

Agradecimentos

Ao Prof. Doutor Altamiro da Costa Pereira e Prof. Armando Teixeira - Pinto por me terem concedido a oportunidade de fazer parte do projecto de investigação, no qual o estágio foi enquadrado, sabendo que o projecto iria estar parado alguns meses, por motivo de maternidade. Acreditaram e apostaram em mim.

Ao Prof. Armando Teixeira -Pinto pela sua orientação e escolha do tema, que permitiu a exploração de técnicas estatísticas muito recentes.

À Orquídea pela sua disponibilidade e simpatia em me apresentar e explicar a base de dados e, muitas vezes me ajudar quando eu não estava a identificar alguma variável,... no meio de tantas.

À Camila pelo incentivo no uso do LateX.

Ao pessoal do CINTESIS, em particular do piso 8, que me acolheu com toda a simpatia.

À Prof.^a Doutora Ana Cristina Braga pela sua orientação e sugestões de melhoria deste relatório.

À minha família, em especial aos meus pais Isabel e Fernando, irmão Pedro, ao meu marido Nuno e à minha filha Catarina, pelo apoio incondicional, motivação, compreensão nos momentos de maior indisponibilidade minha, e por estarem sempre presentes.

Resumo

A Sepsis é uma infecção geral grave, caracterizada por uma resposta inflamatória sistémica (designada por Síndrome de Resposta Inflamatória Sistémica), geralmente causada pela presença de um agente infeccioso na corrente sanguínea. A Sepsis tem sido identificada em diversos estudos epidemiológicos como sendo a principal causa de morte nos doentes críticos, internados nas Unidades de Cuidados Intensivos. A Sepsis pode sistematizar-se através do conceito PIRO (Predisposição, Infecção, Resposta e disfunção do Órgão). Existem factores susceptíveis de serem importantes em cada um dos componentes PIRO. Neste trabalho foram estudados os factores P de Predisposição, tais como o sexo, idade, doenças crónicas, comorbilidades, e os factores R de Resposta, tais como a temperatura, frequência cardíaca, respiratória, número de leucócitos, neutrófilos, PCR.

O objectivo é encontrar os factores que mais contribuem para a mortalidade hospitalar.

Relativamente aos factores de Predisposição foi efectuada a regressão logística com os métodos *stepwise* e o método de selecção de variáveis LASSO. Este último método é particularmente importante quando estamos perante um grande número de variáveis explicativas, pois tem a vantagem de reduzir algumas das variáveis a zero, dependendo do valor do parâmetro que faz essa redução. O problema da selecção de variáveis permite decidir quais as variáveis a incluir no modelo de modo a obter um bom *tradeoff* entre viés e variância. Isto leva-nos a um conjunto de variáveis mais parcimonioso e que esteja associado com a mortalidade hospitalar.

Quanto às variáveis R, após análise estatística univariada e bivariada podemos chegar à conclusão de quais as variáveis que são significativamente diferentes entre os indivíduos que faleceram e que tiveram alta e de como variaram durante os dias de internamento na Unidade de Cuidados Intensivos (UCI). Dado que estamos perante dados longitudinais, esta primeira análise permitir-nos-á decidir qual a melhor metodologia a implementar.

Conhecendo bem o comportamento das variáveis estamos em condições de, mais tarde desenvolver um modelo que nos permita construir um *score* PIRO.

Abstract

Sepsis a serious general infection, characterized by a systemic inflammatory response (referred to as Systemic Inflammatory Response Syndrome), usually caused by the presence of an infectious agent in the bloodstream. The Sepsis has been identified in several epidemiological studies as being the leading cause of death in critically ill patients hospitalized in intensive care units. Sepsis can be systematized through the PIRO concept (predisposition, infection, response and organ dysfunction). There are factors likely to be important in each of the components PIRO. In this work we studied the factors P of predisposition, such as gender, age, chronic illness, comorbidities, and R factors of response, such as temperature, heart rate, respiratory rate, leukocyte count, CRP.

The aim is to find the factors that most contribute to hospital mortality.

As regards predisposing factors, logistic regression was performed with stepwise methods and variable selection method LASSO. The latter method is particularly important when we are dealing with a large number of explanatory variables; it has the advantage of shrink some of the variables to zero, depending on the value of the parameter that causes this shrinkage. The problem of variable selection allows you to decide which variables to include in the model to obtain a good tradeoff between bias and variance. This leads us to a more parsimonious set of variables and is association with hospital mortality.

As for the R variables, after univariate and bivariate statistical analysis we reach the conclusion which variables are significantly different between the patients who died and who went discharged and how was the variation of variables along the days of stay in Intensive Care Unit (ICU). Since we are dealing with longitudinal data, this initial analysis will allow us to decide on the best methodology to implement. Knowing well the behavior of the variables we are able to later develop a model that allows us to build a PIRO *score*.

Conteúdo

Agradecimentos	i
Resumo	iii
Abstract	v
Lista de Abreviaturas	ix
Lista de Figuras	xi
Lista de Tabelas	xiii
1 Introdução	1
1.1 Enquadramento	1
1.2 Objectivos	2
1.3 Estrutura do Relatório	3
2 Breve Descrição do Projecto de Investigação	4
2.1 Base de dados e Gestão dos dados	5
2.2 Dados	6
3 Métodos Estatísticos	9
3.1 Modelo de Regressão Logística	9
3.1.1 Interpretação dos Coeficientes	12
3.2 Inferência e Avaliação da qualidade do Modelo	13
3.2.1 Teste da Razão de Verossimilhanças	13
3.2.2 Inferência sobre os coeficientes do modelo	14
3.2.3 Curvas ROC	15
3.3 Métodos de selecção de variáveis	16
3.3.1 Métodos <i>Stepwise</i>	16
3.3.2 Akaike Information Criterion e Bayesian Information Criterion	17
3.3.3 LASSO	18
3.3.4 LASSO para a regressão logística	22
3.3.5 <i>Group</i> LASSO	22
3.3.6 <i>Group</i> LASSO para a regressão logística	23
4 Resultados	25
4.1 Análise dos factores P de Predisposição	25
4.2 Análise dos factores R de Resposta	32

5	Considerações finais	39
	Referências Bibliográficas	41
	Anexos	44
	Anexo I	45
	Anexo II	54

Lista de Abreviaturas

- AIC - Akaike information criterion
- ATB - Antibioterapia
- BIC - Bayesian information criterion
- CAS - Community Acquired Sepsis
- CCT - Corticoterapia
- CV - Cardiovascular
- CRV- Cross-validation
- DP - Desvio padrão
- EP - Erro padrão
- FC - Frequência Cardíaca
- FR - Frequência Respiratória
- IQR - Interquartile range
- LASSO - Least Absolute Shrinkage and Selection Operator
- LR - Likelihood Ratio
- MSE - Mean Square Error
- OLS - Ordinary Least Squares
- OR - Odds Ratio
- PCR - Polymerase chain reaction
- PIRO - Predisposição, Infecção, Resposta, Órgão
- QT - Quimioterapia
- RT - Radioterapia
- SACiUCI - Sepsis adquirida na Comunidade internada em Unidade de Cuidados Intensivos
- SAPS II - Simplified Acute Physiology Score
- SBIM - Serviço de Bioestatística e Informática Médica
- SIDA - Síndrome da Imunodeficiência Adquirida
- SIRS - Systemic Inflammatory Response Syndrome
- TNM - Tumor, Nódulos linfáticos, Metástase
- UCI - Unidade de Cuidados Intensivos
- VIH- Vírus da Imunodeficiência Humana

Lista de Figuras

2.1	Fluxograma dos pacientes incluídos. Fonte: Dr. António H.Carneiro (Adaptado) .	7
3.1	A função logística $F(\theta) = \frac{e^\theta}{1+e^\theta}$	11
3.2	Curva de ROC	16
3.3	Estimação do LASSO (Tibshirani, 1996)	20
3.4	Soluções $\{\hat{\beta}_\lambda\}_{\lambda \geq 0}$ para um modelo que consiste num intercepto (.) e dois factores com 3 graus de liberdade cada (—, _ _ _ _ _): λ_{max} é o valor de penalidade λ tal que nenhum grupo penalizado está no modelo. Fonte: Adaptado de Meier <i>et all</i> , 2008.	24
4.1	Coefficientes estimados versus valores lambda	31
4.2	Caixas de bigodes das distribuições da Febre	33
4.3	Caixas de bigodes das distribuições da Frequência Cardíaca	34
4.4	Caixas de bigodes das distribuições da Frequência Respiratória	35
4.5	Caixas de bigodes das distribuições da Leucocitose	36
4.6	Caixas de bigodes das distribuições dos Neutrófilos	37
4.7	Caixas de bigodes das distribuições do PCR	38

Lista de Tabelas

2.1	Variáveis do PIRO	8
4.1	Características da amostra relativamente aos factores de Predisposição por mortalidade hospitalar	26
4.2	Regressão logística univariada e multivariada para os factores de Predisposição	28
4.3	Coefficientes estimados para diferentes métodos de selecção e “ <i>shrinkage</i> ” .	30
4.4	Estatísticas descritivas da variável Febre	33
4.5	Estatísticas descritivas da variável Frequência Cardíaca (FC)	34
4.6	Estatísticas descritivas da variável Frequência Respiratória (FR)	35
4.7	Estatísticas descritivas da variável Leucocitose	36
4.8	Estatísticas descritivas da variável Neutrófilos	37
4.9	Estatísticas descritivas da variável PCR	38

Capítulo 1

Introdução

1.1 Enquadramento

O Mestrado em Estatística de Sistemas, da Universidade do Minho tem a duração de 2 anos lectivos. É constituído pela componente lectiva com duração de um ano e um trimestre e um estágio curricular com a duração de três trimestres.

Tendo concluído a componente lectiva, era importante escolher o local de estágio, numa área onde fossem aplicados os conhecimentos teórico-práticos adquiridos ao longo do Mestrado.

Depois de alguma pesquisa, surgiu a oportunidade de concorrer a uma Bolsa de Investigação, na Faculdade de Medicina do Porto no âmbito do Projecto PIC/IC/83312/2007 - “O conceito PIRO da Sepsis- implementação e validação de um novo sistema de classificação em doentes com Sepsis”, financiado pela Fundação para a Ciência e Tecnologia (FCT). Esta bolsa tinha e tem por objectivos a participação nas várias tarefas de investigação do projecto, nomeadamente a responsabilidade da gestão da base de dados e a implementação da maioria da análise estatística. Houve uma selecção de candidatos feita mediante avaliação curricular e entrevista. Após análise curricular a Comissão de Selecção seleccionou o candidato que entendeu possuir o perfil mais adequado à natureza das tarefas a desempenhar.

Assim, tendo sido admitida como Bolseira de Investigação do projecto, o estágio curricular enquadrou-se nas funções de assistente de investigação e decorreu no Centro de Investigação em Tecnologias e Sistemas de Informação em Saúde (CINTESIS, FCT, SAL 753) da Faculdade de Medicina da Universidade do Porto.

O Centro de Investigação em Tecnologias e Sistemas de Informação (CINTESIS) foi criado seguindo a visão de acolher trabalhos de investigação multidisciplinares, a fim de reunir a massa crítica necessária para contribuir na melhoria da saúde e educação. A visão de investigação do CINTESIS foca-se no desenvolvimento de métodos e ferramentas e no desenho e implementação de estudos de avaliação trabalhando com a colecção, recolha,

codificação, classificação, processamento, análise, apresentação, partilha e interpretação de dados biomédicos. A medicina é baseada cada vez mais na evidência científica que emerge da informação e conhecimento produzidos tanto na prática médica como nos estudos de investigação, a análise de decisão em saúde depende dos dados e métodos que resultam das ciências da computação e bioestatística.

1.2 Objectivos

O projecto “O conceito PIRO da Sepsis – implementação e validação de um novo sistema de classificação em doentes com Sepsis [PIC/IC/83312/2007]”, pretende desenvolver e validar o sistema de estadiamento clínico PIRO, estratificando os doentes com Sepsis, Sepsis grave ou choque séptico, através do outcome e do potencial de resposta à terapêutica. Pretende também, criar uma ferramenta de apoio à decisão clínica que permita a previsão do *outcome* e a resposta à terapêutica.

Este relatório baseia-se no trabalho desenvolvido enquanto assistente de investigação deste projecto.

Para isso, a colaboração de alguns elementos da área da Medicina foi fundamental, nomeadamente na escolha das variáveis a estudar e nas relações entre variáveis que eram necessárias realizar. O trabalho multidisciplinar nesta área é indispensável tanto na escolha como na interpretação das variáveis, após o seu tratamento estatístico.

Inicialmente, foi feito o enquadramento no projecto com a exploração da base de dados. Após esta fase, foi proposta a análise das variáveis candidatas a preditores de dois dos componentes do PIRO, os factores de Predisposição (P) e os factores de Resposta (R).

Como variável dependente tem-se o Resultado do Internamento Hospitalar cujo *outcome* é Falecido ou Alta Hospitalar. Esta variável é binária, por isso foi realizada a análise de regressão logística, dos factores de Predisposição, sendo aqui proposto um método de selecção de variáveis denominado por LASSO. Habitualmente, os métodos de selecção de variáveis mais utilizados são os métodos *stepwise*. Apesar de o método LASSO ter sido sugerido em 1995, ainda não existem muitos trabalhos que utilizem este método, uma das dificuldades encontradas ao longo do estágio e elaboração do relatório. Mesmo o *package* utilizado para a análise dos dados está ainda em fase experimental.

O LASSO é uma técnica de selecção de variáveis que tem sido amplamente utilizada na área da genética, dado que é muito vantajoso quando existe um grande número de preditores ou o número de preditores é bastante maior do que o número de observações, tendo a característica desejável de incluir no modelo apenas um conjunto de variáveis explicativas, igualando a zero todas as variáveis que não têm efeitos.

Outras tarefas desenvolvidas durante os meses de estágio tiveram como resultado a

publicação de dois *Abstracts*¹ no *American Journal of Respiratory and Critical Care Medicine* (AJRCCM), que foram apresentados sob a forma de posters na *American Thoracic Society (ATS) International Conference 2011 (Denver-Colorado)*, e a realização do artigo “*Bloodstream infection as a marker of community-acquired sepsis severity. Results from the Portuguese Community-Acquired Sepsis Study (SACiUCI study)*”, a aguardar publicação (Anexo I). Também surgiu a oportunidade de frequentar o curso de Métodos Modernos de Classificação e Regressão, organizado pelo Departamento de Ciências da Informação e Decisão em Saúde, Faculdade de Medicina do Porto, no âmbito da Escola de Verão 2011 - Investigação Clínica e em Cuidados de Saúde: Métodos e Tecnologias de Informação, cujo docente foi o Prof. Doutor Jarozlav Harezlak, da Indiana University School of Medicine.

1.3 Estrutura do Relatório

Este relatório é constituído por cinco capítulos. No primeiro capítulo será referido o enquadramento, os objectivos e a estrutura do relatório.

No segundo capítulo será feita uma breve descrição do projecto de investigação.

O método de selecção LASSO, bem como os modelos teóricos de regressão logística e os métodos de selecção de variáveis *stepwise* serão apresentados no terceiro capítulo.

No quarto capítulo constarão os resultados das análises efectuadas para os factores de Predisposição (P), que incluem o Qui-Quadrado, regressão logística univariada e multivariada, com métodos de selecção *stepwise* e LASSO. Também serão apresentados os resultados das análises dos factores de Resposta (R). Estas variáveis representam medições durante seis dias, demonstrando a evolução dos indivíduos ao longo do seu internamento na UCI sendo, por isso, dados longitudinais. Foram realizadas estatísticas descritivas, testes não paramétricos de Mann-Whitney, para comparar os resultados dos Falecidos e dos que tiveram Alta e testes não paramétricos de Wilcoxon, para determinar se a variação das variáveis foi significativa, dentro de cada um dos grupos. A melhor abordagem e metodologia para o tratamento destes dados dependerá destes resultados e não será apresentada neste relatório.

Algumas considerações finais serão realizadas no quinto capítulo.

Importa ainda referir que todas as análises estatísticas foram efectuadas com recurso aos softwares estatísticos PASW Statistics versão 18 e R 2.13.0.

¹Joao Goncalves-Pereira, Pedro R. Pova, Cristina Lobo, and Antonio H. Carneiro: **Time Of Antibiotic Therapy And Sepsis Resolution. Results From The Portuguese Community-Acquired Sepsis Study (SACiUCI Study)**, *Am. J. Respir. Crit. Care Med.* 183: A3853.

Pedro R. Pova, Joao Goncalves-Pereira, Cristina Lobo, and Antonio H. Carneiro: **Positive Blood Cultures As A Marker Of Sepsis Severity. Results From The Portuguese Community-Acquired Sepsis Study (SACiUCI Study)**, *Am. J. Respir. Crit. Care Med.* 183: A4669

Capítulo 2

Breve Descrição do Projecto de Investigação

A Sepsis é uma condição potencialmente fatal caracterizada por uma infecção generalizada do organismo, denominada por SIRS (*systemic inflammatory response syndrome*) sendo uma das causas mais frequentes de mortalidade nas Unidades de Cuidados Intensivos.

A definição de Sepsis, Sepsis grave e choque séptico da última conferência de consenso não permite a caracterização precisa e o estadiamento dos doentes com sepsis. O conceito PIRO foi proposto em 2001 pela conferência de consenso internacional reunindo simultaneamente a SCCM/ESICM/ACCP/ATS/SIS.

É um sistema de classificação originalmente sugerido por John Marshall. Sendo a Sepsis uma condição heterogénea é possível descrevê-la com base em quatro características da mesma forma que o cancro pode ser descrito com base no sistema TNM de classificação dos tumores malignos – tumor, nódulos linfáticos e metástase - desenvolvido por Pierre Denoix em 1946. A classificação PIRO permitirá o estudo de populações mais homogéneas em termos de prognóstico e resposta à terapêutica.

PIRO significa (P) Predisposição, (I) Infecção, (R) resposta do hospedeiro e grau de disfunção de (O) órgão concomitante. O conceito PIRO surge como uma proposta de investigação e conceito a desenvolver e aperfeiçoar. A elaboração do sistema de classificação PIRO requer a avaliação da sepsis de forma a definir variáveis que podem prever o *outcome*. Até à presente data não existe nenhum modelo PIRO implementado e/ou validado.

Apesar da definição de Sepsis, Sepsis grave ou choque séptico, estes termos não permitem a caracterização precisa e o estadiamento destes doentes nem a previsão do seu *outcome*. Um sistema bem caracterizado que permita o estadiamento e a predição do *outcome* beneficiaria largamente o clínico com um sistema de apoio à decisão e também o doente com estabelecimento dos riscos e decisões baseadas menos nas crenças pessoais

e mais na evidência científica.

Juntamente com o desenvolvimento da *Surviving Sepsis Campaign*, foi criada em 2004 uma rede de unidades de cuidados intensivos portuguesa. Este grupo de estudo designou-se por Sepsis Adquirida na Comunidade internada em Unidade de Cuidados Intensivos (SACiUCI) e incluiu um grande número de unidades de Norte a Sul de Portugal, representando 41% de todas as Unidades de cuidados intensivos.

