Universidade Estadual de Maringá

Departamento de Estatística



Vinicius Garcia Tosta

Modelo conjunto longitudinal e logístico para análise de dados de pneumonia associada à ventilação mecânica

Maringá – Paraná 30/04/2019 Vinicius Garcia Tosta

Modelo conjunto longitudinal e logístico para análise de dados de pneumonia associada à ventilação mecânica

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá como requisito parcial para obtenção do título de Mestre em Bioestatística.

Orientadora: Prof.^a Dr.^a Isolde Previdelli

Universidade Estadual de Maringá - UEM Departamento de Estatística - DEs Programa de Pós-Graduação em Bioestatística

> Maringá – Paraná 30/04/2019

Dados Internacionais de Catalogação na Publicação (CIP) (Biblioteca Central - UEM, Maringá, PR, Brasil)

Tosta, Vinicius Garcia T716m Modelo conjunto longitudinal e logístico para dados de pneumonia associada à ventilação mecânica / Vinicius Garcia Tosta. -- Maringá, 2019. 36 f. : il., figs., tabs. Orientadora: Profa. Dra. Isolde Previdelli. Dissertação (mestrado) - Universidade Estadual de Maringá, Departamento de Estatística, Programa de Pós-Graduação em Bioestatística, 2019. 1. Pneumonia associada à ventilação mecânica (PVA). 2. Modelo conjunto longitudinal e logístico. 3. Biomarcador. 4. Pressão diastólica. I. Previdelli, Isolde, orient. II. Universidade Estadual de Maringá. Departamento de Estatística. Programa de Pós-Graduação em Bioestatística. III. Título. CDD 23.ed.570.15195 Elaine Cristina Soares Lira - CRB 1202/9

VINICIUS GARCIA TOSTA

Modelo conjunto longitudinal e logístico para análise de dados de

pneumonia associada à ventilação mecânica

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

BANCA EXAMINADORA

Prof^a. Dra. Isolde Previdelli

Universidade Estadual de Maringá – PBE/UEM

Prof. Dr. Josmar Mazucheli Universidade Estadual de Maringá - PBE/UEM Prof. Dr. Omar Cléo Neves Pereira

Universidade Estadual de Maringá - DES/UEM

Maringá, 31 de maio de 2019.

Agradecimentos

À Universidade Estadual de Maringá que, através do Programa de Pós-Graduação em Bioestatística, abriu as portas para mim e me proporcionou as melhores condições para que eu pudesse realizar o trabalho;

À minha mãe, Áurea Garcia Tosta, uma pessoa absolutamente diferenciada e que sempre me deu forças nos momentos difíceis com seus valiosos conselhos e me apoiou em todas as circunstâncias.

À Cristina Togashi, uma pessoa igualmente diferenciada, de um caráter irretocável, e que também possibilitou que eu conseguisse alcançar o sucesso.

À minha orientadora, prof^a Dr^a Isolde Previdelli, por ser uma das fomentadoras desse projeto, pelas suas orientações e por sua paciência.

Ao prof. Dr. Josmar Mazucheli, membro da banca, que, com suas críticas fortemente construtivas, discussões ao longo do mestrado e também a disponibilização dos códigos em SAS para o ajuste do modelo, fez o trabalho subir de patamar.

Ao Omar Pereira, membro da banca, e ao Emerson Barili, pessoas da melhor qualidade, que sempre me incentivaram, mesmo nos momentos de grandes dificuldades.

À Beatriz Gurgel e seu pai, Sander Gurgel, os grandes responsáveis pela fomentação e execução desse projeto, tanto em termos de planejamento quanto de coleta de dados. Não fossem eles, essa dissertação não passaria de uma mera ideia. Também devo muito ao Samuel, aluno de iniciação científica, que me ajudou bastante na parte de manipulação de dados e no cálculo dos escores, facilitando bastante minha vida.

À todos os amigos da minha turma, Naiara, Talita, Jéssica, Patrícia, Everton, Roney, Mônica e da turma anterior, Jean e Aline, que, sem dúvida alguma, melhoraram o ambiente de forma infinita. Se eu infelizmente esqueci de alguém, por favor, sinta-se agradecido.

À prof^a Dr^a Eniuce que, num momento de extrema dificuldade que enfrentei, me ajudou de uma forma esplendorosa.

Ao secretário do PBE, Daniel Sbrana, por ter colaborado fortemente no que se refere a manter os alunos informados, cientes de suas obrigações junto ao mestrado. Não fosse ele, com certeza, não me tornaria mestre em bioestatística.

Gostaria de fazer um agradecimento especial à Antônia Menck, a dona da pensão onde estou morando. Uma pessoa simplesmente formidável, de um caráter inigualável e que não se encontra tão facilmente. Tornou minha vida muito mais fácil e agradabilíssima nesses três anos. Queria agradecer também seu marido Henrique, um homem trabalhador, honesto, íntegro, que sempre me tratou com total respeito. Jamais poderia me esquecer também dos meus grandes amigos caninos. Thor(Lhasa + Poodle), Julie e Billy (pinschers) e, o Rex, um Basset que parecia ser imortal, mas que nos deixou no alto de seus 18 anos. Nos momentos de grande estresse que passei, os amigos de quatro patas sempre me passaram uma forte sensação de que tudo daria certo, não sabia exatamente de que modo, mas daria certo. Sou extremamente grato a todos que fizeram parte da minha convivência nessa famosa pensão Menck, construída em 1945, e que resiste bravamente aos efeitos do tempo, com algumas reformas é verdade, mas com sua espinha dorsal mantida, graças aos cuidados valiosos da dona Antônia.

À CAPES, pelo suporte financeiro.

E a todos que, direta ou indiretamente, fizeram parte da minha formação.

Resumo

A Pneumonia Associada à Ventilação Mecânica (PAV) é uma complicação bastante séria nas UTIs, a qual possui altos índices de mortalidade e seu diagnóstico é bastante difícil. Pelo fato de não existir um padrão ouro para a PAV, diversos trabalhos têm sido realizados com o intuito de melhorar a acurácia do diagnóstico de tal complicação. Sob esse contexto, o presente trabalho objetivou encontrar biomarcadores cujo comportamento ao longo do tempo pudesse sinalizar que a PAV está se desenvolvendo em um indivíduo e, para isso, utilizou-se o modelo conjunto longitudinal e logístico. A principal característica do mesmo é o fato do submodelo logístico incorporar a informação do biomarcador do processo longitudinal proveniente dos efeitos aleatórios. Os parâmetros do modelo conjunto foram estimados por intermédio do método de máxima verossimilhança mediante o procedimento NLMIXED, do software SAS. Dentre os 32 biomarcadores considerados no estudo, somente para um deles a suposição de normalidade dos efeitos aleatórios foi satisfeita, houve uma conexão relevante entre o processo longitudinal e logístico, as medidas preditivas sensibilidade, especificidade e área sob a curva foram satisfatórias e os resíduos do submodelo misto tiveram um bom comportamento: pressão diastólica. Como resultado para o modelo conjunto ajustado referente a tal biomarcador, encontrou-se que, em média, a cada redução de uma unidade na média da pressão diastólica, ocorre um aumento de aproximadamente 14% na chance de o indivíduo desenvolver PAV. Finalmente, evidenciou-se que este modelo é adequado através do gráfico do envelope, tanto normal como semi-normal, desenvolvido com base nos resíduos de Anscombe Estudentizados, em que todos os pontos se encontravam na região delimitada pelos quantis.

