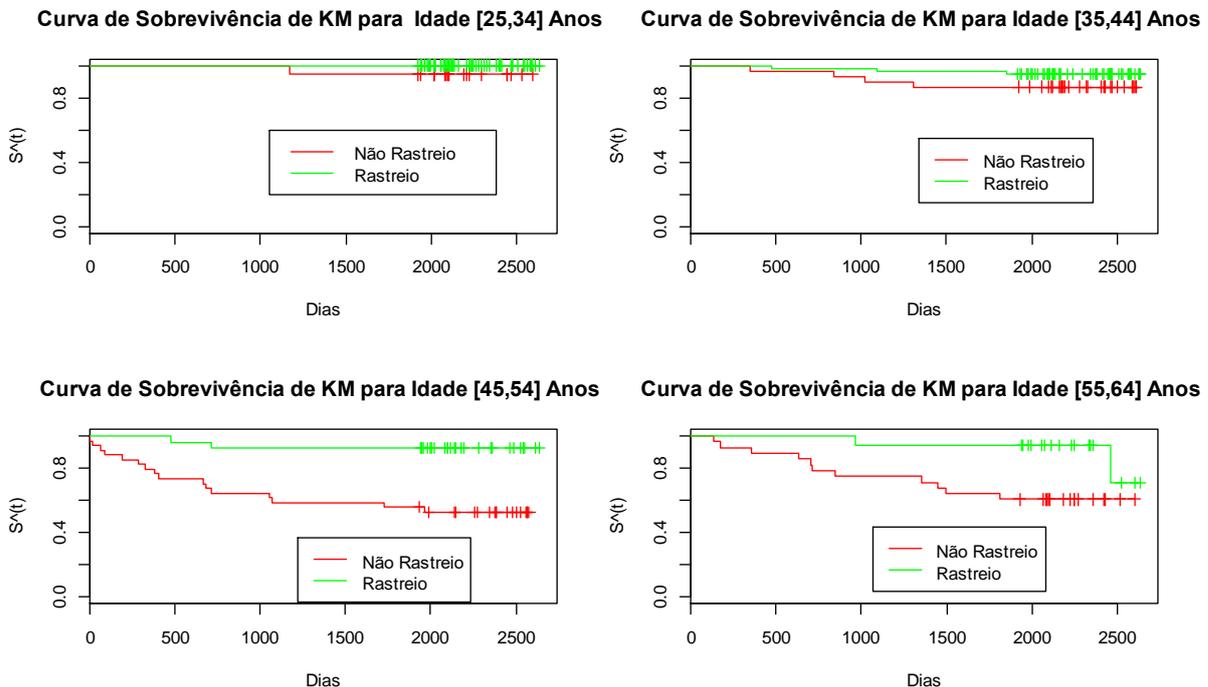


Figura 13 - Curva de sobrevivência estimada de Kaplan-Meier para o cruzamento das faixas etárias e rastreio

O gráfico do cruzamento da variável faixa etária e rastreio apresentado na Figura 13, sugere que na faixa etária dos [25,34] e dos [35,44] anos a sobrevivência não se diferenciam por rastreio. No entanto para as faixas etárias dos [45,54] e [55,64] anos a probabilidade de sobrevivência é

sempre maior nas mulheres submetidas ao rastreio do que nas mulheres não submetidas ao rastreio. Através do teste de log-rank ficou comprovado que existem diferenças estatisticamente significativas na sobrevivência das mulheres submetidas e não submetidas ao rastreio só a partir dos 45 anos.

Tabela 36 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por estadiamento do cancro.

Estadiamento	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
Estádio I	41	100,0	(100,0-100,0)	97,6	(83,9-99,7)	97,6	(83,9-99,7)
Estádio II	11	100,0	(100,0-100,0)	90,9	(50,8-98,7)	81,8	(44,7-95,1)
Estádio III	24	70,8	(48,4-84,9)	29,2	(13,0-47,6)	12,5	(3,1-28,7)

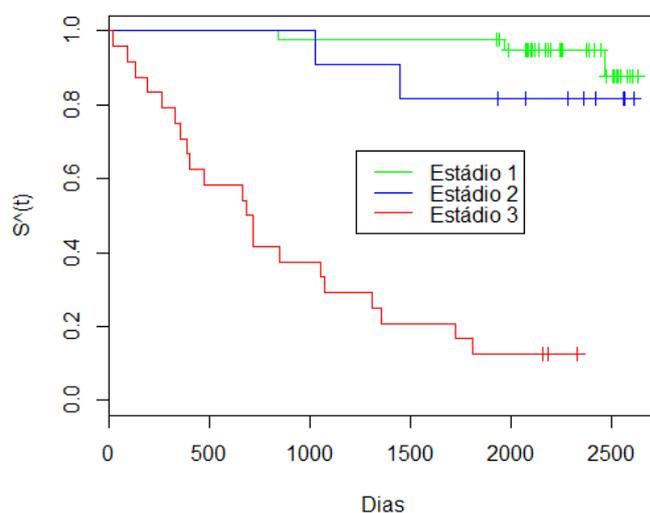


Figura 14 - Curva de sobrevivência estimada de Kaplan-Meier por estadiamento

Da Tabela 36 e Figura 14, observa-se que as mulheres que apresentaram o estadiamento do cancro à data diagnóstico com estágio 3 apresentaram menor probabilidade de sobrevivência em relação aos restantes estadiamento, para 1, 3 e 5 anos após o diagnóstico.

Tabela 37 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por distrito de residência

Distrito de residência	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
Aveiro	49	98,0	(86,4-99,7)	95,9	(84,7-99,0)	93,9	(82,2-98,0)
Castelo Branco	13	92,3	(56,6-98,9)	92,3	(56,6-98,9)	92,3	(56,6-98,9)
Coimbra	82	95,1	(87,5-98,1)	87,8	(78,5-93,2)	82,9	(72,9-89,5)
Guarda	19	100,0	(100,0-100,0)	94,7	(68,1-99,2)	94,7	(68,1-99,2)
Leiria	127	96,9	(91,8-98,8)	92,1	(85,9-95,7)	90,6	(84,0-94,5)
Viseu	34	88,2	(71,6-95,4)	73,5	(55,3-85,3)	70,6	(52,2-83,0)

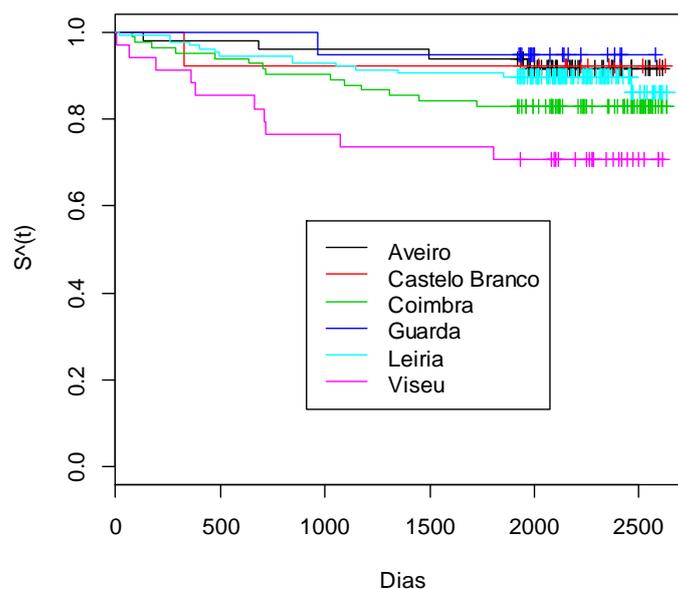


Figura 15 - Curva de sobrevivência estimada de Kaplan-Meier por distrito de residência das pacientes

Constata-se da Tabela 37 e Figura 15, que as mulheres residentes no distrito de Viseu são aquelas que apresentaram menor probabilidade de sobrevivência para 1, 3 e 5 anos após o diagnóstico e que as mulheres residentes no distrito da Guarda são as que apresentaram maior probabilidade de sobrevivência para os mesmos anos.

Tabela 38 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por comportamento do cancro

Comportamento do cancro	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
2 - In situ, não invasivo	197	100,0	(100,0-100,0)	99,0	(96,0-99,7)	99,0	(96,0-99,7)
3 - Maligno, com localização primária	127	89,0	(82,1-93,3)	75,6	(67,1-82,2)	69,3	(60,5-76,5)

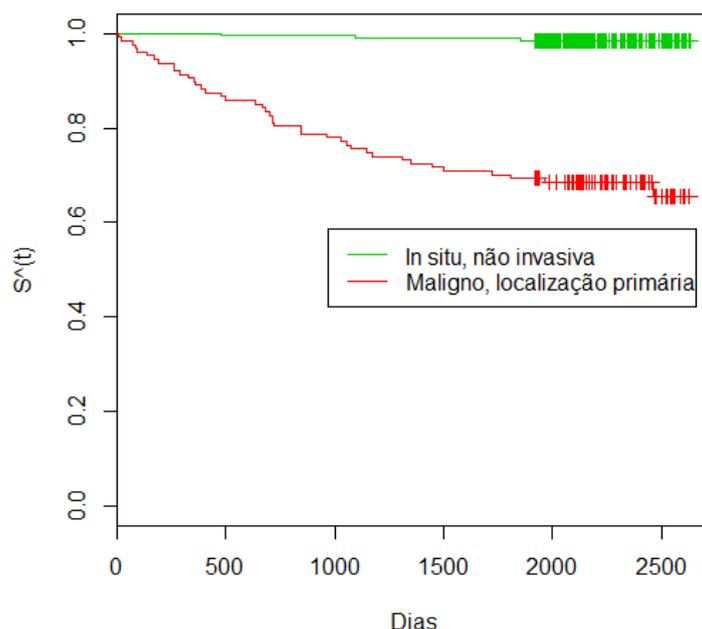


Figura 16 - Curva de sobrevivência estimada de Kaplan-Meier por comportamento do cancro

Em relação ao variável comportamento do cancro apresentado na Tabela 38 e Figura 16, observa-se que as mulheres com o comportamento de cancro in situ, não invasivo têm probabilidade de sobrevivência muito mais elevada do que as mulheres com comportamento do cancro maligno, com localização primária, para 1, 3 e 5 anos após o diagnóstico do CCU.

O tempo médio de sobrevivência para as mulheres com comportamento do cancro in situ, não invasivo é de 2610 dias e o tempo médio de sobrevivência das mulheres com comportamento do cancro maligno, com localização primária é de 2022 dias. Em média as mulheres com comportamento do cancro in situ, não invasivo sobrevivem 588 dias mais do que as mulheres com comportamento do cancro maligno, com localização primária (ver Anexo XIV).

Tabela 39 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por estado civil

Estado civil	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
0 - Solteira	34	91,2	(75,1-97,1)	85,3	(68,2-93,6)	85,3	(68,2-93,6)
1 - Casada	135	97,0	(92,3-98,9)	88,1	(81,4-92,6)	85,9	(78,8-90,8)
2 - Viúva	10	100,0	(100,0-100,0)	100,0	(100,0-100,0)	100,0	(100,0-100,0)
3 - Divorciada	20	100,0	(100,0-100,0)	100,0	(100,0-100,0)	100,0	(100,0-100,0)

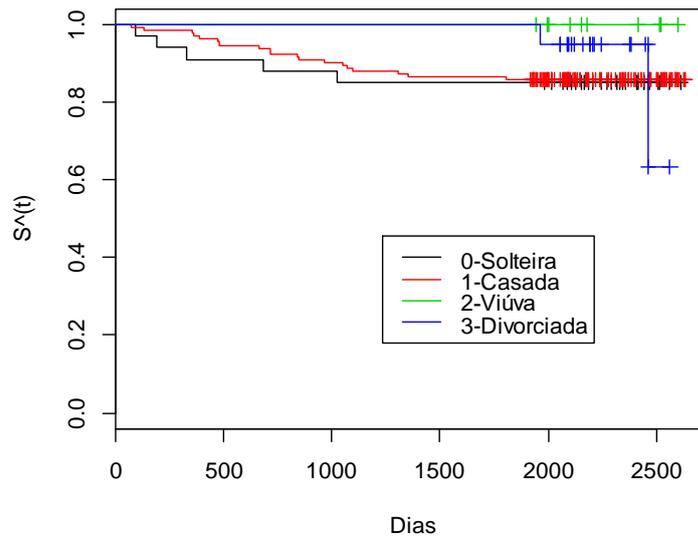


Figura 17 - Curva de sobrevivência estimada de Kaplan-Meier por estado civil

Da Tabela 39 e Figura 17, verifica-se que as mulheres solteiras tiveram pior sobrevivência no 1, 3 e 5 anos após o diagnóstico do CCU em relação às restantes.

