



Joana Paula Saraiva Santos

Regressão Logística aplicada aos fatores de risco de infeções em doentes com neoplasias hematológicas

Dissertação para obtenção do Grau de Mestre em Estatística para a Saúde

Orientador: Professora Doutora Sara Simões Dias, Invited Assistant,
Faculdade de Ciências Médicas da Universidade Nova de Lisboa

Co-orientadora: Maria Rosário Oliveira Martins, Professora Catedrática do
IHMT-NOVA



Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical



2 de maio de 2022

[Regressão Logística aplicada aos fatores de risco de infeções em doentes com neoplasias hematológicas]

Copyright © Joana Paula Saraiva Santos, Instituto de Higiene e Medicina Tropical e Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

O Instituto de Higiene e Medicina Tropical Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

*Este trabalho é dedicado aos doentes da Unidade de Hematologia do Centro
Hospitalar de Setúbal, pela sua confiança*

Agradecimentos

À minha orientadora Prof. Doutora Sara Simões Dias e à coordenadora do curso Prof. Doutora Maria Rosário Oliveira pela imensa disponibilidade ao longo de todo o curso.

Ao Instituto de Higiene e Medicina Tropical por permitir a realização deste curso de Mestrado em Estatística aplicada à Saúde e pela sua primorosa organização num formato de e-learning.

À Comissão de Ética para a Saúde e Administração do Centro Hospitalar de Setúbal por terem permitido a realização deste trabalho e, por último, mas não menos importantes, às minhas colegas Dra. Anabela Neves e Dra. Ana Santo António pela sua colaboração.

Resumo

As infecções durante o tratamento em doentes com neoplasias hematológicas constituem um fator de morbidade e mortalidade importante neste grupo de doentes. O conhecimento epidemiológico das intercorrências infecciosas é uma ferramenta importante na sua abordagem terapêutica. Os métodos de regressão são uma componente fundamental da análise de dados que pretende estabelecer a relação entre um resultado e uma ou mais variáveis preditivas. Este estudo teve como objetivo identificar, recorrendo a métodos de regressão logística, os fatores de risco associados ao internamento por infeção em doentes com doença hematológica tratados num centro não académico. As causas mais frequentes de internamento foram neutropenia febril sem foco (28%), infeção do trato urinário (16%) e pneumonia (8%), sendo os agentes microbiológicos mais frequentemente isolados bactérias gram-negativas. Na análise multivariada foi identificado como fator protetor o valor de hemoglobina (OR=0,6, IC 95% 0,42-0,89) e como fator de risco o tratamento antineoplásico prévio (OR=4,8, IC 95% 3,35-6,98).

Abstract

Infections during treatment of patients with hematological neoplasms are an important morbidity and mortality factor. The knowledge in the epidemiology of infections in this group of patients is an important tool in its therapeutic approach. Regression methods are a fundamental component of data analysis for estimating relationships between an outcome and one or more predictive variables. This study aimed to identify, using logistic regression methods, the risk factors linked with hospitalization for infection in patients with hematological neoplasms treated in a non-academic centre. The most frequent causes of hospitalization were febrile neutropenia (28%), urinary tract infections (16%) and pneumonia (8%) and the most frequently isolated microbiological agents were gram-negative bacteria. In the multivariate analysis, the hemoglobin value on diagnosis had a protective effect (OR 0.6, IC 95% 0,42-0,89) while previous antineoplastic treatment (OR 4.8, IC 95% 3,35-6,98) was identified as a risk factor.

Índice

Resumo	6
Abstract.....	7
Lista de Tabelas	10
Lista de Figuras.....	10
Introdução	1
Regressão Logística	6
Ajuste do modelo de regressão logística.....	6
Testar a significância dos coeficientes	8
Estimação dos intervalos de confiança	9
Interpretação do ajuste do modelo de regressão	10
Seleção de variáveis	12
Qualidade do ajuste do modelo	14
Metodologia.....	19
Tipo de estudo e participantes.....	19
Variáveis	20
Análise estatística.....	21
Resultados.....	23
Caracterização dos participantes	23
Análise de fatores de risco para internamento	27
Caracterização dos internamentos durante tratamento.....	27
Caracterização dos internamentos por intercorrências infecciosas.....	28
Isolamentos microbiológicos	29
Evolução.....	29
Discussão	30

Conclusão.....	33
Apêndice	34
Tratamentos realizados.....	34
R Code.....	35

Lista de Tabelas

Tabela 1 - Dados socio-demográficos. ICC – índice de comorbilidade de Charlson. PS – Performance Status de acordo com o Eastern Cooperative Oncology Group (ECOG)..	24
Tabela 2 - Caracterização dos doentes no momento de inclusão do estudo.	26
Tabela 3 - Regressão logística univariada e multivariada.	27
Tabela 4 - Causas de internamento - todas as causas.	28
Tabela 5 - Motivo de internamento por causa infecciosa.	28

Lista de Figuras

Gráfico 1 - Distribuição dos diagnósticos observados, por grupo.	244
---	-----

Introdução

Nas ciências da saúde existe o interesse em estabelecer relações causais entre um evento e as características observadas num indivíduo ou grupo de indivíduos, com o intuito de prever quais os fatores que determinam um dado acontecimento, como o aparecimento de uma doença ou a resposta a uma determinada intervenção.

Os métodos de regressão são uma componente fundamental da análise de dados que pretende estabelecer a relação entre um resultado (a variável resposta ou variável dependente), e uma ou mais variáveis preditivas (variáveis independentes). Quando a variável dependente é contínua aplicam-se métodos de regressão linear. Os métodos de regressão logística aplicam-se em situações em que a variável resposta é binária ou dicotómica, por exemplo, a presença ou ausência de um evento.

Neste estudo pretendem-se identificar através de regressão logística os fatores de risco associados ao internamento por infeção (evento) em doentes submetidos a tratamento antineoplásico.

Em Portugal, no ano de 2020, a incidência de Linfoma Não Hodgkin em 2020 foi de 2 098 casos, correspondendo a 3,5% do total de incidência de todas as neoplasias; a incidência de leucemia foi de 1547 (2,6% do total); o Mieloma Múltiplo teve uma incidência de 886 casos (1,5% do total) e de Linfoma de Hodgkin foi de 233 casos (0,39% do total) (International Agency for Research on Cancer, 2021).

Apesar de se verificar um aumento da sobrevivência nos doentes com neoplasias hematológicas, reflexo da melhoria dos cuidados de saúde e desenvolvimento de novos tratamentos (Horasan *et al.*, 2011; Kochanek *et al.*, 2019; Menzo *et al.*, 2015), as complicações infecciosas durante o tratamento continuam a ser um fator de morbilidade e

mortalidade acrescida, o que pode ser atribuído à idade crescente da população doente, à presença de múltiplas comorbilidades, e ao estado de imunossupressão associado à doença e aos tratamentos (Babady, 2016; Gustinetti & Mikulska, 2016).

Os fatores de risco associados a complicações infecciosas são múltiplos e frequentemente presentes em simultâneo. Podem dividir-se categoricamente entre aqueles associados ao hospedeiro e ao tratamento. No primeiro grupo incluem-se: quebra de barreiras anatómicas, imunodeficiências humorais, imunodeficiências celulares, disfunção de órgão, doenças concomitantes, infeções anteriores e estado nutricional. Os tratamentos antineoplásicos realizados incluem os agentes clássicos de quimioterapia, anticorpos monoclonais ou inibidores de receptores celulares, que induzem imunossupressão por disfunção do sistema imunitário. A radioterapia predispõe também a infeção por inflamação e lesão estrutural dos tecidos e pode ainda afetar a imunidade celular. O recurso a dispositivos invasivos como catéteres periféricos ou centrais podem ser pontos de partida de infeção, por disrupção de barreiras anatómicas, e a utilização de antibióticos predispõe à colonização por bactérias resistentes (Brioli et al., 2019).

A neutropenia induzida pela quimioterapia, definida como a contagem absoluta de neutrófilos inferior a 500/ μ L, ou <1000/ μ L com descida prevista para 500/ μ L, é um dos efeitos adversos imediatos e potencialmente grave associado ao tratamento antineoplásico. A neutropenia febril, definida pela presença de temperatura oral >38.3°C ou em duas avaliações consecutivas >38.0°C durante 2 h na presença de neutropenia, é considerada uma emergência médica. Nestes doentes a febre pode ser o único sinal de infeção e o início de antibioterapia empírica é recomendado, devendo ser reajustada após os resultados dos exames microbiológicos (Babady, 2016; Strojnik et al., 2016).

A maioria dos estudos publicados sobre a epidemiologia das infeções em doentes com neoplasias hematológicas incluem simultaneamente doentes com neoplasias hematológicas e oncológicas, ou exclusivamente doentes com neutropenia, ou com complicações específicas como pneumonia, bacteriemia ou sepsis. Para além disso, a maioria dos estudos é realizada em centros académicos especializados e estimativas sobre a epidemiologia da sepsis e choque séptico em doentes neutropenicos para além destes centros é escassa (Kochanek *et al.*, 2019; Zorina e Styche, 2015). Adicionalmente, a maioria dos estudos realizados inclui doentes quer com neoplasias hematológicas com

necessidade de tratamento intensivo em internamento e com alto risco de complicações (nomeadamente leucemias agudas e doentes submetidos a transplante de medula óssea), quer doentes de menor risco tratados em ambulatório (por exemplo, doenças linfoproliferativas ou mieloma múltiplo).