Palavras-chave: Pneumonia associada à ventilação mecânica. Modelo conjunto longitudinal e logístico. Biomarcador. Pressão diastólica.

Abstract

Ventilator-Associated Pneumonia (VAP) is a very serious complication in ICUs, which has high mortality rates and its diagnosis is very difficult. Because there is no gold standard for VAP, several studies have been performed to improve the accuracy of the diagnosis of such complication. In this context, the present work aimed to find biomarkers whose behavior over time could signal that the VAP is developing in an individual and, for that, the joint model longitudinal and logistic was used. Its main feature is that the logistic submodel incorporates the longitudinal process biomarker information from random effects. The parameters of the joint model were estimated using the maximum likelihood method using the NLMIXED procedure of the SAS software. Among the 32 biomarkers considered in the study, for only one of them the assumption of normality of random effects was satisfied, there was a relevant connection between the longitudinal and logistic process, the predictive measures sensitivity, specificity and area under the curve were satisfactory and the residuals of the mixed submodels performed well: diastolic pressure. As a result of the adjusted joint model for this biomarker, it was found that, on average, with each reduction of one unit in the mean diastolic pressure, there is an approximately 14 % increase in the individual's chance to develop PAV. Finally, it was evidenced that this model is adequate through the normal and half-normal envelope graph developed based on the Anscombe Studentized residuals, where all points were in the region delimited by the quantiles.

Keywords Mechanical Ventilation Associated Pneumonia. Joint model longitudinal and logistic. Biomarker. Diastolic pressure.

Lista de ilustrações

Figura 1 –	Fluxograma da exclusão dos pacientes	23
Figura 2 –	Fluxograma da exclusão/adição dos biomarcadores	25
Figura 3 –	Gráfico dos resíduos para os modelos conjuntos estimados em relação	
	aos biomarcadores pressão diastólica e escala de <i>Glasgow</i>	29
Figura 4 –	Trajetórias longitudinais para a pressão diastólica segundo o diagnóstico	
	de PAV	30
Figura 5 –	Gráficos do envelope para o modelo conjunto referente ao biomarcador	
	pressão diastólica	31
Figura 6 –	Curva ROC para o modelo conjunto referente à pressão diastólica	32
Figura 7 –	Curva de Kaplan-Meier para os dados da PAV	34

Lista de tabelas

Tabela 1 –	Significado dos biomarcadores	24
Tabela 2 –	Quantidade e proporção de dados faltantes, média amostral, desvio pa-	
	drão amostral e coeficiente de variação de acordo com o diagnóstico de	
	PAV para cada biomarcador	26
Tabela 3 –	Resultados para os modelos ajustados	28
Tabela 4 –	Estimativas dos parâmetros para o modelo conjunto referente à pressão	
	diastólica	29

Sumário

1	Mod	elo conjunto longitudinal e logístico para dados de pneumonia associ-
	ada a	à ventilação mecânica
	1.1	Introdução
	1.2	Planejamento do estudo
	1.3	Modelo conjunto longitudinal e logístico
	1.4	Resultados
	1.5	Conclusão
	1.6	Limitações
	1.7	Recomendações 34
Re	ferên	cias

\sim		· .	1		-
ເລ	n	111			
Ja	JU.	ιι	u	UU.	

Resumo

A Pneumonia Associada à Ventilação Mecânica (PAV) é uma complicação bastante séria nas UTIs, a qual possui altos índices de mortalidade e seu diagnóstico é bastante difícil. Pelo fato de não existir um padrão ouro para a PAV, diversos trabalhos têm sido realizados com o intuito de melhorar a acurácia do diagnóstico de tal complicação. Sob esse contexto, o presente trabalho objetivou encontrar biomarcadores cujo comportamento ao longo do tempo pudesse sinalizar que a PAV está se desenvolvendo em um indivíduo e, para isso, utilizou-se o modelo conjunto longitudinal e logístico. A principal característica do mesmo é o fato do submodelo logístico incorporar a informação do biomarcador do processo longitudinal proveniente dos efeitos aleatórios. Os parâmetros do modelo conjunto foram estimados por intermédio do método de máxima verossimilhança mediante o procedimento NLMIXED, do software SAS. Dentre os 32 biomarcadores considerados no estudo, somente para um deles a suposição de normalidade dos efeitos aleatórios foi satisfeita, houve uma conexão relevante entre o processo longitudinal e logístico, as medidas preditivas sensibilidade, especificidade e área sob a curva foram satisfatórias e os resíduos do submodelo misto tiveram um bom comportamento: pressão diastólica. Como resultado para o modelo conjunto ajustado referente a tal biomarcador, encontrou-se que, em média, a cada redução de uma unidade na média da pressão diastólica, ocorre um aumento de aproximadamente 14% na chance de o indivíduo desenvolver PAV. Finalmente, evidenciou-se que este modelo é adequado através do gráfico do envelope, tanto normal

como semi-normal, desenvolvido com base nos resíduos de Anscombe Estudentizados, em que todos os pontos se encontram na região delimitada pelos quantis.

Palavras-chave: Pneumonia associada à ventilação mecânica. Modelo conjunto Longitudinal e Logístico. Biomarcador. Pressão diastólica.

Abstract

Ventilator-Associated Pneumonia (VAP) is a very serious complication in ICUs, which has high mortality rates and its diagnosis is very difficult. Because there is no gold standard for VAP, several studies have been performed to improve the accuracy of the diagnosis of such complication. In this context, the present work aimed to find biomarkers whose behavior over time could signal that the VAP is developing in an individual and, for that, the longitudinal and logistic joint model was used. Its main feature is that the logistic submodel incorporates the longitudinal process biomarker information from random effects. The parameters of the joint model were estimated using the maximum likelihood method using the NLMIXED procedure of the SAS software. Among the 32 biomarkers considered in the study, for only one of them the assumption of normality of random effects was satisfied, there was a relevant connection between the longitudinal and logistic process, the predictive measures sensitivity, specificity and area under the curve were satisfactory and the residuals of the Mixed submodels performed well: diastolic pressure. As a result of the adjusted joint model for this biomarker, it was found that, on average, with each reduction of one unit in the mean diastolic pressure, there is an approximately 14 % increase in the individual's chance to develop VAP. Finally, it was evidenced that this model is adequate through the normal and half-normal envelope graph developed based on the Anscombe Studentizated residuals, where all points were in the region delimited by the quantiles.

Keywords Mechanical Ventilation Associated Pneumonia. Longitudinal and Logistic joint model. Biomarker. Diastolic pressure.

1.1 Introdução

Nas últimas décadas, o uso do suporte ventilatório invasivo, indubitavelmente, foi um avanço no tratamento da insuficiência respiratória. De acordo com Carvalho (2006), apesar de salvar muitas vidas, a aplicação de uma pressão positiva nos pulmões, através de uma prótese colocada nas vias aéreas, pode gerar uma série de efeitos adversos, tais como: a instabilidade hemodinâmica, principalmente nos pacientes hipovolêmicos, isto é, aqueles que apresentam uma redução no volume sanguíneo; uma maior freqüência de infecções respiratórias, devido à redução dos mecanismos de defesa locais pela presença do tubo; e, mais recentemente, a lesão induzida pela ventilação mecânica, que está associada às altas pressões alveolares que são atingidas em algumas unidades decorrentes de um pulmão doente, bastante heterogêneo. Sob esse contexto, a pneumonia associada à ventilação mecânica (PAV) encontra-se como um dos efeitos adversos mais temíveis no ambiente da terapia intensiva.