Tabela 40 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico variável Fumador

Fumador	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
0 - Não	158	96,2	(91,7-98,3)	89,2	(83,3-93,2)	87,3	(81,1-91,6)
1 - Sim	74	98,2	(90,8-99,8)	91,9	(82,8-96,3)	91,9	(82,8-96,3)

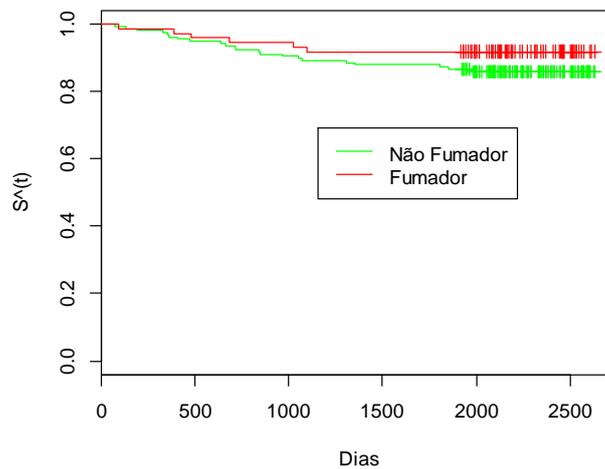


Figura 18 - Curva de sobrevivência estimada de Kaplan-Meier por variável fumador

O gráfico apresentado na Figura 18, sugere que não existem diferenças significativas na sobrevivência para as mulheres fumadoras e não fumadoras.

Tabela 41 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por infeção por HPV

Infeção por HPV	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
0 - Sem HPV	7	100	(100,0-100,0)	100	(100,0-100,0)	100	(100,0-100,0)
1 - Com HPV	95	98,9	(92,8-99,9)	94,7	(87,8-97,8)	94,7	(87,8-97,8)

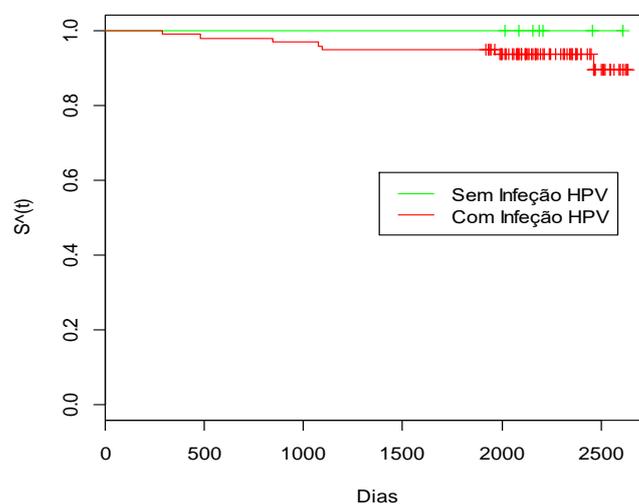


Figura 19 - Curva de sobrevivência estimada de Kaplan-Meier por infeção por HPV

Da Tabela 41 e Figura 19, podemos concluir que as mulheres sem infeção por HPV apresentaram probabilidade de sobrevivência maior do que as mulheres com infeção por HPV, para 1, 3 e 5 anos após o diagnóstico do CCU.

Tabela 42 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por antecedentes familiares

Antecedentes Familiares	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
0 - Sem antecedentes	88	95,5	(88,3-98,3)	84,1	(74,6-90,3)	84,1	(74,6-90,3)
1 - Com antecedentes	105	98,1	(92,6-99,5)	93,3	(86,5-96,8)	90,5	(83,0-94,8)

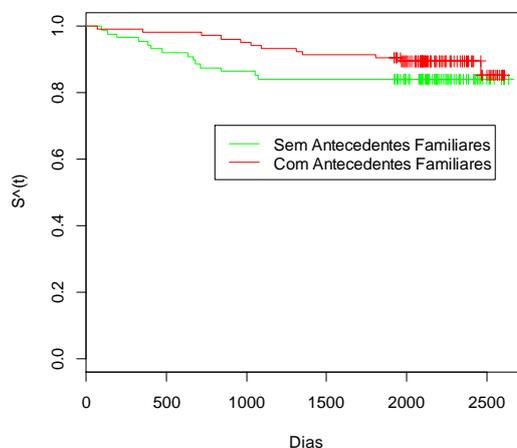


Figura 20 - Curva de sobrevivência estimada de Kaplan-Meier por antecedentes familiares

Da Tabela 42 e Figura 20, observa-se que as mulheres com antecedentes familiares com cancro apresentam maior probabilidade de sobrevivência do que as mulheres sem antecedentes familiares para 1, 3 e 5 anos após o diagnóstico do CCU.

Tabela 43 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por diabetes

Diabete	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
0 - Não	58	98,3	(88,4-99,8)	91,4	(80,5-96,3)	89,7	(78,4-95,2)
1 - Sim	12	83,3	(48,2-95,6)	75,0	(40,8-91,2)	75,0	(40,8-91,2)

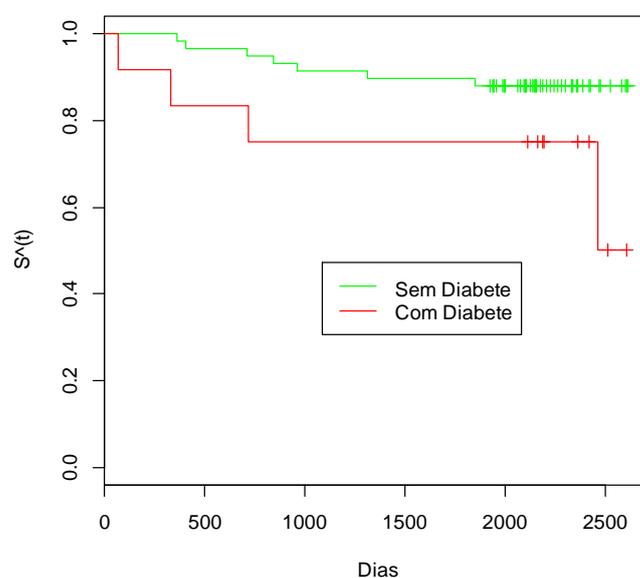


Figura 21 - Curva de sobrevivência estimada de Kaplan-Meier por diabetes

Constata-se da Tabela 43 e Figura 21, que as mulheres sem diabetes apresentam maior probabilidade de sobrevivência do que as mulheres com diabetes para 1, 3 e 5 anos após o diagnóstico do CCU.

Tabela 44 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por obesidade

Obesidade	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
0 - Não	42	97,6	(84,3-99,7)	90,5	(76,6-96,3)	90,5	(76,6-96,3)
1 - Sim	18	88,9	(62,4-97,1)	77,8	(51,1-91,0)	72,2	(45,6-87,4)

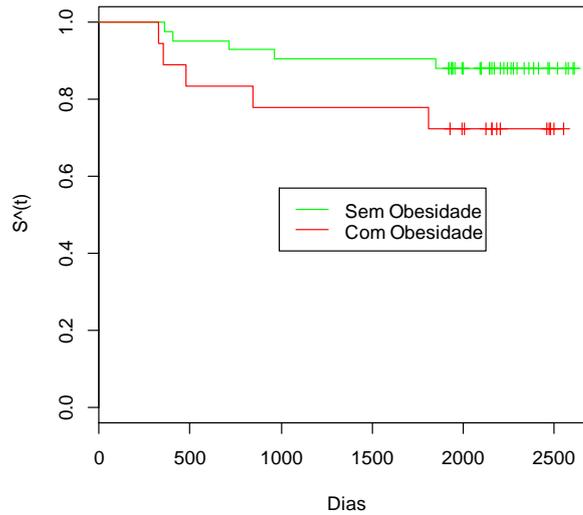


Figura 22 - Curva de sobrevivência estimada de Kaplan-Meier por obesidade

Da Tabela 44 e Figura 22, verifica-se que as mulheres sem obesidade apresentaram maior probabilidade de sobrevivência do que as mulheres com obesidade para 1, 3 e 5 anos após o diagnóstico do CCU.

Tabela 45 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por contraceção

Contraceção	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
0 - Não	25	92,0	(71,6-97,9)	84,0	(62,8-93,7)	80,0	(58,4-91,1)
1 - Sim	178	97,2	(93,4-98,8)	91,0	(85,7-94,4)	90,4	(85,1-94,0)

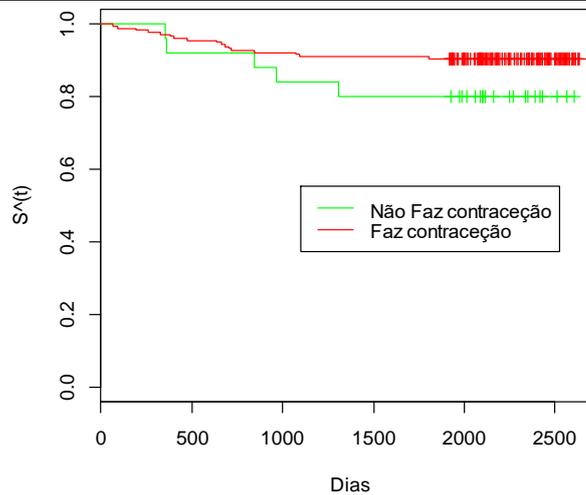


Figura 23 - Curva de sobrevivência estimada de Kaplan-Meier por contraceção

Em relação a variável contraceção, o gráfico apresentado da Figura 23, sugere que não existe diferenças significativas na sobrevivência para as mulheres que fazem ou não fazem contraceção.

Tabela 46 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por nível de escolaridade

Nível de escolaridade	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
1 - 1º Ciclo	57	93,0	(82,4-97,3)	77,2	(64,0-86,1)	77,2	(64,0-86,1)
2 - 2º Ciclo	30	100,0	(100,0-100,0)	100,0	(100,0-100,0)	100,0	(100,0-100,0)
3 - 3º Ciclo	47	97,9	(85,8-99,7)	97,9	(85,8-99,7)	97,9	(85,8-99,7)
4 - Secundário	44	97,7	(84,9-99,7)	90,9	(77,6-96,5)	90,9	(77,6-96,5)
5 - Licenciatura	25	92,0	(71,6-97,9)	84,0	(62,8-93,7)	80,0	(58,4-91,1)

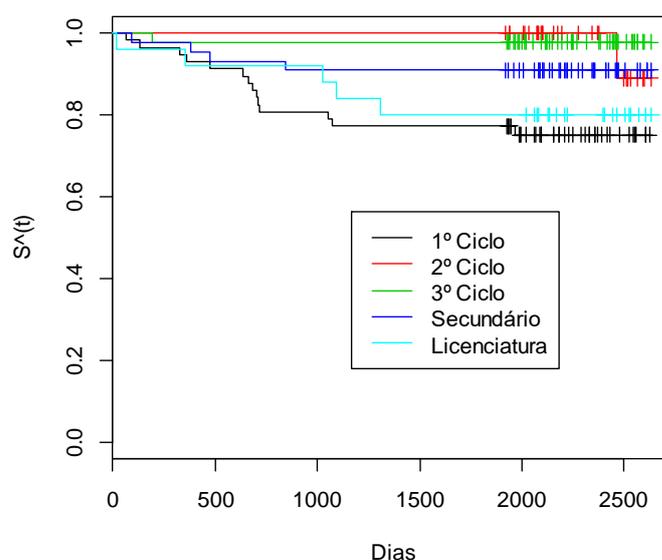


Figura 24 - Curva de sobrevivência estimada de Kaplan-Meier por nível de escolaridade

Da Tabela 46 e Figura 24, constatou-se que são as mulheres com 1º ciclo que apresentaram menor probabilidade de sobrevivência em relação às restantes, com exceção do 1 ano após o diagnóstico do CCU.