O objectivo deste grupo de estudo era avaliar a epidemiologia da Sepsis Adquirida na Comunidade (CAS- *Community-acquired sepsis*) em pacientes admitidos nas unidades de cuidados intensivos portuguesas, avaliar a conformidade com as directrizes da *Surviving Sepsis Campaign* e melhorar a conformidade dessas directrizes.

O SACiUCI é um estudo de corte prospectivo, multi-cêntrico realizado durante um ano (1 de Dezembro de 2004 a 30 de Novembro de 2005), em 17 UCIs portuguesas. Foram incluídos no estudo todos os pacientes adultos (idade ≥ 18) com CAS, consecutivamente admitidos nas unidades de cuidados intensivos participantes. Os pacientes foram acompanhados até à sua alta do Hospital ou até ao seu falecimento.

2.1 Base de dados e Gestão dos dados

Os dados foram recolhidos prospectivamente usando um formulário num *software* de base de dados, criado pelo Serviço de Biostatística e Informática Médica – SBIM da Faculdade de Medicina da Universidade do Porto.

Foram organizadas sessões de esclarecimento no Norte, Centro e Sul do País sobre a recolha dos dados, onde participaram os investigadores responsáveis. Em cada UCI existia um médico responsável pela colheita dos dados. Periodicamente cada UCI recebia um relatório com erros e inconsistências na base de dados, sendo feita uma revisão desses dados. O *software* de base de dados foi desenhado de forma a rejeitar e identificar inconsistências.

Cada formulário incluía 237 itens organizados em 2 secções. A secção 1 para todos os doentes admitidos na UCI, dados sobre a idade, sexo, categoria da admissão, SAPS II, data da admissão e alta do Hospital e UCI. E a secção 2 para pacientes com sepsis

adquirida na comunidade incluía dados que caracterizavam o problema da sépsis, tais como a severidade da sépsis, foco de infecção, hemoculturas colhidas, positivas, agente responsável.

Dados sobre os factores predisponentes tais como a presença de comorbilidades e a evolução dos critérios da sépsis, como a disfunção do órgão e tratamento nos primeiros 5 dias de estadia na UCI, tratamento antimicrobiano, mudanças no tratamento, motivos para isso e duração do tratamento.

Por fim, esta secção também compreendia o registo de procedimentos recomendados pela *Surviving Sepsis Campaign*, para o tratamento da sépsis, de modo a verificar a sua conformidade (Anexo II).

2.2 Dados

Foram admitidos nas unidades de cuidados intensivos 4202 doentes, tendo sido excluídos 60 doentes, 53 por terem idade inferior a 18 anos e 7 porque a informação do formulário estava incompleta.

Assim, dos 4142 doentes, 3245 (78%) apresentavam sepsis não adquirida na comunidade e 897 (22%), apresentavam sepsis adquirida na comunidade. Este último subgrupo foi o objecto de estudo deste trabalho. Dos 897 doentes com sepsis, 265 (30%) faleceram na UCI, apenas 1 estava na UCI e 631 (70%) foram dispensados da UCI. Dos 631 doentes que saíram da UCI, 72 (11%) faleceram no internamento do Hospital, 5 (1%) mantinham-se no Hospital e 554 (88%) tiveram alta. Dos 897 doentes inicialmente internados na UCI, 554 (62%) tiveram alta (Figura 2.1).

Importa referir que os 897 doentes com sepsis, estavam estratificados segundo a gravidade da sepsis. Assim, 83 (9%) tinham sepsis, 362 (40%) tinham sépsis grave e 452 (50%) choque séptico.

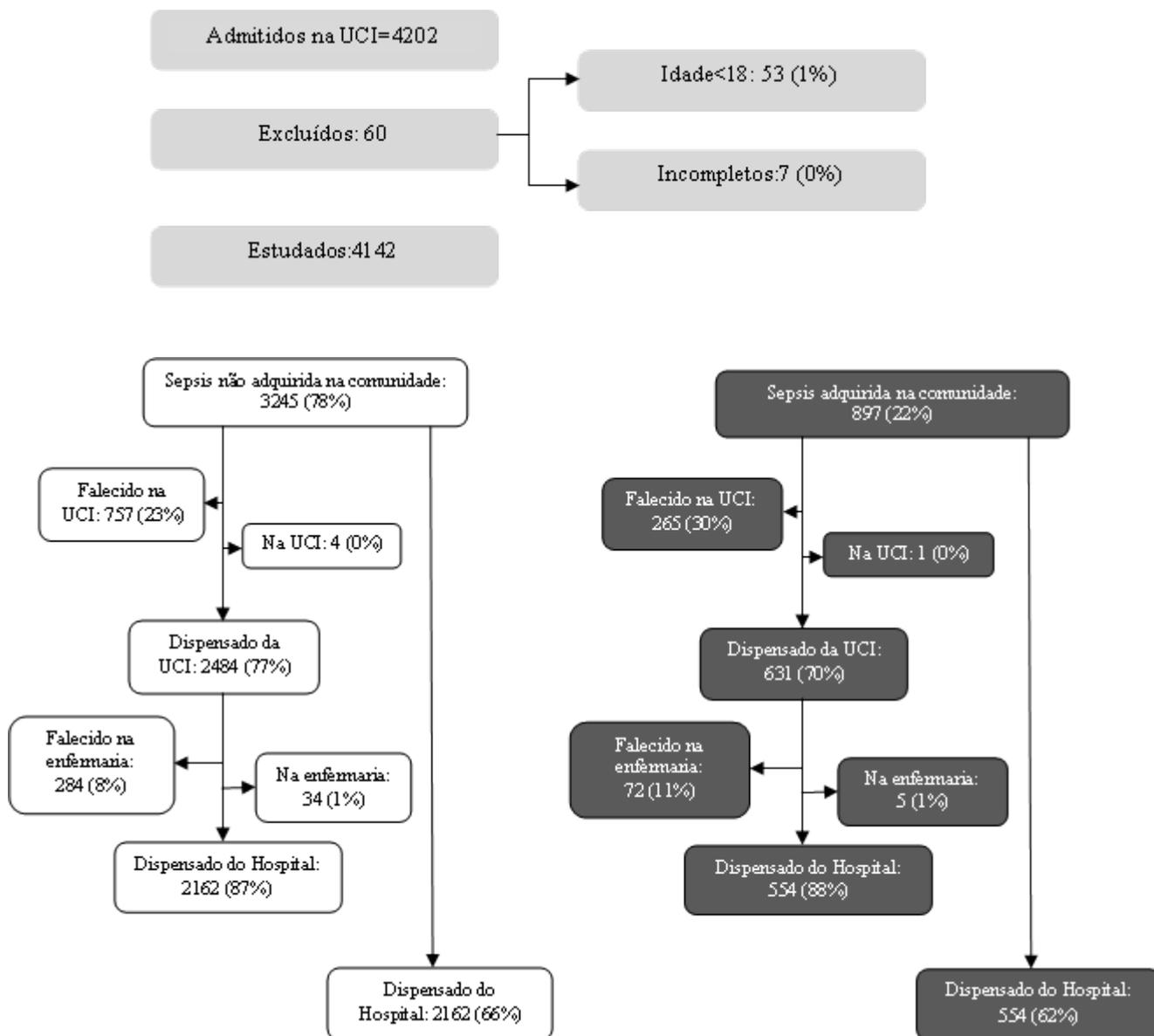


Figura 2.1: Fluxograma dos pacientes incluídos. Fonte: Dr. António H.Carneiro (Adaptado)

As variáveis existentes no formulário (Anexo II) foram exportadas para o programa estatístico SPSS e a partir daí, seleccionadas de acordo com o preconizado pelo conceito PIRO (Tabela 2.1). As variáveis da tabela 2.1 são as candidatas a preditores de cada um

dos componentes do PIRO e são as que, provavelmente estão mais associadas à mortalidade hospitalar por sépsis grave ou choque séptico.

Tabela 2.1: Variáveis do PIRO

P	I	R	O
Predisposição	Infecção	Resposta	Disfunção do Órgão
Idade	Foco Infecção	°C	Neurológica
Sexo	Agente	FC	Respiratória
CCT Curta	Local de identificação do agente	FR	CV
CCT Longa	ATB apropriada	Leucócitos	Renal
QT	ATB adequada	% neutrófilos	Hepática
RT	Tempo até administração da 1ª toma do ATB	PCR	Hematológica
SIDA	ATB descalção		Metabólica
VIH	ATB prévia		Gastro-intestinal
Neoplasia			Sépsis Grave
Doença Hepática			Choque séptico
Doença Renal			Tipo de suporte da disfunção
Doença Cardiovascular			Lactato
Doença Respiratória			
Doença Hematológica			
Nº de Comorbilidades			

Pretende-se efectuar a modelação dos dados usando técnicas estatísticas apropriadas com vista a encontrar as variáveis clínicas relevantes e com significância estatística de modo a criar um *score* PIRO.

Capítulo 3

Métodos Estatísticos

3.1 Modelo de Regressão Logística

Modelos de regressão em que a variável dependente pede uma resposta de “sucesso” ou “insucesso”, como “Falecido” ou “Alta”, são conhecidos como modelos de regressão com variável dependente dicotômica ou modelos de escolha binária. Dentre os modelos de regressão existentes para modelizar tais situações, um dos mais simples de usar será o Modelo de Regressão Linear. Considerando a equação de regressão linear:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + u_i \quad (3.1.1)$$

Admita-se que Y_i designa uma variável binária codificada com o valor 1 ou o valor 0, onde o valor 1 representa a presença de um atributo e o valor 0 representa a ausência desse atributo. Para uma sequência ordenada X_{1i}, \dots, X_{ki} , num modelo estatístico com as hipóteses clássicas, deveria ser possível obter valores diferentes de Y . No entanto, para esta especificação são, apenas, dois os valores possíveis. Seja p_i a probabilidade de se verificar um acontecimento, $p_i = Prob(Y_i = 1 | X_{1i}, \dots, X_{ki})$, e $1 - p_i$ a probabilidade de não se verificar o acontecimento. O valor médio de Y_i , terá que ser igual a p_i :

$$E(Y_i) = 0(1 - p_i) + 1p_i = p_i \quad (3.1.2)$$

Se $E(u_i) = 0$, teremos na equação (1)

$$E(Y_i) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} = p_i \quad (3.1.3)$$

No entanto, este Modelo de Regressão Linear apresenta várias limitações, tais como a não normalidade do termo de erro, a sua heterocedasticidade e a possibilidade de a probabilidade estimada ficar fora do limite 0-1. Uma vez que $E(Y)$ é uma medida de probabilidade, teremos que ter $0 \leq E(Y) \leq 1$. Se $E(Y)$ é uma função linear de X_j , é inevitável que para alguns valores de X_j , $E(Y)$ se situe fora do intervalo $[0,1]$. A estimação pelo método ordinário dos mínimos quadrados (OLS) dos coeficientes em (3) não irá obedecer a esta restrição, sendo possível que possa gerar valores estimados de Y inferiores a 0 ou superiores a 1. A hipótese da normalidade do termo de erro não é possível visto que apenas poderá assumir dois valores, $1 - (\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})$, com probabilidade p_i ou $0 - (\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})$, com probabilidade $1 - p_i$, pelo que terá distribuição binomial. Quanto à hipótese de homocedasticidade do termo de erro, tem-se que $E(u_i) = 0$ e $var(u_i) = E(u_i^2)$. u_i^2 poderá assumir apenas dois valores:

$$[1 - (\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})]^2, \text{ com probabilidade } p_i \text{ e}$$

$$[-(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})]^2, \text{ com probabilidade } 1 - p_i, \text{ pelo que}$$

$$E(u_i^2) = [1 - (\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})]^2 p_i + (\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})^2 (1 - p_i) \quad (3.1.4)$$

Dado que $\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} = p_i$, obtém-se $var(u_i) = E(u_i^2) = p_i(1 - p_i)$, ou seja a variância do termo de erro não é constante, verificando-se heterocedasticidade. Assim, é necessário um modelo que obvie a estas dificuldades, garantindo nomeadamente que as probabilidades estimadas nunca se encontrem fora do intervalo 0-1. Por exemplo, considerando o modelo sobre o resultado do internamento hospitalar, “Falecido” ($Y = 1$) ou “Alta” ($Y = 0$), existem uma série de factores associados a um vector que explicam o *outcome*, tal que:

$$Prob(Y = 1) = F(X, \beta)$$

$$Prob(Y = 0) = 1 - F(X, \beta) \quad (3.1.5)$$

O conjunto de parâmetros reflecte o impacto das mudanças de na probabilidade de falecer. É necessário escolher uma função F tal que a probabilidade produza resultados no intervalo $[0,1]$. Uma dessas funções é a função de distribuição logística. Esta função é também popular dada a sua forma sigmóide (Figura 3.1). A forma sigmóide é particularmente apelativa para os epidemiologistas se a variável θ é vista como representando

um índice que combina as contribuições de vários factores de risco, de modo que $F(\theta)$ representa o risco para um dado valor de θ . Assim, o risco é mínimo para valores baixos de θ , sobe numa faixa de valores intermédios de θ , e mantém-se próximo de 1 quando θ é suficientemente grande. Os epidemiologistas acreditam que esta curva sigmóide se aplica a uma variedade de condições de doença (Kleinbaum, 1998). Se for uma variável aleatória com função distribuição logística, então:

$$F(\theta) = \frac{e^\theta}{1 + e^\theta} \quad (3.1.6)$$

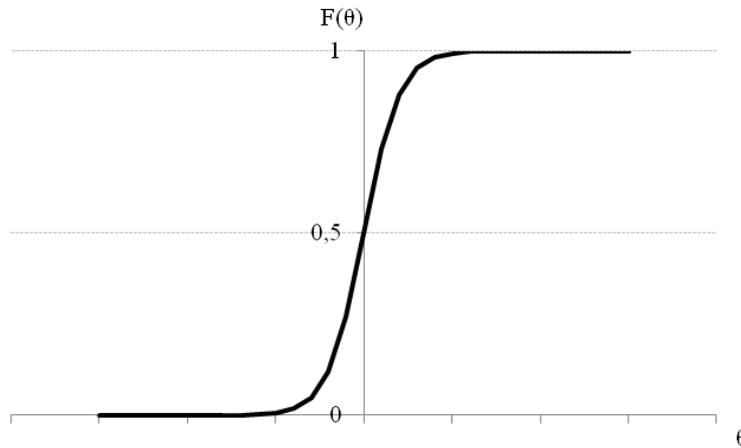


Figura 3.1: A função logística $F(\theta) = \frac{e^\theta}{1+e^\theta}$

Donde:

$$Prob(Y_i) = F(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}) = \frac{e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}} \quad (3.1.7)$$

Esta especificação corresponde ao usualmente designado modelo *logit*. A estimação dos modelos de escolha binária é usualmente baseada no Método de Máxima Verosimilhança cuja função é dada por:

$$\begin{aligned} L(\beta|Y_i, X_{ki}) &= \prod_{i=1}^n f(Y_i) = \prod_{i=1}^n Prob(Y_i = 1)^{Y_i} [1 - Prob(Y_i = 1)]^{1-Y_i} = \\ &= \prod_{i=1}^n F(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})^{Y_i} [1 - F(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki})]^{1-Y_i} = \\ &= \prod_{i=1}^n \left[\frac{e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}} \right]^{Y_i} \left[1 - \frac{e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}} \right]^{1-Y_i} \end{aligned} \quad (3.1.8)$$

De forma a encontrar o máximo desta função, maximiza-se o Logaritmo de $L(\beta|Y_i, X_{ki})$, ou seja

$$\begin{aligned}
 \log L(\beta|Y_i, X_{ki}) &= \sum_{i=1}^n Y_i \log \left[\frac{e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}} \right] - (1 - Y_i) \log(1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}) = \\
 &= \sum_{i=1}^n Y_i (\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}) - Y_i \log(1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}) - (1 - Y_i) \log(1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}) = \\
 &= \sum_{i=1}^n Y_i (\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}) - \log(1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}) \quad (3.1.9)
 \end{aligned}$$

3.1.1 Interpretação dos Coeficientes

Os coeficientes de regressão β do modelo logístico têm um papel importante fornecendo informação acerca da relação dos preditores do modelo e a variável dependente. A quantificação desta relação é dada pelo *odds ratio*. O *odds ratio* (OR) é uma medida bastante usada que compara dois ou mais grupos na previsão do *outcome* (variável dependente) (Kleinbaum, 1998).

Odds é o rácio da probabilidade que algum evento (por exemplo, a mortalidade hospitalar) irá ocorrer dividido pela probabilidade que o mesmo evento não ocorra.

Sendo o evento D, o *odds* é dado por:

$$Odds(D) = \frac{Prob(D)}{1 - Prob(D)}$$

Por definição, qualquer *odds ratio* (OR) é um rácio de dois *odds*, ou seja

$$OR_{A vs B} = \frac{odds(D_A)}{odds(D_B)} = \frac{Prob(D_A)}{1 - Prob(D_A)} / \frac{Prob(D_B)}{1 - Prob(D_B)}$$

Onde A e B são dois grupos de indivíduos que estão a ser comparados.

Por exemplo, supondo A=M os elementos do sexo masculino e B=F os elementos do sexo feminino, então D_F é o grupo de indivíduos do sexo feminino que faleceram e D_M é o grupo de indivíduos do sexo masculino que faleceram. Supondo, $Pr(D_F) = 0.1$ e $Pr(D_M) = 0.3$, tem-se $\frac{0.3}{1-0.3} / \frac{0.1}{1-0.1} = 2.57$, ou seja os indivíduos do sexo masculino têm um risco cerca de 3 vezes maior de falecer do que os indivíduos do sexo feminino.

Um OR de um, significaria que o risco nos dois grupos seria igual, ou seja não haveria o efeito em relação ao sexo, de falecer.

Para obter a expressão do OR de um modelo de regressão logística, tem-se que comparar o *odds* de dois grupos de indivíduos. Os grupos são sexo masculino ($X_1 = 1$) e sexo feminino ($X_1 = 0$). Para este exemplo, o log odds de sexo feminino e sexo masculino pode ser escrito como:

$$\log_e \text{odds}(\text{sexo masculino}) = \beta_0 + \beta_1 \cdot 1 = \beta_0 + \beta_1 \text{ e}$$

$$\log_e \text{odds}(\text{sexo feminino}) = \beta_0 + \beta_1 \cdot 0 = \beta_0, \text{ respectivamente.}$$

Segue-se que

$$OR_{M \text{ vs } F} = \frac{\text{odds}(\text{sexo masculino})}{\text{odds}(\text{sexo feminino})} = \frac{e^{(\beta_0 + \beta_1)}}{e^{\beta_0}} = e^{\beta_1}$$

Assim, o OR que compara as duas categorias do preditor é obtido pela exponencial do coeficiente do preditor no modelo de regressão logística (Kleinbaum, 1998).