Segundo Kalanuria, Zai e Mirski (2014) (KALANURIA; ZAI; MIRSKI, 2014), a PAV é definida como sendo a pneumonia que ocorre 48-72 horas após a intubação endotraqueal, caracterizada pela presença de infiltrado progressivo ou novo (substância mais densa que o ar, por exemplo, pus, edema, sangue, proteínas ou células), sinais de infecção sistêmica (febre, contagem alterada de glóbulos brancos), alterações nas características de expectoração e detecção de um agente causador. De acordo com Chastre e Fagon (2002), tal complicação ocorre em 8 a 28% dos pacientes em ventilação mecânica e, em contraste com infecções de órgãos mais frequentemente envolvidos (por exemplo, trato urinário e pele), cuja mortalidade é baixa, variando de 1 a 4%, a taxa de mortalidade da PAV varia de 24 a 50% e pode chegar a 76% em alguns casos específicos ou quando a infecção pulmonar é causada por patógenos de alto risco.

Conforme as diretrizes da Sociedade Paulista de Infectologia (2006), a PAV é a segunda infecção mais frequente em UTIs americanas e a mais frequente em UTIs europeias. No Brasil, ainda que haja uma ausência de dados nacionais e multicêntricos, evidências clínicas mostram a PAV como sendo a infecção mais frequente dentro das UTIs.

Além de ser bastante séria, a PAV também se apresenta como uma complicação de difícil diagnóstico. De acordo com Dalmora et al. (2013), não existe um padrão ouro para se identificar seguramente a ocorrência da PAV. Klompas (2007), em uma revisão sistemática de 14 estudos que avaliaram os critérios clínicos e microbiológicos para a PAV com os resultados da autópsia, observou uma variação importante no grau de concordância de 23 a 92%. A sensibilidade e a especificidade também sofreram forte variação de acordo com os critérios utilizados nos estudos e, isoladamente, tiveram acurácia bastante baixa, além do fato de que até 50% dos pacientes que preenchiam critérios para a PAV tinham diagnóstico alternativo.

Diante do contexto apresentado, o objetivo desse estudo consiste em identificar biomarcadores cujo comportamento ao longo do tempo influencie de algum modo a chance do paciente desenvolver a PAV durante o período de internação. Conforme Group et al. (2001), o Programa Internacional de Segurança Química, liderado pela Organização Mundial da Saúde (OMS) e em coordenação com as Nações Unidas e a Organização Internacional do Trabalho, definiu um biomarcador como "gualguer substância, estrutura ou processo que pode ser medido no corpo ou em seus produtos e influenciar ou prever a incidência de resultado ou doença". No caso da PAV, de acordo com Balthazar et al. (2001), os critérios clínicos usuais se baseiam em biomarcadores tais como temperatura (avaliação da ocorrência de febre), contagem de leucócitos (avaliação da leucocitose, caracterizada pelo aumento no número de glóbulos brancos) e secreções traqueais purulentas ou opacidades radiográficas pulmonares (avaliação do patógeno causador da PAV). Todavia, muitos pacientes em em ventilação mecânica tem doenças graves de base, aumento da colonização da orofaringe e vários motivos para apresentar febre e leucocitose, o que reduz drasticamente a acurácia de tais critérios para se diagnosticar a PAV. Isso reforça a necessidade de realização do presente estudo para se avaliar a possibilidade de que outros biomarcadores estejam sinalizando com maior precisão e rapidez a ocorrência de tal complicação.

Uma abordagem que se faz bastante válida para se alcançar o objetivo proposto é o modelo conjunto longitudinal e logístico, em que os precursores de tal metodologia foram Wang, Wang e Wang (2000), os quais desenvolveram um modelo para avaliar o impacto de medidas longitudinais na variável resposta de interesse, que no caso, assume uma natureza dicotômica. Como aplicação, ajustaram um modelo conjunto relacionado a dados de obesidade em crianças. Seguindo essa linha, Horrocks, Heuvel et al. (2009) mostraram que medidas longitudinais relacionadas à hiperestimulação ovariana controlada aumentaram o poder de predição do *status* de gravidez das pacientes. Já Zhang, Chen e Zou (2014), utilizando abordagem semelhante, avaliaram o quanto que mudanças nas escalas de estresse conjugal ao longo do tempo poderiam levar mulheres mais velhas ao transtorno de depressão maior. A diferença em relação ao trabalho anterior é que além do modelo para dados longitudinais e binários, modelou-se também o mecanismo de dados faltantes que ocorreram nos escores de estresse conjugal.

Diante do exposto, o modelo conjunto longitudinal e logístico se fez bastante pertinente para se avaliar a ocorrência ou não de PAV em função do comportamento dos biomarcadores, isto é, analisar se cada um dos biomarcadores presentes no estudo podem predizer de modo satisfatório a chance de um paciente desenvolver a PAV.

O trabalho está organizado do seguinte modo: primeiramente, é apresentado o planejamento do estudo da PAV, na seção 1.2. Posteriormente, na seção 1.3 apresenta-se a parte teórica do modelo conjunto longitudinal e logístico, tanto em relação à estrutura do modelo quanto à estimação dos parâmetros. A seção 1.4 apresenta o processamento dos dados tanto em termos de pacientes quanto de biomarcadores, bem como uma breve análise descritiva e, finalmente, os resultados referentes ao ajuste do modelo conjunto. A conclusão, limitações e recomendações se encontram nas seções 1.5, 1.6 e 1.7, respectivamente.

1.2 Planejamento do estudo

Essa pesquisa é resultado de um projeto interdisciplinar desenvolvido entre o Programa de Pós-Graduação em Bioestatística (PBE), Departamento de Estatística e Hospital Universitário Regional de Maringá, pela prof^a Isolde Previdelli, em parceria com o médico responsável pela Unidade de Tratamento Intensivo, Sander Gurgel, e os graduandos em medicina e estatística, Beatriz Gurgel e Samuel Vianna, respectivamente.

O tipo de estudo para essa pesquisa é "tempo de duração fixado", em que não é possível estabelecer o total de participantes, uma vez que depende da entrada dos pacientes na UTI com características relacionadas à PAV. A coleta dos dados ocorreu no período de dezembro de 2017 a maio de 2018.

Após a aprovação pelo Comitê de Ética local (Universidade Estadual Maringá -UEM), prontuários médicos de pacientes internados na UTI do Hospital Universitário Regional de Maringá foram coletados diariamente por um único pesquisador e de maneira longitudinal variáveis da história médica, fisiológicas e laboratoriais.

A UTI do HURM tem oito leitos com capacidade de monitorização 24 horas/por dia. Pacientes internados por mais de 24 horas no período de junho de 2017 até dezembro de 2017 foram incluídos no protocolo. Pacientes com menos de 24 horas de internação foram excluídos.

As variáveis da história médica que foram levantadas são as comorbidades do paciente (doença pulmonar obstrutiva crônica, insuficiência cardíaca, arritmia cardíaca, doença isquêmica do coração, síndrome de apnéia do sono, asma, doenças neuromusculares e da parede torácica, infecções prévias, neoplasia, sepse, AVC, diabetes, e outras doenças endócrinas). As variáveis fisiológicas são os parâmetros vitais (frequência respiratória, pulso, pressão arterial média e temperatura axilar), parâmetros de oxigenação (relação PaO_2/FIO_2 , PCO_2 , $AaDO_2$, pH arterial, ventilação mecânica e ventilação não invasiva) e fatores fisiológicos agudos (uso de vasopressor/inotrópico, SARA, diálise, débito urinário, balanço hídrico, dieta e escala de Glasgow). As variáveis laboratoriais são hemograma, culturas, bilirrubinas, creatinina, ureia, potássio, sódio, bicarbonato, albumina e glicose. Também foram calculados os escores *APACHE II, SAPS II, CPIS* E *SOFA* (consultar <<u>https://www.mdcalc.com></u> para mais informações sobre os escores).