Tabela 47 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por sintomas

Sintoma	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
0 - "Assintomática"	25	100,0	(100,0-100,0)	96,0	(74,8-99,4)	96,0	(74,8-99,4)
1 - "Sintomática"	42	90,5	(76,6-96,3)	73,8	(57,7-84,6)	66,7	(50,3-78,7)

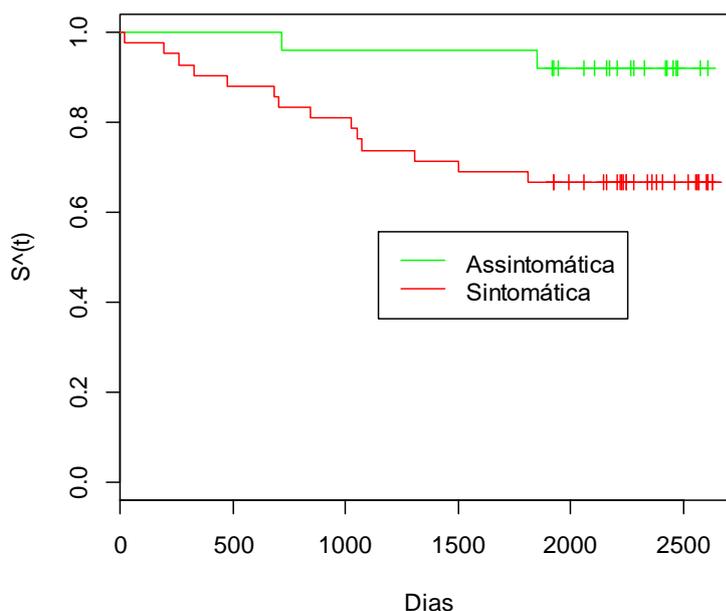


Figura 25 - Curva de sobrevivência estimada de Kaplan-Meier por sintomas

Da Tabela 47 e Figura 25, conclui-se que as mulheres que não tiveram sintomas do CCU apresentaram probabilidade de sobrevivência muito mais elevada do que as mulheres que apresentaram sintomas para 1, 3 e 5 anos após o diagnóstico do CCU.

Tabela 48 - Estimativa de sobrevivência a 1, 3 e 5 anos após o diagnóstico por ano de diagnóstico

Ano diagnóstico	n	1 ano		3 ano		5 ano	
		$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC	$\hat{S}(t)\%$	95% IC
2009	150	94,0	(88,8-96,8)	88,0	(81,6-92,3)	85,3	(78,6-90,1)
2010	174	97,1	(93,2-98,8)	91,4	(86,1-94,7)	89,1	(83,4-92,9)

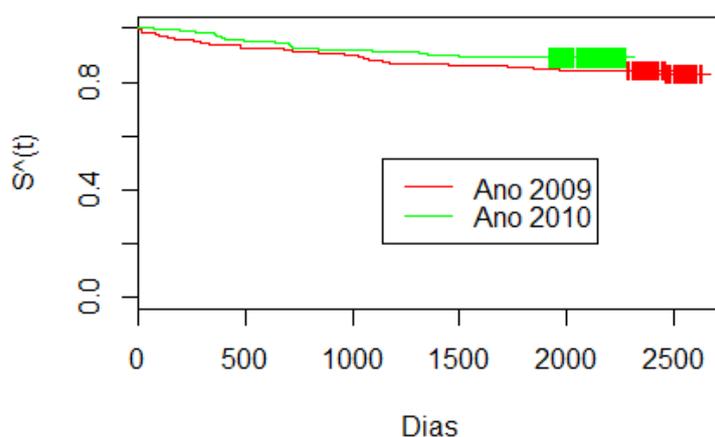


Figura 26 - Curva de sobrevivência estimada de Kaplan-Meier por ano de diagnóstico

Da Tabela 48 e Figura 26, verificou-se que o ano 2010 apresentou probabilidade de sobrevivência muito mais elevada do que o ano 2009.

De seguida, vamos estimar a função de sobrevivência, utilizando o estimador de Nelson-Aalen, que se obtém executando os seguintes comandos no Anexo XV.

Apresenta de seguida a curva de sobrevivência utilizando o estimado de Nelson-Aalen.

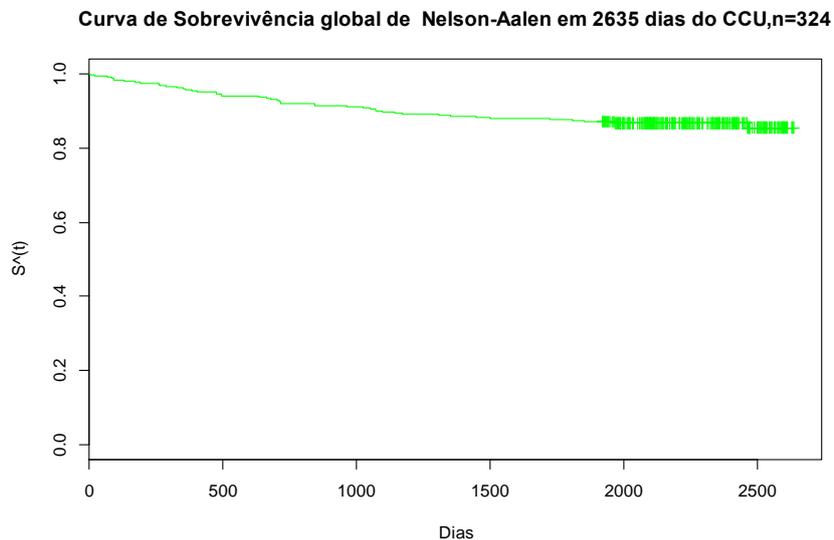


Figura 27 - Estimativa da curva de sobrevivência global do CCU de Nelson-Aalen

Tal como a curva de sobrevivência obtida utilizando o estimador de Kaplan-Meier, conclui-se que à medida que o tempo pós diagnóstico passa, a estimativa de probabilidade de sobrevivência decresce lentamente, tendendo a estabilizar em valores próximos de 0.80. Devemos salientar que no nosso caso é preferível o estimador de Kaplan-Meier, em vez do estimador de Nelson-Aalen, uma vez que o estimador de Nelson-Aalen é mais adequado para amostras de pequenas dimensões.

4.3.4 TESTE LOG-RANK

Deve-se recordar que o teste de log-rank compara a distribuição da ocorrência dos eventos observados em cada estrato com a distribuição que seria esperada se a incidência fosse igual em todos os estratos. Se a distribuição observada for equivalente à distribuição esperada, dizemos que a curva de sobrevivência das mulheres pertencentes ao estrato é equivalente à curva de sobrevivência das mulheres em geral, ou seja, a covariável não tem efeito na sobrevivência (Maria Sá Carvalho et al., 2005).

Para realizar o teste, calcula-se a estatística em duas etapas: primeiramente estima-se o número de eventos esperados para cada estrato k , segundo a hipótese nula de incidência igual em todos

os estratos. Chama-se esse número esperado no R de “Expected”. Em seguida, calcula-se a estatística de teste. Esta estatística segue uma distribuição X^2 , com k-1 graus de liberdade, quando a hipótese nula é verdadeira (Maria Sá Carvalho et al., 2005).

Interessa-nos verificar se as diferenças que observamos relativamente à sobrevivência das mulheres, tendo em conta o grupo a que pertencem, são de facto significativas. Assim sendo, utilizou-se o teste de log-rank para testar:

H_0 : As curvas de sobrevivência são todas iguais

vs

H_1 : Existe pelo menos um par em que as curvas de sobrevivência se diferenciam

Para efetuar este teste deve-se proceder da seguinte forma no R (exemplo utilizando a variável rastreio do CCU).

Tabela 49 - Teste de log rank para a variável rastreio do CCU

```
> survdiff(Surv(temp_followp, status) ~ factor(Rastreio_uter0), data=dados)
Call:
survdiff(formula = Surv(temp_followp, status) ~ factor(Rastreio_uter0),
  data = dados)

n=271, 53 observations deleted due to missingness.

              N Observed Expected (O-E)^2/E (O-E)^2/V
factor(Rastreio_uter0)=0 110      32    14.4    21.5    34.1
factor(Rastreio_uter0)=1 161       7    24.6    12.6    34.1

Chisq= 34.1 on 1 degrees of freedom, p= 5.17e-09
```

No teste de log-rank, encontramos para cada estrato o número de pessoas em risco no início do estudo (N), o número total de eventos observados (observed) e esperados (Expected). A última coluna da saída da função `survdiff()` dá-nos a estatística de teste log-rank. O resultado de teste de X^2 é dado no final da tabela e aponta para a rejeição da hipótese nula (p -valor=5.17e-09) para o caso do variável rastreio do CCU. O *software* R também calcula a estatística $(O - E)^2/E$. Esta é uma estatística de teste aproximada, sendo o seu resultado mais conservador. A estatística de log-rank é exata, e por isso, é preferível.

Efetuando o mesmo processo para cada uma das outras variáveis construiu-se a seguinte tabela:

Tabela 50 - Teste log-rank

Variável	χ^2 (df)	p-valor
Rastreio do CCU	34.1 (1)	5.17e-09
Faixa etária	44.4 (3)	1.25e-09
Estadiamento do cancro	70.5 (2)	4.44e-16
Distrito de residência	12.6 (5)	0.0277
Comportamento do cancro	65.3 (1)	6.66e-16
Estado civil	1.9 (3)	0.6
Fumador	1.5 (1)	0.214
Infeção por HPV	0.5 (1)	0.478
Antecedentes familiares	1.1 (1)	0.292
Diabete	3.2 (1)	0.0727
Obesidade	2.5 (1)	0.115
Contraceção	2.4 (1)	0.122
Nível de escolaridade	16.7 (4)	0.00216
Sintoma	5.5 (1)	0.0189
Ano do diagnóstico	1.8 (1)	0.181

Através da Tabela 50, considerando um nível de significância de 5% ($\alpha = 5\%$), verifica-se que as variáveis qualitativas, rastreio do CCU, faixa etária, estadiamento do cancro à data do diagnóstico, distrito de residência das mulheres, comportamento do cancro, nível de escolaridade e sintomas do cancro quando consideradas isoladamente, têm influência estatisticamente significativa na sobrevivência das mulheres com CCU.

4.3.5 AJUSTAMENTO DO MODELO DE REGRESSÃO DE COX

No âmbito do modelo de regressão de Cox, vamos procurar obter o modelo que melhor se adequa aos nossos dados.

Vamos agora proceder ao ajustamento do modelo de Cox aos dados. O primeiro passo consistiu em avaliar a influência de cada uma das variáveis por si só, no tempo de vida dos indivíduos.

4.3.5.1. MODELO DE REGRESSÃO DE COX SIMPLES

Nesta secção, vamos construir o modelo de Cox simples para todas as covariáveis existentes na nossa base de dados, com o intuito de perceber quais as covariáveis isoladamente podem influenciar o tempo de sobrevivência dos indivíduos.

Este procedimento no R realiza-se na seguinte forma:

Tabela 51 - Modelo de regressão de Cox simples

```
> M1<-coxph(Surv(temp_followp,status)~factor(Rastreio_uter0,levels=c(1,0)),data=dados)
> summary(M1)
Call:
coxph(formula = Surv(temp_followp, status) ~ factor(Rastreio_uter0,
  levels = c(1, 0)), data = dados)

n= 271, number of events= 39
(53 observations deleted due to missingness)

              coef exp(coef) se(coef)      z Pr(>|z|)
factor(Rastreio_uter0, levels = c(1, 0))0 2.0581    7.8310  0.4175 4.929 8.25e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
factor(Rastreio_uter0, levels = c(1, 0))0  7.831    0.1277    3.455    17.75

Concordance= 0.735 (se = 0.039 )
Rsquare= 0.116 (max possible= 0.794 )
Likelihood ratio test= 33.54 on 1 df,  p=6.979e-09
Wald test              = 24.3 on 1 df,  p=8.249e-07
Score (logrank) test = 34.13 on 1 df,  p=5.165e-09
```

A coluna coef representa os valores de β estimados pela verossimilhança parcial. Valores positivos indicam variáveis que contribuem para o aumento de risco de morte por CCU e valores negativos de β indicam variáveis que contribuem para a redução do risco de morte.