Epidemiologia

A frequência das infeções reportada na literatura é variável diferindo entre o grupo de doentes que é analisado. Num artigo de revisão sobre os episódios de febre e infeção em doentes com cancro (Nesher e Rolston, 2014) em aproximadamente 45-50% dos doentes não havia evidência clínica ou microbiológica de infeção, em apenas 20-25% dos episódios de febre era identificado um agente causador e 20-25% dos doentes tinham um foco de infeção identificado (por exemplo, pneumonia, enterocolite, celulite) mas com resultados culturais negativos. Os locais mais comuns de infeção encontrados em doentes com neutropenia são: infeções respiratórias (35-40%), bacteriémia (15-35%), tracto urinário (5-15%), infeções da pele (5-10%), e infeções da cavidade oral e tracto intestinal (5-10%).

Num estudo retrospectivo em doentes com neutropenia induzida por quimioterapia e infeção grave admitidos em Unidade de Cuidados Intensivos que incluiu 75,4% de doentes com linfoma (Strojnink et al., 2016) as infeções mais frequentes foram: bacteriémia (36,2%), infeção do trato respiratório inferior (36,2%), trato gastrointestinal (18,8%), trato respiratório superior (2,9%), tracto urinário (1,5%), pele e tecidos moles (1,5%) e febre inexplicada (2,9%). Num outro estudo (Eyre *et al.*, 2021) que incluiu apenas doentes com diagnóstico de LNH B difuso grandes células em tratamento de 1ª linha houve 33% de admissões hospitalares por pelo menos um episódio de infeção, sendo as causas mais frequentes de internamento neutropenia febril (5,5%) e febre ou infeção não neutropénica (4,9%). Já noutro estudo prospetivo incluindo doentes com diversos tipos de linfoma não Hodgkin (Pettengell *et al.*, 2009) a incidência de neutropenia febril, calculada como a percentagem de doentes com neutropenia febril durante o tratamento, foi de 22%.

Em doentes com Mieloma Múltiplo (Brioli *et al.*, 2019), submetidos a diferentes tipos de tratamento as infeções mais frequentes foram pneumonia (8%), bacteriémia (6-8%), sepsis (2%) e infeção por varicela zoster (1%).

Com o objetivo de avaliar a incidência de pneumonia, bacteriemia e infecção urinária adquiridas em meio hospitalar foram analisados todos os doentes hospitalizados às 48h e identificados 31% de casos de infecções hospitalares (Huoi *et al.*, 2013); as incidências de pneumonia, bacteriemia e infecção urinária calculadas como o número de eventos por 1000 pacientes-dia em risco, foram, respetivamente, de 3.3, 12.0 e 2.9 por 1000 pacientes-dia.

Vários estudos epidemiológicos têm documentado a predominância de infecções causadas por bactérias gram-positivas relativamente a gram-negativas como causas de infecção documentada microbiologicamente em doentes neutropénicos atingindo os 75-80% nalguns centros, relacionado com a utilização de cateter central, a profilaxia antibiótica dirigida contra bacilos gram-negativos e com fármacos associados a mucosite. Contudo, a maioria destas publicações são focadas na bacteriemia monobacteriana que constitui apenas 20-30% dos casos microbiologicamente documentados (Nesher e Rolston, 2014). Em infecções noutras locais como pulmão, trato gastrointestinal e urinário, existe um predomínio de bactérias gram-negativas. Para além disso, 60-80% das infecções das infecções polimicrobianas incluem gram-negativos e 30-35% são causadas exclusivamente por gram-negativos.

De facto, alguns centros começam a reportar um crescimento na frequência de gram-negativos mesmo em doentes com bacteriemia, resultante da descontinuação da profilaxia antibacteriana e ao surgimento de bactérias gram-negativas multirresistentes (Menzo *et al.*, 2015).

Avaliação do risco

Atualmente existem índices de risco para complicações médicas no contexto de neutropenia febril com o objetivo de identificar os doentes com neutropénia febril. Uma das ferramentas mais utilizadas na avaliação dos doentes com neutropenia é o índice de risco MASCC (Multinational Association for Supportive Care in Cancer). Neste estudo (Klastersky *et al.*, 2000) foram identificados através da análise de regressão logística os fatores independentes no momento de instalação de febre, associados aos seguintes resultados do episódio de neutropenia febril: 1) resolução da febre nos 5 dias consecutivos sem complicações graves, ou 2) resolução da febre nos 5 dias seguintes com complicações. Primeiramente, foi feita uma análise univariada de modo a selecionar as

variáveis associadas ao resultado, e foram selecionadas todas as covariáveis com valor p inferior a 0,1 para serem incluídos no modelo multivariado. Foram identificadas as seguintes características para identificar doentes com baixo risco de complicações: sem sintomas ou sintomas mínimos/moderados; carga sintomática da doença, sem hipotensão, sem DPOC, tumor sólido ou neoplasia hematológica sem história de infecção fúngica prévia, sem desidratação com necessidade de fluidoterapia, instalação da febre em ambulatório, idade inferior a 60 anos (Klastersky *et al.*, 2000).

As complicações previstas pelo índice MASCC incluem: hipotensão, insuficiência respiratória, admissão em unidade de cuidados intensivos, coagulação intravascular disseminada, confusão, delírio ou alteração do estado mental, insuficiência cardíaca congestiva, diátese hemorrágica com necessidade de suporte transfusional, arritmia ou alterações do electrocardiograma com necessidade de tratamento, insuficiência renal com necessidade de investigação e/ou tratamento ou outras complicações clinicamente significativas.

Contudo os índices de risco não devem substituir critérios clínicos de gravidade a serem avaliados pelo médico no momento da avaliação inicial, tais como a presença de neutropenia profunda (definida por neutropenia $\leq 100/\mu\text{L}$ com duração esperada > 7 dias), disfunção de órgão ou a presença de outras comorbilidades significativas (Taplitz *et al.*, 2018).

Neste estudo é feito o seguimento de doentes com neoplasia hematológica em início de tratamento com quimioterapia com o objetivo de caracterizar epidemiologicamente os internamentos ocorridos durante o tratamento e identificar os fatores de risco associados ao internamento por causa infecciosa utilizando o método de regressão logística e respetivo cálculo do *odds ratio*.

Regressão Logística

Num modelo de regressão, assume-se que a mudança numa variável independente, x , irá diretamente conduzir a uma mudança na variável dependente, y . Por convenção, a variável dependente é representada graficamente num eixo vertical, eixo y , e a variável independente, num eixo horizontal, ou eixo x , e a equação $y = \beta_0 + \beta_1 x$ é definida como a equação de regressão linear onde β_0 é o coeficiente de interceção com o eixo y , e β_1 é o coeficiente de regressão. Assim, β_0 é o valor da equação quando $x=0$ e β_1 é o declive da linha, e quando a variável independente x aumenta uma unidade, a variável dependente y irá mudar por β_1 unidades (Michael J. Campbell, David Machin e Stephen J. Walters, 2007).

No caso de uma variável dicotómica, o valor de y para cada variação de x situa-se entre $0 \leq y \leq 1$. Num gráfico deste tipo, a variação de y por cada variação de x torna-se progressivamente mais pequena à medida que o valor de y se aproxima de 0 ou 1, semelhante à distribuição cumulativa de uma variável continua. Por este motivo, os modelos de regressão logística baseiam-se numa distribuição logística (David W. Hosmer et al., 2013).

Ajuste do modelo de regressão logística

Assumindo que temos uma amostra com n observações independentes do par (x_i, y_i) , $i=1,2, \dots, n$, onde y_i é o valor da variável dicotómica e x_i é o valor da variável independente para o sujeito i . A variável dicotómica, ou resultado, y_i , é codificada como

0 ou 1, representando a ausência ou presença de uma determinada característica, respetivamente (David W. Hosmer et al., 2013).

O modelo é escrito considerando um resultado positivo para a variável dependente, ou seja, a presença de determinada característica ou evento, e assumindo que a probabilidade esperada de um resultado positivo para um indivíduo com o fator de risco x é π . Assim, a forma do modelo de regressão logística é definida como, $\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$, e a transformação *logit* de $\pi(x)$ é:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x.$$

Para o ajuste do modelo de regressão logística para um conjunto de dados é necessário que se estimem os valores de β_0 e β_1 . Em regressão linear, o método mais usado para estimar estes parâmetros é o método dos mínimos quadrados, em que são selecionados para os valores de β_0 e β_1 que minimizam a soma dos desvios ao quadrado dos valores observados de y em relação aos esperados. Contudo, quando o modelo dos mínimos quadrados é aplicado a um modelo com um resultado dicotómico, os estimadores não reúnem as mesmas propriedades estatísticas. O método geral de estimação que conduz à função dos mínimos quadrados no modelo de regressão linear, em que os erros apresentam uma distribuição normal, é chamada máxima verosimilhança.

Este método estima os valores para os parâmetros desconhecidos que maximizam a probabilidade de obter o conjunto de dados observados. Este método aplica a função de máxima verosimilhança, que expressa a probabilidade dos dados observados como uma função dos parâmetros desconhecidos. Os estimadores de máxima verosimilhança são os valores que maximizam esta função.