Dados demográficos, diagnóstico de admissão dos pacientes, duração de ventila-

ção mecânica, tempo de internação na UTI e hospitalar total, sobreviventes, não sobreviventes, tempo de desenvolvimento de pneumonia associada à ventilação mecânica (PAV), patógenos responsáveis, padrão de cultura e antibiograma foram obtidos e classificados dentro do estudo.

Durante o período de estudo, todos os dados foram coletados prospectivamente e os pacientes foram acompanhados até a alta hospitalar ou óbito. Foi coletado por um único pesquisador, diariamente, o aparecimento das seguintes complicações: infecções pós-operatórias na UTI (pneumonia, sepse abdominal e urinária bem como infecções de ferida operatória); complicações respiratórias (embolia pulmonar, lesão pulmonar aguda e suporte respiratório > 24 horas devido a lesão pulmonar aguda); complicações cardio-vasculares (arritmia, hipotensão, edema pulmonar agudo, infarto agudo do miocárdio, acidente vascular cerebral, parada cardíaca com desfecho fatal); complicações abdominais (diarreia por *Clostridium difficile*, obstrução intestinal aguda, hemorragia digestiva alta e fistulas); complicações hematológicas (plaquetopenia <100.000 ou tempo de protrombina >1,5 vezes o controle) e complicações renais (debito urinário< 500ml ou creatinina sérica >170 mumol/l ou diálise por insuficiência renal aguda).

Nesse estudo, a PAV foi diagnosticada quando o paciente, em ventilação mecânica por mais de 48 horas, apresentasse infiltrado pulmonar novo, associado à febre (temperatura axilar acima de 38,3 °C) e/ou leucocitose (leucócitos totais acima de 10000/mm³) e/ou leucopenia (leucócitos totais abaixo de 4000/mm³) e/ou secreção purulenta no aspirado traqueal com cultura positiva. Dados de cultura foram obtidos conforme rotina própria do serviço.

1.3 Modelo conjunto longitudinal e logístico

O modelo conjunto proposto, baseado em Horrocks, Heuvel et al. (2009), foi utilizado para analisar a estrutura e a intensidade da associação entre as medidas longitudinais dos biomarcadores e a probabilidade de o paciente ser diagnosticado com PAV. O submodelo do processo longitudinal foi definido em termos de um modelo de efeitos mistos. A probabilidade do paciente ser diagnosticado com PAV foi modelada por intermédio de uma regressão logística, de modo que esta incorpora a informação do biomarcador do processo longitudinal advinda dos efeitos aleatórios. O que se pretende avaliar é como os biomarcadores pode contribuir na predição da ocorrência da PAV. Considerando-se que $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{im_i})$ representa as medidas longitudinais contínuas para o indivíduo i, $i = 1, 2, \ldots, n$, nos tempos t_{i1}, \ldots, t_{im_i} , o modelo conjunto é expresso por:

$$\begin{cases} \boldsymbol{Y_i} &= \boldsymbol{X_i}^{(1)}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{\varepsilon_i} \\ \log\left(\frac{p_i}{1-p_i}\right) &= \boldsymbol{X_i}^{(2)}\boldsymbol{\alpha} + \boldsymbol{b_i}\boldsymbol{\gamma}, \end{cases}$$

em que β é o vetor $p \ge 1$ dos parâmetros de efeitos fixos, $X_i^{(1)} \in Z_i$ são as matrizes de planejamento $m_i \ge p \in m_i \ge q$, respectivamente, e b_i são os efeitos aleatórios. Para o submodelo logístico, $X_i^{(2)}$ representa a matriz de planejamento e α o vetor de parâmetros correspondente. É importante salientar que o vetor de parâmetros γ mede a força de associação entre os dois processos, isto é, longitudinal e logístico. Se $\gamma = 0$, então as medidas longitudinais não são providas de uma habilidade preditiva para o evento de interesse.

As seguintes suposições são assumidas:

$$\boldsymbol{b_i} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$$
 e $\boldsymbol{\varepsilon_i} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I_{m_i}})$

de modo que Σ é a matriz de covariâncias para os efeitos aleatórios e σ^2 é a variância residual.

Para a estimação dos parâmetros do modelo conjunto, foi utilizada a inferência frequentista e, assim sendo, deve-se definir a função de verossimilhança para o modelo conjunto. Seja θ o vetor contendo todos os parâmetros que devem ser estimados em relação aos dois processos e $L(\theta|\mathcal{D})$ a função de verossimilhança correspondente, em que $\mathcal{D} = \{y_i^{(1)}, y_i^{(2)}; i = 1, ..., n\}$ representa os dados observados, constituídos pelos componentes binário $y_i^{(2)}$ e longitudinal $y_i^{(1)} = (y_{i1}, ..., y_{ij}), j = 1, ..., m_i$. Desse modo, $L_i(\theta|\mathcal{D})$ denota a contribuição do *i*-ésimo indivíduo para a verossimilhança, $L(\theta|\mathcal{D})$, definida como

$$L_i(\boldsymbol{\theta}|\mathcal{D}) = L_{1i}(\boldsymbol{b}_i, \boldsymbol{\Sigma}, \boldsymbol{\beta}, \sigma^2 | \mathcal{D}) \times L_{2i}(\boldsymbol{b}_i, \boldsymbol{\alpha}, \gamma | \mathcal{D}),$$

em que

$$L_{1i}(\boldsymbol{b}_{i},\boldsymbol{\Sigma},\boldsymbol{\beta},\sigma^{2}|\boldsymbol{\mathcal{D}}) = (2\pi\sigma^{2})^{-m_{i}/2} \exp\left\{-\frac{||\boldsymbol{y}_{i}^{(1)}-\boldsymbol{X}_{i}\boldsymbol{\beta}-\boldsymbol{Z}_{i}\boldsymbol{b}_{i}||^{2}}{2\sigma^{2}}\right\}$$
$$\times (2\pi)^{-q_{b}/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left(-\boldsymbol{b}_{i}^{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}_{i}/2\right)$$

е

$$L_{2i}(\boldsymbol{b}_{i},\boldsymbol{\alpha},\boldsymbol{\gamma}) = p_{i}^{y_{i}^{(2)}} \times (1-p_{i})^{1-y_{i}^{(2)}}$$
$$= \left(\frac{\exp\left(\boldsymbol{X}_{i}^{(2)}\boldsymbol{\alpha} + \boldsymbol{b}_{i}\boldsymbol{\gamma}\right)}{\exp\left(\boldsymbol{X}_{i}^{(2)}\boldsymbol{\alpha} + \boldsymbol{b}_{i}\boldsymbol{\gamma}\right) + 1}\right)^{y_{i}^{(2)}} \times \left(1 - \left(\frac{\exp\left(\boldsymbol{X}_{i}^{(2)}\boldsymbol{\alpha} + \boldsymbol{b}_{i}\boldsymbol{\gamma}\right)}{\exp\left(\boldsymbol{X}_{i}^{(2)}\boldsymbol{\alpha} + \boldsymbol{b}_{i}\boldsymbol{\gamma}\right) + 1}\right)\right)^{1-y_{i}^{(2)}},$$

de modo que $||x|| = \{\sum_i x_i^2\}^{1/2}$ denota a norma do vetor Euclideano e q_b representa a dimensão dos efeitos aleatórios b_i . A parte da verossimilhança referente aos dados longitudinais, $L_{1i}(b_i, \Sigma, \beta, \sigma^2 | D)$, foi baseada em Rizopoulos (2012).