A coluna se(coef) indica o erro padrão dessas estimativas, a coluna z representa a estatística de Wald e a coluna p o p-valor do teste hipótese nula de $\beta = 0$, assumido que a estatística de Wald segue uma distribuição normal. O *software* R também calcula exp(coef). Este valor é interpretado como uma razão de riscos (Hazard Ratio), onde valores acima de 1 indica sobre-risco, valores abaixo de 1 indicam proteção. O intervalo de confiança para essa razão de risco é apresentado nas colunas lower .95 upper .95.

De seguida, encontra apresentado na Tabela 52 os parâmetros mais importantes do modelo de regressão de Cox simples para todas as covariáveis existentes na base de dados, utilizando os mesmos códigos do R apresentado na Tabela 51, mudando apenas os nomes das variáveis.

Tabela 52 - Análise de modelo de regressão de Cox simples

Variáveis	Categorias	HR=exp ($\hat{\beta}$)	95% IC para HR	P-valor
Rastreio do CCU	Não (0)	7,831	(3,455-17,75)	8,25e-07
	Sim (1)	1		
Faixa etária	[25,34] anos	0,1755	(0,02159-1,427)	0,103593
	[35,44] anos	1		
	[45,54] anos	5,5771	(2,35753-13,193)	9,15e-05
	[55,64] anos	5,1188	(2,10424-12,452)	0,000318
Estadiamento	Estádio 1	1		
	Estádio 2	2,693	(0,4498-16,13)	0,278
	Estádio 3	35,821	(9,6429-133,07)	9,06e-08
Distrito de residência	Aveiro	0,7253	(0,23874-2,204)	0,57109
	Castelo Branco	0,6996	(0,09200-5,320)	0,73000
	Coimbra	1,5757	(0,75077-3,307)	0,22931
	Guarda	0,4688	(0,06163-3,566)	0,46427
	Leiria	1		
	Viseu	3,0528	(1,35565-6,874)	0,00705
Estado civil	0 - Solteira	1,086	(0,4053-2,908)	0,870
	1 - Casada	1		
	2 - Viúva	3,677e-08	(0,0000- Inf)	0,997
	3 - Divorciada	0,6695	(0,1559-2,875)	0,589
Fumador	0 - Não	1		
	1 - Sim	0,5686	(0,2306-1,402)	0,22
Infeção por HPV	0 - Não	1,36e-08	(0,00-Inf)	0,999
	1 - Sim	1		
Antecedentes familiares	0 - Não	1		
	1 - Sim	0,6624	(0,3062-1,433)	0,295
Nível de escolaridade	1 - 1º Ciclo	1		
	2 - 2º Ciclo	0,11637	(0,01529-0,8858)	0,0378
	3 - 3º Ciclo	0,07643	(0,01005-0,5815)	0,013
	4 - Secundário	0,34053	(0,11204-1,0350)	0,0575
	5 - Licenciatura	0,75688	(0,27216-2,1049)	0,5935
Sintomas	0 - Assintomático	1		
	1 - Sintomático	4,955	(1,125-21,82)	0,0344
Ano do diagnóstico	2009	1		
	2010	0,6648	(0,3642-1,214)	0,184

Continuação da Tabela 52.

Variáveis	Categorias	HR=exp ($\hat{\beta}$)	95% IC para HR	P-valor
Comportamento do cancro	2 - In situ, não invasiva	1		
	3 - Maligno, localização primária	25,44	(7,873-82,17)	6,34e-08
Diabetes	0 - Não	1		
	1 - Sim	2,955	(0,8559-10,21)	0,0866
Obesidade	0 - Não	1		
	1 - Sim	2,613	(0,7561-9,032)	0,129
Contraceção	0 - Não	1		
	1 - Sim	0,4639	(0,1711-1,258)	0,131
Nº de Gestações		1,329	(1,164-1,517)	2,65e-05
Nº de partos		1,273	(1,091-1,485)	0,00211
Nº de abortos		1,663	(1,233-2,244)	0,000868
Nº de parceiros		0,9477	(0,7049-1,274)	0,722
Idade na 1º gravidez		0,9579	(0,8458-1,085)	0,498
Idade da 1º atividade sexual		0,9362	(0,7917-1,107)	0,44
Idade da menarca		1,027	(0,8339-1,264)	0,803

Através da Tabela 52, verifica que a coluna de P-valor que foi calculado com base na estatística de Wald, onde o teste de Wald consiste em testar a seguinte hipótese: $H_0: \beta = 0$ vs $H_1: \beta \neq 0$, ou seja $\beta = 0$ significa que a covariável isoladamente não tem influência significativa no tempo de sobrevivência por CCU. Para um nível de significância $\alpha = 5\%$, verifica-se que as covariáveis rastreio do CCU, faixa etária, estadiamento do cancro à data do diagnóstico, comportamento do cancro, distrito de residência, nível de escolaridade, sintomas, número de gestações, número de partos e números de abortos, quando considerado isoladamente, têm influência estatisticamente significativa na sobrevivência das mulheres com CCU.

Vamos exemplificar como se interpreta o $\exp(\hat{\beta})$, primeiramente só faz sentido interpretar o valor das $HR = \exp(\hat{\beta})$ se a covariável for estatisticamente significativa, isto é, se apresentar o $p\text{-valor} < 0,05$, por exemplo vamos interpretar uma variável qualitativa, rastreio do CCU apresentado na Tabela 52. Significa que uma mulher que não fez o rastreio do CCU tem 7,831 vezes mais risco de morte por CCU quando comparado com uma mulher que fez o rastreio do CCU. Convém notar o facto do intervalo de confiança para HR não conter o valor 1, indica-nos igualmente que esta variável tem influência significativa na sobrevivência dos indivíduos.

Vamos exemplificar como interpretar uma variável contínua, por exemplo, a variável número de gestações, também apresentada na Tabela 52, significa que para cada aumento de uma gestação o risco de morte por CCU aumenta 32,9%.

4.3.5.2 MODELO DE REGRESSÃO DE COX MÚLTIPLO

Nesta seção partindo das covariáveis que se revelaram estatisticamente significativas quando ajustadas isoladamente, vamos tentar ajustar um modelo de Cox múltiplo válido.

A estratégia para encontrar um modelo de Cox válido foi a de ajustar um modelo de Cox válido a partir das covariáveis que foram estatisticamente significativas no modelo de Cox simples. Deve-se recordar que as covariáveis que foram estatisticamente significativas isoladamente são: rastreio do CCU, faixa etária, estadiamento do cancro à data do diagnóstico, distrito de residência dos pacientes, nível de escolaridade, comportamento do cancro, sintomas do cancro, número de gestações, número de partos e número de abortos.

Primeiramente, para o ajustamento do modelo de Cox múltiplo foram excluídas, as variáveis sintomas do cancro e estadiamento do cancro à data do diagnóstico pelo facto de apresentarem 257 e 248 observações em falta, respetivamente. O passo seguinte, consistiu em começar o modelo com uma única covariável e adicionar as covariáveis uma a uma até que as covariáveis do modelo deixassem de ser estatisticamente significativas ($p\text{-valor} > 0,05$). No nosso caso, em particular, iniciamos o modelo apenas com a variável rastreio do CCU, que se revelou estatisticamente significativa. De seguida consideramos o modelo de Cox com as duas covariáveis, rastreio do CCU e faixa etária, onde ambas foram estatisticamente significativas. Efetuando os mesmos procedimentos, o modelo selecionado é o modelo com apenas as três covariáveis que são: rastreio do CCU, faixa etária e comportamento do CCU.

O modelo de Cox válido encontrado a partir do modelo de Cox simples, foi o modelo com as seguintes covariáveis que são: rastreio do CCU, faixa etária e o comportamento do cancro. Devemos salientar que a variável faixa etária tem 4 níveis, logo tivemos de criar 3 variáveis dummy.

A escolha de um modelo válido não foi fácil, devido ao facto de algumas covariáveis apresentarem vários valores em falta, por essa razão não foi possível utilizar o critério de AIC,

nem o teste de Razão verossimilhança para escolher o melhor modelo. De seguida encontra-se apresentado o modelo selecionado, mas, no entanto, não temos a garantia que esse modelo é o melhor modelo, visto se poder encontrar a partir desse estudo um outro modelo válido.

O modelo final é então dado por:

$$h(t, z) = h_0(t) \exp(\beta_1 \text{Rastreio_utero} + \beta_2 \text{Idade_grupo1} + \beta_3 \text{Idade_grupo3} + \beta_4 \text{Idade_grupo4} + \beta_5 \text{Comp_cancro})$$

No Anexo XVI, encontra apresentado o código do R e o output do modelo final.

De seguida encontra apresentado na Tabela 53 os parâmetros do modelo final.

Tabela 53 - Estimativa dos parâmetros do modelo final

Variáveis	Categorias	HR=exp($\hat{\beta}$)	95% IC para HR	P-valor
Rastreio do CCU	Não (0)	2,4734	(1,03899 -5,888)	0,04072
	Sim (1)	1		
Faixa etária	1 - [25,34] Anos	0,2888	(0,03515-2,372)	0,24764
	2 - [35,44] Anos	1		
	3 - [45,54] Anos	2,8863	(1,19967-6,944)	0,01796
	4 - [55,64] Anos	2,2232	(0,87998-5,617)	0,09111
Comportamento do cancro	2 - In situ, não invasiva	1		
	3 - Maligno, localização primária	8,0971	(2,29631-28,552)	0,00114

Agora o passo seguinte consiste em validar o pressuposto de proporcionalidade de risco do modelo de regressão de Cox selecionado.

4.3.5.3. VALIDAÇÃO DO PRESSUPOSTO DE RISCOS PROPORCIONAIS

Nesta secção, para validar o pressuposto de risco proporcional vão ser utilizados os resíduos de Schoenfeld.

4..5.3.1. RESÍDUOS DE SCHOENFELD

Os resíduos de Schoenfeld permitem responder, se o efeito de uma covariável é sempre o mesmo durante todo o tempo de observação, ou seja, permite verificar se o pressuposto de

riscos proporcionais para determinada covariável é aceitável. Se assim for, não deve existir nenhuma tendência sistemática no gráfico, onde a curva lowess ajudará na interpretação (Maria Sá Carvalho et al., 2005).

No Anexo XVII, encontra apresentado o código do R utilizado para obter os resíduos de Schoenfeld para o modelo final.