3.2 Inferência e Avaliação da qualidade do Modelo

3.2.1 Teste da Razão de Verossimilhanças

A análise estatística do modelo de regressão logística é, geralmente, conduzida em termos análogos aos que se conhecem para o modelo de regressão linear. Para testar a hipótese nula de que os coeficientes são simultaneamente iguais a zero, o equivalente ao teste F no modelo de regressão linear é a estatística da razão de verossimilhanças (*Likelihood Ratio*-LR). Tem-se que,

$$H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0 \text{ vs } H_1 : \exists_{(j=0, \dots, k)} : \beta_j \neq 0 \text{ (sendo } k \text{ o número de variáveis)}$$

Dada a hipótese nula, a estatística LR segue uma distribuição Qui-Quadrado com número de graus de liberdade igual ao número de variáveis independentes. A estatística de teste da razão de verossimilhanças é calculada como

$$LR = -2[\ln L(\text{modelo incompleto}) - \ln L(\text{modelo completo})]$$

Sob a hipótese H_0 , tem-se, assintoticamente,

$$LR = -2[\ln L(\text{modelo incompleto}) - \ln L(\text{modelo completo})] \sim \chi^2_{(K)}$$

Quando se rejeita a hipótese nula conclui-se que pelo menos um dos coeficientes é diferente de zero, ou seja o modelo é estatisticamente significativo.

3.2.2 Inferência sobre os coeficientes do modelo

Para testar a significância dos coeficientes do modelo, ou seja, quais as variáveis independentes que influenciam significativamente o modelo, tem-se o Teste de Wald.

Neste teste, testa-se se um determinado coeficiente é nulo, condicionado pelos valores estimados dos outros coeficientes.

$$H_0 : \beta_j = 0 | \beta_0, \beta_1, \beta_{j-1}, \beta_{j+1}; \beta_k$$

vs

$$H_1 : \beta_j \neq 0 | \beta_0, \beta_1, \beta_{j-1}, \beta_{j+1}; \beta_k (j = 1, \dots, k)$$

A estatística de teste é:

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

Este valor, quando elevado ao quadrado leva-nos à estatística de Wald com distribuição Qui- Quadrado com 1 grau de liberdade.

A estatística de Hosmer and Lemeshow (2000) avalia a qualidade do ajustamento do modelo criando 10 grupos ordenados de sujeitos (decis) e comparando o número realmente existente em cada grupo (observado) com o número esperado pelo modelo de regressão logística (predicted). A estatística de teste é uma estatística do Qui- Quadrado cujo resultado desejável é a não significância, indicando que a previsão do modelo não difere significativamente do observado. Os 10 grupos ordenados são criados baseados na sua estimativa de probabilidade. Os que têm uma estimativa de probabilidade abaixo de 0.1 formam um grupo, e assim sucessivamente, até aqueles com probabilidade 0.9 a 1.0. Cada uma destas categorias é dividida em dois grupos baseados no resultado sucesso, insucesso.

As frequências esperadas para cada uma das células são obtidas do modelo. Se o modelo é bom, então a maioria dos sujeitos com sucesso estão classificados nos decis maiores de risco e os sujeitos com insucesso estão classificados nos decis com menor risco.

A estatística de Hosmer-Lemeshow é dada por:

$$X_{HL} = \sum_{i=1}^g \frac{(O_i - n'_i \hat{\pi}_i)^2}{n'_i \hat{\pi}_i (1 - \hat{\pi}_i)}$$

Onde O_i são os eventos observados, n'_i é o número total de observações no i -ésimo grupo e $\hat{\pi}_i$ é a probabilidade estimada (Hosmer and Lemeshow, 2000).

A estatística segue assintoticamente uma distribuição de χ^2 com $g - 2$ graus de liberdade.

3.2.3 Curvas ROC

Uma forma de determinar a precisão de um modelo é dada pela área abaixo da curva ROC (Receiver Operating Characteristic). Esta curva originária da teoria de detecção do sinal, demonstra a relação entre sinal e ruído. O gráfico mostra a probabilidade de detectar o sinal (sensibilidade - verdadeiros positivos) e o falso sinal (ruído) ((1-especificidade)-falsos positivos) para uma série de possíveis valores de corte. A área abaixo da curva de ROC varia de 0 a 1 fornecendo uma medida para a capacidade de discriminar entre sujeitos com a característica de interesse e sujeitos sem a característica de interesse.

Na Figura 3.2 está representada uma curva de ROC, cujos eixos são a sensibilidade versus 1-Especificidade para todos os valores possíveis de corte.

A área abaixo da curva fornece uma medida de discriminação que quanto mais próxima for de 1, maior é a capacidade do modelo para discriminar os indivíduos falecidos relativamente aos indivíduos que têm alta. A área abaixo da curva de ROC da Figura 3.2 é de 0.675.

Curvas de ROC com áreas de 0.5 não têm poder discriminativo, entre 0.7 e 0.8 têm uma discriminação aceitável, entre 0.8 e 0.9 têm uma discriminação boa e com área maior ou igual a 0.9 têm um poder de discriminação excelente (Hosmer and Lemeshow, 2000).

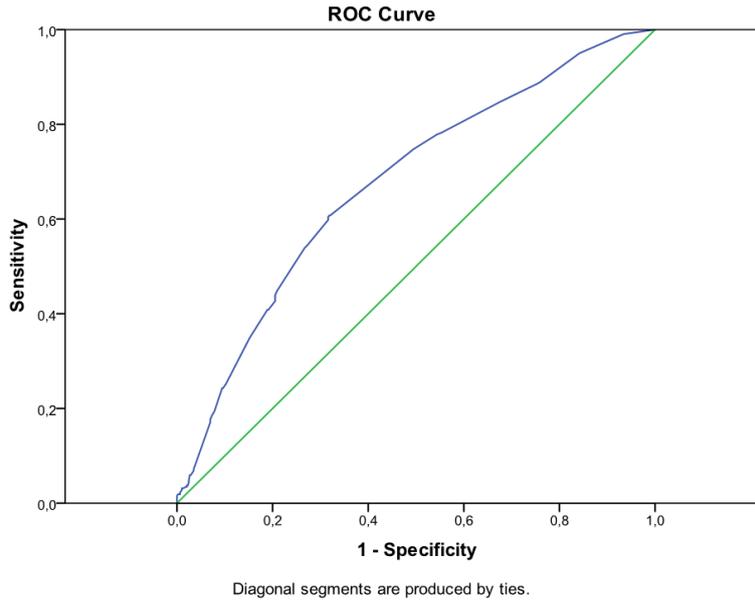


Figura 3.2: Curva de ROC

3.3 Métodos de selecção de variáveis

3.3.1 Métodos *Stepwise*

Perante um modelo de regressão com um elevado número de preditores, o objectivo será reter apenas algumas dessas variáveis e eliminar as restantes do modelo. Existindo um grande número de preditores pode-se ter problemas entre estas variáveis, tal como a multicolinearidade. Também é desejável seleccionar um conjunto menor de variáveis que ajuste tão bem quanto o modelo com todas as variáveis e que contenha as variáveis predictoras mais importantes (Kyung et al., 2010). Um modelo com muitas variáveis explicativas será mais difícil de interpretar já que os efeitos de algumas das variáveis serão pequenos, por isso será melhor a determinação de um modelo mais pequeno com efeitos mais fortes e que tenha um erro preditivo menor do que o modelo com todas as variáveis.

Algumas das estratégias usadas para seleccionar variáveis de modo a chegar ao “melhor modelo” são a selecção *forward stepwise* e a selecção *backward stepwise*. A selecção *forward stepwise* começa com a constante β_0 e sequencialmente adiciona ao modelo o preditor X_i mais correlacionado com Y que melhora o ajustamento.

A melhoria do ajustamento é baseada na estatística $F = \frac{RSS(\hat{\beta}) - RSS(\tilde{\beta})}{RSS(\tilde{\beta}) / (N - k - 2)}$, onde $\hat{\beta}$ é o parâmetro das estimativas e $\tilde{\beta}$ são as estimativas que resultam quando é adicionado um preditor, para um modelo com k variáveis e o $RSS(\beta) = \sum_{i=1}^N (y_i - x_i^T \beta)$.

Quando não existirem variáveis explicativas que produzam uma estatística F maior do

que o percentil 90 ou 95 da distribuição $F_{1,N-k-2}$, o processo de introdução de variáveis pára.

A selecção *backward stepwise* começa pelo modelo com todas as variáveis independentes e sequencialmente vai excluindo variáveis usando a estatística F para escolher os preditores a eliminar. O preditor que tem uma menor estatística F é o eliminado e o processo pára quando cada preditor eliminado no modelo tem um valor F maior que percentil 90 ou 95 da distribuição $F_{1,N-k-2}$ (Hastie et al, 2009).

Na regressão logística a significância é assegurada através do Teste da Razão de Verosimilhança. Assim, em cada passo do procedimento a variável mais importante, em termos estatísticos, é aquela que produz a maior mudança no logaritmo da verosimilhança em relação ao modelo que não contém a variável (Hastie et al, 2009).

Este tipo de processos é discreto, pois as variáveis ou entram ou saem do modelo e, por isso podem exibir uma variância elevada não reduzindo o erro preditivo.

Os métodos “*shrinkage*” são contínuos e não têm uma variabilidade muito alta (Hastie et al, 2009). As estimativas dos mínimos quadrados normalmente têm pouco enviesamento mas uma variância grande. O erro preditivo pode ser melhorado reduzindo o valor dos coeficientes ou levando estes coeficientes a serem zero. Assim, sacrifica-se um pouco do enviesamento e reduz-se a variância dos valores estimados, o que pode melhorar o erro preditivo (Tibshirani, 1996). Surgem então, os métodos de penalidade de forma a descobrir as variáveis explanatórias mais relevantes com um erro preditivo menor.

3.3.2 Akaike Information Criterion e Bayesian Information Criterion

Uma das formas de seleccionar um modelo estatístico é através do *Akaike information criterion* (AIC). Dado um conjunto de dados e perante vários candidatos a modelos, estes podem ser ordenados de acordo com os valores AIC.

Para seleccionar um modelo através do AIC, escolhe-se simplesmente o modelo que dá o menor valor de AIC dentre um conjunto de modelos considerados.

Se os modelos considerados ajustam os dados igualmente bem então, de acordo com o critério AIC o melhor modelo é o que tem menor número de parâmetros.

No caso geral, o AIC é dado por

$$AIC = -2(\ln(L)) + 2K$$

onde K é o número de parâmetros do modelo e L é função de máxima verosimilhança para o modelo estimado.

Outro critério proposto é o *Bayesian information criterion* (BIC) dado por:

$$BIC = -2\ln L + 2K\ln N$$

onde N é o número de observações.

O melhor modelo a partir de um conjunto de modelos considerados é o que minimiza o BIC.

Apesar de algumas semelhanças entre o AIC e o BIC eles vêm de diferentes abordagens. A diferença é que o BIC tende a penalizar modelos complexos de um modo mais pesado através do termo de penalidade. Este termo é igual a 2 no caso do AIC e igual a $\ln N$ no caso do BIC (este depende do número de observações).

Para a selecção de modelos, não há uma escolha clara entre o AIC e o BIC. Dado um conjunto de modelos incluindo o verdadeiro modelo, a probabilidade do BIC seleccionar o modelo mais correcto aproxima-se quando N tende para infinito, por outro lado o AIC tende a escolher modelos cada vez mais complexos à medida que N cresce. Para amostras finitas o BIC tende a escolher modelos demasiado simples, por causa da sua penalidade (Hastie et al, 2009).

Ambos os critérios incluem o princípio de *Occam's* que afirma que se todos os modelos explicam os dados de igual maneira então devemos preferir o mais simples e económico.

Dado um conjunto de modelos $f_\alpha(x)$ indexado a um parâmetro “*tuning*” α define-se

$$AIC(\alpha) = erro(\alpha) + 2 \cdot \frac{d(\alpha)}{N} \hat{\sigma}_\varepsilon^2$$

onde $erro(\alpha) = \frac{1}{N} \sum_{i=1}^N L(y_i, \hat{f}(x_i))$ é o “*training error*”, $d(\alpha)$ é o número de parâmetros para cada modelo e $\hat{\sigma}_\varepsilon^2$ é a variância estimada.

Esta função fornece uma estimativa da curva de erro e o parâmetro “*tuning*” $\hat{\alpha}$ que a minimiza. O modelo final será $f_{\hat{\alpha}}(x)$ (Hastie et al, 2009).

3.3.3 LASSO

O LASSO (*Least Absolute Shrinkage and Selection Operator*) é um método de selecção e “*shrinkage*” proposto para os modelos de regressão linear, introduzido por Robert Tibshirani em 1995. Dado um conjunto de variáveis preditoras X_{1i}, \dots, X_{ki} e um *outcome* y_i , o LASSO ajusta um modelo linear $\hat{y}_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$ onde os preditores estão standardizados e os valores resposta centrados.

O critério é minimizar a soma dos quadrados

$$\sum_{i=1}^N (y_i - \sum_k \beta_k X_{ki})^2$$

com uma restrição nos coeficientes da forma $\sum |\beta_k| \leq t$. O valor t é um parâmetro “*tuning*”, quando este é grande a restrição não tem efeito e a solução é a regressão linear múltipla. Para valores pequenos, mas maiores ou iguais a zero as soluções são versões “reduzidas” das estimativas dos mínimos quadrados. Os coeficientes poderão tomar o valor zero, a escolha de t influencia o número de preditores a usar no modelo de regressão linear. O melhor valor de t pode ser estimado através de *cross-validation* (validação cruzada), usado em alguns *packages* estatísticos.

Reformulando o problema de otimização com uma restrição, através dos multiplicadores de Lagrange passamos a ter:

$$\sum_{i=1}^N (y_i - \sum_k \beta_k X_{ki})^2 + \lambda \sum_{k=1}^p |\beta_k|$$

onde $\lambda \sum_{k=1}^p |\beta_k|$ é o termo de penalidade e λ é o parâmetro “*tuning*” que controla a força da penalização.

O LASSO usa a penalidade L1 que faz o “*shrinkage*” dos coeficientes, bem como a selecção das variáveis, o que resulta em coeficientes iguais a zero.

A Figura 3.3 representa o LASSO quando há apenas dois preditores. Os contornos elípticos são a função da soma dos quadrados dos resíduos, centrados na estimativa dos mínimos quadrados, $\hat{\beta}$. A região de restrição é um quadrado $|\beta_1| + |\beta_2| \leq t$. O primeiro ponto é encontrado quando os contornos elípticos tocam a região de restrição. Como o quadrado tem cantos, se a solução ocorrer num canto, então ter-se-á um parâmetro β_k igual a zero (Hastie et al, 2009).

No R, estão disponíveis os *packages* *lasso2* (Turlach et al, 2011) e *lars* (Hastie et al, 2011), para a implementação deste método.

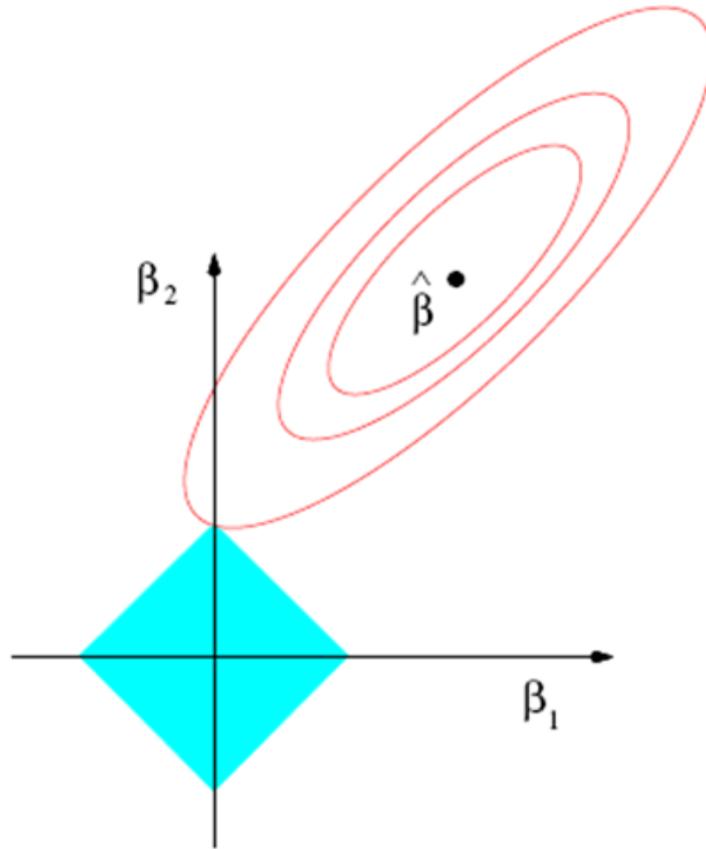


Figura 3.3: Estimação do LASSO (Tibshirani, 1996)

A validação cruzada (“*cross validation*” (CRV)) é um método estatístico que divide os dados em dois segmentos, dados de treino (“*training data*”) e dados de validação (“*validation data*”), e avalia se um modelo estimado num conjunto de dados de treino (“*training sets*”) pode ser generalizado a um conjunto de dados independentes. Uma das formas de validação cruzada é o “*k-fold cross-validation*”. Aqui a amostra original é particionada aleatoriamente em k subamostras. Das k subamostras, apenas uma vai ser usada para validar os dados, testando o modelo. As restantes $k-1$ subamostras são usadas como dados de treino (*training data*). O processo é repetido k vezes com cada uma das k subamostras usadas exactamente uma vez nos dados de validação. Calcula-se a média dos k resultados de modo a produzir uma única estimação. Para o valor de $k=N$, também conhecido como “*leave-one-out cross-validation*”, a validação cruzada é aproximadamente não enviesada para o verdadeiro erro preditivo, mas pode ter variância elevada porque os N conjuntos de dados de treino são similares uns aos outros. O esforço computacional também é maior. Para $k=5$, CRV tem variância baixa, mas o enviesamento pode ser um problema, no entanto as escolhas típicas são $k=5$ ou 10 .

Dado um conjunto de modelos $f(x, \alpha)$ indexado a um parâmetro “*tuning*” α , denotado

por $\hat{f}^{-k}(x, \alpha)$ define-se

$$CRV(\alpha) = \frac{1}{N} \sum_{i=1}^N L(y_i, \hat{f}^{-k(i)}(x_i, \alpha))$$

Esta função fornece uma estimativa da curva de erro e o parâmetro “*tuning*” $\hat{\alpha}$ que a minimiza. O modelo final será $f(x, \hat{\alpha})$, que se ajustará a todos os dados (Hastie et al, 2009).

Tradeoff entre viés e variância

A performance de um estimador $\hat{\theta}$ de um parâmetro θ é medida pelo seu erro quadrático médio (MSE – *Mean Square Error*) dado por:

$$MSE = E(\theta - \hat{\theta})^2 = var(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2 = var(\hat{\theta}) + Viés(\hat{\theta})^2$$

Apesar da falta de viés ser um óptimo aspecto de um estimador, não garante um valor baixo do MSE. O valor mínimo é obtido quando é feito um *tradeoff* apropriado entre o viés de um estimador e a sua variância. Observa-se que introduzindo uma determinada quantidade de viés num estimador, isso levará a uma redução significativa da sua variância, tanto que o MSE é reduzido e a performance do estimador é melhorada.