A expressão para $\pi(x)$ estima a probabilidade condicional de y ser igual a 1 dado o valor x , e $1 - \pi(x)$ representa a probabilidade condicional de y ser igual a 0 dado x : $\Pr(Y=0|x)$. Assim, para os pares (x_i, y_i) , onde $y_i = 1$, a contribuição para a função de verosimilhança é $\pi(x_i)$ e para os pares onde $y_i = 0$, a contribuição para a função de verosimilhança é $1 - \pi(x_i)$. Para o par (x_i, y_i) a contribuição para a função de verosimilhança expressa-se por: $\pi(x_i)^{y_i} [1 - \pi(x_i)]^{1 - y_i}$. Assumindo que as observações são independentes, a função de verosimilhança é obtida através do produto dos termos da equação anterior:

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i}.$$

O princípio da máxima verosimilhança pressupõe que se utilize o valor estimado de β que maximiza a expressão na equação anterior. A transformação logarítmica da função de verosimilhança é:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\}.$$

Para determinar o valor de β que maximiza $L(\beta)$ diferencia-se a equação em relação a β_0 e β_1 e igualizam-se ambas as equações a 0. Estas equações são conhecidas como equações de verosimilhança, para β_0 e β_1 respetivamente:

$$\sum [y_i - \pi(x_i)] = 0 \text{ e } \sum x_i [y_i - \pi(x_i)] = 0.$$

Na regressão linear, as equações de máxima verosimilhança são obtidas pela função da soma dos desvios quadrados, são lineares para os parâmetros β_0 a β_1 e podem ser facilmente resolvidas. No caso das expressões de máxima verosimilhança para β_0 e β_1 usadas em regressão logística, não são lineares e requerem a utilização de software programado em regressão logística (David W. Hosmer et al., 2013).

Testar a significância dos coeficientes

Uma vez estimados os coeficientes, pretende-se avaliar a significância das variáveis no modelo. Este processo envolve a formulação e testagem de hipóteses estatísticas para determinar se as variáveis independentes no modelo estão significativamente relacionadas com a variável resposta.

Uma abordagem para testar a significância de um coeficiente de uma variável num modelo consiste em formular a seguinte questão: o modelo que inclui a variável em questão é mais informativo sobre a variável dependente do que o modelo que não inclui a variável? Esta questão é respondida comparando os valores observados da variável resposta com aqueles preditos por cada um dos dois modelos.

Em regressão logística, a comparação entre os valores observados e preditos é feita com base na função de *log-verosimilhança*, baseada na seguinte expressão:

$$D = -2 \ln \left[\frac{\text{verosimilhança do modelo ajustado}}{\text{verosimilhança do modelo saturado}} \right].$$

O valor entre os parêntesis é a razão de verossimilhança e fazendo menos 2 vezes o seu logaritmo, dá o valor de estatística D, designando o desvio.

Para avaliar a significância de uma variável independente, são comparados os valores de D com e sem a variável independente na equação. A mudança do desvio, D, devido à inclusão da variável independente no modelo é dada pela estatística G:

$$G = D(\text{modelo sem a variável}) - D(\text{modelo com a variável})$$

Os softwares de regressão logística fornecem quer o valor de *log-verossimilhança* ou o valor de D, pelo que é possível avaliar a significância da adição de novas variáveis no modelo ou de verificar um valor reportado de G. No caso de uma única variável independente, primeiro é ajustado um modelo contendo o coeficiente β_0 constante. Depois, é ajustado um modelo com a variável independente juntamente com a constante. Isto dá origem a outro valor de *log-verossimilhança*. O teste da razão de verossimilhança é obtido multiplicando a diferença entre estes dois valores por - 2. O resultado é dado juntamente com valor *p* associado à distribuição chi-quadrado.

Outro teste estatisticamente equivalente é o teste de Wald. O teste de Wald é igual à razão do valor estimado do coeficiente β_1 pela estimativa do seu desvio padrão, habitualmente apresentados nos programas de software. Sob a hipótese nula e suposições de tamanho da amostra, esta razão segue uma distribuição normal. Alguns programas de software também apresentam a estatística $W^2 = z^2$, que tem a distribuição chi-quadrado com 1 grau de liberdade.

Alguns autores defendem que estes dois testes, teste de razão de verossimilhança e teste de Wald, são ambos aplicáveis, contudo, em situações em que os valores são semelhantes, e um teste tem $p < 0.05$ e outro tem $p > 0.05$, deve ser usado o valor *p* do teste de razão de verossimilhança (David W. Hosmer et al., 2013).

Estimação dos intervalos de confiança

O intervalo de confiança estimados para os coeficientes β_0 e β_1 são frequentemente referidos como intervalos de confiança de Wald. Os limites $100 \times (1 - \alpha)\%$ para o

coeficiente β_1 são $\widehat{\beta}_1 \pm z_{1-\alpha/2} \widehat{SE}(\widehat{\beta}_1)$, e para o coeficiente β_0 são $\widehat{\beta}_0 \pm z_{1-\alpha/2} \widehat{SE}(\widehat{\beta}_0)$ (David W. Hosmer, Stanley Lemeshow e Rodney X. Sturdivant, 2013).

Interpretação do ajuste do modelo de regressão

A interpretação do ajuste do modelo deve ser feita assumindo que as variáveis no modelo são significativas do ponto de vista clínico ou estatístico. Posteriormente, a questão que se coloca é, que informação os coeficientes das variáveis independentes do modelo fornecem, ou seja, qual a variação da variável dependente por unidade de mudança da variável independente.

Por exemplo, numa regressão linear do peso sobre a altura, em centímetros, de adolescentes masculinos em que o coeficiente β_1 seja 5, significa que cada aumento de 1 cm em altura está associado a um aumento de 5 Kg de peso.

No modelo de regressão logística, o coeficiente estimado é a variação em *logit* correspondente a uma mudança de uma unidade na variável independente [i.e., $\beta_1 = g(x+1) - g(x)$]. A interpretação dos coeficientes depende das escalas da variável independente (David W. Hosmer et al., 2013).

Variável dicotômica independente

A variável independente é nominal ordenada e dicotômica, codificada como 0 ou 1. A diferença em *logit* para um indivíduo com $x = 1$ e $x = 0$ é:

$$g(1) - g(0) = (\beta_0 + \beta_1 \times 1) - (\beta_0 + \beta_1 \times 0) = (\beta_0 + \beta_1) - (\beta_0) = \beta_1.$$

Assim, o coeficiente β_1 , ou diferença *logit*, é a diferença entre o logaritmo da probabilidade quando $x=1$ e o logaritmo da probabilidade quando $x=0$. Contudo, em termos práticos, a mudança na escala logarítmica é difícil de explicar e pode não ser explícita para interpretação. De modo a fornecer uma interpretação mais significativa é necessário introduzir o *odds ratio* (OR), ou razão de chances, como medida de associação.

As chances de um resultado positivo (i.e. $y=1$) entre os indivíduos com $x=1$ é $\pi(1)/[1-\pi(1)]$ e as chances do resultado entre os indivíduos com $x=0$ é $\pi(0)/[1-\pi(0)]$.

A razão de chances (OR) é a razão entre a chance para $x=1$ e a chance para $x = 0$, e é dada pela equação:

$$OR = \frac{\frac{\pi(1)}{[1-\pi(1)]}}{\frac{\pi(0)}{[1-\pi(0)]}}.$$

Assim, e depois de substituir as expressões pelas probabilidades do modelo de regressão logístico obtém-se a relação entre a razão de chances e os coeficientes de regressão: $OR = e^{\beta_1}$.

Por exemplo, se a variável dependente y for definida como “presença ou ausência de cancro do pulmão” e se x for a variável independente que define que um indivíduo é fumador, um OR de 2 significa que a chance de cancro do pulmão entre fumadores é duas vezes maior que a probabilidade de cancro do pulmão entre não fumadores, na população em estudo. Outro exemplo, sendo a variável y definida como a presença ou ausência de doença cardíaca e x a variável que define se um indivíduo pratica exercício físico intenso, um OR = 0.5, significa que a o risco de doença cardíaca entre os indivíduos que praticam exercício é metade daquela dos indivíduos que não praticam exercício.

Juntamente com os valores estimados devem ser sempre referidos os intervalos de confiança para fornecer informação adicional sobre o valor dos parâmetros. As inferências sobre os intervalos de confiança são habitualmente baseados na distribuição logarítmica de OR, $\ln(\widehat{OR}) = \widehat{\beta}_1$, que tende a seguir uma distribuição normal. Obtém-se um intervalo de confiança $100 \times (1 - \alpha)\%$ para a razão de chances calculando os limites de um intervalo de confiança para o *log - odds ratio* e depois calculando a sua exponencial. Os limites dos intervalos de confiança são dados pela expressão: $\exp[\widehat{\beta}_1 \pm z_{1-\alpha/2} \times \widehat{SE}(\widehat{\beta}_1)]$ (David W. Hosmer et al., 2013).

Variável continua independente

Quando um modelo de regressão logística contém uma variável continua independente, a interpretação do coeficiente estimado depende da forma como é introduzido no modelo e das unidades da variável. Assumindo que a função *logit* é linear na covariável continua, x , a equação para *logit* é $g(x) = \beta_0 + \beta_1 x$.

Supondo que estamos interessados na razão de chances para o incremento de uma unidade na covariável (i.e., $x + 1$ versus x), na equação *logit*, $x + 1$ é: $g(x+1) = \beta_0 + \beta_1(x+1)$. Assim, o estimador para a diferença da função *logit* é $\hat{g}(x+1) - \hat{g}(x) = \hat{\beta}_1$ e o estimador da razão de chances é $\widehat{OR} = \exp(\hat{\beta}_1)$.