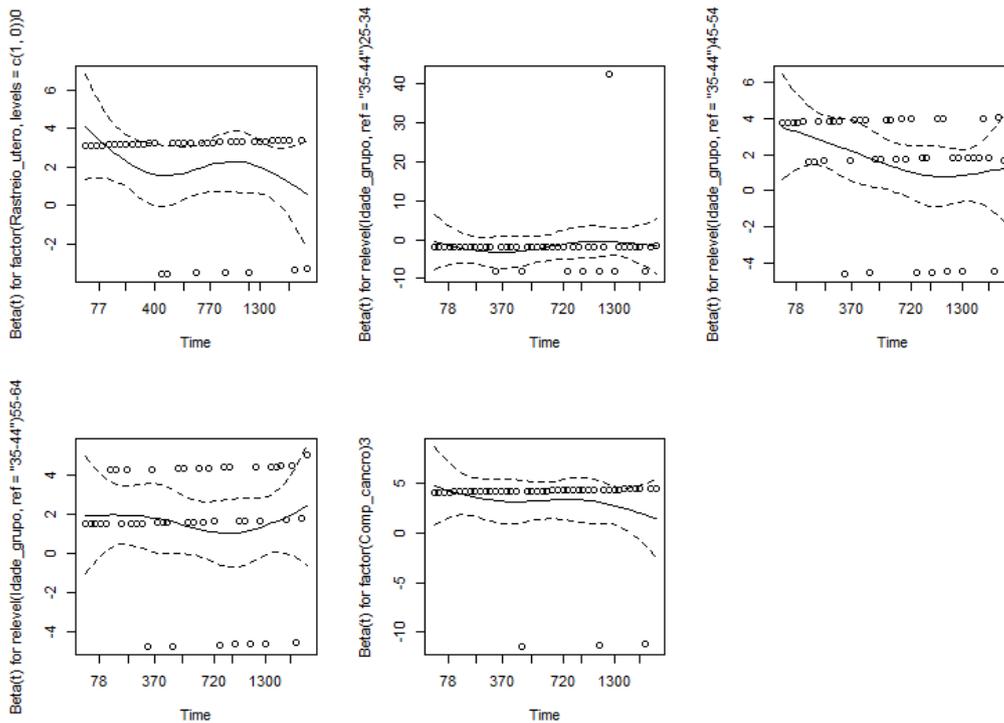


Figura 28 - Resíduos de Schoenfeld para o modelo final (apresentado na Tabela 53)

Os gráficos de resíduos de Schoenfeld são interpretados da seguinte forma, se a premissa de riscos proporcionais não é violada, então espera-se que essa reta esteja dentro dos intervalos de confiança do alisamento lowess dos resíduos (Maria Sá Carvalho et al., 2005)

Para o nosso caso em particular, todas as covariáveis em estudo é verificado o pressuposto de riscos proporcionais, uma vez que a reta se encontra dentro dos respetivos intervalos de confiança e as covariáveis parecem permanecer constantes ao longo do tempo.

Para ver a proporcionalidade do risco global do modelo final utiliza-se o seguinte código do R:

Tabela 54 - Proporcionalidade global do risco

```

> residuos<-cox.zph(M_final)
> residuos

```

	rho	chisq	p
factor(Rastreio_uterio, levels = c(1, 0))0	-0.1642	0.815	0.367
relevel(Idade_grupo, ref = "35-44")25-34	0.0538	0.121	0.728
relevel(Idade_grupo, ref = "35-44")45-54	-0.2429	2.021	0.155
relevel(Idade_grupo, ref = "35-44")55-64	0.0242	0.021	0.885
factor(Comp_cancro)3	-0.0782	0.175	0.676
GLOBAL	NA	6.402	0.269

A última coluna proporciona o p-valor da estatística de qui-quadrado (p). Dessa coluna observa-se que não se rejeita a proporcionalidade de risco para nenhuma das covariáveis. Também não é rejeitada a hipótese de proporcionalidade global do risco do modelo selecionado, uma vez que o p-valor global (0,269) é superior ao nível de significância ($\alpha=5\%$).

Como a premissa da proporcionalidade de risco foi verificada para o modelo de Cox em causa, então esse modelo de Cox é válido e estamos em condições de interpretar os parâmetros do modelo final.

4.3.5.4. INTERPRETAÇÃO DOS PARÂMETROS DO MODELO FINAL

Recordar que na Tabela 53, encontra apresentado os parâmetros do modelo selecionado para qual a premissa da proporcionalidade de risco foi verificada e lembrar que só devemos interpretar os parâmetros estatisticamente significativos.

Da Tabela 53, conclui-se que, as mulheres não submetidas ao rastreio do CCU têm 2,4734 vezes mais risco de morte por CCU quando comparado com as mulheres submetidas ao rastreio do CCU, mantendo as outras covariáveis constante, ou seja, para indivíduos da mesma faixa etária e com o mesmo tipo de comportamento do cancro.

Apenas a faixa etária dos [45,54] anos é que revelou estatisticamente significativa, por isso, só faz sentido interpretar essa faixa etária. Significa que as mulheres na faixa etária dos [45,54] anos têm 2,8863 vezes mais risco de morte por CCU do que as mulheres na faixa etária dos [35,44] anos, mantendo as outras variáveis constantes.

As mulheres com comportamento do cancro maligno, com localização primária têm 8,0971 vezes mais risco de morte por CCU do que as mulheres com comportamento do cancro in situ, não invasivo, mantendo as outras variáveis constantes.

CAPÍTULO 5

DISCUSSÃO

Da análise descritiva conclui-se que das 324 mulheres com diagnóstico do CCU no RORC, de 1 de janeiro de 2009 até 31 de dezembro de 2010, 150 tiveram o diagnóstico do CCU em 2009 e 174 em 2010. Das 324 mulheres registaram-se no total de 44 óbitos por CCU que correspondem os 13,58% (44/324). 72,73% (32/44) dos óbitos foram registados nas mulheres não submetidas ao rastreio do CCU, o que parece ser natural uma vez que as mulheres não submetidas ao rastreio do CCU apresentam sempre menor probabilidade de sobrevivência e apresentam o estadiamento do cancro à data do diagnóstico no estado muito mais avançado do que as mulheres submetidas ao rastreio do CCU. 91,18% (41/44) dos óbitos ocorreram nas mulheres com comportamento do cancro maligno, com localização primária, e está de acordo com o esperado, uma vez que as mulheres com comportamento do cancro maligno, com localização primária têm menos probabilidade de sobrevivência do que as mulheres com comportamento do cancro in situ, não invasivo. A proporção de mulheres rastreadas foi de 49,69 (161/324), a proporção das não rastreadas foi de 33,95% (110/324) e a proporção das mulheres sem informação do rastreio foi de 16,36% (53/324). 73,64 % (81/110) das mulheres não submetidas ao rastreio do CCU tiveram comportamento do cancro maligno, com localização primária. 80,75% (130/161) das mulheres submetidas ao rastreio do CCU tiveram o comportamento do cancro in situ, não invasivo. Desta análise conclui-se que o comportamento do cancro parece estar influenciado pelo rastreio, uma vez que as mulheres submetidas ao rastreio do CCU tendem a ter o comportamento in situ, não invasivo e as mulheres não submetidas ao rastreio do CCU tendem a ter comportamento do cancro maligno, com localização primária. Ainda da análise descritiva conclui-se que a mulher mais nova envolvida no estudo tem 25 anos e a mulher mais velha tem 64 anos. 75% das mulheres envolvidas no referido estudo têm idades inferiores ou iguais a 50 anos e a idade média das mulheres é de 42,62 anos. A faixa etária predominante é a dos [35,44] anos e a com menor frequência é a dos [55,64] anos. O número de óbitos mais elevado foi registado para as mulheres residentes no distrito de Coimbra e Leiria, registaram-se no total de 14 óbitos para ambos os distritos, enquanto que o mais baixo foi registado no distrito de Castelo Branco e Guarda, em que foi registado apenas um óbito para ambos os distritos. O tempo mínimo de seguimento foi de 2 dias e o tempo máximo de seguimento foi de 2635 dias.

Do teste de associação de qui-quadrado (ver Tabela 27) concluímos que existe associação estatisticamente significativa entre o rastreio do CCU e o comportamento do CCU. As mulheres não submetidas ao rastreio do CCU têm aproximadamente 4 vezes mais possibilidades de ter

comportamento do cancro maligno, com localização primária do que as mulheres submetidas ao rastreio do CCU (ver Tabela 28), este resultado está de acordo com o esperado.

Da análise de sobrevivência conclui-se que o tempo médio de sobrevivência após o diagnóstico do CCU é de 2381,3 dias. O tempo médio de sobrevivência das mulheres submetidas ao rastreio é de 2556 dias e das mulheres não submetidas ao rastreio é de 2078 dias. As mulheres submetidas ao rastreio do CCU sobrevivem em média 478 dias mais do que as mulheres não submetidas ao rastreio do CCU, o que parece ser natural, uma vez que no grupo das mulheres submetidas ao rastreio do CCU, ocorreram poucos óbitos. O tempo médio de sobrevivência das mulheres com comportamento do cancro in situ, não invasivo é de 2610 dias e das mulheres com comportamento maligno, com localização primária é de 2022 dias. Em média as mulheres com comportamento do cancro in situ, não invasivo sobrevivem 588 dias mais do que as mulheres com comportamento maligno, com localização primária, isto parece ser natural, uma vez que morreram menos mulheres com comportamento do cancro in situ, não invasivo do que as mulheres com comportamento do cancro maligno, com localização primária. De acordo com a publicação do RORC no ano 2015, podemos concluir que houve uma melhoria na sobrevivência global a 1, 3 e 5 anos após o diagnóstico do CCU nas mulheres com diagnóstico do CCU em 2007 e 2008 comparado com as mulheres com diagnóstico em 2009 e 2010. A sobrevivência global na região centro a 1, 3 e 5 anos após o diagnóstico do cancro é de 87,3%, 69,0% e 64,2% respetivamente, no total de 268 mulheres com diagnóstico do CCU em 2007 e 2008 (ROR Centro, 2015), enquanto que a probabilidade de sobrevivência global na região centro a 1, 3 e 5 anos após o diagnóstico do CCU é de 95,7%, 89,8% e 87,7% respetivamente, no total de 324 mulheres com diagnóstico em 2009 e 2010. A sobrevivência a 5 anos após o diagnóstico do CCU nas mulheres submetidas ao rastreio é 97,6% enquanto que a sobrevivência das mulheres não submetidas ao rastreio do CCU é de 71,8%. Existe uma diferença de 25,8% na sobrevivência das mulheres submetidas ao rastreio quando comparado com as mulheres não submetidas ao rastreio. Essa diferença traduz que, ao fim de 5 anos seriam poupadas 67 vidas, caso essas mulheres fossem submetidas ao rastreio do CCU, por isso é importante que as mulheres se submetam ao rastreio do CCU. As mulheres com comportamento do cancro in situ, não invasivo, após 5 anos de diagnóstico tiveram uma sobrevivência de 99,0%, enquanto as mulheres com comportamento do cancro maligno, com localização primária tiveram uma sobrevivência de 69,3% no mesmo período, e está de acordo com a literatura existente, uma vez que as mulheres com comportamento do cancro maligno, com localização primária têm sempre

menor probabilidade de sobrevivência do que as mulheres com comportamento do cancro in situ, não invasivo. As mulheres na faixa etária dos [45,54] apresentaram menor probabilidade de sobrevivência após o diagnóstico do CCU do que as mulheres na faixa etária dos [55,64] anos, para 1, 3 e 5 anos, no entanto esta diferença não é estatisticamente significativa. A sobrevivência a 5 anos após o diagnóstico do CCU, nas mulheres que apresentaram o estágio I à data de diagnóstico é de 97,6%, enquanto que o estágio II essa sobrevivência é de 81,8% e para o estágio III é de 12,5%. Das 324 mulheres apenas 76 tinham informação em relação ao estadiamento do cancro à data do diagnóstico. Apesar dessa variável apresentar um número reduzido de informações, este resultado está de acordo com a literatura existente, quanto maior é o estadiamento, menor é a probabilidade de sobrevivência por CCU.

Do modelo de regressão de Cox simples, concluiu-se que as covariáveis que por si só influenciaram a sobrevivência após o diagnóstico do CCU, foram nomeadamente: rastreio do CCU, faixa etária, estadiamento do cancro à data diagnóstico, sintomas do cancro, comportamento do cancro, distrito de residência das mulheres, nível de escolaridade, número de partos, número de gestações e número de abortos. Esses fatores aparecem também em outros estudos de sobrevivência do CCU, como fatores de risco para o CCU. Em estudos idênticos, as covariáveis como o tabagismo, infeção por HPV, idade da primeira relação sexual, obesidade, diabetes e número de parceiros sexuais aparecem como fatores de risco, no nosso caso, essas variáveis não se revelaram estatisticamente significativas, pode ser pelo facto dessas variáveis apresentarem vários valores em falta. Da mesma análise concluiu-se que as mulheres não submetidas ao rastreio do CCU, com sintomas do CCU, com estadiamento do cancro mais avançado (estadiamento IIII), com comportamento do cancro maligno, com localização primária, na faixa etária dos [45,54] anos, pertencentes ao distrito de Viseu, com 1º ciclo, com mais partos, abortos e gestações, apresentaram maior risco de morte por CCU. Ainda desta análise concluiu-se que uma mulher não submetida ao rastreio do CCU tem 7,831 vezes mais risco de morte por CCU do que uma mulher submetida ao rastreio do CCU. Uma mulher com comportamento do cancro maligno, com localização primária tem 25,44 vezes mais risco de morte por CCU do que uma mulher com comportamento do cancro in situ, não invasivo.