Se se estiver perante um número muito grande de variáveis, então talvez se tenha melhores predições omitindo algumas delas. Às vezes as variáveis são quase colineares o que dificulta a inversão da matriz $X^T X$, pois o seu determinante é quase zero. Ter-se-á parâmetros do modelo instáveis, com grande variância e por isso difíceis de interpretar. A multicolinearidade é apenas um exemplo extremo do *tradeoff* entre viés e variância com que se pode depar sempre que se faz regressão. Ao incluir muitas variáveis as predições serão pobres devido ao aumento da variância. Assim, modelos com muitas variáveis têm viés baixo mas variância alta, modelos com poucas variáveis têm viés alto e variância baixa. O modelo com as melhores capacidades preditivas aparece equilibrando estes dois extremos conduzindo ao melhor *tradeoff* entre viés e variância. O problema da selecção de variáveis permite decidir quais as variáveis a incluir no modelo de modo a obter um bom *tradeoff*.

Os métodos de penalidade, nomeadamente o LASSO são uma variante da regressão linear múltipla cujo objectivo é contornar o problema da colinearidade dos preditores.

Para isso é introduzido um parâmetro extra que determinará quanto é que a Regressão LASSO se afasta da Regressão Linear. Se este parâmetro for muito pequeno a regressão

LASSO não vai ultrapassar os problemas de colinearidade eficientemente. Se for muito grande o viés dos parâmetros será demasiado grande, assim como os parâmetros e previsões dos MSE's. Terá então que ser escolhido o parâmetro óptimo, usualmente por *cross-validation*. A matriz será artificialmente modificada de modo a que o seu determinante seja diferente de zero. Ao fazer isto, os parâmetros do novo modelo serão enviesados, mas as variâncias destes novos parâmetros são menores do que os da regressão linear, assim como os seus MSE's.

Erros padrão

Os erros padrão, em princípio podem ser calculados usando *bootstrap*. No entanto, os *packages* do R não os fornecem. A razão é que os erros padrão não são muito expressivos para estimativas muito enviesadas como as que surgem dos métodos de penalidade. Os métodos de penalidade reduzem a variância dos estimadores através da introdução de viés. O viés de cada estimador é por isso um componente importante do seu erro quadrático médio, enquanto a variância contribui apenas numa pequena parte. Na maioria das aplicações dos métodos de penalidade é impossível obter uma estimativa precisa do viés. Qualquer cálculo baseado no *bootstrap* só dará uma ideia da variância das estimativas.

Reportar um erro padrão de uma estimativa penalizada pode dar uma impressão errada de grande precisão, ignorando a falta de exactidão causada pelo viés (Goeman, 2011).

3.3.4 LASSO para a regressão logística

A penalidade L1 usada no LASSO pode ser usada para selecção de variáveis e “*shrinkage*” em qualquer modelo linear (Hastie et al, 2009).

Para a regressão logística, maximiza-se a versão penalizada de 3.1.9

$$\max_{\beta_0, \beta_k} \left\{ \sum_{i=1}^n Y_i(\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}) - \log(1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}) - \lambda \sum_{k=1}^p |\beta_k| \right\}$$

Os *packages* *glmnet* (Friedman et al, 2010) e *glmnet* (Park et al, 2011), disponíveis no R, para além da regressão linear permitem implementar o modelo de regressão logística.

3.3.5 Group LASSO

O *Group LASSO* é uma extensão do LASSO que faz selecção de variáveis em grupos de variáveis predefinidos nos modelos de regressão linear (Yuan and Lin, 2006, Meier, Van de Geer, Buehlmann, 2008).

A solução LASSO não é satisfatória no caso particular de existirem variáveis preditoras categóricas (factores), dado que selecciona as variáveis *dummies* individualmente em vez de considerar todos os factores. Além disso as soluções do LASSO dependem de como as variáveis *dummy* estão codificadas. Teremos diferentes soluções, escolhendo diferentes contrastes para uma variável categórica. Estes problemas podem ser ultrapassados através do *Group LASSO* que introduz uma extensão da penalidade LASSO. O estimador é definido como

$$\hat{\beta}_\lambda = \operatorname{argmin}_\beta \left\{ \sum_{i=1}^N (y_i - \sum_k \beta_k X_{ki})^2 + \lambda \sum_{g=1}^G \|\beta_{\mathfrak{J}_g}\| \right\}$$

onde \mathfrak{J}_g é o conjunto de índices pertencentes ao g -ésimo grupo de variáveis, $g=1, \dots, G$. Tem a propriedade de fazer selecção da variável no nível do grupo e é invariante a transformações ortogonais (Meier et al, 2008).

3.3.6 *Group LASSO* para a regressão logística

Assumindo que se tem observações (x_i, y_i) independentes e identicamente distribuídas, $i = 1, \dots, n$, de um vector $x_i \in \mathbb{R}^p$, p -dimensional de G preditores e uma resposta binária $y_i \in \{0, 1\}$. As variáveis categóricas e contínuas são ambas permitidas.

Denotando por df_g os graus de liberdade do g -ésimo preditor podemos reescrever $x_i = (x_{i,1}^T, \dots, x_{i,G}^T)^T$ com o grupo de variáveis $x_{i,g} \in \mathbb{R}^{df_g}$, $g = 1, \dots, G$. Um factor com três níveis terá 2 graus de liberdade e uma variável contínua terá 1 grau de liberdade.

O estimador $\hat{\beta}_\lambda$ da regressão logística *Group LASSO* é dado pelo minimizante da função convexa:

$$S_\lambda(\beta) = -\left(\sum_{i=1}^n Y_i (\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}) - \log(1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}}) \right) - \lambda \sum_{g=1}^G s(df_g) \|\beta_g\|_2$$

O parâmetro $\lambda \geq 0$ controla a quantidade da penalização. A função $s(\cdot)$ é usada para re-escalar a penalidade com respeito à dimensionalidade do vector parâmetro β_g . Usa-se $s(df_g) = df_g^{1/2}$ para assegurar que o termo de penalidade é da ordem do número de parâmetros df_g (Meier et al, 2008).

O algoritmo que resolve este problema de optimização está disponível no *package grplasso* do R, permitindo implementar a regressão logística quando existem variáveis categóricas no modelo (Meier, 2009).

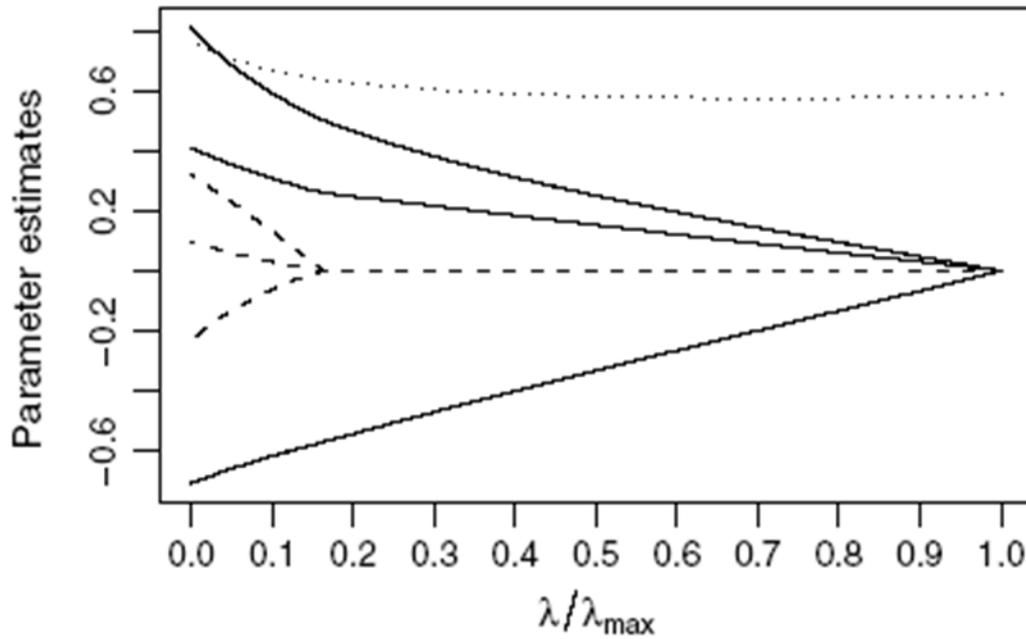


Figura 3.4: Soluções $\{\hat{\beta}_\lambda\}_{\lambda \geq 0}$ para um modelo que consiste num intercepto (.....) e dois factores com 3 graus de liberdade cada (—, - - - -): λ_{max} é o valor de penalidade λ tal que nenhum grupo penalizado está no modelo. Fonte: Adaptado de Meier *et all*, 2008.

Para calcular as soluções $\hat{\beta}_\lambda$ numa série de valores do parâmetro de penalidade $0 \leq \lambda_k < \dots < \lambda_1 \leq \lambda_{max}$, começa-se pelo λ_{max} onde apenas o intercepto está na equação. Usa-se $\hat{\beta}_{\lambda_k}$ como o valor inicial de $\hat{\beta}_{\lambda_{k+1}}$ e processa-se iterativamente até $\hat{\beta}_{\lambda_k}$ com λ_k igual ou próximo de zero. Pode ser usada validação cruzada de modo a escolher o parâmetro λ .

Capítulo 4

Resultados

4.1 Análise dos factores P de Predisposição

Relativamente aos factores de Predisposição, depois da análise descritiva de cada uma das variáveis relativamente à mortalidade hospitalar e da análise bivariada através do Qui-Quadrado, foi realizada a análise logística univariada. Desta análise foram escolhidas as variáveis cujo valor $p < 0.2$, sendo então feita a regressão logística múltipla com estas variáveis. Outra abordagem foi a análise dos dados usando os métodos *stepwise* e o método de selecção de variáveis LASSO, com os *packages* existentes no R. O objectivo principal é, através do modelo de regressão logística conjuntamente com LASSO, seleccionar as variáveis mais importantes para predizer o “resultado do internamento hospitalar” (morte ou alta hospitalar).

Existem *packages* estatísticos (alguns ainda em desenvolvimento) no R (versão 2.13) de forma a implementar o LASSO. Com a implementação do *package grplasso* foram seleccionadas as variáveis consideradas mais importantes para o modelo.

Foi construída uma base de dados com as variáveis descritas na Tabela 2.1. Dos 897 doentes com sepsis adquirida na comunidade, seleccionaram-se apenas os que apresentavam sépsis grave e choque séptico, tendo sido excluídos 83 (9%) doentes com sépsis e 6 doentes cujo resultado final (morte ou alta hospitalar) era desconhecido. A base de dados ficou constituída por 808 doentes.

Os doentes em que era desconhecida a presença de corticoterapia de longa duração (2.7%), curta duração (2.2%), quimioterapia (1.4%), radioterapia (1.6%), VIH (19.1%), SIDA (19.7%), doença crónica hepática (2.5%), renal (3.6%), cardiovascular (2.5%), respiratória (2.5%), hematológica (2.6%), neoplásica (5.7%) e n.^o de comorbilidades (1.1%), foram adicionados à classe do “não”. Assim, na classe do “não”, ficaram os doentes que não tinham doença ou era desconhecida a sua presença.

A taxa de mortalidade hospitalar foi de 40%. O teste de independência do Qui- Qua-

drado encontrou associação estatisticamente significativa entre a mortalidade e o género (42% no sexo masculino *vs* 35% no sexo feminino, $p=0.033$), a mortalidade e a idade (as classes de idade maior do que 70 anos têm uma taxa de mortalidade mais elevada do que as classes menores do que 70 anos, $p < 0.001$), a mortalidade e a quimioterapia (39% *vs* 55%, $p=0.043$), a mortalidade e a doença crónica hepática (38% *vs* 54%, $p=0.009$), a mortalidade e a doença hematológica (39% *vs* 68%, $p=0.010$), a mortalidade e a doença neoplásica (38% *vs* 62%, $p=0.002$) e a mortalidade e as comorbilidades (os pacientes com mais do que 2 comorbilidades têm uma taxa de mortalidade de 50%, enquanto os que não têm comorbilidades têm uma taxa de mortalidade de 35% e os que têm uma comorbilidade têm uma taxa de mortalidade de 39%, $p=0.002$).

Não foi encontrada associação para a estação do ano ($p=0.990$), corticoterapia crónica ($p=0.478$), radioterapia ($p=0.360$), infecção VIH ($p=0.506$), SIDA ($p=0.554$), doença cardiovascular ($p=0.249$), doença renal ($p=0.110$), doença respiratória ($p=0.682$) (Tabela 4.1).

Tabela 4.1: Características da amostra relativamente aos factores de Predisposição por mortalidade hospitalar

	Total		Mortalidade		Alta Hospitalar		valor p*
	n(808)	%	n(321)	%(40)	n(487)	%(60)	
Sexo							
Feminino	290	36	101	35	189	65	<i>0.003</i>
Masculino	518	64	220	42	298	58	
Idade							
<39	105	13	21	20	84	80	<i><0.001</i>
40-59	228	28	88	39	140	61	
60-69	169	21	55	33	114	67	
70-74	107	13	48	45	59	55	
75-79	104	13	53	51	51	49	
>80	95	12	56	59	39	41	
Estação							
Primavera	182	23	74	41	108	59	0.990
Verão	180	22	72	40	108	60	
Outono	191	24	75	39	116	61	
Inverno	255	32	100	39	155	61	
Corticoterapia de longa duração							
Não/Desconhecido	758	94	299	39	459	61	0.524
Sim	50	6	22	44	28	56	
Corticoterapia de curta duração							
Não/Desconhecido	785	97	311	40	474	60	0.709
Sim	23	3	10	44	13	57	
Corticoterapia global							
Não/Desconhecido	737	91	290	39	447	61	0.478
Sim	71	9	31	44	40	56	

	Total	Mortalidade	Alta Hospitalar				
Quimioterapia							
Não/Desconhecido	768	95	299	39	469	61	0.043
Sim	40	5	22	55	18	45	
Radioterapia							
Não/Desconhecido	797	99	315	40	482	60	0.360**
Sim	11	1	6	55	5	45	
VIH							
Não/Desconhecido	763	94	301	39	462	61	0.506
Sim	45	6	20	44	25	56	
SIDA							
Não/Desconhecido	772	95	305	40	467	60	0.554
Sim	36	5	16	44	20	56	
Doença crónica hepática							
Não/Desconhecido	736	91	282	38	454	62	0.009
Sim	72	9	39	54	33	46	
Doença crónica renal							
Não/Desconhecido	754	93	294	39	460	61	0.110
Sim	54	7	27	50	27	50	
Doença crónica cardiovascular							
Não/Desconhecido	660	82	256	39	404	61	0.249
Sim	148	18	65	44	83	56	
Doença crónica respiratória							
Não/Desconhecido	660	82	260	39	400	61	0.682
Sim	148	18	61	41	87	59	
Doença hematológica							
Não/Desconhecido	789	98	308	39	481	61	0.010
Sim	19	2	13	68	6	32	
Doença neoplásica							
Não/Desconhecido	763	94	293	38	470	62	0.002
Sim	45	6	28	62	17	38	
N.º de Comorbilidades							
0	405	50	140	35	265	65	
1	193	24	76	39	126	59	0.002
2 ou mais	210	26	105	50	96	50	

Dado que se está na presença de um *outcome* (resultado do internamento hospitalar – morte ou alta hospitalar) com resposta binária, o modelo a estimar será o de regressão logística. Este modelo pretende determinar a influência dos factores idade, sexo, estação do ano, corticoterapia de longa duração, de curta duração e global, quimioterapia, radioterapia, infecção VIH, SIDA, doença crónica hepática, renal cardiovascular, doença hematológica e neoplásica e comorbilidades na probabilidade de falecer.

A análise logística univariada revelou uma associação entre a mortalidade e o sexo (o sexo masculino tem um maior risco de morrer relativamente ao sexo feminino, OR=1.4, p=0.033), idade (<39=ref | 40-59 OR=2.5, p=0.001 | 60-69 OR=1.9, p=0.025 | 70-74 OR=3.3, p <0.001 | 75-79 OR=4.2, p <0.001 | > 80 OR=5.7, p <0.001), quimio-

rapia (OR=1.9, p=0.046), doença hepática crónica (OR=1.9, p=0.010), doença renal crónica (OR=1.6, p=0.010), doença hematológica (OR=3.4, p=0.015) e doença neoplásica (OR=2.6, p=0.002). Os melhores preditores de mortalidade hospitalar foram o sexo (OR=1.4, p=0.022), idade (< 39 = ref | 40-59 OR=2.4, p=0.002 | 60-69 OR=1.9, p=0.029 | 70-74 OR=3.3, p<0.001 | 75-79 OR=4.5, p<0.001 | >80 OR=6.7, p<0.001), doença hepática (OR=2.1, p=0.004), doença renal (OR=2.0, p=0.024), doença hematológica (OR=2.7, p=0.063) e doença neoplásica (OR=2.5, p=0.007) (Tabela 3).

Tabela 4.2: Regressão logística univariada e multivariada para os factores de Predisposição

	Raw OR	valor p	OR Ajustado	valor p
Sexo				
Feminino	-Ref-		-Ref-	
Masculino	1.4	<i>0.033</i>	1.4	<i>0.022</i>
Idade				
<39	-Ref-		-Ref-	
40-59	2.5	<i>0.001</i>	2.4	<i>0.002</i>
60-69	1.9	<i>0.025</i>	1.9	<i>0.029</i>
70-74	3.3	<i><0.001</i>	3.3	<i><0.001</i>
75-79	4.2	<i><0.001</i>	4.5	<i><0.001</i>
>80	5.7	<i><0.001</i>	6.7	<i><0.001</i>
Estação				
Primavera	-Ref-			
Verão	1.0	0.898		
Outono	0.9	0.784		
Inverno	0.9	0.761		
Corticoterapia de longa duração				
Não/Desconhecido	-Ref-			
Sim	1.2	0.520		
Corticoterapia de curta duração				
Não/Desconhecido	-Ref-			
Sim	1.2	0.709		
Corticoterapia global				
Não/Desconhecido	-Ref-			
Sim	1.2	0.479		
Quimioterapia				
Não/Desconhecido	-Ref-			
Sim	1.9	<i>0.046</i>		
Radioterapia				
Não/Desconhecido	-Ref-			
Sim	1.8	0.319		
VIH				
Não/Desconhecido	-Ref-			
Sim	1.2	0.506		
SIDA				
Não/Desconhecido	-Ref-			

Sim	1.2	0.555		
Doença crónica hepática				
Não/Desconhecido	-Ref-		-Ref-	
Sim	1.9	0.010	2.1	0.004
Doença crónica renal				
Não/Desconhecido	-Ref-		-Ref-	
Sim	1.6	0.010	2.0	0.024
Doença crónica cardiovascular				
Não/Desconhecido	-Ref-		-Ref-	
Sim	1.2	0.249		
Doença crónica respiratória				
Não/Desconhecido	-Ref-		-Ref-	
Sim	1.1	0.682		
Doença hematológica				
Não/Desconhecido	-Ref-		-Ref-	
Sim	3.4	0.015	2.7	0.063
Doença neoplásica				
Não/Desconhecido	-Ref-		-Ref-	
Sim	2.6	0.002	2.5	0.007
N.º de Comorbilidades				
0	-Ref-			
1	1.2	0.253		
2 ou mais	1.9	<0.001		

Os factores de risco, ou seja os factores que poderão influenciar a mortalidade hospitalar, incluídos no modelo final poderão ter um forte impacto nos resultados do estudo. Se um investigador inclui muitas variáveis ou não tem dados suficientes, os coeficientes de um modelo de regressão logística poderão ser enviesados, sobrestimados o que pode levar a resultados falsos -positivos de associação dos factores de risco. Por isso, a selecção dos factores associados à mortalidade é uma parte importante da construção de um modelo.