Para uma interpretação útil das variáveis contínuas é necessário desenvolver um método para definir a razão de chances para uma alteração arbitrária de “ c ” unidades na covariável. O estimador do *log-odds ratio* para uma alteração de “ c ” unidades em x é $\hat{g}(x+1) - \hat{g}(x) = c\hat{\beta}_1$ e o estimador de *odds ratio* é $\widehat{OR}(x+c, x) = \exp(c\hat{\beta}_1)$ ou mais concisamente $\widehat{OR}(c)$.

O estimador do desvio padrão de $c\hat{\beta}_1$ é $\widehat{SE}(c\hat{\beta}_1) = |c|\widehat{SE}(\hat{\beta}_1)$ onde $|c|$ representa o valor absoluto de c . Os limites estimados para um intervalo de confiança $100 \times (1 - \alpha)\%$ são $\exp[c\hat{\beta}_1 \pm z_{1-\alpha/2}|c|\widehat{SE}(\hat{\beta}_1)]$.

Uma vez que o valor estimado e os seus intervalos de confiança dependem do valor de c escolhido, o valor de c deve ser sempre especificado. Por exemplo, num caso em que se pretenda avaliar a chance de desenvolver insuficiência cardíaca com a idade (anos), se o valor estimado de $\hat{\beta}_1$ para a variável dependente idade for 0,111 e quisermos saber qual a probabilidade de desenvolver insuficiência cardíaca por cada 10 anos de idade, temos $\widehat{OR}(10) = \exp(10 \times 0,111) = 3,03$, e é interpretado como se por cada aumento de 10 anos em idade, a chance de desenvolver insuficiência cardíaca aumenta 3 vezes (David W. Hosmer, Stanley Lemeshow e Rodney X. Sturdivant, 2013).

Seleção de variáveis

A seleção de variáveis a serem inseridas no modelo pode ser feita recorrendo a metodologias que permitem identificar as variáveis preditivas do modelo. Os métodos mais utilizados são: a seleção intencional de variáveis e seleção *stepwise* (David W. Hosmer, Stanley Lemeshow e Rodney X. Sturdivant, 2013).

A seleção intencional das variáveis segue os seguintes passos:

1. Análise univariada de todas as variáveis independentes. Para as variáveis categóricas recorrendo a tabelas de contingência para a variável dependente ($y =$

0,1), e para as variáveis contínuas ajustando um modelo de regressão logística univariado. Identificam-se como candidatas a um primeiro modelo multivariado, qualquer variável que apresente um valor de p inferior a 0,25, juntamente com todas as variáveis com importância clínica.

2. Ajusta-se um modelo contendo todas as variáveis e avalia-se a importância de cada covariável através do valor p e são eliminadas do modelo. É ajustado um novo modelo, contendo as variáveis significativas, que é comparado com o modelo anterior.
3. Após o ajuste do modelo mais pequeno, comparam-se os valores dos coeficientes estimados com os coeficientes estimados no modelo maior. Se a variação dos coeficientes estimados for $> 20\%$ pode significar que uma ou mais das variáveis excluídas podem ser importantes no ajuste do efeito das variáveis que permanecem no modelo e devem ser adicionadas de novo ao modelo.
4. Adicionam-se ao modelo, uma a uma, as variáveis não selecionadas do passo 1 e avalia-se a sua significância, com o objetivo de identificar as variáveis que, embora não diretamente relacionadas com a variável dependente, podem ter uma contribuição importante na presença de outras variáveis.
5. Uma vez obtido o modelo com as variáveis principais, procuram-se interações entre as variáveis no modelo.
6. Por fim, deve ser avaliada a qualidade do ajuste do modelo.

A seleção *stepwise* das variáveis é uma metodologia amplamente disponível em programas de análise estatística. É um método particularmente útil em situações em que o resultado a ser estudado é novo e se pretende analisar de maneira rápida e eficaz várias covariáveis possíveis, de modo a estabelecer associações e interações preditivas.

O procedimento *stepwise* para a seleção de variáveis é baseado num algoritmo estatístico que escolhe o modelo por comparação de alguma medida de ajustamento, normalmente o critério de informação de Akaike (AIC).

Qualidade do ajuste do modelo

As técnicas utilizadas na avaliação da qualidade do modelo tentam avaliar quão bem um dado modelo se adequa aos dados, em vez de avaliar se um modelo é ou não mais apropriado que outro modelo.

Assumindo que o resultado (Y) é binário, a qualidade do ajuste requer uma medida sumária sobre todos os sujeitos que compare o resultado observado (Y_i) para o indivíduo i com o resultado previsto ($\hat{Y}_i$) para este indivíduo obtido através do modelo ajustado, i.e., $\hat{Y}_i = \hat{P}(X_i)$.

Quando determinado coletivamente sobre todos os sujeitos, se a medida for pequena ou não significativa dizemos que o modelo está bem ajustado ou que não há evidência suficiente para concluir que é um mau ajuste.

Uma medida bastante usada para avaliar a qualidade do modelo é o desvio. Contudo, para modelos de regressão logística a utilização do desvio para avaliar a qualidade do ajuste pode ser problemática. Uma alternativa popular é a estatística de Hosmer-Lemeshow (D.G. Kleinbaum & M. Klein, 2010).

Modelo saturado e modelo totalmente parametrizado

Diz-se que existe um ajuste perfeito se $Y_i - \hat{Y}_i = 0$ para todos os indivíduos i . Apesar das características matemáticas de qualquer modelo logístico requerem que o valor previsto para qualquer indivíduo deva estar entre 0 ou 1, raramente acontece que os valores previstos se situem entre 0 ou 1 para todos os indivíduos. De facto, os valores previstos para a maioria dos indivíduos situar-se-á acima de 0 e abaixo de 1.

Desta forma, alcançar o ajuste perfeito não é um objetivo prático, mas um conceito ideal. Uma vez que se pretende usar este modelo ideal como referência, é conveniente designá-lo por modelo saturado.

Num modelo de regressão logística a unidade de análise é o grupo de indivíduos que partilham o mesmo padrão de covariável num grupo e o objetivo do modelo é prever a proporção observada de pessoas num grupo que apresenta o resultado. Assim, em

regressão logística, diz-se que um modelo é saturado se para cada covariável o modelo for capaz de prever a proporção observada.

Um modelo completamente parametrizado é o maior modelo possível que pode ser ajustado usando as variáveis que queremos autorizar no modelo. Tal modelo pode ser usado em alternativa ao modelo saturado como referência *gold standard*.

Um modelo pode ser escolhido através da razão de verossimilhança que compara dois modelos.

A abordagem tradicional para avaliar a qualidade do ajuste do modelo considera o modelo saturado como o modelo ajustado perfeito, o que faz sentido quando as unidades da análise são sujeitos individuais, uma vez que os resultados observados são 0 ou 1 em vez de algum valor no meio. Contudo, a utilização do modelo saturado como referência para avaliar o ajuste do modelo na regressão logística pode ser problemático.

A função de verossimilhança para o modelo saturado é:

$$L_{ss} = \prod_{i=1}^n P(X_i)^{y_i} [1 - P(X_i)]^{1-y_i}.$$

Uma vez que o modelo saturado se ajusta perfeitamente aos dados, a estimativa de máxima verossimilhança $\hat{Y}_i = \hat{P}(X_i)$, por definição, tem de ser igual ao valor Y_i observado para cada indivíduo i .

A fórmula para a função de verossimilhança para o modelo saturado envolve substituir Y_i por $P(X_i)$. Desta forma, sempre que Y_i for uma variável (0,1), a máxima verossimilhança do modelo saturado será sempre 1 (D.G. Kleinbaum & M. Klein, 2010).

Função desvio

Uma outra medida para avaliar a qualidade do ajuste é o desvio. A fórmula geral para o desvio é:

$$\text{Dev}(\hat{\beta}) = -2\ln(\hat{L}_c / \hat{L}_{max}).$$

Onde $\hat{\beta}$ se refere aos coeficientes de regressão no modelo a ser avaliado, $\hat{L}_c$ refere-se à máxima verossimilhança do modelo atual e $\hat{L}_{max}$ refere-se à máxima verossimilhança do modelo saturado.

Assim, o desvio avalia a verossimilhança do modelo atual com a verossimilhança do modelo que prediz de modo perfeito os resultados observados. Quanto mais próximas forem as verossimilhanças, melhor o ajuste e menor o desvio. Se $\hat{L}_c = \hat{L}_{max}$, então o desvio é 0, o seu valor mínimo. Por outro lado, se $\hat{L}_c$ for muito menor que $\hat{L}_{max}$, então a razão é uma pequena fração, de modo que o logaritmo da razão é um número negativo elevado e -2 vezes este número negativo elevado dará um número positivo elevado.

Quando é utilizada a função desvio para avaliar a qualidade do ajuste é comum comparar o desvio com valores da distribuição chi-quadrado com $n-k-1$ graus de liberdade quando o modelo atual contém $k+1$ parâmetros.

A hipótese nula é a de que o modelo é ajustado, e a hipótese alternativa é que o modelo não se ajusta. Desta forma, se o desvio é significativamente grande, o modelo é considerado pouco ajustado.

Contudo, a utilização do desvio para avaliar a qualidade do ajuste para um modelo de regressão logística nem sempre é a mais adequada.

Quando o número de covariáveis (G) é consideravelmente menor que o número de observações (n), pode assumir-se que a fórmula de desvio tem uma distribuição chi-quadrado com $n-k-1$ graus de liberdade.

Contudo, quando G é quase tão grande como n , em particular quando G é igual a n , o desvio não pode assumir ter uma distribuição chi-quadrado. Neste caso, o número de indivíduos, n_g , para cada covariável permanece pequeno à medida que o tamanho da amostra aumenta.