Do cruzamento da sobrevivência por rastreio e faixa etária, concluiu-se que existem diferenças estatisticamente significativas na sobrevivência por rastreio apenas para as mulheres com idade igual ou superior a 45 anos. Deste resultado, concluiu-se que o rastreio só tem influência na sobrevivência por CCU a partir dos 45 anos. A sobrevivência a 5 anos após o diagnóstico do

CCU das mulheres na faixa etária dos [45,54] submetidas ao rastreio do CCU foi de 92,6%, enquanto que das mulheres da mesma faixa etária, mas não submetidas ao rastreio apresentaram uma sobrevivência de 55,9%. As mulheres na faixa etária dos [55,64] anos submetidas ao rastreio do CCU apresentaram uma sobrevivência de 94,4% e as mulheres na mesma faixa etária, mas não submetidas ao rastreio do CCU apresentaram uma sobrevivência de 60,7%.

Do modelo de Cox múltiplo conclui-se que as covariáveis que influenciaram a sobrevivência por CCU são, rastreio do CCU, faixa etária e o comportamento do cancro. Verificou-se ainda que as mulheres não submetidas ao rastreio do CCU têm 2,4734 vezes mais risco de morte por CCU do que as mulheres submetidas ao rastreio do CCU, para as mulheres da mesma faixa etária e com o mesmo tipo de comportamento do cancro. As mulheres com comportamento do cancro maligno, com localização primária têm 8,0971 vezes mais risco de morte por CCU do que as mulheres com comportamento do cancro in situ não invasivo, mantendo as outras variáveis constantes. Também se conclui que apenas as faixas etárias dos [45,54] anos é que revelaram estatisticamente significativos. Ainda do modelo de Cox múltiplo selecionado, conclui-se que como os parâmetros $\hat{\beta}$ são todos positivos, com exceção da faixa etária dos [25,34] anos, (ver Anexo XVI), isso significa que as mulheres não submetidas ao rastreio do CCU, com comportamento do cancro do colo do útero maligno, com localização primária e com idades pertencentes a qualquer uma das faixas etárias, exceto a faixa etária dos [25,34] anos (faixa etária protetor) têm um risco acrescido de ter óbito por CCU, por apresentarem o $HR > 1$. No entanto apenas a faixa etária dos [45,54] é que apresentam risco estatisticamente significativa, por apresentarem o $p\text{-valor} > 0.05$.

Ainda, com esse estudo pretende-se chamar a atenção das mulheres para a necessidade da realização do rastreio do CCU, uma vez que, de acordo com os resultados do estudo obtidos, as mulheres submetidas ao rastreio têm uma probabilidade de sobrevivência muito mais elevada do que as mulheres não submetidas ao rastreio, visto que, o quanto antes for feito o rastreio, menor será o estadiamento da doença e conseqüentemente mais fácil será o tratamento da doença. Também verificamos que o comportamento do cancro é influenciado por rastreio, isto é, as mulheres submetidas ao rastreio do CCU tendem a ter comportamento do cancro in situ e apresentaram maior probabilidade de sobrevivência e as mulheres não submetidas ao rastreio do CCU tendem a ter comportamento do CCU maligno, com localização primária e apresentaram menor probabilidade de sobrevivência.

CAPÍTULO 6

CONCLUSÕES, LIMITAÇÕES E RECOMENDAÇÕES

Neste capítulo iremos apresentar as conclusões do presente estudo, tendo também em consideração os resultados e a respetiva discussão, descrito no capítulo anterior. Após as conclusões, serão também apresentadas algumas limitações encontradas e feitas algumas recomendações para futuros estudos a realizar nesta área.

6.1. CONCLUSÕES

O presente estudo teve como objetivo principal, verificar se existem diferenças estatisticamente significativas na sobrevivência das mulheres submetidas e não submetidas ao rastreio do CCU e também verificar quais os fatores que podem influenciar a sobrevivência por CCU. Este estudo foi realizado com base em 324 arquivos médicos das mulheres com idade compreendidas entre os 25 e 64 anos e com diagnóstico do CCU em RORC do IPO de Coimbra por qualquer uma das fontes, com exceção de ROENO e RORSUL, a partir de 1 de janeiro de 2009 até 31 de dezembro de 2010. Podemos concluir que houve uma melhoria na sobrevivência global da Região Centro, das doentes com diagnóstico em 2007 e 2008 comparado com as doentes com diagnóstico em 2009 e 2010 para 1, 3 e 5 anos após o diagnóstico do CCU, esta comparação foi feita com base na publicação RORC no ano 2015. Pode-se determinar que no final deste estudo, para um nível de significância de 5% existem diferenças estatisticamente significativas na sobrevivência das mulheres submetidas e não submetidas ao rastreio do CCU, ou seja, as mulheres submetidas ao rastreio do CCU apresentaram maior probabilidade de sobrevivência após o diagnóstico do CCU. Após 5 anos do diagnóstico do CCU as mulheres submetidas ao rastreio do CCU apresentaram uma probabilidade de sobrevivência de 96,9% e as mulheres não submetidas ao rastreio do CCU no mesmo período apresentaram uma sobrevivência de 71,8%, ainda se conclui deste estudo que as mulheres submetidas ao rastreio do CCU sobrevivem em média 478 dias mais do que as mulheres não submetidas ao rastreio do CCU. Do cruzamento de sobrevivência por rastreio e faixa etária concluímos que existem diferenças estatisticamente significativas na sobrevivência por rastreio apenas para idade maior ou igual a 45 anos. Também concluímos que as cováriáveis que isoladamente influenciaram a sobrevivência por CCU foram rastreio do CCU, faixa etária, estadiamento do cancro à data do diagnóstico, comportamento do cancro, sintomas do cancro, nível de escolaridade, distrito de residência dos pacientes, número de partos, número de gestações e número de abortos. Do modelo de regressão de Cox múltiplo verificamos ainda que as mulheres não submetidas ao rastreio do CCU, com comportamento do

cancro maligno com localização primária e pertencentes a faixa etária dos [45,54] apresentaram um risco acrescido estatisticamente significativo para ter óbito por CCU.

Conclui-se que as mulheres não submetidas ao rastreio do CCU têm aproximadamente 4 vezes mais possibilidades de ter cancro com comportamento maligno, com localização primária do que as mulheres submetidas ao rastreio do CCU.

As técnicas de rastreio são atualmente, um dos componentes chave nos programas de prevenção do CCU. O diagnóstico precoce constitui ainda a ferramenta mais eficaz no combate a essa patologia.

Verifica-se que, de todos os tumores malignos, o CCU é o que pode ser controlado com maior efetividade, com possibilidade de redução de cerca de 80% da incidência, através de programas de rastreio citológico cervical organizados, de base populacional com periodicidade de 3 ou 5 anos. Conclui-se ainda que, em Portugal, apesar das recomendações oficiais e da comunidade científicas, o rastreio citológico da mulher ainda continua a ser oportunístico, o que poderá ser apontado como a causa dos elevados números de novos casos do CCU que surgem a cada ano.

6.2. LIMITAÇÕES

Na continuidade do que foi concluído, é importante referir as limitações que foram encontradas na realização do referido estudo:

- ✓ Inexistência de estudo a nível Regional e Nacional que compara sobrevivência das mulheres rastreadas e não rastreadas;
- ✓ Como foram analisadas doentes diagnosticadas com CCU em 2009 e 2010 e nessa altura as informações não se encontravam informatizadas, a recolha de dados foi feita nos arquivos médicos um a um, o que se revelou bastante trabalhoso;
- ✓ Várias variáveis recolhidas apresentaram vários valores em falta, isso contribui para que algumas variáveis que aparecem em outros estudos como fator de risco para o CCU, não se revelaram estatisticamente significativas para o nosso estudo;
- ✓ Devido as variáveis apresentarem valores em falta ("NA"), não foi possível usar o critério de AIC, nem o teste de razão de verosimilhança para encontrar o melhor modelo;
- ✓ Na maioria dos casos das mulheres submetidas ao rastreio do CCU não foi possível saber se fizeram rastreio organizado de base populacional ou se fizeram rastreio por

iniciativa do médico de família, logo não podemos distinguir o tipo de rastreio presente no referido estudo.

Mesmo, diante das limitações citadas, a presente investigação tem a sua relevância porque mostrou a real situação da sobrevivência das mulheres com CCU na Região Centro. Também, com esse estudo, conseguimos concluir que existem diferenças estatisticamente significativas na sobrevivência das mulheres submetidas e não submetidas ao rastreio do CCU. Pode dizer-se que só existem diferenças estatisticamente significativas na sobrevivência das mulheres submetidas e não submetidas ao rastreio do CCU, para as mulheres com idades igual ou superior a 45 anos. A importância deste estudo deve-se ao facto de não se encontrar outro realizado, nem a nível Regional nem a nível Nacional.

6.3. RECOMENDAÇÕES

- ✓ Novos estudos semelhantes, necessitam a serem realizados para melhor se explorar a probabilidade de sobrevivência por CCU das mulheres submetidas e não submetidas ao rastreio do CCU e também para explorar quais os fatores que podem influenciar a sobrevivência por CCU, tanto a nível regional como a nível nacional;
- ✓ Implementação de políticas públicas para amenizar os casos do CCU, principalmente com a administração da vacina do HPV;
- ✓ Nunca esquecer que o papel do médico de família, no processo de programa de rastreio do CCU, é fundamental. Os médicos de família devem adquirir conhecimentos básicos acerca dos aspetos epidemiológicos do programa de rastreio, bem como fazer uma atualização constante dos conhecimentos em diversas áreas, para assim melhor servir o seu doente e conseguir melhores ganhos na saúde.

BIBLIOGRAFIA

- Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6, 701–726.
- Alves, T. (2003). Prevenção do cancro do colo do útero. *Revista Portuguesa de Clínica Geral*, 19, 455–460.
- Andersen, P., & Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research*, 11, 91–115.
- Andrade JM, NH, Y., AB, O., M, P., ST, P., JV, P., & MJ, A. (2001). Rastreamento, Diagnóstico e Tratamento do Carcinoma do Colo do Útero. *Projeto Diretrizes Associação Médica Brasileira E Conselho Federal de Medicina*, 1–18.
- Arbyn, M., Anttila, A., Jordan, J., Ronco, G., Schenck, U., Segnan, N., ... von Karsa, L. (2010). European Guidelines for Quality Assurance in Cervical Cancer Screening. Second edition—summary document. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 21(3), 448–458. <http://doi.org/10.1093/annonc/mdp471>
- ARS Algarve, IP. (2016). Retrieved November 12, 2016, from <http://www.arsalgarve.min-saude.pt/portal/?q=node/3828>
- Atlas da Saúde. (2015). Luta contra o cancro do colo do útero. Retrieved November 13, 2016, from <http://www.atlasdasaude.pt/publico/content/em-media-mil-novos-casos-de-cancro-do-colo-do-utero-em-portugal>
- Ault, A. (2006). Advances with HPV vaccines for the prevention of cervical cancer and other HPV - related diseases. *Physician 's Weekly*, 23(33), 1–4.
- Ault, A. (2007). Effect of prophylactic human papillomavirus L1 virus-like-particle vaccine on risk of cervical intraepithelial neoplasia grade 2, grade 3, and adenocarcinoma in situ: a combined analysis of four randomised clinical trials. *The Lancet*, 369(6), 1861–1868.
- Baseman, J., & Koutsky, L. (2005). The epidemiology of human papillomavirus infections. *Journal of Clinical Virology*, 325, 516–524.
- Bolfarine, H., Rodrigues, J., & Achcar, J. A. (1991). Análise de sobrevivência. // *Escola de*

Regressão. IM-UFRJ, RJ.