Usualmente é usada a regressão logística ou a regressão logística com os métodos *stepwise* para seleccionar os factores associados a um determinado *outcome*. No entanto estes métodos tendem a seleccionar coeficientes elevados, isto porque estes métodos de selecção envolvem um teste à significância estatística das variáveis ignorando a incerteza introduzida pelo processo de selecção do modelo.

Outro método de selecção de variáveis é o método de penalidade que reduz a sobreestimação dos coeficientes e o erro padrão. De modo a comparar os métodos de selecção de variáveis *stepwise* e LASSO, foram realizadas análises usando estas diferentes metodologias (Tabela 4.3). Os métodos *Stepwise* (*Backward LR* e *Forward LR*) foram efectuados no PASW versão 18.0. Neste estudo existiam variáveis binárias e categóricas (Idade, estação e número de comorbilidades), por isso o procedimento utilizado foi o *group* LASSO. Este procedimento inclui um termo de penalidade que à medida que aumenta, aproxima de zero as estimativas. Para isso, foi usado o *package* estatístico do R 2.13.0, *grplasso*.

Aqui as variáveis *dummy* de um factor serão automaticamente tratadas como um grupo (Meier, 2009).

Testando diferentes níveis de intensidade da penalidade será obtido um subconjunto adequadamente pequeno de variáveis. Através de *k-fold cross-validation* (o valor k escolhido foi $k=8$, visto que para este valor todos os grupos terão aproximadamente tamanhos iguais) escolheu-se o parâmetro de penalidade que produzia menor erro preditivo para o modelo logístico. Foram testados diferentes graus da penalidade (λ) de modo a chegar a um conjunto de 9 variáveis. A Figura 4.1 mostra os resultados do *package grplasso*: à medida que o lambda aumenta, o número de variáveis ou coeficientes iguais a zero, representados por diferentes linhas/cores na Figura 4.1, aumenta, apresentando um conjunto de variáveis mais conveniente. Todas as variáveis são standardizadas, para que as variáveis medidas em diferentes unidades possam ser comparadas umas com as outras.

Tabela 4.3: Coeficientes estimados para diferentes métodos de selecção e “*shrinkage*”

	Backward	Forward	LASSO ($\lambda = 29$)
	$\beta(EP)$	$\beta(EP)$	$\beta(EP)$
Sexo			
Feminino			
Masculino	0.363 (0.161)	0.345 (0.161)	0.361 (0.160)
Idade			
<39			
40-59	0.938 (0.291)	0.919 (0.289)	0.857 (0.285)
60-69	0.780 (0.311)	0.757 (0.309)	0.644 (0.300)
70-74	1.327 (0.330)	1.302 (0.327)	1.191 (0.319)
75-79	1.639 (0.332)	1.636 (0.329)	1.477 (0.320)
>80	2.033 (0.340)	2.013 (0.338)	1.837 (0.328)
VIH			
Não/Desconhecido			
Sim	0.649 (0.343)	0.673(0.343)	
Doença crónica hepática			
Não/Desconhecido			
Sim	0.732(0.264)	0.745(0.263)	0.735 (0.261)
Doença crónica renal			
Não/Desconhecido			
Sim	0.720 (0.304)	0.763 (0.301)	
Doença hematológica			
Não/Desconhecido			
Sim	0.960 (0.536)		1.096 (0.534)
Doença neoplásica			
Não/Desconhecido			
Sim	0.917 (0.335)	1.019 (0.328)	0.870 (0.334)

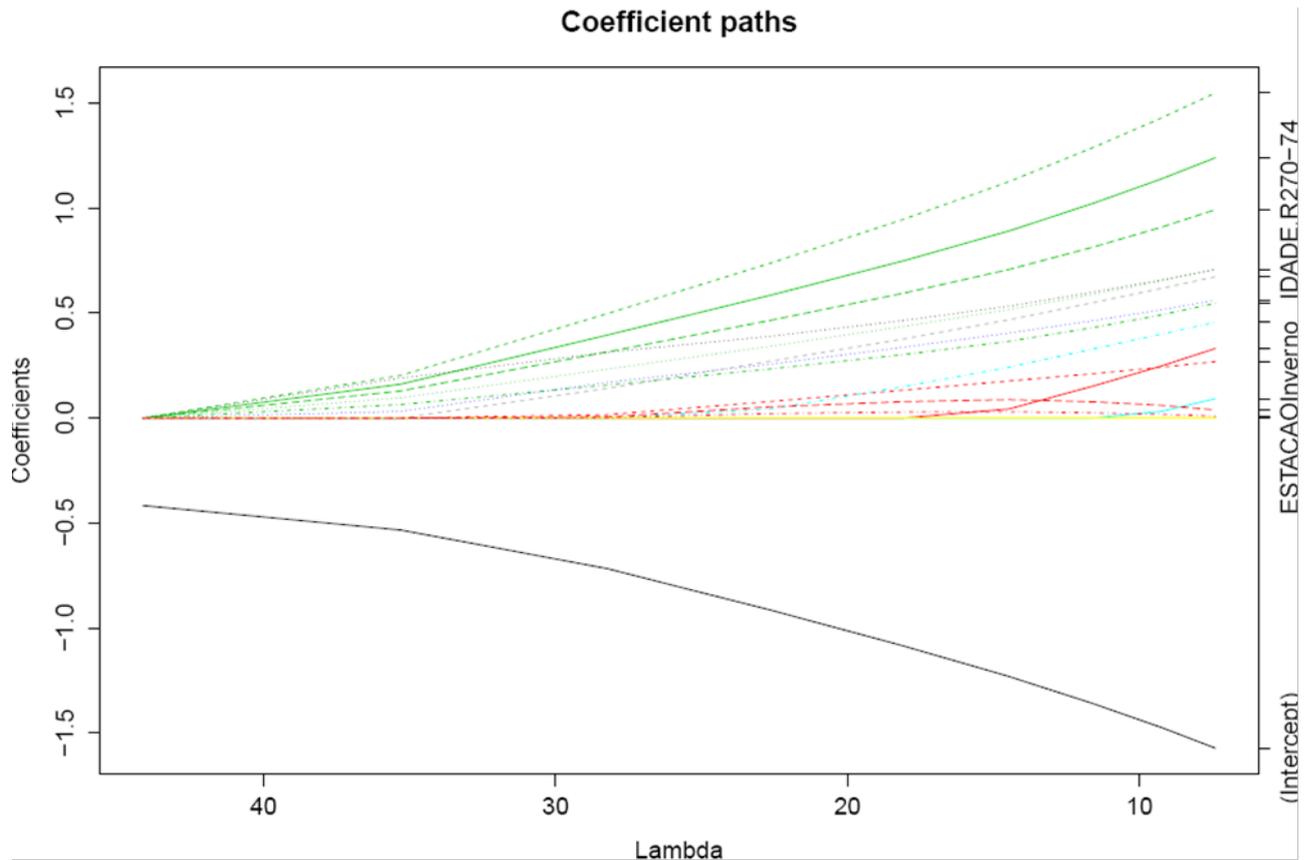


Figura 4.1: Coeficientes estimados versus valores lambda

O valor máximo obtido de lambda é de 44.14, para este valor todos os coeficientes são iguais a zero. O valor mínimo de lambda é de 7.41, mesmo para este valor a estação do ano, a corticoterapia de curta duração, a corticoterapia global, a quimioterapia, radioterapia, SIDA, doença crónica cardiovascular e doença crónica respiratória não são factores associados ao *outcome* mortalidade hospitalar. Da Figura 4.1 observa-se que a idade, a doença hepática e neoplásica estão fortemente associadas ao resultado do internamento hospitalar mesmo para valores de lambda maiores que 35. Para lambda igual a 29 tem-se que as variáveis seleccionadas foram o sexo, a idade, doença hepática, doença hematológica e neoplásica. Relativamente aos métodos *stepwise*, o método *backward* seleccionou mais uma variável do que o método *forward* – a variável doença hematológica. O LASSO não seleccionou as variáveis VIH e doença crónica renal como os outros métodos.

Outros métodos utilizados para determinar o lambda óptimo, variando o lambda do valor mínimo ao valor máximo, foram os critérios AIC e BIC. A escolha do lambda óptimo através do BIC, recaía sobre o lambda igual a 30, enquanto que pelo critério AIC, seria o lambda igual a 29. Para lambda igual a 30, ter-se-ia um modelo com 8 variáveis, sendo que a variável Sexo era a excluída.

4.2 Análise dos factores R de Resposta

As variáveis do factor R de resposta são do tipo quantitativo (contínuo) sendo por isso, apresentada uma análise preliminar destes dados através de estatísticas descritivas e análises bivariadas (teste não paramétrico de Mann-Whitney e teste não paramétrico de Wilcoxon). Visto que foram recolhidos dados relativos ao primeiro dia de hospitalização (H1) do doente e da sua evolução desde o primeiro ao quinto dia de internamento na UCI (U1 a U5), estamos perante medições longitudinais. A modelagem de dados longitudinais requer que o comportamento destas variáveis seja estudado de forma a observar qual a melhor abordagem a ser realizada. Neste capítulo não se chegará à modelação destes dados.

Foram realizadas as estatísticas descritivas de todas as variáveis. Para a comparação entre os indivíduos que faleceram e os que tiveram Alta Hospitalar realizou-se o teste não paramétrico de Mann-Whitney, dado que as variáveis não seguiam uma distribuição normal. A distribuição das variáveis foi verificada através de gráficos e do teste Kolmogorov-Smirnov. Para determinar se a variação/evolução das medições do primeiro dia para o segundo dia, do segundo dia para o terceiro dia e assim sucessivamente, dentro dos indivíduos que faleceram e dentro dos indivíduos que tiveram Alta Hospitalar foi significativa, efectuou-se o teste não paramétrico de Wilcoxon.

Em todas as medições da temperatura ao longo do internamento na UCI, não existem diferenças significativas entre os dois grupos. Foi encontrada diferença estatisticamente significativa na variação da temperatura entre o primeiro dia de internamento na UCI (U1) e o segundo dia na UCI (U2) ($p < 0.001$), sendo a temperatura mais elevada no segundo dia, e a variação da temperatura entre o segundo e terceiro dia ($p = 0.005$), sendo a temperatura mais elevada no segundo dia (U2) relativamente ao terceiro dia (U3), dentro do grupo dos Falecidos. Dentro dos indivíduos que tiveram Alta hospitalar, a variação da temperatura foi significativa entre U1 e U2 ($U1 < U2$), U2 e U3 ($U2 > U3$), e U4 e U5 ($U4 > U5$) ($p < 0.001$).

Tabela 4.4: Estatísticas descritivas da variável Febre

	Resultado do Internamento Hospitalar										
	Falecido (n=321)					Alta Hospitalar (n=487)					
	Média	DP	Mediana	IQR		Média	DP	Mediana	IQR		valor p
FEBH1	37.7	1.2	37.9	37.0	38.5	37.6	1.2	37.7	36.8	38.5	0.672
FEBU1	37.5	1.3	37.5	36.8	38.3	37.5	1.1	37.5	36.8	38.2	0.934
FEBU2	37.8	1.1	37.8	37	38.5	37.8	0.9	37.7	37	38.5	0.524
FEBU3	37.6	1	37.5	37	38.4	37.6	0.9	37.5	37	38.2	0.785
FEBU4	37.6	1	37.5	37	38.1	37.6	0.8	37.5	37	38	0.634
FEBU5	37.6	1	37.5	37	38.2	37.5	0.8	37.5	37	38	0.615

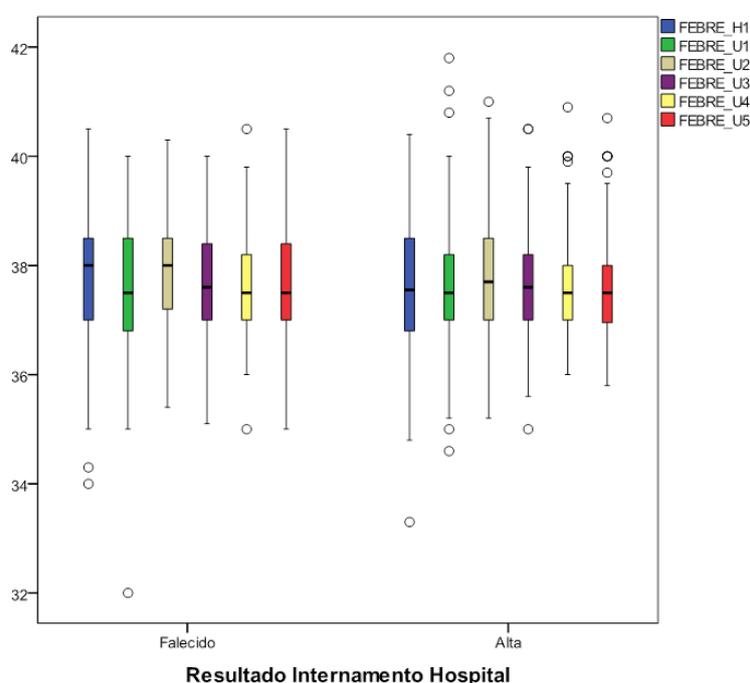


Figura 4.2: Caixas de bigodes das distribuições da Febre

Quanto à variável Frequência Cardíaca (FC), existem diferenças significativas entre os dois grupos nas medições da FC em U1, U2, U3, U4 e U5, sendo a FC significativamente mais elevada nos Falecidos. Foi encontrada diferença estatisticamente significativa na variação da FC entre o primeiro dia de internamento na UCI (U1) e o segundo dia na UCI (U2) ($p < 0.018$), sendo a FC mais elevada no primeiro dia, e a variação da FC entre o terceiro (U3) e quarto dia (U4) ($p < 0.001$), sendo a FC mais elevada no terceiro dia (U3) relativamente ao quarto dia (U4), dentro do grupo dos Falecidos. Dentro dos indivíduos que tiveram Alta Hospitalar, a variação da temperatura foi significativa entre U1 e U2 ($U1 > U2$), U2 e U3 ($U2 > U3$), e U3 e U4 ($U3 > U4$) ($p < 0.001$).

Tabela 4.5: Estatísticas descritivas da variável Frequência Cardíaca (FC)

Resultado do Internamento Hospitalar											
	Falecido (n=321)				Alta Hospitalar (n=487)						
	Média	DP	Mediana	IQR	Média	DP	Mediana	IQR	valor p		
FCH1	112.4	24.2	111	95	125	111.6	23.6	110	95.0	130	0.775
FCU1	122.1	26.6	120	105	140	113.7	22.3	113	100.0	128	<0.001
FCU2	117.8	26.1	120	100	140	107.9	23.9	107	90.0	124	<0.001
FCU3	113.7	24.6	114	98	130	103.5	23.0	100	89.0	119	<0.001
FCU4	108.2	25.1	105	90	124	100.3	21.5	100	87.0	110	<0.001
FCU5	109.2	27.2	108	90	122	100.6	20.4	100	88.0	110	<0.001

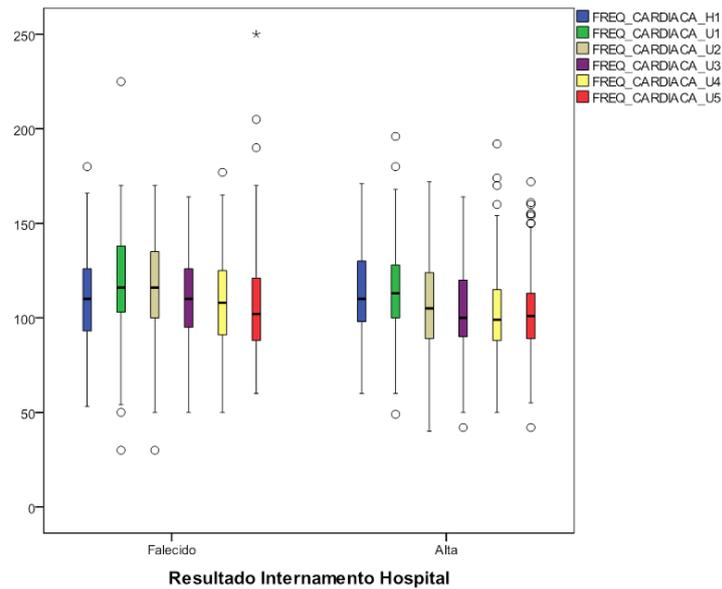


Figura 4.3: Caixas de bigodes das distribuições da Frequência Cardíaca

Em relação à variável Frequência Respiratória (FR), não existem diferenças significativas entre os dois grupos nas medições da FR. Foi encontrada diferença estatisticamente significativa na variação da FR entre U2 e U3 ($U2 > U3$) ($p=0.045$), U3 e U4 ($U3 > U4$) ($p=0.004$) e U4 e U5 ($U4 > U5$) ($p=0.001$), dentro do grupo dos que tiveram Alta Hospitalar.

Tabela 4.6: Estatísticas descritivas da variável Frequência Respiratória (FR)

Resultado do Internamento Hospitalar											
	Falecido (n=321)					Alta Hospitalar (n=487)					valor p
	Média	DP	Mediana	IQR		Média	DP	Mediana	IQR		
FRH1	32.2	8.4	32	26	36	30.1	7.9	30	24.0	35	0.198
FRU1	32.9	9.7	32	28	37	31.7	9.9	30	24.0	36	0.276
FRU2	28.5	8.9	28	22	32	28.5	7.2	28	24.0	32	0.974
FRU3	23.8	6.2	22.5	19	29	26.0	6.7	27	21.0	30	0.282
FRU4	24.3	5.6	24	18	28	26.9	6.5	26	22.0	30	0.178
FRU5	23.9	5.9	24	18	29	25.3	6.1	25	21.0	29	0.468

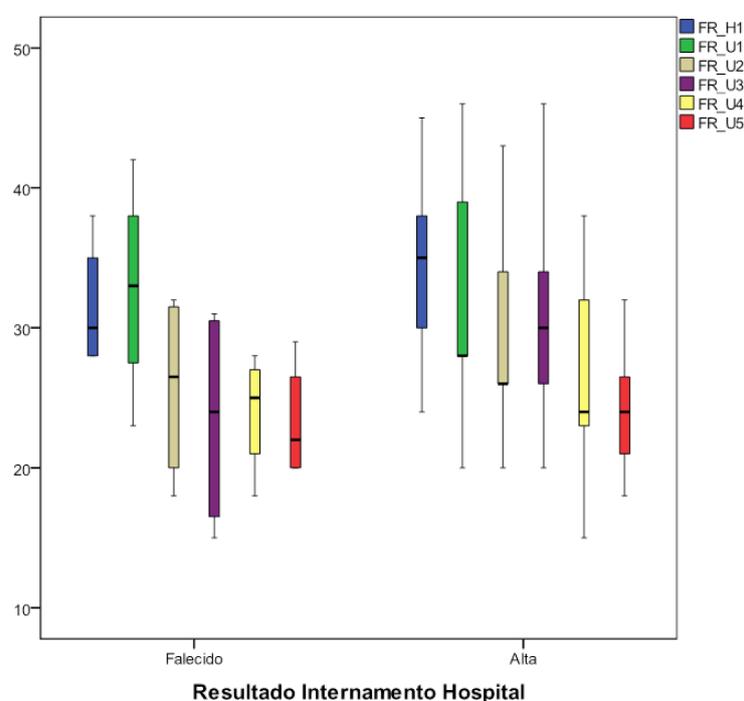


Figura 4.4: Caixas de bigodes das distribuições da Frequência Respiratória

Relativamente à variável Leucocitose, existem diferenças significativas entre os dois grupos nas medições da Leucocitose do terceiro dia (U3) e quarto dia (U4). Foi encontrada diferença estatisticamente significativa na variação da Leucocitose entre U1 e U2 ($U1 > U2$) ($p < 0.001$), U2 e U3 ($U2 > U3$) ($p = 0.006$) e U3 e U4 ($U3 > U4$) ($p < 0.001$), dentro do grupo dos que tiveram Alta.