Assim sendo, deve-se encontrar as estimativas dos parâmetros que maximizem a função de verossimilhança. Entretanto, por ser provida de razoável complexidade, não é possível obter as estimativas dos parâmetros analiticamente. Assim, é necessário utilizar um método iterativo para se alcançar tal objetivo, como é descrito em Rizopoulos (2012).

Para se ajustar os modelos, utilizou-se o procedimento *NLMIXED* (SAS Institute Inc. (2015b)), disponível no *software SAS* e, para se obter os resíduos, utilizou-se o procedimento *IML* (SAS Institute Inc. (2015a)), também do *SAS*. Para se avaliar a adequabilidade do submodelo misto, utilizou-se os resíduos condicionais padronizados, expressos por:

$$oldsymbol{arepsilon}_i = oldsymbol{y}_i^{(1)} - oldsymbol{X} oldsymbol{\hat{eta}} - oldsymbol{Z} oldsymbol{ ilde{b}}_i$$

em que \tilde{b}_i denota os efeitos aleatórios preditos pelo modelo conjunto e $\hat{\beta}$ as estimativas do vetor de parâmetros β . Para o modelo conjunto, considerou-se os resíduos de Anscombe Estudentizados, e na sequência, visando-se verificar a qualidade do ajuste do mesmo, o gráfico do envelope (normal e semi-normal) foi desenvolvido no software *R* (R Core Team (2018)) e se encontra na seção 1.4.

$$r_{A_i} = \sqrt{m_i} \frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i)\sqrt{V(\hat{\mu}_i)}},$$
(1.1)

em que

$$A(\mu) = \int_{-\infty}^{\mu} V^{-1/3}(t) dt,$$
(1.2)

 m_i é o número de ensaios para a *i*-ésima observação, $\mu_i = m_i p_i$ para i = 1, ..., n e V é a função de variância correspondente. Para a distribuição binomial, $V(\mu_i) = \mu(1 - \mu_i)$ e, fazendo-se alguns cálculos envolvendo as expressões 1.1 e 1.2, tem-se:

$$r_{A_i} = \frac{\sqrt{m_i}(B(y_i, 2/3, 2/3) - B(\hat{\mu}_i, 2/3, 2/3))}{(\hat{\mu}_i(1 - \hat{\mu}_i))^{1/6}},$$

em que $B(z, a, b) = \int_0^z t^{(a-1)} (1-t)^{(b-1)} dt$ é a função *Beta* incompleta. Esta pode ser escrita em termos do produto entre a função acumulada da distribuição Beta(a, b), F(z, a, b), e a função *Beta* completa, B(a, b). Portanto, os resíduos de *Anscombe* podem ser reescritos como:

$$r_{A_i} = \frac{\sqrt{m_i}(F(y_i, 2/3, 2/3) \times B(2/3, 2/3) - F(\hat{\mu}_i, 2/3, 2/3) \times B(2/3, 2/3))}{(\hat{\mu}_i(1 - \hat{\mu}_i))^{1/6}}$$

No contexto do modelo conjunto longitudinal e binário, pelo fato de a variável resposta do submodelo logístico ser dicotômica e assim, seguir uma distribuição $Bernoulli(p_i)$,

tem-se que $m_i = 1$ e, portanto, $\hat{\mu}_i = \hat{p}_i$. Assim, os resíduos de *Anscombe* para o modelo conjunto são expressos por:

$$r_{A_i} = \frac{F(y_i^{(2)}, 2/3, 2/3) \times B(2/3, 2/3) - F(\hat{p}_i, 2/3, 2/3) \times B(2/3, 2/3)}{(\hat{p}_i(1 - \hat{p}_i))^{1/6}}$$

Objetivando-se a obtenção dos resíduos de *Anscombe* Estudentizados, primeiramente, se faz necessário calcular a matriz H, a qual, sob a ótica do modelo conjunto longitudinal e logístico, é expressa por:

$$m{H} = m{W}^{1/2} m{X}^{(2)} [(m{X}^{(2)})^t m{W} m{X}^{(2)}]^{-1} (m{X}^{(2)})^t m{W}^{1/2} m{X}^{1/2}$$

em que $W = diag(\omega_1, \ldots, \omega_n)$ é a matriz diagonal de pesos. Para a distribuição *Bernoulli*, considerando-se a função de ligação *logito*, $\omega_i = p_i(1 - p_i)$. É importante ressaltar que a matriz $X^{(2)}$, além das colunas referentes a cada uma das covariáveis presentes no submodelo logístico, também deve conter uma coluna adicional referente aos efeitos aleatórios preditos, a qual, como se sabe, está relacionada ao parâmetro de associação γ .

Uma vez calculada a matriz H, obtem-se h_i , o *i*-ésimo elemento da diagonal principal de H. De posse de tais quantidades, os resíduos de Anscombe Estudentizados, r_{Aest_i} , são expressos por:

$$r_{Aest_i} = \frac{r_{A_i}}{\sqrt{1 - h_i}}$$

Uma outra característica importante a ser avaliada é a qualidade da predição, que foi realizada por intermédio da curva *ROC*. Tal procedimento foi realizado no software *R* (R Core Team (2018)).

1.4 Resultados

O conjunto de dados, originalmente, era composto por 60 pacientes internados na UTI do Hospital Universitário Regional de Maringá, dentre os quais 15 (25%) se encontravam num quadro de PAV, 12 (20%) de outras pneumonias e 33 (55%) não apresentavam quaisquer uma dessas enfermidades. Entretanto, todos os pacientes que tiveram outras pneumonias já haviam sido internados com esta condição, diferentemente daqueles que tiveram a PAV, em que tal complicação se desenvolveu durante a internação. Desse modo, os 12 pacientes acometidos por outras pneumonias foram excluídos do estudo, em virtude de que o objetivo do mesmo é avaliar biomarcadores que possam influenciar na ocorrência da PAV, e assim tais pacientes representariam uma fonte de variabilidade bastante relevante e que poderia afetar drasticamente a qualidade e a confiabilidade dos resultados.

Uma variável que, biologicamente, pode ser importante para a predição da ocorrência da PAV é o foco infeccioso, a qual é categórica e provida de 5 níveis: 1 - pulmonar, 2 abdominal, 3 - renal, 4 - neurológico e 5 - outro. Todavia, pelo fato de somente 5 pacientes pertencerem ao nível 5, e dentre estes, três serem diagnosticados com PAV e dois não, há um problema denominado esparsidade dos dados (Agresti (2007)), em que ocorre uma má distribuição dos dados pelas categorias, resultando em frequências extremamente baixas ou até mesmo nulas, gerando dificuldades na estimação dos parâmetros do modelo. Assim sendo, estes pacientes também foram excluídos. O fluxograma da Figura 1 apresenta o processamento dos dados referentes aos pacientes descrito.



Figura 1 – Fluxograma da exclusão dos pacientes

No que se refere aos biomarcadores, também foi necessário realizar um processamento. Pelo fato de alguns biomarcadores serem denotados por siglas ou elementos químicos, foi construída a Tabela 1 a fim de se apresentar os significados dos mesmos.