- Botelho, F., Silva, C., & Cruz, F. (2009). Epidemiologia explicada—análise de sobrevivência. *Acta Urológica*.
- Bragueto, T., & Suzuki, L. E. (2008). Vacinas contra o Papilomavírus Humano-(HPV). *NewLab*, 87, 58–68.
- Brenna, S. M. F., Hardy, E., Zeferino, L. C., & Namura, I. (2001). Conhecimento, atitude e prática do exame de Papanicolaou em mulheres com câncer de colo uterino. *Cadernos de Saúde Pública*, 17(4), 909–914.
- Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30, 89–99.
- Breslow, N., & Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, (2), 437–453.
- Burd, E. (2003). Human Papillomavirus and Cervical Cancer. *Clinical Microbiology Reviews*, 16(1), 1–17.
- Bustamante-Teixeira, M. T., Faerstein, E., & Latorre, M. do R. (2002). Técnicas de análise de sobrevida. *Cadernos de Saúde Pública*, 18(3), 579–594. <http://doi.org/10.1590/S0102-311X2002000300008>
- Carvalho, M. S., Andreozzi, V. L., Codeço, C. T., Barbosa, M. T. S., & Shimakura, S. E. (2005). *Análise de Sobrevida Teoria e Aplicação em Saúde*. (F. O. Cruz, Ed.) (20th ed.). Rio de Janeiro.
- Carvalho, M. S., Andreozzi, V. L., Codeço, C. T., Campos, D. P., Barbosa, M. T. S., & Shimakura, S. E. (2011). *Análise de Sobrevivência: teoria e aplicações em saúde* (2ª).
- Castellsagué, X., Díaz, M., Sanjosé, S. de, Muñoz, N., Herrero, R., Franceschi, S., ... Bosch, F. X. (2006). Worldwide Human Papillomavirus Etiology of Cervical Adenocarcinoma and Its Cofactors: Implications for Screening and Prevention. *Journal of National Cancer Institute*, 98(5), 303–315. <http://doi.org/10.1093/jnci/djj067>
- CDC. (2009). Centers for Disease Control and Prevention. Chapter 9 - Human Papillomavirus (HPV). *The Pink Book: Chapters - Epidemiology and Prevention of Vaccine - Preventable Diseases*, 1–12.
- CervexBrush. (2016). CervexBrush. Home page. Retrieved November 13, 2016, from

<http://www.cervexbrush.com/index.html>

Collett, D. (2015). Modelling survival data in medical research.

Colosimo, E. . (1995). Análise de Sobrevivência Aplicada. *Reunião Anual Da Região Brasileira Da Sociedade Internacional de Biometria (RBRAS)*.

COLOSIMO, E. A., & GIOLO, S. R. (2005). *ANÁLISE DE SOBREVIVÊNCIA APLICADA*. (EDGARD BLUCHER, Ed.).

Colosimo, E. A., & Giolo, S. R. (2006). Análise de Sobrevivência Aplicada.

Cox, D. (1972). Regression models and life tables. *Journal of the Royal Statistical Society*, 2(34), 187–220.

Cox, D. (1975). Partial likelihood. *Biometrika*, 62(6), 269–276.

Cox, D., & Oakes, D. (1984). *Analysis of survival data*. Chapman and Hall. London – New York.

Cox, J. T. (1995). Epidemiology of cervical intraepithelial neoplasia: the role of human papillomavirus. *Baillière's Clinical Obstetrics and Gynaecology*, 9(1), 1–37.
[http://doi.org/doi:10.1016/S0950-3552\(05\)80357-8](http://doi.org/doi:10.1016/S0950-3552(05)80357-8)

Crow, E., & Shimizu, K. (1988). Lognormal distributions: Theory and applications. *New York: Marcel Dekker*.

Crowley, Hu, M. (1977). Covariance Analysis of Heart Transplant Survival Data. *Journal of the American Statistical Association*, 72, 277–281.

Cutts, F., Franceschi, S., Goldie, S., Castellsague, X., Sanjose, S. de, GarnettV, G., ... Markowitz, L. (2007). Human papillomavirus and HPV vaccines: a review. *Bulletin of the World Health Organization*, 85(9), 719–726. <http://doi.org/http://dx.doi.org/10.1590/S0042-96862007000900018>

Direcção Geral de Saúde. (2008). Plano Nacional de Vacinação (PNV). Introdução da vacina contra infecções por vírus do Papiloma Humano (Texto de apoio à Circular Normativa nº22/DSCS/DPCD).

ECCA. (2016). The European Cervical Cancer Association. Retrieved November 13, 2016, from <http://www.ecca.info/>

FIGO. (2016). International Federation of Gynecology and Obstetrics. Home page. Retrieved from

<http://www.figo.org/>

Grambsch, P., & Therneau, T. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. Retrieved from <http://biomet.oxfordjournals.org/content/81/3/515.short>

Greenwood, M. (1926). Reports on Public Health and Medical Subjects. *London: Her Majesty's Stationary Office*, 33, 1–26.

Info Cancro. (2016). CANCRO DO COLO DO ÚTERO: FACTORES DE RISCO. Retrieved November 13, 2016, from <http://www.roche.pt/sites-tematicos/infocancro/index.cfm/tipos/cancro-do-colo-do-utero/ccu-factores-de-risco/>

Instituto Nacional de Estatística. (2016). Causas de morte 2014.

JAMA Oncology. (2015). The Global Burden of Cancer 2013, 505–527. <http://doi.org/10.1001/jamaoncol.2015.0735>

Kahn, J. A. (2009). HPV Vaccination for the Prevention of Cervical Intraepithelial Neoplasia. *The New England Journal of Medicine*, 361, 271–278. <http://doi.org/10.1056/NEJMct0806938>

Kalbfleisch, & Prentice. (1980). The Statistical Analysis of Failure Time Data. New York.

Kaplan, E., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457–481.

Lawless, J. (1982). Statistical models and methods for lifetime data. *New York: Wiley*.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.

Meier, P. (1975). Estimation of a distribution function from incomplete observations. *In Perspectives in Probability and Statistics*, Ed. J. Gani, 67–87.

Moscicki, A.-B., Hills, N., Shiboski, S., Powell, K., Jay, N., Hanson, E., ... Palefsky, J. (2001). Risks for incident human papillomavirus infection and low-grade squamous intraepithelial lesion development in young females. *Journal of the American Medical Association*, 285(23), 2995–3002. <http://doi.org/doi:10.1001/jama.285.23.2995>

Muñoz, N., Bosch, F. X., Sanjosé, S. de, Herrero, R., Castellsagué, X., Shah, K. V., ... Meijer, C. J. L. M. (2003). Epidemiologic Classification of Human Papillomavirus Types Associated

- with Cervical Cancer. *The New England Journal of Medicine*, 348(2), 518–527. <http://doi.org/10.1056/NEJMoa021641>
- Muñoz, N., Manalastas, R., Pitisuttithum, P., Tresukosol, D., Monsonego, J., Ault, K., ... Saah, A. (2009). Safety, immunogenicity, and efficacy of quadrivalent human papillomavirus (types 6, 11, 16, 18) recombinant vaccine in women aged 24-45 years: a randomised, double-blind trial. *Lancet (London, England)*, 373(9679), 1949–57. [http://doi.org/10.1016/S0140-6736\(09\)60691-7](http://doi.org/10.1016/S0140-6736(09)60691-7)
- Naucler, P., Ryd, W., Törnberg, S., Strand, A., Wadell, G., Elfgren, K., ... Dillner, J. (2007). Human Papillomavirus and Papanicolaou Tests to Screen for Cervical Cancer. *New England Journal of Medicine*, 357(16), 1589–1597. <http://doi.org/10.1056/NEJMoa073204>
- Nelson, W. (1972). Theory and Applications of Hazard Plotting for Censored Failure Data. *Technometrics*, 14, 945–966.
- Nelson, W. (1990). Accelerated Life Testing: Statistical Models Test Plan and Data Analysis. *J. Wiley, New York, USA*.
- O País. (2013). Cancro do colo do útero no mundo. Retrieved November 13, 2016, from <http://opais.sapo.mz/index.php/sociedade/45-sociedade/26355-cancro-do-colo-do-utero-no-mundo.html>
- Organização Panamericana da Saúde, & Organização Mundial da Saúde. (2000). *Classificação Internacional de Doenças para Oncologia* (3^o). Genebra.
- Paavonen, J., & Patricia. (2007). Efficacy of human papillomavirus (HPV - 16/18) AS04 - adjuvanted vaccine against cervical infection and pre-cancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomized study in young women. *The Lancet*, 369, 2161–2170.
- Palefsky, J. M., & Holly, E. A. (2003). Chapter 6: Immunosuppression and Co-infection with HIV. *Journal of the National Cancer Institute Monographs*, 31, 41–46.
- Poli, M., & Silveira, G. (1994). Ginecologia preventiva. *Porto Alegre: Artes Médicas*.
- ROBERTO A. MEDRONHO. (2009). *Epidemiologia*. (Atheneu, Ed.) (2.^a). São Paulo.
- ROR Centro. (2015). *SOBREVIVÊNCIA Doentes diagnosticados na Região Centro em 2007 e 2008*. (I. P. de O. de Coimbra, Ed.). Coimbra. Retrieved from <http://www.rorcentro.com.pt>

- Schiffman, M., & Castle, P. (2003). *Human papillomavirus: epidemiology and public health. Archives of Pathology and Laboratory Medicine* (Vol. 127).
- Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1), 239–241.
- Seeley, Stephens, & Tate. (2001). Anatomia & fisiologia. In Lusodidacta (Ed.), (3rd ed.). Porto. Retrieved from <http://hdl.handle.net/10849/204>
- Sociedade Portuguesa de Ginecologia Oncológica. (2007). CANCRO GINECOLÓGICO REUNIÃO DE CONSENSO NACIONAL. Coimbra.
- Solomon, D., & Nayar, R. (2005). O sistema Bethesda para o relato de diagnóstico citológico cervico-vaginal. In *Rio de Janeiro: Revinter* (2nd ed.).
- Sousa, A. P. de. (2011). *Cancro do Colo do Útero: Tendências e Estudos Recentes*. Universidade Fernando Pessoa.
- SPG. (2010). Sociedade Portuguesa de Ginecologia e SPGO - Secção Portuguesa de Ginecologia (Org.). *Vacinas Contra O HPV: Reunião Do Consenso Nacional. Cascais*.
- Stanley, M. (2008). HPV vaccines: are they the answer? *British Medical Bulletin*, 88, 59–74.
- Stewart, B. W., & Wild, C. P. (2014). *World Cancer Report 2014*. (I. A. for R. on Cancer, Ed.).
- Trottiera, H., & Franco, E. L. (2006). The epidemiology of genital human papillomavirus infection. *Vaccine*, 24(1), 4–15. <http://doi.org/http://dx.doi.org/10.1016/j.vaccine.2005.09.054>
- Wee, C. C., Huang, A., Huskey, K. W., & McCarthy, E. P. (2008). Obesity and the likelihood of sexual behavioral risk factors for HPV and cervical cancer. *National Institute of Health: Obesity*, 16(11), 2552–2555. <http://doi.org/10.1038/oby.2008.394>
- Weibull, W. (1939). A statistical theory of the strength of materials. *Ingeniors Vetenskaps Akademien*, 151(45–55).
- Weibull, W. (1949). A statistical representation of fatigue failures in solids. *Trans. Royal Institute of Technology*.
- Weibull, W. (1951). A Statistical Distribution Function of Wide Applicability. *Journal of Applied Mechanics*, 18, 293–297.
- Widdice, L. E., & Moscicki, A.-B. (2008). Updated Guidelines for Papanicolaou Tests, Colposcopy,

and Human Papillomavirus Testing in Adolescents. *Journal of Adolescent Health*, 43(4), S41–S51. <http://doi.org/http://dx.doi.org/10.1016/j.jadohealth.2008.04.007>