Tabela 4.7: Estatísticas descritivas da variável Leucocitose

Resultado do Internamento Hospitalar											
	Falecido (n=321)					Alta Hospitalar (n=487)					
	Média	DP	Mediana	IQR		Média	DP	Mediana	IQR		valor p
LEUH1	15.8	10.2	14.0	8.7	19.6	15.5	8.2	13.9	9.6	19.9	0.632
LEUU1	15.1	9.7	13.5	8.9	19.6	15.0	8.1	13.1	9.1	19.5	0.714
LEUU2	15.3	9.3	13.1	9.5	19.0	14.0	7.6	12.2	9.0	16.9	0.073
LEUU3	14.7	7.9	12.6	9.2	18.1	13.4	7.6	11.4	8.5	16.5	0.021
LEUU4	14.4	8.1	13.1	8.7	17.6	13.0	7.6	11.1	8.1	15.5	0.014
LEUU5	14.4	8.7	12.1	8.6	18.3	12.7	7.0	11.3	8.5	14.9	0.072

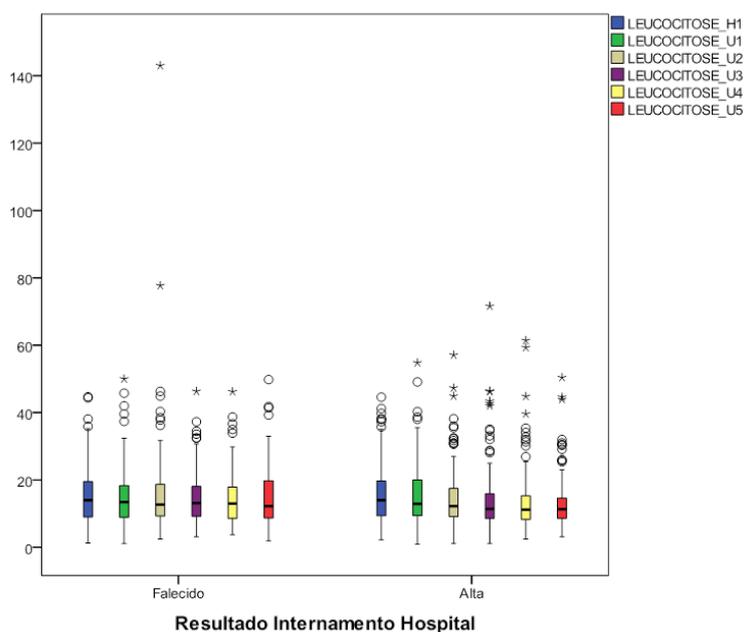


Figura 4.5: Caixas de bigodes das distribuições da Leucocitose

Quanto à variável Neutrófilos, existem diferenças significativas entre os dois grupos nas medições dos Neutrófilos do segundo dia (U2), do terceiro dia (U3), do quarto dia (U4) e do quinto dia (U5). Dentro do grupo dos Falecidos a variação dos neutrófilos foi estatisticamente significativa entre U3 e U4 ($U3 > U4$) ($p=0.019$) e U4 e U5 ($U4 > U5$) ($p=0.028$). Foi encontrada diferença estatisticamente significativa na variação dos Neutrófilos entre U1 e U2 ($U1 > U2$) ($p=0.029$), U2 e U3 ($U2 > U3$) ($p < 0.001$), U3 e U4 ($U3 > U4$) ($p < 0.001$) e U4 e U5 ($U4 > U5$) ($p < 0.001$), dentro do grupo dos que tiveram Alta.

Tabela 4.8: Estatísticas descritivas da variável Neutrófilos

	Resultado do Internamento Hospitalar										
	Falecido (n=321)					Alta Hospitalar (n=487)					
	Média	DP	Mediana	IQR		Média	DP	Mediana	IQR		valor p
NEUH1	84.1	12.0	88	80	91.0	84.2	10.9	87	81	91	0.627
NEUU1	86.1	10.8	89	84	92.8	86.1	9.7	89	82	93	0.680
NEUU2	87.7	8.6	89.5	85	93.0	86.0	9.6	88	82	92	0.003
NEUU3	87.0	9.3	89	84	93.0	84.6	9.0	86	81	91	<0.001
NEUU4	85.8	9.7	89	82	92.0	82.3	10.0	84	76.5	89	<0.001
NEUU5	85.4	9.6	88	82	92.0	81.2	9.8	83	75	89	<0.001

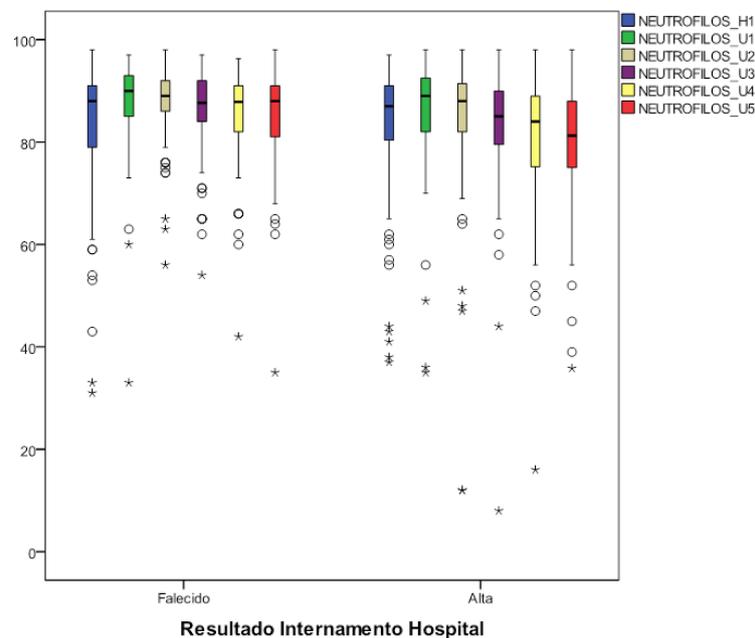


Figura 4.6: Caixas de bigodes das distribuições dos Neutrófilos

Finalmente, quanto à variável PCR, existem diferenças significativas entre os dois grupos nas medições da PCR do terceiro dia (U3), do quarto dia (U4) e do quinto dia (U5). Dentro do grupo dos Falecidos a variação do PCR foi estatisticamente significativa entre U3 e U4 ($U3 > U4$) ($p < 0.001$) e U4 e U5 ($U4 > U5$) ($p < 0.001$). Foi encontrada diferença estatisticamente significativa na variação dos Neutrófilos entre U2 e U3 ($U2 > U3$) ($p < 0.001$), U3 e U4 ($U3 > U4$) ($p < 0.001$) e U4 e U5 ($U4 > U5$) ($p < 0.001$), dentro do grupo dos que tiveram Alta.

Tabela 4.9: Estatísticas descritivas da variável PCR

Resultado do Internamento Hospitalar											
	Falecido (n=321)					Alta Hospitalar (n=487)					valor p
	Média	DP	Mediana	IQR		Média	DP	Mediana	IQR		
PCR_H1	20.3	15.8	18.0	9.2	28.4	20.7	14.3	18.8	9.1	29.6	0.501
PCR_U1	21.0	12.7	19.7	10.9	30.4	20.6	12.7	19.5	10.0	29.3	0.615
PCR_U2	22.2	11.4	21.9	12.9	30.0	20.6	11.8	19.1	11.2	28.2	0.069
PCR_U3	20.8	10.8	20.4	12.7	28.0	17.7	11.2	15.9	8.9	25.0	<0.001
PCR_U4	18.6	10.9	17.3	10.0	26.6	14.2	10.3	12.2	6.2	20.0	<0.001
PCR_U5	16.4	11.0	13.8	7.0	25.9	11.7	8.8	9.4	4.7	17.0	<0.001

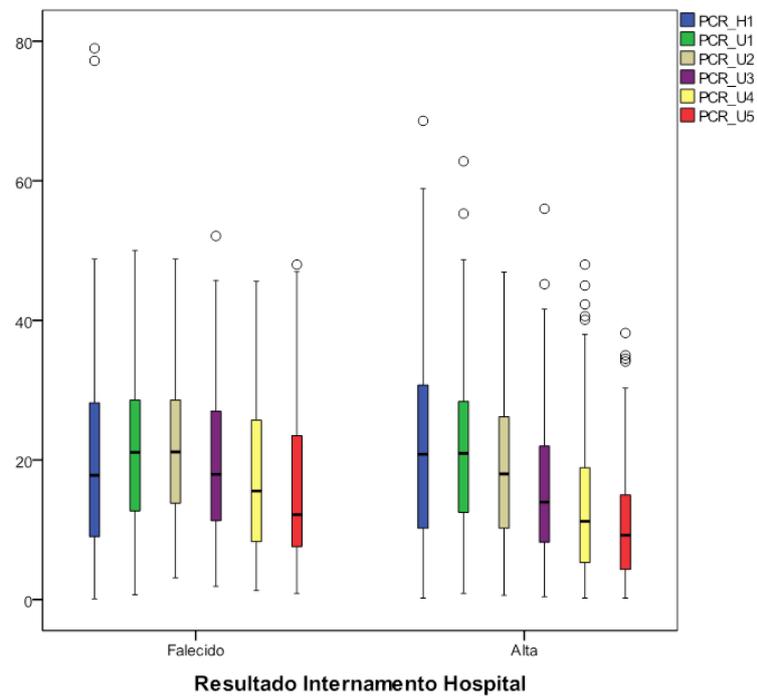


Figura 4.7: Caixas de bigodes das distribuições do PCR

Capítulo 5

Considerações finais

O objectivo deste trabalho, foi sobretudo explorar a metodologia LASSO para a selecção de variáveis explicativas.

Sendo o *outcome* a mortalidade hospitalar, a regressão logística foi o método usado, pois trata-se de uma variável dependente binária.

A necessidade de criar um *score* que possa prever o resultado do internamento hospitalar, de um doente com sepsis adquirida na comunidade tem extrema importância na tomada de decisão relativamente à estratégia a usar na terapêutica dos doentes. Para isso foram registados uma série de dados relacionados com o doente e que fazem parte do sistema de classificação PIRO.

Neste trabalho, apenas foram estudadas as variáveis relativas aos factores P de Predisposição e as variáveis R de Resposta, sendo apresentada uma aplicação da regressão logística LASSO, com as variáveis dos factores de Predisposição.

Os métodos *stepwise* (*backward e forward*) são os mais usuais e amplamente divulgados, mas perante um grande número de variáveis explicativas existem vantagens no uso do LASSO, nomeadamente a criação de um modelo com melhor capacidade preditiva que faça um bom equilíbrio entre viés e variância.

A escolha do parâmetro de penalidade é de suma importância pois determina o número de variáveis explicativas a introduzir no modelo. A escolha óptima deste parâmetro foi feita através de *cross-validation*, sendo também verificados os critérios AIC e BIC.

Os resultados do Capítulo 4 mostram a capacidade do modelo de regressão logística LASSO excluir algumas variáveis e fornecer um conjunto mais parcimonioso de variáveis.

No entanto, note-se que esta ferramenta de modelação é apenas uma ajuda e que para se ter um modelo final é preciso a opinião de um especialista na área da medicina, de confirmar a coerência dos resultados com outros métodos estatísticos e ver se a associação destes factores com a mortalidade hospitalar é aceitável.

Comparando o método de selecção LASSO com os métodos *stepwise* verificou-se que o LASSO não seleccionou as variáveis VIH e doença crónica renal. Na regressão logística

univariada a doença crónica renal era um bom preditor de mortalidade hospitalar, bem como na regressão logística multivariada. No entanto, de acordo com o teste Qui-Quadrado não havia associação entre esta variável e a mortalidade hospitalar. Novamente, através do teste do Qui-Quadrado verificou-se que a variável VIH não estava associada à mortalidade hospitalar.

Em conclusão, apesar de ter sido utilizado um *package* que está ainda em fase experimental e, por isso ainda não existirem muitos trabalhos sobre este assunto, a regressão logística LASSO demonstra ser um método de fácil aplicação e interpretação, a ser aplicado em estudos futuros.

Quanto às variáveis dos factores R de resposta, a análise dos dados revelou existirem diferenças significativas:

- Na frequência cardíaca nos cinco dias de internamento na UCI, sendo esta mais elevada nos indivíduos que faleceram.
- No número de leucócitos no terceiro e quarto dia de internamento na UCI, sendo estes mais elevados nos indivíduos que faleceram.
- Na percentagem de neutrófilos no segundo, terceiro, quarto e quinto dia de internamento na UCI, sendo estes mais elevados nos indivíduos que faleceram.
- Na PCR no terceiro, quarto e quinto dia de internamento na UCI, sendo estes mais elevados nos indivíduos que faleceram.

Relativamente à variação das variáveis dos factores de Resposta, notou-se que houve uma descida dos valores, principalmente a partir do terceiro dia.

A modelação destes dados não foi efectuada, primeiro devido à escassez de tempo e em segundo lugar, porque era necessário aqui um trabalho multidisciplinar que envolvesse a opinião médica, de modo a tratar as variáveis que realmente tenham significado e influenciem a mortalidade hospitalar. Uma possível abordagem para estes dados será usar a GEE (Generalized estimating equation), no entanto devido à existência de *missings* ter-se-á que ter algum cuidado na sua utilização.

De modo a construir um *score* PIRO será necessário estudar também as variáveis I de Infecção e O de disfunção do Órgão. A ideia final será atribuir pontuações/pesos às variáveis que contribuem para a mortalidade hospitalar e fazer a validação deste sistema de classificação.

O trabalho desenvolvido durante o estágio curricular permitiu o acesso à exploração de uma grande base de dados, à operacionalização de algumas variáveis, transformando variáveis contínuas em variáveis categóricas e agrupando variáveis categóricas num número menor classes. Permitiu ainda efectuar vários tipos de análises estatísticas, desde as análises univariadas (estatísticas descritivas e análise de frequências), bivariadas (testes de Qui-Quadrado, testes paramétricos t-Student, testes não paramétricos de Mann-Withney,

Wilcoxon), às multivariadas (regressão logística), abrangendo assim muitos dos conhecimentos adquiridos durante a parte curricular do Mestrado em Estatística de Sistemas.

Também demonstrou a importância do trabalho multidisciplinar. A análise dos dados e as técnicas estatísticas podem estar a cargo dos estatísticos, mas para a selecção das variáveis e a interpretação dos resultados é fundamental a colaboração de elementos da área em que se está a trabalhar, neste caso particular, a Medicina.

Referências Bibliográficas

Agresti A. (2007): An Introduction to Categorical Data Analysis, 2nd Edition, Wiley, New Jersey.

Arriagada S. D., Díaz R. F., Donoso F. A. and Cruces R.P.(2010): Clasificación PIRO en sepsis grave y shock séptico pediátrico: Nuevo modelo de estratificación y su utilidad en pronóstico, Rev Chil Infect 27 (1): 17-23.

Alberti C., Brun-Buisson C., Goodman S. V., Guidici D., Granton J., Moreno R., Smithies M., Thomas O., Artigas A. and Le Gall J.R. (2003), Influence of Systemic Inflammatory Response Syndrome and Sepsis on

Outcome of Critically III Infected Patients, American Journal of Respiratory and Critical care Medicine Vol. 168, 77-84.

Caster O. (2007): Mining the WHO Drug Safety Database Using LASSO Logistic Regression, UUDM Project Report, <http://www.math.uu.se/research/pub/Caster1.pdf>.

Cox D.R. (1970): Analysis of Binary Data, Methuen & Co Ltd, London.

Friedman J., Hastie T., Tibshirani R. (2010): LASSO and elastic-net regularized generalized linear models, Technical Document for CRAN,

<http://cran.rproject.org/web/packages/glmnet/glmnet.pdf>.

Goeman J.J. (2011): L1 and L2 Penalized Regression Models, Technical Document for CRAN, <http://cran.r-project.org/web/packages/penalized/penalized.pdf>.

Goeman J.J. (2011): L1 Penalized Estimation in the Cox Proportional Hazards Model, Biometrical Journal 52 (1), pp. 70-84.

Gogos C.A., Lekkou A. , Papageorgiou O., Siagris D., Skoutelis A. and Bassaris P. H.(2003): Clinical prognostic markers in patients with severe sepsis: a prospective analysis of 139 consecutive cases, Journal of Infection, 47, 300–306 Elsevier.

Granja C., Lopes A., Dias C., Costa -Pereira A., Carneiro A. (2005): Patients' recollections of experiences in the intensive care unit may affect their quality of life, Critical Care, Vol 9 No 2.

Granja C., Teixeira-Pinto A. and Costa -Pereira A.(2002): Quality of life after intensive care – evaluation with EQ-5D questionnaire, Intensive Care Med 28:898–907.

Granja C., Morujão E. and Costa -Pereira A. (2003): Quality of life in acute respiratory distress syndrome survivors may be no worst than in other ICU survivors, *Intensive Care Med* 29:1744–1750.

Granja C., Dias C., Costa -Pereira A., Sarmento A. (2004): Quality of life of survivors from severe sepsis and septic shock may be similar to that of others who survive critical illness, *Critical Care* 8:R91-R98.

Hastie T., Tibshirani R. and Friedman J. (2009): *The Elements of Statistical Learning: Data Mining, Inference and Prediction*; 2nd Edition, Springer, Canada.

Hastie T., Efron B. (2011): Technical Document for CRAN,
<http://cran.r-project.org/web/packages/lars/lars.pdf>.

Howell M.D., Talmor D., Schuetz P., Hunziker S., Jones A. and Shapiro N. (2011): Proof of principle: The predisposition, infection, response, organ failure sepsis staging system, *Critical Care*, Vol 39, N.º 2.