Os biomarcadores *ARDS*, FRA, secreção traqueal, aspecto da secreção, diagnóstico de DPOC, sépsis, *MDR*, presença dos patógenos *Pseudomonas aeruginosa* e *Acinetobacter baumannii*, *MRSA*, terapia antibiótica adequada, insuficiência ou falência respiratória e insuficiência ou falência cardíaca foram excluídos por duas razões: sendo biomarcadores categóricos, isto é, são avaliados ao longo do tempo características binárias, nominais ou ordinais, o ajuste do modelo conjunto se torna mais difícil de ser realizado em virtude da complexidade do mesmo. Outra razão é que, para alguns desses biomarcadores categóricos, há uma quantidade relativamente grande de dados faltantes, impossibilitandose assim o ajuste.

Biomarcadores	Significado
$AaDO_2$	Diferença alvéolo-arterial de oxigênio, em mmHg
APACHE II	Acute Physiology and Chronic Health disease Classification System II
ARDS	Síndrome da Angústia Respiratória Aguda
CPIS	Clinical Pulmonary Infection Score
DPOC	Doença Pulmonar Obstrutiva Crônica
FiO_2	Fração inspirada de oxigênio, em %
FRA	Falência Renal Aguda
Hb	Hemoglobina, em g/100mL
HCO_3	Concentração de bicarbonato no sangue, em mEq/L (miliequivalentes por litro)
K	Concentração de potássio no sangue, em mEq/L
MDR	Resistência multidroga
MRSA	Staphylococcus aureus resistente à meticilina
Na	Concentração de sódio no sangue, em mEq/L
P.dia	Pressão arterial diastólica, em mmHg
P.sis	Pressão arterial sistólica, em mmHg
PAM	Pressão Arterial Média, em mmHg
PaO_2	Pressão arterial parcial de O_2 , em mmHg
PaO_2/FiO_2	Relação entre a pressão arterial parcial e a fração inspirada de oxigênio, em mmHg
pCO_2	Pressão parcial de gás carbônico, em mmHg
рН	Potencial hidrogeniônico, em escala logarítmica
SAPS II	Simplified Acute Physiology Score II
SO_2	Saturação arterial de oxigênio, em %
SOFA	Sequential Organ Failure Assessment

Tabela 1 – Significado dos biomarcadores

Os escores utilizados para se calcular o risco de mortalidade também foram obtidos: *SOFA*, *APACHE II* e *SAPS II*. Para se avaliar o risco do paciente desenvolver a PAV, calculou-se também o escore *CPIS*. A Figura 2 apresenta o processamento dos biomarcadores descrito.

Pelo fato de diversos biomarcadores apresentarem valores razoavelmente grandes, aplicou-se uma transformação nos mesmos, visando-se desse modo evitar problemas relacionados a convergência referentes a seus respectivos modelos conjuntos ajustados sub-sequentemente. Para os biomarcadores $AaDO_2$, débito urinário, leucócitos e plaquetas, foi aplicada a transformação logarítmica. Para o biomarcador balanço hídrico, uma vez que este apresentou valores bastantes grandes tanto positivos quanto negativos (se a quantidade de líquido que o paciente recebeu foi superior à eliminada, então o valor é positivo, caso contrário, negativo), utilizou-se a seguinte transformação:

$$x = \begin{cases} \log(x), & se \ x > 0\\ -\log(-x), & se \ x < 0 \end{cases}$$

Com exceção do pH, o único biomarcador que permaneceu inalterado devido ao fato de sua escala ser originalmente logarítmica, todos os outros biomarcadores foram centrados na média zero.



Figura 2 - Fluxograma da exclusão/adição dos biomarcadores

Os biomarcadores utilizados na análise se encontram na Tabela 2. Pode-se observar que a proporção de dados faltantes, para todos os biomarcadores, não ultrapassa 20%, considerando-se um total de 687 observações. Pode-se notar também que os biomarcadores para os quais parece haver uma diferença relevante entre os pacientes com PAV e sem, no que se refere à média amostral, são a frequência respiratória, glicose, HCO_3 , potássio, PaO_2 , pressão diastólica, PAM e pulso. Em termos de variabilidade, há uma disparidade bastante considerável para a maioria dos biomarcadores em relação ao diagnóstico da PAV, visto que, para PaO_2 , por exemplo, o coeficiente de variação para as medidas referentes aos pacientes com PAV é aproximadamente 232 vezes o coeficiente de variação para aqueles sem tal complicação, em valores absolutos.

Os pacientes ficaram internados na UTI, em média, durante aproximadamente 16 dias, sendo que o tempo mínimo de internação foi de 3 dias e o máximo de 53 dias, com desvio-padrão de aproximadamente 13 dias, o que indica um forte desbalanceamento sob o aspecto longitudinal.

Capítulo 1. Modelo conjunto longitudinal e logístico para dados de pneumonia associada à ventilação mecânica

Tabela 2 –	Quantidade e	proporção de	e dados	faltantes,	média	amostral,	desvio	padrão
	amostral e coe	eficiente de v	ariação d	de acordo	com o	diagnóstic	o de PA	V para
	cada biomarca	ldor						

		\bar{x}	(s)	CV		
Biomarcadores	Dados faltantes	Sem PAV	Com PAV	Sem PAV	Com PAV	
$AaDO_2$	64 (9,32%)	8,65 (0,91)	8,73 (0,7)	0,1	0,08	
APACHE II	0 (0%)	-0,56 (7,24)	-1,04 (7,66)	-12,95	-7,38	
Balanço Hídrico	79 (11,5%)	5,06 (4,51)	4,83 (4,93)	0,89	1,02	
Bilirrubina total	60 (8,73%)	-0,09 (0,83)	-0,28 (0,32)	-9,2	-1,14	
CPIS	0 (0%)	-0,1 (1,28)	0,2 (1,36)	-12,61	6,9	
Creatinina	57 (8,3%)	0,02 (9,24)	-0,48 (0,88)	511,76	-1,84	
Débito Urinário	109 (15,87%)	7,22 (0,98)	7,28 (0,88)	0,14	0,12	
FiO_2	19 (2,77%)	-0,02 (5,03)	-0,07 (4,32)	-249,91	-60,48	
Frequência Respiratória	48 (6,99%)	0,35 (7,66)	2,59 (7,05)	22,03	2,72	
Glasgow	53 (7,71%)	0,19 (4,1)	-0,55 (3,38)	22,11	-6,18	
Glicose	61 (8,88%)	-7,31 (55,48)	-2,99 (59,5)	-7,59	-19,87	
Hb	57 (8,3%)	0,43 (4,29)	-0,11 (1,9)	9,97	-17,51	
HCO_3	57 (8,3%)	0,15 (4,63)	1,66 (8,69)	30,64	5,24	
Hematócrito	63 (9,17%)	0,67 (6,44)	0,32 (4,58)	9,61	14,43	
K	58 (8,44%)	0,24 (8,74)	-0,41 (0,71)	36,69	-1,74	
Lactato	59 (8,59%)	-0,06 (1,07)	-0,03 (1,14)	-19,36	-37,85	
Leucócitos	63 (9,17%)	9,49 (0,44)	9,38 (0,4)	0,05	0,04	
Na	58 (8,44%)	0,33 (56,56)	0,18 (7,36)	168,89	39,95	
P.dia	48 (6,99%)	1,04 (16,65)	-5 (15,85)	15,99	-3,17	
P.sis	48 (6,99%)	0,08 (33,06)	-0,15 (30,91)	430,91	-202,36	
PAM	48 (6,99%)	0,72 (20,67)	-3,38 (19,28)	28,71	-5,7	
PaO_2	57 (8,3%)	0,1 (38,98)	-14,89 (24,61)	383,23	-1,65	
PaO_2/FIO_2	64 (9,32%)	1,92 (33,44)	-2,46 (28,78)	17,46	-11,69	
pCO_2	57 (8,3%)	1,93 (46,59)	1,36 (8,72)	24,09	6,42	
рН	57 (8,3%)	7,41 (0,08)	7,4 (0,22)	0,01	0,03	
Plaquetas	66 (9,61%)	12,41 (0,67)	12,37 (0,61)	0,05	0,05	
Pulso	47 (6,84%)	-1,09 (21,25)	1,2 (18,57)	-19,42	15,47	
SAPS II	0 (0%)	-1,3 (17,04)	-1,92 (17,33)	-13,09	-9,04	
SO_2	57 (8,3%)	-0,12 (5,97)	-0,89 (2,7)	-50,77	-3,05	
SOFA	0 (0%)	-0,32 (2,46)	0,17 (2,4)	-7,58	13,95	
Temperatura	49 (7,13%)	0,04 (1,17)	-0,05 (1,03)	26,78	-20,78	
Vasopressores	58 (8,44%)	-6,57 (7,53)	-5,85 (7,86)	-1,15	-1,34	