Wolschick, N. M., Consolaro, M. E. L., Suzuki, L. E., & Boer, C. G. (2007). Câncer do colo do útero: tecnologias emergentes no diagnóstico, tratamento e prevenção da doença. *Revista Brasileira de Análises Clínicas*, 39(2), 123–129. Retrieved from <http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src=google&base=LILACS&lang=p&nextAction=lnk&exprSearch=477002&indexSearch=ID>

Zeferino, L. C. (2008). O desafio de reduzir a mortalidade por câncer do colo do útero. *Revista Brasileira de Ginecologia E Obstetrícia*, 30(5), 213–215. <http://doi.org/10.1590/S0100-72032008000500001>

ANEXOS

Anexo I. REPRESENTAÇÃO GRÁFICA PARA COMPARAR CURVA DE SOBREVIVÊNCIA DE KAPLAN-MEIER

```
> fit<-survfit(Surv(time,status)~x,data=aml)
> plot(fit,col=c(3,2),mark.time=T,ylab="S^(t)"
+ ,main="Curva de sobrevivência de Kaplan-Meier")
> legend(locator(1) ,c("Maintained", "Nonmaintained")
+ ,lty=c(1,1) , col=c(3,2))
```

Anexo II. COMPARAÇÃO DE CURVA DE SOBREVIVÊNCIA USANDO O TESTE LOG-RANK

```
> survdiff(Surv(time,status)~x,data=aml, rho=0)
Call:
survdiff(formula = Surv(time, status) ~ x, data = aml, rho = 0)


```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
x=Maintained	11	7	10.69	1.27	3.4
x=Nonmaintained	12	11	7.31	1.86	3.4

```

Chisq= 3.4 on 1 degrees of freedom, p= 0.0653

```

Anexo III. LEITURA DA BASE DE DADOS A UTILIZAR

```
> dados<-read.csv2("Base de Dados_final.csv",header=T)
```

Anexo IV. CATEGORIZAÇÃO DAS VARIÁVEIS NO R

```
> dados$Estadio<-factor(dados$Estadio)
> dados$Rastreo_uterio<-factor(dados$Rastreo_uterio)
> dados$Estado_civil<-factor(dados$Estado_civil)
> dados$Contacecao<-factor(dados$Contacecao)
> dados$Nivel_escol<-factor(dados$Nivel_escol)
> dados$Diabete<-factor(dados$Diabete)
> dados$Infecao_HPV<-factor(dados$Infecao_HPV)
> dados$Antecedentes_familiar<-factor(dados$Antecedentes_familiar)
> dados$Fumador<-factor(dados$Fumador)
> dados$Obesidade<-factor(dados$Obesidade)
> dados$Sintomas<-factor(dados$Sintomas)
> dados$status<-factor(dados$status)
> dados$Comp_cancro<-factor(dados$Comp_cancro)
> dados$Distrito<-factor(dados$Distrito)
> dados$Ano diag<-factor(dados$Ano diag)
```

Anexo V. VERIFICAÇÃO DOS PRESSUPOSTOS DE TESTE DE ASSOCIAÇÃO DE QUI-QUADRADO

```
> a<-table(dados$Rastreio_uter0,dados$Comp_cancro,dnn=c("Rastreio","Comportamento"))
> a
      Comportamento
Rastreio  2    3
0      29   81
1     130   31
> chisq.test(a)$expected##### permite ver se as freq esperada maior que 5 e nenhum menor que 1
      Comportamento
Rastreio  2    3
0  64.53875 45.46125
1  94.46125 66.53875
```

Anexo VI. CÁLCULO DO RISCO RELATIVO

```
> library(epitools)
> epitab(dados$Rastreio_uter0,dados$Comp_cancro,method="riskratio",pvalue="chi2",rev="rows")
$tab
      Outcome
Predictor  2      p0  3      p1 riskratio  lower  upper  p.value
1  130 0.8074534 31 0.1925466  1.00000  NA    NA    NA
0   29 0.2636364 81 0.7363636  3.82434 2.734325 5.34888 4.349907e-19

$measure
[1] "wald"

$conf.level
[1] 0.95

$spvalue
[1] "chi2"
```

Anexo VII. BOXPLOT E ESTATÍSTICA DESCRITIVA PARA VARIÁVEL IDADE

```
> boxplot(dados$Idade,main="Distribuição da Idade",ylab="Idade")
> mean(dados$Idade)##### dá a média
[1] 42.62037
> median(dados$Idade)##### dá a mediana
[1] 41
> min(dados$Idade)##### dá o mínimo
[1] 25
> max(dados$Idade)##### dá o máximo
[1] 64
> quantile(dados$Idade,c(0.25,0.50,0.75))##### dá os quartil
25% 50% 75%
34  41  50
```

Anexo VIII. CRIAÇÃO DA VARIAVEL FAIXA ETÁRIA

```
> dados$Idade_grupo[dados$Idade >= 25 & dados$Idade <=34] <- 1
> dados$Idade_grupo[dados$Idade >= 35 & dados$Idade <=44] <- 2
> dados$Idade_grupo[dados$Idade >= 45 & dados$Idade <=54] <- 3
> dados$Idade_grupo[dados$Idade >= 55 & dados$Idade <=64] <- 4
> dados$Idade_grupo<- factor(dados$Idade_grupo,labels=c("25-34","35-44","45-54","55-64"))
```

Anexo IX. FREQUÊNCIAS ABSOLUTAS E RELATIVAS PARA FAIXA ETÁRIA

```
> table(dados$Idade_grupo)##### frequência absoluta
25-34 35-44 45-54 55-64
   89   111    68    56
> round(prop.table(table(dados$Idade_grupo))*100,2)##### freq relativo
25-34 35-44 45-54 55-64
27.47 34.26 20.99 17.28
```

Anexo X. ESTIMAÇÃO DE FUNÇÃO DE SOBREVIVÊNCIA GLOBAL DE KAPLAN-MEIER E A SUA REPRESENTAÇÃO GRÁFICA

```
> fit1<-survfit(Surv(temp_followp,status)~1,data=dados
+ ,conf.type=c("log-log"))
> plot(fit1,main="Curva de Sobrevivência global de Kaplan-Meier em 2635 dias do CCU,n=324"
+ ,conf.int=F,col=3,mark.time=T,xlab="Dias",ylab="S^(t)")
> round(summary(fit1,time=c(365,1095,1825))$surv*100,1)#sobrevivência a 1, 3, 5 anos
[1] 95.7 89.8 87.3
```

Anexo XI. TEMPO MÉDIO DE SOBREVIVÊNCIA POR CCU

```
> fit1<-survfit(Surv(temp_followp,status)~1,data=dados,conf.type=c("log-log"))
> print(fit1,print.rmean=T)##### tempo médio de sobrevivência
Call: survfit(formula = Surv(temp_followp, status) ~ 1, data = dados,
  conf.type = c("log-log"))
```

n	events	*rmean	*se(rmean)	median	0.95LCL	0.95UCL
324.0	44.0	2381.3	37.3	NA	NA	NA

* restricted mean with upper limit = 2635

Anexo XII. ESTIMAÇÃO DE FUNÇÃO DE SOBREVIVÊNCIA DE KALAN-MEIER POR RASTREIO E A SUA REPRESENTAÇÃO GRÁFICA

```
> fit2<-survfit(Surv(temp_followp,status)~factor(Rastreio_uteru),
+ conf.type=c("log-log"),data=dados)
> plot(fit2,col=c(2,3),mark.time=T,xlab="Dias",ylab="S^(t)")
> legend(locator(1),c("Não Rastreio","Rastreio"),lty=c(1,1),col=c(2,3))
```

Anexo XIII. TEMPO MÉDIO DE SOBREVIVÊNCIA DO CCU POR RASTREIO DO CCU

```
> fit2<-survfit(Surv(temp_followp,status)~factor(Rastreio_uter0),
+ conf.type=c("log-log"),data=dados)
> print(fit2,print.rmean=T)#####média por rastreio
Call: survfit(formula = Surv(temp_followp, status) ~ factor(Rastreio_uter0),
  data = dados, conf.type = c("log-log"))

 53 observations deleted due to missingness
      n events *rmean *se(rmean) median 0.95LCL 0.95UCL
factor(Rastreio_uter0)=0 110     32   2078      85.9    NA      NA      NA
factor(Rastreio_uter0)=1 161      7  2556      26.5    NA      NA      NA
* restricted mean with upper limit = 2622
```

Anexo XIV. TEMPO MÉDIO DE SOBREVIVÊNCIA DO CCU POR COMPORTAMENTO DO CANCRO

```
> fit24<-survfit(Surv(temp_followp,status)~factor(Comp_cancro),
+ data=dados,conf.type=c("log-log"))
> print(fit24,print.rmean=T)
Call: survfit(formula = Surv(temp_followp, status) ~ factor(Comp_cancro),
  data = dados, conf.type = c("log-log"))

      n events *rmean *se(rmean) median 0.95LCL 0.95UCL
factor(Comp_cancro)=2 197      3   2610      13.9    NA      NA      NA
factor(Comp_cancro)=3 127     41   2022      83.3    NA      NA      NA
* restricted mean with upper limit = 2633
```

Anexo XV. ESTIMAÇÃO DE FUNÇÃO DE SOBREVIVENCIA GLOBAL USANDO O ESTIMADO DE NELSON-AALEN

```
> M0<-survfit(coxph(Surv(temp_followp,status)~1,data=dados))
> plot(M0,conf.int=F,mark.time=T,xlab="Dias"
+ ,main="Curva de Sobrevivência global de Nelson-Aalen em 2635 dias do CCU,n=324",
+ ylab="S^(t)",col="green")##### mark.time=T vai mostrar no grafico as censuras
```

Anexo XVI. MODELO FINAL DE COX MULTIVARIADO

```

> M_final<-coxph(Surv(temp_followp,status)~factor(Rastreio_uter0,levels=c(1,0))
+ +relevel(Idade_grupo, ref="35-44")+factor(Comp_cancro),data=dados)
> summary(M_final)
Call:
coxph(formula = Surv(temp_followp, status) ~ factor(Rastreio_uter0,
  levels = c(1, 0)) + relevel(Idade_grupo, ref = "35-44") +
  factor(Comp_cancro), data = dados)

n= 271, number of events= 39
(53 observations deleted due to missingness)

              coef exp(coef) se(coef)      z Pr(>|z|)
factor(Rastreio_uter0, levels = c(1, 0))0  0.9056  2.4734  0.4425  2.046 0.04072 *
relevel(Idade_grupo, ref = "35-44")25-34 -1.2422  0.2888  1.0745 -1.156 0.24764
relevel(Idade_grupo, ref = "35-44")45-54  1.0600  2.8863  0.4479  2.366 0.01796 *
relevel(Idade_grupo, ref = "35-44")55-64  0.7989  2.2232  0.4729  1.690 0.09111 .
factor(Comp_cancro)3                      2.0915  8.0971  0.6430  3.253 0.00114 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
factor(Rastreio_uter0, levels = c(1, 0))0  2.4734  0.4043  1.03899  5.888
relevel(Idade_grupo, ref = "35-44")25-34  0.2888  3.4631  0.03515  2.372
relevel(Idade_grupo, ref = "35-44")45-54  2.8863  0.3465  1.19967  6.944
relevel(Idade_grupo, ref = "35-44")55-64  2.2232  0.4498  0.87998  5.617
factor(Comp_cancro)3                      8.0971  0.1235  2.29631  28.552

```

Anexo XVII. RESÍDUOS DE SCHOENFELD

```

> m1<-coxph(Surv(temp_followp,status)~factor(Rastreio_uter0,levels=c(1,0))
+ ,data=dados)
> m2<-coxph(Surv(temp_followp,status)~relevel(Idade_grupo,ref="35-44"),data=dados)
> m3<-coxph(Surv(temp_followp,status)~factor(Comp_cancro),data=dados)
> par(mfrow=c(2,3))
> res_rastreio<-cox.zph(m1)
> plot(res_rastreio)
> res_Idade<-cox.zph(m2)
> plot(res_Idade)
> res_Comp<-cox.zph(m3)
> plot(res_Comp)

```