Hosmer, D. W.; Lemeshow, S. (2000): *Applied logistic regression*, 2nd Edition, Wiley, New York.

Kleinblaum D. G., Kupper L. L., Muller K.E., Nizam A. (1998): *Applied Regression Analysis and Other Multivariable Methods*, 3rd Edition, Duxbury Press, Pacific Grove.

Kyung M., Gill J., Ghosh M. and Casella G. (2010): Penalized Regression, Standard Errors and Bayesian Lassos, *Bayesian Analysis* 5(2), pp. 369 – 412.

Lee S., Lee H., Abeel P. and Ng A. (2006): Efficient L1 – regularized logistic regression, *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*.

Meier L., Geer S. and Buhlmann P. (2008): The group LASSO for logistic regression, *Journal of the Royal Statistical Society, Series B* 70(1), pp. 53 – 71.

Meier L. (2009): Technical Document for CRAN,
<http://cran.r-project.org/web/packages/grplasso/grplasso.pdf>.

Menard S. (1995): *Applied Logistic Regression Analysis*, Sage Publications, Thousand Oaks. Moreno R., Afonso S. and Fevereiro T. (2002): Organ failure, *SpringerLink*, Part 1, 29-46.

Moreno R., Afonso S. and Fevereiro T.(2006): Incidence of Sepsis in Hospitalized Patients, *Current Infectious Disease Reports*, Volume 8, Number 5, 346-350.

Moreno R., Diogo A.C. and Afonso S.(2009): Risk Stratification in Severe Sepsis: Organ Failure Scores or PIRO?, *Springer-Verlag Berlin Heidelberg* p. 11-22.

Moreno R., Metnitz B., Adler L., Hoechtl A., Bauer P., Metnitz P.G.H. (2008), Sepsis mortality prediction based on predisposition, infection and response, *Intensive Care Med* 34:496–504.

Moreno R., Vincent J.L., Matos R., Mendonça A., Cantraine F., Thijs L., Takala J., Sprung C., Antonelli M., Bruining H., Willatts S. (1999): The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective multicentre study, *Intensive Care Med* 25: 686-696.

Opal S.M., (2005): Concept of PIRO as a new conceptual framework to understand Sepsis, *Pediatr Crit Care Med*, Vol. 6, No. 3.

Park M., Hastie T. (2011): Technical Document for CRAN,
<http://cran.r-project.org/web/packages/glmpath/glmpath.pdf>.

Park T. and Casella G. (2008): The Bayesian LASSO, *Journal of the American Statistical Association* 103(482), pp. 681 – 686.

Rello J.(2008): Demographics, guidelines, and clinical experience in severe community-acquired pneumonia, *Critical Care Department*, 12(Suppl 6):S2.

Rosolem M.M., Rabello L.S.C.F., Leal J.V., Soares M., Lisboa T. and Salluh J.I.F. (2010): Entendendo o conceito PIRO: da teoria à prática clínica – Parte 2, *Rev Bras Ter Intensiva*, 22 (1): 64-68.

Tibshirani R. (1996): Regression shrinkage and selection via the LASSO, *Journal of the Royal Statistical Society, Series B* 58(1), pp. 267 – 288.

Tibshirani R. (2011): Regression shrinkage and selection via the LASSO: a retrospective, *Journal of the Royal Statistical Society, Series B* 73 (3), pp. 273 – 282.

Turlach B., Lokhorst J., Venables B. (2011): Technical Document for CRAN,
<http://cran.r-project.org/web/packages/lasso2/lasso2.pdf>.

Wheeler G. (2010): The LASSO Logistic Regression Model: Modifications to aid causality assessment for Adverse Events Following Immunization, Master's Thesis (MSc Medical Statistics), London School of Hygiene & Tropical Medicine, University of London.

Yuan M. and Lin Y. (2006): Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society Series B* 68, pp. 49 – 67.

Zou H. and Hastie T. (2005): Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society Series B* 67, pp. 301 – 320.

Anexos

Anexo I

Abstract 1 Poster

Title: Time of Antibiotic Therapy and sepsis resolution. Results from the Portuguese Community-Acquired Sepsis Study (SACiUCI study) Gonçalves-Pereira J1, Póvoa PR1, Lobo C2, Carneiro AH3 1-UCIP, São Francisco Xavier Hospital, Lisbon, Portugal; 2- Dept Biostatistics and Medical Informatics, CINTESIS, Faculty of Medicine, University of Porto, Porto, Portugal; 3-Santo António Hospital, Porto, Portugal

Rationale: To assess the impact of the time until the start of antibiotic therapy in the outcome of patients with Community-Acquired Sepsis (CAS) admitted to Intensive Care Units (ICU).

Methods: During a 12-month period a cohort multiple centre observational study was conducted in 17 Portuguese ICU. Adult patients consecutively admitted with CAS were included. Time elapsed since their hospital admission until the start of the first antibiotic dose was recorded. The day of ICU discharge or death and hospital outcome were noted. Comparison between survivors and non-survivors was performed.

Results: Eight hundred ninety seven patients (age 60 ± 17 yrs, 64% men, and hospital mortality 38 %) were studied. Antibiotics were started within 3 hrs after hospital admission in 29% of patients (group 1), between 3 to 12 hrs in 42% (group 2) and > 12 hrs in 29% (group 3). The ICU mortality of group 1, group 2 and group 3 were 24%, 31% and 34%, respectively ($p=0.022$). In comparison with group 1, the prescription of antibiotics between 3 to 12 hrs and >12hrs after hospital admission were associated with poor outcome (group 2, OR 1.43; 95% CI 0.956-2.139; group 3, OR 1.651; 95% CI 1.076-2.536). In the subgroup of patients with microbiological documented CAS (N=364) time of antibiotic therapy was also significantly associated with outcome (>3hrs, OR 1.895; 95% CI 1.041-3.448).

Conclusions: An early start of antibiotic therapy, < 3 hrs, was associated with a lower mortality rate in CAS patients admitted in ICU.

Abstract 2 Poster

Title: Positive Blood cultures as a marker of sepsis severity. Results from the Portuguese Community-Acquired Sepsis Study (SACiUCI study) Gonçalves-Pereira J1, Póvoa PR1, Lobo C2, Carneiro AH3 1-UCIP, São Francisco Xavier Hospital, Lisbon, Portugal; 2- Dept Biostatistics and Medical Informatics, CINTESIS, Faculty of Medicine, University

of Porto, Porto, Portugal; 3-Santo António Hospital, Porto, Portugal

Rationale: To assess the prognostic value of positive blood cultures (BC) in Community-Acquired Sepsis (CAS) admitted in Intensive Care Units (ICU).

Methods: During 12-month period a cohort multiple centre observational study was conducted in 17 Portuguese ICU. Adult patients consecutively admitted with CAS were included and followed during their first 5 ICU days. The day of ICU discharge or death, hospital outcome and the results from BC collected on the day of hospital admission were recorded. Comparison between survivors and non-survivors was performed.

Results: Eight hundred ninety seven patients (age 60 ± 17 yrs, 64% men, and hospital mortality 38 %) were studied. Blood cultures were performed on the day of hospital admission on 803 patients. Patients with positive BC had a mean (SD) SAPS II of 48.3 (SD=18.7) and a mean (SD) age of 56.8 (SD=17.6), which were not significantly different from those with negative BC (47.7 and 58.8 $p=0.814$ and $p=0.419$, respectively), however their ICU mortality rate was significantly higher (39% vs 26%, $p=0.001$, respectively). Positive BC was significantly associated with ICU mortality (OR=1.8, 95% CI: 1.3 – 2.6). One hundred and thirty eight out of 160 positive BC were monomicrobial: 47% with Gram-positive bacteria and 40% with Gram-negative bacteria. Patients with respiratory ($n=486$) or intra-abdominal infection ($n=135$) were also more likely to die if they presented positive BC at hospital admission (42% vs 27%; $p=0.016$ and 47% vs 26%; $p=0.028$, respectively).

Conclusions: In CAS patients, positive BC at hospital admission was independently associated with a poor outcome.

Artigo

Title: Bloodstream infections as a marker of community-acquired sepsis severity. Results from the Portuguese Community-Acquired Sepsis Study (SACiUCI study)

Running title: Bloodstream infection in community-acquired sepsis

Keywords: Bloodstream infection; community acquired sepsis; blood cultures; septic shock; intensive care unit

Authors Goncalves-Pereira J, MD 1,2, Pova PR, MD PhD 1,2, Lobo C 3, Carneiro AH, MD 4 on behalf of the Portuguese Community-Acquired Sepsis Study Group

Affiliations 1Polyvalent Intensive Care Unit, São Francisco Xavier Hospital, CHLO, Lisbon, Portugal 2CEDOC, Faculty of Medical Sciences, New University of Lisbon, Lisbon, Portugal 3Faculty of Medicine, Department of Biostatistics and Medical Informatics, CINTESIS, University of Porto, Porto, Portugal 4Santo António Hospital, Porto, Portugal

Address for correspondence: João G. Pereira, MD Unidade de Cuidados Intensivos Polivalente Hospital de São Francisco Xavier Centro Hospitalar Lisboa Ocidental Estrada

do Forte do Alto do Duque 1449-005 Lisboa Portugal Phone - +351 21 043 1104/5 Fax - +351 21 043 1301 Email - joaogpster@gmail.com

Word count

Abstract: 249 words Manuscript: 2856 words References: 26 Figures: 2 Tables: 4

Abstract

Background: The bloodstream infection (BSI) impact on the outcome of patients with Community-Acquired Sepsis (CAS) admitted in Intensive Care Units (ICU) is presently unknown.

Methods: During a 12 month period a cohort multiple centre observational study was conducted in 17 Portuguese ICU. Adult patients consecutively admitted with CAS, in whom blood cultures (BC) were collected before initiation of antibiotic therapy, were selected. The results of the BC, clinical data and the ICU and in-hospital mortality were recorded. Comparison between survivors and non-survivors was performed.

Results: Blood cultures were collected in 803 patients with CAS on hospital admission and BSI was diagnosed in 160 (19.9%). Patients with and without BSI had similar mean Simplified Acute Physiology Score (SAPS) II and age. The presence of BSI was independently associated with mortality, both in ICU [Adjusted Odds Ratio (AOR)=1.88, 95% confidence interval (CI): 1.22 – 2.92; p=0.001] and hospital (AOR 1.51; 95%CI, 1.00-2.30; p=0.015). Cumulative mortality was significantly higher in BSI patients since the ninth day onwards. On day 4 of ICU stay, BSI patients were found to be significantly more dependent of vasopressor (p=0.002) but not of ventilator support. A seasonal variation of BSI isolates was noted: Gram-negative BSI was more common during summer, whilst Gram-positive isolates were more frequent in the winter, without mortality differences.

Conclusions: In ICU admitted CAS patients, BSI at hospital admission was independently associated with ICU and hospital mortality. Bloodstream infections were more commonly due to Gram positive during winter and to Gram negative in the summer.

Keywords – bloodstream infections; community-acquired sepsis; mortality; prognosis

Introduction

Community or hospital-acquired infection and sepsis are responsible for 25% of Intensive Care Unit (ICU) admissions. Despite great advances in the understanding of its pathophysiology, severe sepsis remains associated with high mortality, morbidity and hospital resource use [1]. Microbiological documentation occurs in only a fraction of patients with suspected infection, especially in community acquired sepsis (CAS) [2]. Bloodstream infections (BSI) diagnosis may be even more difficult to detect, because many critically ill patients receive empiric antibiotic therapy before the collection of blood cultures. The Surviving Sepsis Campaign guidelines [3] reinforce the need of obtaining blood cultures before starting antibiotic therapy in patients with CAS and with a high risk of dying in order to identify the causative microorganism and target antibiotic therapy. Our group, SACiUCI (Sepsis Adquirida na Comunidade e internada em Unidade de Cuidados Inten-

sivos) study group, had already shown that performing blood cultures in patients with severe sepsis and septic shock was independently associated with improved survival [4]. Nevertheless, antibiotic therapy should not be withheld by the performance of blood cultures, because delay in starting antibiotics markedly increases mortality [5]. This seems to be particularly important in patients with septic shock or BSI [5, 6]. Hospital-acquired BSI is associated with excess length of stay, extra costs, and excess mortality in critically ill patients [7]. However, data on BSI in patients with CAS is scarce. Furthermore, it is not known if the presence of a BSI in infected patients increases the risk of death. In this study, we evaluate the impact of BSI on mortality of patients with severe CAS.

Methods

The SACiUCI study is a prospective, multiple-centre, observational study designed to evaluate the epidemiology of CAS in patients admitted in Portuguese ICUs. A detailed description of the study has been previously published [4, 8, 9]. Briefly, all patients older than 18 yrs, newly admitted to the participating 17 ICUs, were consecutively enrolled during a 12 month period and followed up until death or hospital discharge. Only the first ICU admission was included. The study design was approved by the local Hospital Ethics Committees. Informed consent was waived due to its observational nature. Patients with CAS, defined as the onset of infection before hospital admission or not present at admission but that became evident in the first 48 hrs, were eligible for further analyses. Presence of sepsis, severe sepsis or septic shock were defined according to the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference criteria [10]. Community-acquired sepsis episodes were divided according to mutually exclusive primary infection source: respiratory, neurologic, urologic, intra-abdominal, skin and soft tissues, gynaecologic and obstetrics, endovascular and others. In each participating ICU, a medical doctor was locally responsible for inclusion and revision of all clinical data included in the study protocol. All data were managed by the Department of Biostatistics and Medical Informatics at the University of Porto, where a dedicated database for SACiUCI had been created. All patients consecutively admitted with CAS that had blood cultures collected before initiation of antibiotic therapy constituted this study population. Data collection included demographic data and comorbid diseases, clinical and laboratory data at hospital admission, at ICU admission and from Hospital to ICU admission. Simplified Acute Physiology Score (SAPS) II were computed for all included patients [11]. The days of ICU and Hospital discharge were recorded. Microbiological and clinical infectious data were reported, along with the antibiotics prescribed. Blood cultures collection at the day of hospital admission and its results were closely scrutinized. Blood cultures isolated microorganisms considered most likely to be contaminants were considered negative for further analyses. Organ dysfunction was evaluated during the first 5 ICU days according to the Sequential Organ Failure Assessment score [12], along with clinical and laboratory variables, including temperature, C-reactive protein (CRP) and white blood cell count.

Patient ICU and hospital outcome, either discharge or death were our primary outcome measures. Comparison between survivors and non-survivors was performed.

Statistical Analysis.

A single investigator in each participating center performed data entry. Consistency of data was assessed with a rechecking procedure of a 10% random sample of patients (see Acknowledgements). Data were screened in detail (see Acknowledgements) for missing information, implausible and outlying values. Continuous variables were expressed as median [interquartile range] or mean \pm standard deviation according to data distribution. Comparisons between groups were performed with two-tailed unpaired Student's t test or Mann-Whitney U test for continuous variables according to data distribution. Fisher's exact test and chi-square test were used to test association between categorical variables as appropriate. A multiple logistic regression was fitted to assess the impact of BSI on mortality in patients with CAS, adjusting for age, SAPS II, sepsis severity and primary infection focus. Adjusted odds ratios (AOR) with 95% confidence interval (CI) were computed. The Hosmer and Lemeshow test was used to check goodness-of-fit. Cumulative mortality for each day in the ICU was calculated for CAS patients either with or without BSI. Data were analyzed using PASW Statistics v. 18.0 for Mac (SPSS, Chicago, IL). All statistics were two-tailed, and significance level was defined as $p < 0.05$.

Results

General description and outcomes

During the 12-month study period a total of 897 patients with CAS were included. In 94 patients, blood cultures were not performed on the day of hospital admission resulting in 803 patients analysed. Their mean age was 58.4 ± 17.8 yrs and 65.3% were male. At ICU admission their mean SAPS II score was high, 47.9 ± 17.9 and 51% of the patients were in septic shock (Table 1). The overall ICU and Hospital mortality was 29% and 37%, respectively. In 160 patients with CAS (19.9%) blood cultures returned positive. Patients with BSI had a mean age of 56.8 ± 17.6 yrs and a mean SAPS II of 48.3 ± 18.7 , not significantly different from patients without BSI ($p=0.814$ and $p=0.419$, respectively). Nevertheless, the rate of septic shock at ICU admission was higher in patients with BSI (Table 1), although not reaching statistical significance (58.8% vs. 50.5%, $p=0.09$). The primary focus of infection in our patient population was the lung (60.6%); however, among the patients with BSI, only 36.9% had a lower respiratory tract primary infection focus. On the opposite, despite the fact that the urinary tract was the primary infection focus in only 7.8% of CAS patients, it was present in 16.3% of the BSI episodes ($p<0.001$) (Table 1). In the multivariable analysis the presence of BSI at hospital admission was found to be independently associated with the ICU mortality, 39.4% vs. 26.4%, (AOR 1.88; 95% CI; 1.22-2.92; $p=0.001$). Similarly, in-hospital mortality was significantly higher in patients with BSI at hospital admission (AOR 1.51; 95%CI, 1.00-2.30; $p=0.015$). In the sub-groups of patients with lung ($n=486$) or intra-abdominal primary focus of infection

(n=135), the concomitant presence of BSI at hospital admission was also associated with an increased risk of dying still in the ICU (42.4% vs. 27.2%; p=0.016 and 46.7% vs. 25.7%; p=0.028, respectively).

Clinical course

Both CRP concentration and body temperature were significantly higher in patients with BSI. No differences were noted in white blood cell count (figure 1). In Table 2 we present the mortality rate and the persistence of organ support dependence in CAS patients at day 4 of ICU stay. Assessing patients still alive at day 4 in the ICU (N=754), we found that the presence of BSI was associated with a higher rate of vasopressor support dependence (40.9% vs. 27.1%, p=0.002) but not of ventilator support. Mortality rate was also not different at that time point. Nevertheless the cumulative mortality was higher in patients with BSI at hospital admission; however this difference only became significant from the ninth day of ICU stay onwards (figure 2). Among BSI patients (N=160) the ICU mortality increased according to the sepsis severity: 12.5% in patients with sepsis, 27.6% in severe sepsis and 49.5% in septic shock (in-hospital mortality 12.5%, 35.1% and 54.8%, respectively). The only variable found to be independently associated with mortality was SAPS II (AOR per point, 1.07, 95% CI; 1.04-1.10), p<0.001).

BSI microbiology

The microorganisms isolated in blood cultures are listed in Table 3. One single agent was found in 80.6% of BSI. One hundred Gram-positive bacteria were isolated in blood cultures, mainly *Streptococcus pneumoniae* and methicillin-sensible (MS) *Staphylococcus aureus*. Roughly the same number, 93 Gram-negative bacteria, were identified. Enterobacteriaceae, essentially *Escherichia coli* and also *Klebsiella pneumoniae*, were the commonest. All patients with BSI received initial adequate antibiotic therapy as a result of the low resistance rate found. No significant association was found between the type of microorganism responsible for the BSI and mortality. The ICU mortality of Gram-negative BSI was 33.3%, of Gram-positive BSI 49.3%, and of mixed microorganisms BSI 38.9% (p=0.176).