Uma vez realizado o processamento dos dados e uma breve análise descritiva, o próximo passo foi ajustar o modelo conjunto para cada um dos biomarcadores em questão. A estrutura do modelo que foi utilizada para todos os biomarcadores é expressa por:

$$\begin{cases} \hat{\mathbf{Y}}_{i} &= (\hat{\beta}_{0} + \tilde{b}_{0i}) + \hat{\beta}_{1} t_{im_{i}} \\ \log \left(\frac{\hat{p}_{i}}{1 - \hat{p}_{i}}\right) &= \hat{\alpha}_{0} + \hat{\alpha}_{1} foco2_{i} + \hat{\alpha}_{2} foco4_{i} + \hat{\gamma} \tilde{b}_{0i} \end{cases}$$

Assim, no submodelo misto, somente o tempo de internação foi utilizado como covariável e o efeito aleatório foi colocado somente no paciente. Para o submodelo logístico,

26

a variável categórica *foco* foi introduzida como covariável. Como já foi mencionado anteriormente, tal variável possui cinco categorias e, pelo fato de somente 5 pacientes pertencerem ao foco 5 (outros), estes foram excluídos do conjunto de dados. A categoria 1 foi colocada como referência e, uma vez que nenhum paciente pertence à categoria 3, o parâmetro correspondente não foi estimado. Assim, foram estimados os parâmetros relacionados às categorias 2 e 4, que correspondem aos focos abdominal e neurológico, respectivamente.

Por intermédio do procedimento *NLMIXED* (SAS Institute Inc. (2015b)) do *software SAS* (versão *University*), ajustou-se tal modelo para cada um dos 32 biomarcadores utilizando-se o método da máxima verossimilhança. A Tabela 3 apresenta os p-valores relacionados tanto ao parâmetro de associação do modelo conjunto como para o teste de normalidade de *Shapiro-Wilk* para os efeitos aleatórios, bem como as medidas preditivas sensibilidade, especificidade e a área sob a curva (*AUC*). Pode-se notar que, para os biomarcadores PCO_2 e sódio, a capacidade preditiva foi perfeita, no entanto, a suposição de normalidade para os efeitos aleatórios foi violada, bem como os coeficientes γ , que representam a associação entre os processos longitudinal e logístico, também não foram significativos a um nível de 10%. Somente para dois biomarcadores não houve transgressão da hipótese de normalidade para os efeitos aleatórios, ocorreu significância para a estimativa do parâmetro γ e também as medidas preditivas se mostraram satisfatórias: escala de *Glasgow* e pressão diastólica. Diante disso, deu-se prosseguimento na modelagem estatística para estes dois biomarcadores.

Biomarcadores	γ (p-valor)	Normalidade de b_0 (p-valor)	Sensibilidade (%)	Especificidade (%)	AUC
$AaDO_2$	0,71	<0,0001	50,00	83,90	0,66
APACHE II	0,38	0,67	58,30	83,90	0,68
Balanço Hídrico	0,35	0,01	91,70	71,00	0,78
Bilirrubina total	0,12	<0,0001	75,00	83,90	0,80
CPIS	0,63	0,98	66,70	74,20	0,67
Creatinina	0,75	<0,0001	100,00	87,10	0,91
Débito urinário	0,42	0,30	50,00	87,10	0,68
FiO_2	0,33	<0,0001	58,30	80,60	0,68
Frequência Respiratória	0,93	0,15	50,00	83,90	0,63
Glasgow	0,06	0,29	83,30	77,40	0,80
Glicose	0,78	<0,0001	41,70	87,10	0,63
Hb	0,58	0,64	50,00	83,90	0,65
HCO_3	0,92	0,16	50,00	80,60	0,64
Hematócrito	0,81	0,42	66,70	61,30	0,62
К	0,72	<0,0001	100,00	87,10	0,97
Lactato	0,39	0,01	83,30	61,30	0,73
Leucócitos	0,42	0,79	75,00	61,30	0,67
Na	0,75	<0,0001	100,00	100,00	1,00
P.dia	0,06	0,78	75,00	80,60	0,81
P.sis	0,39	0,01	58,30	74,20	0,66
PAM	0,12	0,25	66,70	77,40	0,74
PaO_2	0,20	0,09	75,00	87,10	0,81
PaO_2/FIO_2	0,70	<0,0001	75,00	61,30	0,70
pCO_2	0,39	<0,0001	100,00	100,00	1,00
pH	0,55	<0,0001	91,70	51,60	0,69
Plaquetas	0,49	0,38	50,00	80,60	0,64
Pulso	0,70	0,37	58,30	67,70	0,63
SAPS II	0,33	0,33	58,30	83,90	0,69
SO_2	0,50	0,88	83,30	71,00	0,77
SOFA	0,87	0,95	66,70	67,70	0,63
Temperatura	0,70	0,05	41,70	87,10	0,64
Vasopressores	0,17	0,13	50,00	87,10	0,69

Tabela 3 – Resultados para os modelos ajustados

Contudo, ao se observar a Figura 3, nota-se claramente uma forte tendência de decrescimento nos resíduos do submodelo misto para a escala de *Glasgow*. Neste caso, a Escala de *Glasgow* não foi considerada pois não houve respaldo estatístico no que se refere aos resíduos. Todavia, é importante ressaltar que, de acordo com Miller (2018), pacientes intubados possuem um nível de consciência reduzido que, consequentemente, prejudica a limpeza voluntária de secreções, podendo então se acumular na orofaringe levando à aspiração (micro e macro) de secreções contaminadas ricas em patógenos nocivos. Desse modo, por medir o grau de consciência do paciente, a escala de *Glasgow* se torna importante no sentido de avaliar uma possível vulnerabilidade do paciente devido ao baixo nível de consciência, o que pode contribuir decisivamente para o desenvolvimento da *PAV*.

Para o biomarcador da pressão diastólica, nota-se que os resíduos se comportam aleatoriamente em torno de zero, evidenciado assim um ajuste adequado no que tange ao submodelo misto. Portanto, a modelagem estatística prosseguirá considerando-se somente o biomarcador da pressão diastólica.