Para além disso, se pelo menos uma das variáveis no modelo for contínua, por exemplo, a variável idade, o valor de G tenderá a aproximar-se de n , se variação da idade na amostra for suficientemente grande. Uma vez que os modelos de regressão logística podem incluir variáveis contínuas, existem muitas situações em que a distribuição chi-quadrado não

pode ser usada quando se utiliza o desvio para avaliar a qualidade do ajuste (D.G. Kleinbaum & M. Klein, 2010).

Estatística de Hosmer-Lemeshow

Para evitar a utilização questionável do desvio para avaliar a qualidade do ajuste, a estatística de Hosmer-Lemeshow foi desenvolvida e está disponível na maioria dos programas de software estatístico.

A estatística de Hosmer-Lemeshow é amplamente usada independentemente do número de covariáveis ser próximo do número de observações. Contudo, esta estatística requer que o modelo considere pelo menos 3 covariáveis, raramente resulta em significância quando G é menor que 6, e tem melhor desempenho quando G é semelhante a n (como nos casos em que alguns dos preditores são variáveis contínuas).

Para além disso, tem a propriedade de que será sempre 0 para um modelo totalmente parametrizado. Noutras palavras, o modelo saturado para a estatística de HL é essencialmente um modelo completamente parametrizado.

A estatística de HL é realizada a partir de uma tabela com os eventos e não eventos observados e esperados, categorizados em subgrupos de percentil (D.G. Kleinbaum e M. Klein, 2010).

1. São estimados os riscos previstos $\hat{P}(X_i)$ para todos os indivíduos da amostra.
2. São ordenados os riscos previstos do maior para o mais pequeno (ou do mais pequeno para o maior).
3. Os riscos previstos são ordenados em grupos de percentil, habitualmente em décimos, $Q = 10$.
4. Constrói-se uma tabela que contém os eventos observados e esperados e os não eventos observados e esperados em cada décimo.
5. Os eventos observados e os não eventos observados em cada décimo são obtidos pela contagem do número de indivíduos em cada décimo que são eventos (i.e. $Y_i = 1$) ou não eventos (i.e. $Y_i = 0$), respetivamente.

6. Os eventos esperados em cada décimo são obtidos pela soma dos riscos previstos $\hat{P}(X_i)$ para todos os indivíduos nesse grupo. O número esperado de não eventos é obtido por subtração.
7. A estatística de HL é calculada usando uma fórmula que envolve a soma dos valores Q dos eventos $(O_q - E_q)^2/E_q$ e os valores Q dos não eventos.
8. Por último a estatística de HL é testada comparando o valor do teste HL com o ponto de percentagem de chi-quadrado com $Q - 2$ graus de liberdade. Como $Q = 10$, a estatística HL aproxima-se ao chi-quadrado com 8 graus de liberdade.

Metodologia

Tipo de estudo e participantes

Realizou-se um estudo observacional de doentes seguidos em consulta de Hemato-Oncologia na Unidade de Hematologia do Centro Hospitalar de Setúbal, com diagnóstico de neoplasia hematológica a iniciar tratamento antineoplásico.

O estudo foi aprovado pela Comissão de Ética para a Saúde (CES) do Centro Hospitalar de Setúbal e foi pedido o consentimento informado dos doentes para a recolha de dados. O recrutamento realizou-se entre janeiro de 2020 até final de fevereiro de 2021.

A maioria dos doentes seguidos em consulta de Hemato-Oncologia têm o diagnóstico de doenças linfoproliferativas, síndromes mielodisplásicas e mieloproliferativas e leucemias agudas em doentes não candidatos a quimioterapia de indução, cujo tratamento não obriga a internamento hospitalar. Todos os doentes seguidos nesta consulta realizam tratamento antineoplásico passível de ser realizado em ambulatório com exceção dos doentes que por necessitarem de internamento na fase do diagnóstico por doença sintomática, o iniciam ainda em internamento, podendo ser continuado depois em ambulatório após a alta hospitalar.

Os doentes com patologia aguda, nomeadamente leucemias agudas em doentes candidatos a quimioterapia, linfoma de Burkitt, anemia aplásica ou doentes que tenham de ser submetidos a transplante autólogo ou alogénico são referenciados aos serviços de Hematologia de hospitais centrais.

Não foram considerados elegíveis os doentes com: trombocitose essencial, policitemia vera ou mielofibrose primária (exceto se estiverem sob terapêutica com inibidores JAK2), doentes em que o intuito terapêutico seja estritamente paliativo, doentes acamados sem

potencial de recuperação ou que não sejam considerados elegíveis pelo médico assistente por outro motivo.

Variáveis

No momento de inclusão no estudo foram registados os seguintes dados:

- Idade à data de participação do estudo (em anos)
- Género (feminino, masculino)
- Tipo de residência (habitação própria, instituição de saúde ou lar de acolhimento)
- Diagnóstico
- Índice de comorbilidades de Charlson
- Nível de funcionalidade de acordo com escala PS ECOG
- Tratamento antineoplásico atual e prévio
- Internamento nas últimas 48H antes da inclusão do estudo (sim/não)
- Hábitos tabágicos (sim/não)
- Profilaxias anti-infecciosas realizadas (sim/não)
- Serologias virais (VHC, VHB e HIV 1 ou 2)
- Hemograma

O evento a ser seguido é o internamento hospitalar em doentes com neoplasia hematológica sob tratamento antineoplásico (sim ou não). Identificam-se os casos de internamento por complicação infecciosa os casos em que o internamento foi motivado por infeção, de acordo com a avaliação realizada pelo médico assistente na altura da admissão. Múltiplos episódios de internamento num doente foram considerados eventos separados. Assume-se como ponto de partida de infeção principal aquele que justifica o internamento ou gravidade dos sintomas do doente. Identificam-se os casos de infeção associadas aos cuidados de saúde de acordo com os critérios da *European Centre for Disease Prevention and Control* (European Centre for Disease Prevention and Control., 2019) início dos sintomas no 3º dia de internamento ou posteriormente, sendo o dia de admissão o primeiro dia, e os casos em que haja uma readmissão em menos de 48 horas,

com infecção no 1º ou no 2º dia. Após a alta do internamento, os doentes são reavaliados até completarem tratamento antineoplásico ou fim do estudo.

No caso de internamento, foram registados os seguintes dados:

- Dias decorridos após início de tratamento
- Local de admissão (Serviço de Urgência, Consulta não programada em Hospital de Dia de Oncologia, Consulta programada)
- Local de internamento (Serviço de Observação, internamento em enfermaria, Unidade de Cuidados intermédios, Unidade de Cuidados Intensivos)
- Motivo de internamento (infecção, internamento por causa não infecciosa)
- Transferência para nível de cuidados superior (Unidade de Cuidados Intermédios, Unidade de Cuidados Intensivos).
- Tratamento antineoplásico realizado
- Corticoterapia no último mês (sim/não)
- Infecção associada a cuidados de saúde (sim/não)
- Exames microbiológicos realizados e resultados
- Presença de dispositivo invasivo (ex. cateter venoso central, cateter urinário)
- Mucosite (sim/não)
- Doença em progressão (sim/não)
- Choque séptico (sim/não)
- Número de dias de internamento
- Alta hospitalar/óbito

Análise estatística

As variáveis qualitativas categóricas são apresentadas em contagens absolutas e percentagens. As variáveis quantitativas são caracterizadas calculando as medidas de localização e dispersão (média ou mediana, desvio padrão, intervalo interquartis, conforme adequado).

É analisado o grupo de doentes internados por causa infecciosa comparando-o com o grupo de doentes sem internamento por intercorrência infecciosa utilizando o teste de o teste de t-Student ou Mann-Whitney para as variáveis contínuas e o teste de Qui-

quadrado ou teste de Fisher para variáveis categóricas. Pretende-se usar o método da regressão logística e respetivo cálculo do odds ratio para identificar os fatores associados ao internamento de doentes por causa infecciosa associada ao tratamento antineoplásico.

É utilizado um valor de significância estatística de 0,05, e são apresentados os intervalos de confiança de 95%. Os dados recolhidos serão analisados no programa de estatística R, usando a *library* “glm”. O teste de Hosmer-Lemeshow foi usado recorrendo à *library* “generalhoslem” (Jay e Maintainer, 2022).

Resultados

Caracterização dos participantes

Foram recrutados 50 doentes num período de 14 meses, com idade média de 64 anos ($SD \pm 12,4$), dos 34 aos 87 anos. Vinte e nove (58%) eram do sexo masculino e 21 (42%) do sexo feminino. Durante o período em que decorreu o estudo, 15 doentes (30%) necessitaram de internamento durante o período de tratamento. Destes, 10 doentes (20%) foram internados por causa infecciosa.

Todos os doentes residiam em habitação própria. Os participantes tinham um Índice de Comorbilidades de Charlson (ICC) médio de 2,9 ($DP=2,4$) variável entre 0 e 10. Dezanove doentes (38%) tinham Performance Status (PS ECOG) superior ou igual a 2 – tabela 1.

Vinte e quatro (48%) doentes eram seguidos em consulta por Neoplasias Linfóides de Células B maduras, 15 (30%) tinham diagnóstico de Mieloma Múltiplo, 4 (8%) tinham Linfoma de Hodgkin, 2(4%) tinham diagnóstico de Neoplasias Linfóides de Células T maduras, 2(4%) tinham diagnóstico de Síndrome Mielodisplásico, 1(2%) com diagnóstico de Leucemia Mielomonocítica crónica (gráfico 1).