Seasonal variation of BSI microbiology

Patients with a monomicrobial BSI admitted on winter were more likely to have a Gram-positive infection, whilst those admitted on summer were more prone to Gram-negative agents (p=0.024) – Table 4. This seems to be largely due to the different primary infection focus, namely more lung infections in the winter and a relative increase in the proportion of both intra-abdominal and urinary tract infections in the rest of the year.

Discussion

In our study we found that the presence of BSI in patients with CAS was independently associated with both ICU and in-hospital mortality, despite the use of appropriate antibiotic therapy in all patients. However, this difference became significant relatively late in clinical course, only after the ninth day of ICU stay. We also unveil a seasonal

variation of microbiological isolates in CAS patients, being Gram-positive microorganisms more common during the winter time, while Gram-negatives were more common in the summer. In a population-based study, the incidence of BSI in the community was estimated as 15.7/100000 patients per year [13]. The risk factors for those 340 BSI identified included an age over 65 yrs, male gender, cancer, alcoholism, diabetes, lung disease and admission to an urban hospital. In another cohort of 3901 patients with CAS, the incidence of BSI at hospital admission was found to be 8.2% [14]. In that study the major clinical predictors of BSI were a high temperature ($>39.5^{\circ}\text{C}$), the presence of a central venous line and the suspicion of endocarditis. The authors concluded that it was possible to safely reduce the collection of blood cultures in 26.7% of patients with low risk of BSI without compromising their safety. In fact another study performed in patients admitted to an emergency room with suspected sepsis, found BSI rate to be 5.5%. Despite 30 of the low risk patients discharged home (25 of whom with an antibiotic prescription) ultimately proved to have a BSI, only seven had deteriorated until they returned for a second hospital evaluation [15]. Also in a cohort of 209 patients with community acquired pneumonia, blood cultures only proved to be useful, providing guidance to antibiotic therapy, in patients with a Pneumonia Severity Score class [16] not lower than III [17]. Again BSI was much more common in the high severity patients, increasing from 5.3% in Pneumonia Severity Score Class I patients to 26.7% in class V. These findings are in accordance with our results: In our population all CAS patients needed ICU admission. Moreover 51% were in septic shock when admitted. Therefore a high rate of positive blood cultures, 19.9% was not surprising. The overall mortality of patients with community-acquired BSI range between 39-42% [13, 18, 19], slightly lower than ours (45.6%), reinforcing the severity of our patient population. However it is not clear if the presence of BSI increases disease severity itself or if it is only a marker of disease severity. In a cohort of 2076 infected surgical patients [20], the presence of BSI (17.5%) was not found to be independently associated with mortality. The authors [20] matched the patients with or without BSI according to their primary site of infection, APACHE II, age and class of infecting organism. According to that model, the presence of a hospital-acquired BSI was not associated with increase mortality, neither in pneumonia or abdominal infections. On the opposite, in our study addressing only patients with CAS, the presence of BSI was independently associated with ICU and in-hospital mortality, even after adjusting for disease severity. Similar findings were noted in a recent study of patients with necrotizing fasciitis [21]. The presence of BSI, especially if caused by group A Streptococcus, was associated with increased mortality. Another study noted an independent association between BSI and the risk of developing septic shock, in a cohort of patients admitted to a medical ward with new onset fever (OR 2.09, $p=0.18$) [22]. Initial early and adequate antibiotic therapy is considered of utmost importance for the treatment of critically ill patients with sepsis [5]. In fact Ibrahim et al. [6] found a close relationship between ini-

tial antibiotic inadequacy and mortality in patients with BSI admitted to an ICU (AOR 6.9, $p < 0.001$). In our study, all patients with BSI at hospital admission received initial adequate AB therapy. This finding seems to have resulted from the very low antibiotic resistance rate of the identified bacteria, a common finding in patients with CAS (Table 3) and also to an initial antibiotic policy adjusted to patient risk (including the presence of health care related procedures) and to the severity of the disease. In fact the most common isolated pathogens in patients with CAS and BSI in our study (Table 3) and in others [13, 19] were, as would be expectable, MS *Staphylococcus aureus*, *Escherichia coli* and *Streptococcus pneumoniae*. Nevertheless, even with this excellent AB adequacy, the observed in-hospital mortality was still over 40%, increasing with sepsis severity. Another prospective multicenter study, addressing intensive care patients, 431 with community-acquired and 1266 with hospital-acquired BSI, found an overall mortality rate of 38.7%. Of note this study population was different from ours, including 22.2% of patients with a central venous catheter related infection and 24.2% of patients had a coagulase-negative *Staphylococcus*. Age, illness severity and the presence of immunosuppression, but not inadequate initial antibiotic therapy (OR 0.89; 95%CI 0.61–1.3; $p = 0.55$), were the identified mortality risk factors [18]. In Table 4 we present the seasonal variation of the agents responsible for BSI. In fact *Streptococcus pneumoniae* was found to be more common during the winter time, which is probably related to the lower temperature and also to the increased circulation of respiratory winter virus (especially respiratory syncytial virus and influenza virus) [24]. On the contrary, *Escherichia coli* BSI were shown to be more prevalent in hot weather, during summer time [25, 26]. Accordingly, in our study, a high rate of both intra-abdominal and urinary focus of BSI was seen in that time of the year as well as an increase in the isolation of Gram-negative BSI. In our study the presence of a BSI in CAS patients at hospital admission was associated with a worse prognosis. However, the cumulative mortality of patients with and without BSI was not significantly different till day 8 of ICU stay. But, after day 9 onwards BSI patients present a significant excess mortality (figure 2). The present data suggests that the excess mortality in CAS patients with BSI occur later in clinical course. Nevertheless even at day 4, patients with BSI who were still alive in the ICU were more often dependent of vasopressors, probably reflecting a high clinical severity and or a slower resolution of infection (Table 2). This dependence of vasopressors had been classified elsewhere as response failure and shown to be associated with increased mortality [23]. Despite the fact that both temperature and CRP were significantly different at ICU admission, it was evident the substantial overlap of values between patients with and without BSI (figure 1), precluding their use to diagnose patients with BSI. However, daily monitoring of CRP concentration may prove to be useful. A recent study from our study group clearly showed that the CRP trend help to identify CAS patients not responding to antibiotic therapy, who had an increase risk of dying, as soon as in the first days of antibiotic therapy [9]. Our study has several

important strengths. It is one of the largest multiple-centre epidemiologic studies evaluating BSI in critically ill patients with CAS. It prospectively evaluated patients admitted with CAS for a 12-month period, allowing the evaluation of effects related with seasonal variation. Patients were all followed until death or hospital discharged unveiling differences in cumulative mortality along the time. Data quality has been evaluated through an external audit randomly reviewing selected patients' protocols. However, we recognize that our study has limitations. Its nonrandomized, observational design may have induced some unknown bias in the treatment of patients with CAS especially since ICU admission criteria was defined by local policy and not by protocol. Furthermore, we have no data concerning CAS not admitted to ICU, also an important problem in the process of care.

Conclusion

In CAS patients receiving early appropriate antibiotic therapy, BSI at hospital admission was independently associated with late resolution of sepsis and increased ICU and in-hospital mortality. This difference in mortality was only noted after day nine of ICU admission. There was a seasonal variation in the agents responsible for BSI, which might have implications in the initial antibiotic prescription.

Anexo II

NIVEL 1 - Caracterização dos doentes	
I- TODOS OS DOENTES ADMITIDOS NA UCI:	
Data da primeira inserção do registo:	2005-02-17
Identificação do Registo:	ss / ss ⓘ
Numeração do Registo:	2005 . 01 . 01 ⓘ <small>(ano-mes-nº único)</small>
Data de Nascimento: <input type="text"/> / <input type="text"/> / <input type="text"/>	Sexo: <input type="text" value="Feminino"/>
Idade: <input type="text"/> ⓘ	
Data de Admissão no Hospital:	<input type="text" value="11"/> / <input type="text" value="11"/> / <input type="text" value="2000"/> (dd / mm / yyyy) (horas) <input type="text"/> : <input type="text"/> Hora Desconhecida
Categoria do Diagnóstico de Admissão na UCI:	<input type="text"/>
Data de Admissão na UCI:	<input type="text" value="11"/> / <input type="text" value="11"/> / <input type="text" value="2000"/> (dd / mm / yyyy) (horas) <input type="text"/> : <input type="text"/> Hora Desconhecida
Data de alta na UCI:	<input type="text"/> / <input type="text"/> / <input type="text"/> (dd / mm / yyyy)
Resultado do Internamento na UCI:	<input type="radio"/> Alta <input type="radio"/> Falecido
Data de Alta Hospitalar:	<input type="text"/> / <input type="text"/> / <input type="text"/> (dd / mm / yyyy)

Resultado do Internamento no Hospital:	<input type="radio"/> Alta <input type="radio"/> Falecido
Resultado do Internamento Hospitalar se doente transferido para outra Instituição Hospitalar:	<input type="radio"/> Alta <input type="radio"/> Falecido <input type="radio"/> Desconhecido <input type="radio"/> Não Aplicável
A estadia do doente foi superior a 6h?	<input type="radio"/> Não <input type="radio"/> Sim
SAPS: <input type="text" value="11"/> 	

O doente tem sepsis adquirida na comunidade? 

Não
 Sim

O doente tem uma infecção associada a cuidados de saúde? 

Não
 Sim

Se sim, indique a causa para o doente ter uma infecção associada a cuidados de saúde: 

O doente reúne critérios de SEPSIS à data de admissão na UCI? 

Não
 Sim

NÍVEL 1- Caracterização do problema da Sepsis

II- TODOS OS DOENTES ADMITIDOS NA UCI POR SEPSIS ADQUIRIDA NA COMUNIDADE:

Estratificação de gravidade da sepsis:

- Sepsis !
 Sepsis grave !
 Choque Séptico !

Origem da Sepsis:

Fez colheita de produtos para estudo microbiológico antes de iniciar a antibioterapia?

- Não
 Sim
 Desconhecido

Hemoculturas:

- Colhidas?**
 Não
 Sim
- Positivas?**
 Não
 Sim
- Data: / /
 Hora: : Hora Desconhecida
 Desconhecido

Micróbios Isolados:

Local de Isolamento:
 Local de Isolamento:
 Local de Isolamento:
 Local de Isolamento:
 Local de Isolamento:

SIRS/SEPSIS

NÍVEL 2

III. FACTORES PREDISPONETES

Situações médicas pré existentes:	Sim	Não	Desconhecido
! Corticoterapia de longa duração:	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
! Corticoterapia de curta duração:	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
! Quimioterapia:	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
! Radioterapia :	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
! Seropositivo para o VIH:	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
! SIDA :	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Doença crónica pré existente:	Sim	Não	Desconhecido
! Hepática:	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
! Renal:	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
! Córdio-Circulatória:	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
! Respiratória:	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
! Hematológica:	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
! Neoplásica:	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>

Faltas

H₁ e U₀ coincidem?
 Não
 Sim

H₁, U₀ e U₁ coincidem?
 Não
 Sim

Gravar

Dado que o período de internamento é de 0 dias apenas, não deve preencher a(s) coluna(s) dos dias seguintes.

	Unidades	H ₁	U ₁	U ₂	U ₃	U ₄	U ₅
Temperatura °C		<input type="radio"/> Desc					
Hipotermia °C		<input type="radio"/> Desc <input type="radio"/> Não					
Frequência Cardíaca (BPM)		<input type="radio"/> Desc					
Taquipneia / min		<input type="radio"/> Desc <input type="radio"/> Vent					
Glicemia mg/dL	mg/dL	<input type="radio"/> Desc					
Hipoglicemia		<input type="radio"/> Desc					
Leucocitose x10 ³ /mL		<input type="radio"/> Desc					
% de Neutrófilos		<input type="radio"/> Desc					
Leucopenia x10 ³ /mL		<input type="radio"/> Desc <input type="radio"/> Não					
% de Leucócitos imaturos		<input type="radio"/> Desc					
Proteína C reactiva no plasma mg/dL	mg/dL	<input type="radio"/> Desc					
Lactato sérico mmol/L	mmol/L	<input type="radio"/> Desc					

Limite Inferior do Valor Normal de Glicemia na Instituição mg/dL

Data e Hora de obtenção do 1º valor de lactato na Instituição / / : : Hora Desconhecida

H₁ e U₀ coincidem?
 Não
 Sim

H₁, U₀ e U₁ coincidem?
 Não
 Sim

Gravar

Dado que o período de internamento é de 0 dias apenas. Não deve preencher a(s) coluna(s) dos dias seguintes.

	Unidades	H ₁	U ₁	U ₂	U ₃	U ₄	U ₅
SOFA respiratório PaO ₂ /FIO ₂		<input type="radio"/> Desc					
Ventilação mecânica		<input type="radio"/> Desc					
SOFA cardio-circulatório		<input type="radio"/> Desc					
Dopamina		<input type="radio"/> Desc					
Noradrenalina		<input type="radio"/> Desc					
Dobutamina		<input type="radio"/> Desc					
Sedado		<input type="radio"/> Desc					
SOFA Neurológico-Glasgow (Verificado)		<input type="radio"/> Desc <input type="radio"/> Não					
SOFA Neurológico-Glasgow (Estimado)		<input type="radio"/> Desc <input type="radio"/> Não					
SOFA renal-Creatinina sérica mg/dL	mg/dL	<input type="radio"/> Desc					
Em diálise		<input type="radio"/> Desc					
Em hemofiltração		<input type="radio"/> Desc					
SOFA hepático - bilirubina sérica mg/dL		<input type="radio"/> Desc					
SOFA hematológico plaquetas x 10 ³ /L		<input type="radio"/> Desc					

Houve tratamento Antimicrobiano antes deste episódio de hospitalização no Hospital onde está sediada a UCI?

Não
 Sim

Se sim indique a respectiva Data e Hora desse tratamento:

/ / : :

Desconhecido Data Desconhecida Hora Desconhecida

H₁ e U₀ coincidem?
 Não
 Sim

H₁, U₀ e U₁ coincidem?
 Não
 Sim

Tratamento Antimicrobiano pré UCI

H1

U0

Indique a Data e hora do primeiro tratamento antimicrobiano neste Hospital:

/ / : : Hora Desconhecida

Tratamento Antimicrobiano à admissão na UCI

U1

Mudança de tratamento:

Motivo da Mudança de Tratamento:

Novo Tratamento:

SIRS/SEPSIS

NÍVEL 2

V. TERAPÉUTICA ANTIMICROBIANA

Quadro II

Tratamento Antimicrobiano à admissão na UCI

U2	U3	U4	U5
<input type="button" value="Escolher"/> <input type="text"/>	<input type="button" value="Escolher"/> <input type="text"/>	<input type="button" value="Escolher"/> <input type="text"/>	<input type="button" value="Escolher"/> <input type="text"/>
Mudança de tratamento: <input type="checkbox"/>			
Motivo da Mudança de Tratamento: <input type="text"/>			
Novo Tratamento:			
<input type="button" value="Escolher"/> <input type="text"/>	<input type="button" value="Escolher"/> <input type="text"/>	<input type="button" value="Escolher"/> <input type="text"/>	<input type="button" value="Escolher"/> <input type="text"/>
Duração total do tratamento antimicrobiano deste episódio de sépsis: <input type="text"/>			

NÍVEL 3 - Participação na dinâmica Surviving sepsis Campaign: incentivo à aplicação das recomendações da "SSC" (doentes com sepsis adquirida na comunidade)

VIII. REGISTO DE PROCEDIMENTOS RECOMENDADOS PELA "SURVIVING SEPSIS CAMPAIGN (SSC)" PARA O TRATAMENTO DA SEPSIS

Nível de Penetração das recomendações da SSC

As questões a cinzento já foram respondidas anteriormente pelo que não podem ser alteradas aqui.

Questão	Sim	Não	Desc.	N/A
1. Foi feito o diagnóstico de presunção nas primeiras 6h de contacto com esta instituição	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
2. Foi doseado o lactato sérico, nas primeiras 6h a seguir à hospitalização	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
3. No caso de não ter sido feito o doseamento do lactato sérico nas primeiras 6h após contacto com esta instituição foi identificada a presença de acidose metabólica mediante caracterização do "anion gap" e gasometria	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
4. Em caso de hipotensão nas primeiras 6h foram administrados fluidos (500-1000ml de cristalóides ou 300-500ml de colóides em 30min, a repetir se não houver sinais de sobrecarga de volume) i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
5. Foram administrados antimicrobianos na hora que se seguiu ao diagnóstico de presunção de sépsis i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
6. Foram administrados vasopressores durante e a seguir à administração de fluidos sempre que a PAM <65mmHg, durante as primeiras 6 horas após contacto com esta instituição	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Foi medida a pressão venosa central nas primeiras 6h a seguir ao diagnóstico de sépsis i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Foi medida a pressão venosa central nas primeiras 24h a seguir ao diagnóstico de sépsis	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
9. Foi medida a pressão venosa central durante o internamento na UCI	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
10. Foi medida a SvO ₂ nas primeiras 6h a seguir ao diagnóstico de sépsis i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
11. Foi medida a SvO ₂ nas primeiras 24h a seguir ao diagnóstico de sépsis i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
12. Foi medida a SvO ₂ durante o internamento na UCI	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
13. Foram administrados inotrópicos positivos, depois das primeiras 6h após suspeita do diagnóstico da sépsis i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Foram transfundidos concentrados de glóbulos rubros com a intenção de manter o hematócrito >38 nas primeiras 6h de tratamento da sépsis i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Fez colheitas de produtos para estudo microbiológico antes de iniciar a antibioterapia i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
16. Fez administração de antimicrobianos ajustados ao diagnóstico? i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
17. O tratamento anti-infeccioso foi ajustado com orientação microbiológica? i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Fez tratamento antimicrobiano mais de 10 dias i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
19. Usou Corticóides para tratamento do choque i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. Usou proteína C activada (drotrectogina) i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Fez transfusão de concentrados de glóbulos rubros depois do período inicial de reanimação	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
22. A transfusão de concentrados de glóbulos rubros foi de acordo com as recomendações da SSC i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
23. Fez transfusão plasma	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
24. A transfusão plasma foi de acordo com as recomendações SSC i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
25. Fez transfusão de concentrado de plaquetas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
26. A transfusão de concentrados de plaquetas foi de acordo com as recomendações SSC i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
27. Nos doentes ventilados a ventilação foi programada com limite da pressão de planalto i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28. Nos doentes ventilados a ventilação foi programada com volume corrente baixo i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29. Nos doentes ventilados com ARDS foi utilizada PEEP (10-16cmH ₂ O na fase exsudativa) i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30. Foi feito posicionamento em decúbito ventral ("prone position") i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31. A sedação / analgesia é feita com protocolo com redução periódica da profundidade da sedação	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
32. O doente foi curarizado i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
33. Manteve a glicemia <150 mg/dl depois da estabilização inicial i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
34. Fez diálise	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
35. Fez hemofiltração i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
36. Fez prevenção de trombose venosa profunda i	CI <input type="checkbox"/> *	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
37. Fez profilaxia de úlceras de stress i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

* Contra-Indicada