Tabela 1 - Dados socio-demográficos. ICC – índice de comorbilidade de Charlson. PS – Performance Status de acordo com o Eastern Cooperative Oncology Group (ECOG).

Variável N (%)	Grupo de doentes sem internamento por infeção N=40	Grupo de doentes com internamento por infeção N=10	Total N=50	p value	Teste estatístico
Sexo, n (%)					
F	16 (32%)	5 (10%)	21 (42%)	0,72	Teste de Fisher
M	24 (48%)	5 (10%)	29 (58%)		
Idade (média ± SD)	64 (12,6)	66 (12,5)	64 (±12,4)	0,68	T-Student
ICC (média ± SD)	2,7 (±2,2)	3,7 (±2,9)	2,9 (±2,4)	0,33	Teste de Mann-Whitney
PS ECOG					
0 - 2	26 (52%)	5 (10%)	31 (62%)	0,47	Teste de Fisher
≥ 2	14 (28%)	5 (10%)	19 (38%)		

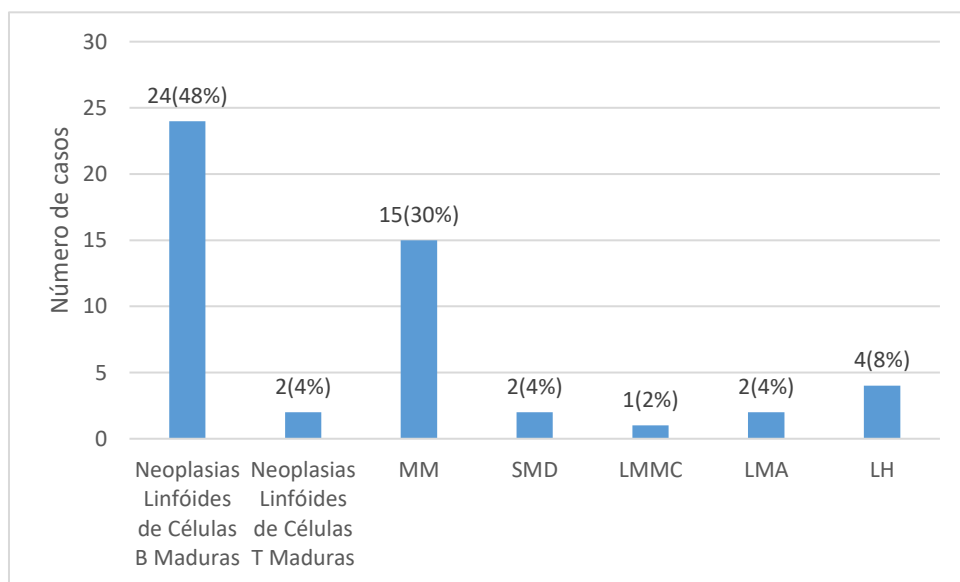


Gráfico 1 - Distribuição dos diagnósticos observados, por grupo. MM – Mieloma Múltiplo; SMD – Síndrome Mielodislásica; LMMC – Leucemia mielomonocítica crónica; LMA – Leucemia mielóide aguda; LH - Linfoma de Hodgkin.

Os tratamentos a que os doentes foram submetidos durante o período de estudo foram heterogêneos e encontram-se discriminados do Apêndice 1.

Dez doentes (20%) tinham feito tratamento antineoplásico prévio: um doente por antecedentes de neoplasia da mama, outro por neoplasia da mama e neoplasia hematológica prévia, e 8 (16%) por doença hematológica recidivada/refratária em segunda ou terceira linha de tratamento (n=7 e n=1, respetivamente).

Houve 3 doentes com serologias positivas (VHC, VIH e infeção concomitante VHC e VIH). Cinco doentes (10%) tinham hábitos tabágicos ativos.

Cinco doentes (10%) estavam internados nas 48H antes do início de tratamento, todos os restantes iniciaram tratamento em ambulatório. Vinte e três doentes (46%) receberam profilaxia anti-infecciosa de acordo com as recomendações aplicáveis a cada protocolo de tratamento.

A caracterização dos doentes no momento de inclusão no estudo encontra-se discriminada na tabela 2.

Tabela 2 - Caracterização dos doentes no momento de inclusão do estudo.

Variável N (%)	Grupo de doentes sem internamento por infecção N=40	Grupo de doentes com internamento por infecção N=10	Total n=50	valor-p	Teste estatístico
Hábitos tabágicos					
Sim	4 (8%)	1 (2%)	5 (10%)	1,0	Teste de Fisher
Não	36 (72%)	9 (18%)	45 (90%)		
Internamento 48h prévias					
Sim	2 (4%)	3 (6%)	5 (10%)	0,04	Teste de Fisher
Não	38 (76%)	7 (14%)	45 (90%)		
Profilaxia					
Sim	18 (36%)	5 (10%)	23 (46%)	0,35	Teste de Fisher
Não	17 (34%)	10 (20%)	27 (54%)		
Tratamento prévio					
Sim	4 (8%)	6 (12%)	10 (20%)	0,04	Teste de Fisher
Não	31 (62%)	9 (18%)	40 (80%)		
Serologia					
Positivas	3 (6%)	0	3 (6%)	0,79	Teste de Fisher
Negativas	30 (60%)	14 (28%)	44 (88%)		
ND ¹	2 (4%)	1 (2%)	3 (6%)		
Hemoglobina (g/dL)					
(média ± DP)	11,3 (±2,0)	9,3 (±2,1)	10,9 (±2,1)	0.007	T-Student
Neutrófilos (x10 ⁹ /L)					
(média ± DP)	4,5 (±2,8)	5,4 (±4,7)	4,7 (±3,3)	0.86	Teste de Mann- Whitney
Linfócitos (x10 ⁹ /L)					
(média ± DP)	3,1(±9,5)	3,3 (±6,5)	3,1 (±9,0)	0.76	Teste de Mann- Whitney
Plaquetas (x10 ⁹ /L)					
(média ± DP)	206 (±124,6)	175 (±91,6)	200 (±118)	0.46	T-Student

¹ ND – não determinado

Análise de fatores de risco para internamento

Calculou-se a regressão logística simples para todas as variáveis. As variáveis cujo valor p estimado era inferior ou igual a 0,25 foram introduzidas no modelo multivariado. As variáveis *Diagnóstico* e *Tratamento Antineoplásico Realizado* não foram incluídas nesta análise pela heterogeneidade dos seus valores e reduzido tamanho da amostra.

Foi então realizada uma análise *stepwise* com as variáveis *Sexo*, *Idade*, *Tratamento prévio* e *Hemoglobina*. Foi selecionado um modelo final incluindo as variáveis *Hemoglobina* ($p=0.0129$) e *Tratamento prévio* ($p=0.0779$) - tabela 3. O teste de Hosmer Lemeshow mostrou que este modelo está bem ajustado aos dados (valor $p=0,24$).

Assim, pode concluir-se que por cada aumento no valor de hemoglobina em 1g/dl o risco de internamento é 37% menor (IC 95% 0,42-0,89) e que entre os doentes submetidos a tratamento prévio, o risco de internamento por causa infecciosa é 4,84 vezes maior (IC 95% 3,35-6,98), mantendo as restantes variáveis constantes.

Tabela 3 - Regressão logística univariada e multivariada.

Variável	Univariada		Multivariada	
	Odds ratio com IC 95%	Valor p	Odds ratio com IC 95%	Valor p
Sexo (M)	0,66 [0,46;0,96]	0.5680	* ²	*
Idade	1,01 [0,70;1,46]	0.677	*	*
Internamento prévio	8.08 [5,6;11.6]	0,036	*	*
Tratamento prévio	3,77 [2,6;5,4]	0,089	4,84[3,35;6,98]	0.0779
Hemoglobina	0,63[0,43;0,91]	0,015	0,61[0,42;0,89]	0.0129

Caracterização dos internamentos durante tratamento

Houve um total de 25 internamentos durante o tratamento antineoplásico, correspondentes a 15 doentes (30%). Dezassete (68%), foram internamentos por causa infecciosa, correspondendo a 10 doentes (20%). Houve 1 doente que sofreu 5 internamentos ao longo do tratamento, outro teve 3 e outro 2. Os restantes sofreram apenas 1 internamento. Houve 5 episódios (20%) de infeções nosocomiais.

² * Variáveis não incluídas no modelo multivariado.

As causas de internamento encontram-se discriminadas na tabela 4.

Tabela 4 - Causas de internamento - todas as causas.

Causas de internamento	N (%)
Infeção	17 (68%)
Conforto sintomático	2 (8%)
Tromboembolismo pulmonar	2 (8%)
Acidente Vascular Cerebral	1 (4%)
Outro	3 (12%)

Mais de metade dos internamentos, 13 (52%), provieram de admissões do Serviço de Urgência (SU), 8 (32%) casos provieram de admissões não programadas em Hospital de Dia de Oncologia e 3 (12%) da Consulta de Hematologia. Os internamentos ocorreram em mediana 11,5 dias após último ciclo de tratamento (IQ = 13,2 - 7,00).

Caracterização dos internamentos por intercorrências infecciosas

Dos 17 internamentos, em 14 (82%) o diagnóstico hematológico era de linfoma não Hodgkin, os restantes eram relativos a SMD (6%), LMA (6%) e MM (6%). Houve 7 (28%) internamentos por neutropenia febril no momento da admissão (definida pela presença de febre e contagem de neutrófilos $< 0,5 \times 10^9/L$).

Tabela 5 - Motivo de internamento por causa infecciosa.

Motivo de internamento	N (%)
Neutropenia febril sem foco	7 (28%)
Infeção do trato urinário	4 (16%)
Pneumonia	2 (8%)
Infeção sem foco, em doente não neutropénico	1 (4%)
Bacteriémia de ponto de partida urinário	1 (4%)
Meningite	1 (4%)
Pneumonia e Infeção do trato urinário	1 (4%)
Total	17 (100%)

Quatorze doentes dos 15 que foram internados (93%) tinham recebido tratamento com corticoide no último mês. Em mais de metade dos internamentos, 13 (52%), os doentes apresentavam algum tipo de dispositivo invasivo: 9 (36%) eram portadores de catéter venoso central (CVC), 3 (12%) apresentavam nefrostomia, e um caso (4%) apresentava cateter urinário. Em 13 internamentos (76,4%) foi utilizado fator de crescimento de granulócitos por neutropenia. Apenas 1 doente (4%) apresentava mucosite na admissão. Em 7 (28%) dos internamentos os doentes encontravam-se em situação de progressão de doença.

Isolamentos microbiológicos

No total, em 11 casos de infeção associada a quimioterapia, incluindo os casos de infeção nosocomial, foram isolados os seguintes agentes microbiológicos: *Escherichia coli* (n=5), *Pseudomonas aeruginosa* (n=3), *Enterococcus faecium* (n=1), *Klebsiella Pneumoniae* (n=1), co-infeção por *Staphylococcus haemolyticus*, *Citrobacter braakii* e COVID-19 (n=1). Nos restantes internamentos não foram isolados quais agentes microbiológicos.

Evolução

Dois doentes (4%) necessitaram de ser transferidos para enfermaria com nível de cuidados superior, um doente em Unidade de Cuidados Intermédios e outro em Unidade de Cuidados Intensivos. Não se registou nenhum caso de infeção complicada por choque séptico durante o internamento. Verificou-se apenas 1 óbito, de uma doente do sexo feminino, de 44 anos, com diagnóstico de Linfoma. Apresentava um ICC de 0 e nefrostomia unilateral. Foi internada após admissão em hospital de Dia de Oncologia por infeção associada aos cuidados de saúde, sem neutropenia. Foi transferida para unidade de cuidados intensivos por pneumonia a COVID-19, e bacteriémia a *Citrobacter braakii* e *Staphylococcus Haemolyticus* e faleceu ao fim de 29 dias de internamento.

Todos os outros doentes tiveram alta hospitalar. A duração mediana dos internamentos foi de 8,5 dias (IQ = 13,25 – 7,0).

Discussão

As infecções durante o tratamento continuam a ser um fator de morbidade e mortalidade importante em doentes com neoplasias hematológicas e o conhecimento epidemiológico das infecções é uma ferramenta importante na abordagem terapêutica destas complicações. Este estudo teve como objetivo identificar, através de regressão logística, os fatores de risco associados ao internamento por infecção de doentes com doença hematológica sob tratamento.

Durante o período de 14 meses foram recrutados para o estudo 50 doentes com idades compreendidas entre 34 e 87 anos, a mediana de idade era 67 anos. O diagnóstico hematológico dos participantes era heterogêneo: a maioria, 52%, correspondia a linfoma não Hodgkin, 30% tinham Mieloma Múltiplo, e as neoplasias mielóides estavam representadas em apenas 6%.

Houve 15 (30%) doentes que sofreram internamento, 10 dos quais (20%) por motivo de infecção. A maioria dos doentes encontrava-se em ambulatório durante o início de tratamento: apenas 5 doentes (10%) estavam internados nas 48H anteriores ao seu início. Esta variável foi utilizada para permitir identificar os casos de infecções associadas aos cuidados de saúde associadas a um maior risco de infecção por bactérias multirresistentes e mortalidade, (Huoi *et al.*, 2013; Liu *et al.*, 2014; Scheich *et al.*, 2018), bem como para identificar os casos com maior gravidade de doença ao diagnóstico (Klastersky *et al.*, 2000). Contudo, neste estudo não houve isolamento microbiológico de qualquer organismo multirresistente e houve apenas 1 óbito.

Não foi possível verificar uma associação entre os diagnósticos hematológicos e o risco de internamento, para o qual pode ter contribuído o tamanho relativamente reduzido da amostra, contudo, na maioria dos internamentos (82%) o diagnóstico hematológico era de linfoma não Hodgkin. Num ensaio clínico em que se pretendia avaliar a eficácia de profilaxia antibiótica que incluiu doentes com linfoma não Hodgkin, a taxa de hospitalização neste grupo de doentes durante o tratamento foi de 20,8% (Cullen *et al.*, 2007); num outro estudo prospetivo em doentes com LNH B difuso de grandes células tratados com R-CHOP, verificaram-se 28% de hospitalizações por uma infeção ou mais (Eyre *et al.*, 2021). Desta forma, pode concluir-se que a taxa de internamentos neste estudo é semelhante aquela já reportada em estudos que incluíram diagnósticos semelhantes.

Dez doentes (20%) tinham feito tratamento prévio com quimioterapia, 8 dos (16%) por doença recidivada/refratária, considerada um fator de risco para complicações médicas nos doentes com cancro comparativamente com os doentes que estão numa fase inicial de tratamento (Talcott *et al.*, 1992; Zimmer & Freifeld, 2019). Neste estudo, os doentes submetidos a tratamento antineoplásico prévio tinham uma razão de chances de 4,8 (IC 95% 3,35-6,98) de internamento por infeção, semelhante ao resultado apresentado por Talcott *et al.*, em que os doentes com doença não controlada apresentavam uma razão de chances de complicações médicas graves no contexto de neutropenia febril no contexto de quimioterapia de 5,43 (IC 95% 1,46-20,26, $p=0,01$). Também no estudo de R. Pettengell *et al.*, os doentes submetidos a quimioterapia prévia tinham um risco de neutropenia febril de 6,39 (IC 95% 1.72-23.68, $p=0,006$).

No presente estudo foi ainda identificado na análise multivariada como fator protetor do internamento o nível de hemoglobina ao diagnóstico, com um *odds ratio* de 0,6 (IC 95% 0,42-0,89). Não foi encontrada na literatura nenhum outro estudo que demonstre esta variável como protetora.

Houve um total de 5 casos de infeções nosocomiais, correspondendo a 20% do total de internamentos. As taxas de infeções nosocomiais em doentes com neoplasias hematológicas diferem entre estudos prospetivos previamente publicados, contudo, numa análise que incluiu apenas doentes com neoplasias linfóides (Liu *et al.*, 2014), verificaram-se 11% de infeções nosocomiais, colocando a taxa observada neste estudo

acima desse valor. Num outro estudo (Engelhart *et al.*, 2002) foram registadas 27,6% de infeções nosocomiais em doentes hospitalizados com neoplasias hematológicas, contudo, neste estudo 66% dos casos eram relativos a leucemias agudas.

Também neste estudo, e à semelhança de estudos prévios a causa mais frequente de internamento foi neutropenia febril, correspondente a 28% dos internamentos prévios (Eyre *et al.*, 2021; Nesher e Rolston, 2014; Pettengell *et al.*, 2009). Não foram documentadas infeções por bactérias multirresistentes e na maioria dos casos as bactérias identificadas eram gram-negativas, o que está de acordo com a tendência reportada de crescimento na frequência de gram-negativos (Menzo *et al.*, 2015; Nesher e Rolston, 2014).

Conclusão

Foi realizado um estudo prospectivo observacional com o objetivo de identificar fatores de risco associados ao internamento por infecção de doentes com doença hematológica em regime de tratamento de ambulatório num período de 14 meses. Foram recrutados 50 doentes, com idades compreendidas entre os 34 anos e 87 anos. Vinte por cento dos doentes sofreu internamento por infecção no contexto de quimioterapia, maioritariamente doentes com diagnóstico de linfoma não Hodgkin. Em 28% dos internamentos a causa foi neutropenia febril sem foco. Foram identificados como grupo de risco para internamento por infecção os doentes que tinham realizado tratamento antineoplásico prévio, com um risco de 4,8 (IC 95% 3,35-6,98), por outro lado o valor de hemoglobina ao diagnóstico surgiu como fator protetor (OR=0,6, IC 95% 0,42-0,89).

Apêndice

Tratamentos realizados

Tratamentos realizados. ABVD (adriamicina, bleomicina, vimblastina, dacarbazina), CHOP (ciclofosfamida, doxorubicina, vincristina, prednisolona), R (Rituximab), DRD (daratumumab, lenalidomida e dexametasona), DRC (dexametasona, Rituximab, Ciclofosfamida), ESHAP (etoposido, metilprednisolona, citarabina, prednisolona), KRD (carfilzomib, lenalidomida, dexametasona), MPT (melfalano, prednisolona, talidomida), DHAP (Dexametasona, citarabina, cisplatina), VMP (bortezomib, melfalano, prednisolona), VP (bortezomib, prednisolona), VTD (bortezomib, talidomida, dexametasona).

Tratamento antineoplásico N (%)	Grupo de doentes sem internamento por infeção N=40	Grupo de doentes com internamento por infeção N=10	Total N=50
ABVD	4 (8%)	0	4 (8%)
Azacitidina	3 (6%)	1 (2%)	4 (8%)
Azacitidina + venetoclax	0	1 (2%)	1 (2%)
DRC	1 (2%)	0	1 (2%)
DRD	2 (4%)	0	2 (4%)
ESHAP	0	1 (2%)	1 (2%)
R-ESHAP	0	1 (2%)	1 (2%)
KRD	1 (2%)	0	1 (2%)
MPT	1 (2%)	0	1 (2%)
R-Bendamustina	5 (10%)	0	5 (10%)
R-CHOP	9 (18%)	2 (4%)	11 (22%)
R-CHOP/R-DHAP	1 (2%)	1 (2%)	2 (4%)
CHOP	0	1 (2%)	1 (2%)
R-CVP	1 (2%)	0	1 (2%)
R-Clorambucilo	0	1 (2%)	1 (2%)
Rituximab	1 (2%)	0	1 (2%)
Gemcitabina/cisplatina	1 (2%)	0	1 (2%)
VMP	8 (16%)	0	8 (16%)
VTD	2 (4%)	0	2 (4%)
VP	0	1 (2%)	1 (2%)

R Code

```
reglog20<-glm(dados2$Evento~dados2$Hb+dados2$`Tratamento  
antineoplásico prévio`,family = binomial(logit))
```

```
summary(reglog20)
```

Call:

```
glm(formula = dados2$Evento ~ dados2$Hb +  
dados2$`Tratamento antineoplásico prévio`,  
     family = binomial(logit))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3016	-0.6291	-0.3948	-0.2421	2.4940

Coefficients:

	Estimate	Std.
Error		
(Intercept)	3.1603	
1.9188		
dados2\$Hb	-0.4788	
0.1925		
dados2\$`Tratamento antineoplásico prévio`	1.5772	
0.8946		
	z value	Pr(> z)
(Intercept)	1.647	0.0996
.		
dados2\$Hb	-2.487	0.0129
*		
dados2\$`Tratamento antineoplásico prévio`	1.763	0.0779
.		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '		
1		

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 50.040 on 49 degrees of freedom
Residual deviance: 39.879 on 47 degrees of freedom
AIC: 45.879

Number of Fisher Scoring iterations: 5

logitgof(dados2\$Evento, fitted(reglog20))
Hosmer and Lemeshow test (binary model)
data: dados2\$Evento, fitted(reglog20)
X-squared = 10.323, df = 8, p-value = 0.2431

Bibliografia

BABADY, N. Esther - Laboratory diagnosis of infections in cancer patients: Challenges and opportunities. **Journal of Clinical Microbiology**. . ISSN 1098660X. 54:11 (2016) 2635–2646. doi: 10.1128/JCM.00604-16.

BRIOLI, Annamaria *et al.* - The risk of infections in multiple myeloma before and after the advent of novel agents: a 12-year survey. **Annals of Hematology**. . ISSN 14320584. 98:3 (2019) 713–722. doi: 10.1007/s00277-019-03621-1.

CULLEN, Michael H. *et al.* - Rational selection of patients for antibacterial prophylaxis after chemotherapy. **Journal of Clinical Oncology**. . ISSN 0732183X. 25:30 (2007) 4821–4828. doi: 10.1200/JCO.2006.08.7395.

DAVID W. HOSMER, Jr.; STANLEY LEMESHOW; RODNEY X. STURDIVANT - **Applied Logistic Regression**. 3. ed. New Jersey : John Wiley & Sons, 2013

D.G. KLEINBAUM; M. KLEIN - **Logistic Regression, Statistics for Biology and Health**. [S.l.] : Springer Science and Business Media, 2010

ENGELHART, Steffen *et al.* - Surveillance for Nosocomial Infections and Fever of Unknown Origin Among Adult Hematology–Oncology Patients. **Infection Control & Hospital Epidemiology**. . ISSN 0899-823X. 23:5 (2002) 244–248. doi: 10.1086/502043.

EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL. - **Point prevalence survey of healthcare-associated infections and antimicrobial use in European acute care hospitals : ECDC PPS validation protocol version 3.1.2**. Stockholm : European Centre for Disease Prevention and Control, 2019. ISBN 9789294982766.

EYRE, Toby A. *et al.* - Infection-related morbidity and mortality among older patients with DLBCL treated with full- or attenuated-dose R-CHOP. **Blood Advances**. . ISSN 24739537. 5:8 (2021) 2229–2236. doi: 10.1182/BLOODADVANCES.2021004286.

GUSTINETTI, Giulia; MIKULSKA, Malgorzata - Bloodstream infections in neutropenic cancer patients: A practical update. **Virulence**. . ISSN 21505608. 7:3 (2016) 280–297. doi: 10.1080/21505594.2016.1156821.

HORASAN, Elif Sahin *et al.* - Bloodstream infections and mortality-related factors in febrile neutropenic cancer patients. **Medical science monitor : international medical journal of experimental and clinical research**. . ISSN 1643-3750. 17:5 (2011) CR304–CR309. doi: 10.12659/msm.881773.

HUOI, Catherine *et al.* - Incidence of Hospital-Acquired Pneumonia, Bacteraemia and Urinary Tract Infections in Patients with Haematological Malignancies, 2004–2010: A Surveillance-Based Study. **PLoS ONE**. . ISSN 1932-6203. 8:3 (2013) e58121. doi: 10.1371/journal.pone.0058121.

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER - **Portugal Source: Globocan Incidence, Mortality and Prevalence by cancer site**, atual. 2021.

JAY, Matthew; MAINTAINER, J - Package «generalhoslem» Title Goodness of Fit Tests for Logistic Regression Models. 2022). doi: 10.1080/00949655.2016.1156682.

KLASTERSKY, Jean *et al.* - The Multinational Association for Supportive Care in Cancer Risk Index: A Multinational Scoring System for Identifying Low-Risk Febrile Neutropenic Cancer Patients. **Journal of Clinical Oncology**. . ISSN 0732-183X. 18:16 (2000) 3038–3051. doi: 10.1200/JCO.2000.18.16.3038.

KOCHANNEK, Matthias *et al.* - Management of sepsis in neutropenic cancer patients: 2018 guidelines from the Infectious Diseases Working Party (AGIHO) and Intensive Care Working Party (iCHOP) of the German Society of Hematology and Medical Oncology (DGHO). **Annals of hematology**. . ISSN 1432-0584. 98:5 (2019) 1051–1069. doi: 10.1007/s00277-019-03622-0.

LIU, Hui *et al.* - Nosocomial Infection in Adult Admissions with Hematological Malignancies Originating from Different Lineages: A Prospective Observational Study. **PLoS ONE**. . ISSN 1932-6203. 9:11 (2014) e113506. doi: 10.1371/journal.pone.0113506.

MENZO, Sara Lo *et al.* - New Insight on Epidemiology and Management of Bacterial Bloodstream Infection in Patients with Hematological Malignancies. **Mediterranean journal of hematology and infectious diseases.** . ISSN 2035-3006. 7:1 (2015) e2015044–e2015044. doi: 10.4084/MJHID.2015.044.

MICHAEL J. CAMPBELL; DAVID MACHIN; STEPHEN J. WALTERS - **Medical Statistics: A Textbook for the Health Sciences, Fourth Edition.** 4. ed. United kingdom : Wiley, 2007

NESHER, Lior; ROLSTON, Kenneth V. I. - The current spectrum of infection in cancer patients with chemotherapy related neutropenia. **Infection.** . ISSN 14390973. 42:1 (2014) 5–13. doi: 10.1007/s15010-013-0525-9.

PETTENGELL, Ruth *et al.* - Multivariate analysis of febrile neutropenia occurrence in patients with non-Hodgkin lymphoma: Data from the INC-EU Prospective Observational European Neutropenia Study. **British Journal of Haematology.** . ISSN 00071048. 144:5 (2009) 677–685. doi: 10.1111/j.1365-2141.2008.07514.x.

SCHEICH, Sebastian *et al.* - Bloodstream infections with gram-negative organisms and the impact of multidrug resistance in patients with hematological malignancies. **Annals of Hematology.** . ISSN 14320584. 97:11 (2018) 2225–2234. doi: 10.1007/s00277-018-3423-5.

STROJNIK, Ksenija *et al.* - Outcome of severe infections in afebrile neutropenic cancer patients. **Radiology and oncology.** . ISSN 1318-2099. 50:4 (2016) 442–448. doi: 10.1515/raon-2016-0011.

TALCOTT, James A. *et al.* - **Risk Assessment in Cancer Patients With Fever and Neutropenia: A Prospective, Two-Center Validation of a Prediction Rule**

TAPLITZ, Randy A. *et al.* - JOURNAL OF CLINICAL ONCOLOGY Outpatient Management of Fever and Neutropenia in Adults Treated for Malignancy: American Society of Clinical Oncology and Infectious Diseases Society of America Clinical Practice Guideline Update. **J Clin Oncol.** 36:2018) 1443–1453. doi: 10.1200/JCO.

ZIMMER, Andrea J.; FREIFELD, Alison G. - Optimal Management of Neutropenic Fever in Patients With Cancer. **Journal of Oncology Practice**. . ISSN 1554-7477. 15:1 (2019) 19–24. doi: 10.1200/JOP.18.00269.

ZORINA, Tatiana; STYCHE, Alexis - Infectious Diseases in Cancer Patients: An Overview. Em SHURIN, MICHAEL R.; THANAVALA, YASMIN; ISMAIL, NAHED (Eds.) - **Infection and Cancer: Bi-Directorial Interactions** [Em linha]. Cham : Springer International Publishing, 2015 Disponível em WWW:<URL:https://doi.org/10.1007/978-3-319-20669-1_14>. ISBN 978-3-319-20669-1. p. 295–311.