Capítulo 1. Modelo conjunto longitudinal e logístico para dados de pneumonia associada à ventilação mecânica

Figura 3 – Gráfico dos resíduos para os modelos conjuntos estimados em relação aos biomarcadores pressão diastólica e escala de *Glasgow*

Conforme a Tabela 4, observa-se que o tempo de internação se mostrou influente em relação à pressão diastólica, assim como parâmetro de associação γ se mostrou significativo, evidenciando assim que há uma conexão importante entre o processo longitudinal do biomarcador pressão diastólica e a chance de um indivíduo desenvolver a PAV. Em termos de interpretação, tem-se que, a cada redução de uma unidade na média da pressão diastólica, ocorre, em média, um aumento de aproximadamente 14% na chance de o indivíduo desenvolver PAV. Tal interpretação está bem representada pela Figura 4. Todos os pacientes que desenvolveram PAV o fizeram antes do 20° dia e, aparentemente, possuíram uma pressão diastólica inferior ao longo do tempo em relação àqueles não tiveram PAV, corroborando-se assim os resultados obtidos no modelo conjunto ajustado.

Submodelo misto									
Parâmetro	Estimativa	Erro padrão	Valor t	P(T > t)					
β_0	0,876	1,642	0,530	0,594					
β_1	-0,131	0,063	-2,090	0,037					
σ	14,233	0,421	33,790	<0,0001					
σ_{b_0}	8,843	1,208	7,320	<0,0001					

Tabela 4 –	Estimativas	dos parâmetros	para o	modelo	conjunto	referente à	à pressão	diastó-
	lica							

Submodelo logístico									
Parâmetro	Estimativa	Erro padrão	Valor t	P(T > t)	Odds Ratio	5%	95%		
α_0	-0,092	0,839	-0,110	0,913	0,912	0,229	3,626		
α_1	-1,320	1,125	-1,170	0,241	0,267	0,042	1,700		
α_2	-1,470	1,021	-1,440	0,150	0,230	0,043	1,233		
γ	-0,133	0,070	-1,900	0,058	1,142 (-)	0,780	0,982		



Capítulo 1. Modelo conjunto longitudinal e logístico para dados de pneumonia associada à ventilação mecânica

Figura 4 – Trajetórias longitudinais para a pressão diastólica segundo o diagnóstico de PAV

Finalmente, com o intuito de se averiguar a qualidade do ajuste do modelo conjunto, foi desenvolvido o gráfico do envelope (Atkinson (1985)), baseado nos seguintes passos:

- 1. Uma vez ajustado o modelo considerando-se os dados originais, obter $d_{(i)}$, os valores ordenados para uma determinada estatística de diagnóstico (resíduos, *leverage*, distância de *Cook* e outras).
- 2. Simular 10000 amostras para $\mathbf{Y}_{i}^{(1)}$, i = 1, ..., n, baseando-se em $\hat{\mathbf{Y}}_{i}^{(1)}$, isto é, $\mathbf{Y}_{i}^{(1)} \sim N(\hat{\mathbf{Y}}_{i}^{(1)}, \hat{\sigma}^{2})$ (parte longitudinal), e para $Y_{i}^{(2)}$ levando-se em consideração \hat{p}_{i} , ou seja, $Y_{i}^{(2)} \sim Bernoulli(\hat{p}_{i})$ (parte logística), em que $\hat{\mathbf{Y}}_{i}^{(1)}$, \hat{p}_{i} são os valores preditos dos processos longitudinal e logístico, respectivamente, e $\hat{\sigma}^{2}$ é a variância residual estimada para o modelo original;
- 3. Para cada uma das 10000 amostras referentes a $Y_i^{(1)}$ e $Y_i^{(2)}$, ajustar o modelo conjunto considerando-se os mesmos preditores inerentes ao modelo original e obter os valores ordenados de uma determinada estatística de diagnóstico, $d_{j(i)}^*$, $j = 1, \ldots, 10000$ e $i = 1, \ldots, n$;
- Para cada *i*, calcular a média (ou a mediana), o mínimo (ou o quantil α/2) e o máximo (ou o quantil 1 α/2) de d^{*}_{j(i)};
- 5. Para o envelope normal, calcular $z_i = \Phi^{-1}\left(\frac{i-0,375}{n+0,25}\right)$ e, para o envelope seminormal, calcular $z_i^* = \Phi^{-1}\left(\frac{i+n-0,125}{2n+0,5}\right)$, em que $\Phi^{-1}(\cdot)$ representa o quantil da distribuição normal padrão.

6. Para o envelope normal, fazer o gráfico dos quantis calculados no item 4 e de $d_{(i)}$ *versus* z_i e, para o envelope semi-normal, fazer o gráfico dos quantis e de $|d_{(i)}|$ *versus* z_i^* .

No presente estudo, considerou-se d_i , a estatística de diagnóstico, como sendo os resíduos de *Anscombe* Estudentizados, apresentados na seção 1.3. Por intermédio do procedimento *IML* (SAS Institute Inc. (2015a)), do *software SAS*, os resíduos do modelo original bem como aqueles resultantes dos ajustes para cada um dos 10000 conjuntos de dados simulados foram obtidos e exportados para o *software R* (R Core Team (2018)), no qual os gráficos de envelope foram desenvolvidos. Com base na Figura 5, pode-se observar que todos os resíduos se encontram na região delimitada pelo envelope, tanto para o normal quanto para o semi-normal, evidenciando assim um ajuste adequado.



Figura 5 – Gráficos do envelope para o modelo conjunto referente ao biomarcador pressão diastólica

Ao se observar a Figura 6, pode-se notar claramente que a inclusão do biomarcador pressão diastólica melhora de forma notável a capacidade preditiva do modelo no que se refere a diagnosticar corretamente pacientes quanto à ocorrência de PAV ou não. Considerando-se somente o modelo logístico, embora a especificidade seja bastante alta, a sensibilidade é baixa, evidenciando-se assim a dificuldade que tal modelo possui para discernir entre os pacientes com PAV e sem. Incorporando-se as medidas realizadas ao longo do tempo para a pressão diastólica, constituindo-se assim o modelo conjunto, notase que, apesar de haver uma redução de 7% na especificidade, ocorre um aumento de 42% na sensibilidade e de 0,2 na área sob a curva. Desse modo, o modelo conjunto é provido de uma capacidade preditiva bastante superior à do logístico, conseguindo identificar de uma forma mais consistente indivíduos com ou sem PAV. Entretanto, é importante frisar que, para ambos os modelos, não houve qualquer processo de separação dos dados em amostra de treinamento e de teste. A amostra utilizada para se construir o modelo foi a mesma para se avaliar a capacidade preditiva e, assim sendo, a sensibilidade e a especificidade podem estar otimistas.



Figura 6 – Curva ROC para o modelo conjunto referente à pressão diastólica

1.5 Conclusão

Neste estudo foi apresentada a fundamentação teórica relacionada ao modelo conjunto longitudinal e logístico, bem como sua aplicação na modelagem da probabilidade de um paciente desenvolver a pneumonia associada à ventilação mecânica com relação ao comportamento dos biomarcadores.

Inicialmente, avaliou-se o modelo conjunto para cada um dos 28 biomarcadores, e após a obtenção dos escores *APACHE II, SAPS II, SOFA* e *CPIS*, fez-se o mesmo para estes, totalizando-se assim 32 biomarcadores. Por razões de dados faltantes e dificuldades em se ajustar um modelo conjunto para biomarcadores categóricos, já que o ajuste para esse tipo de variável se torna razoavelmente complexo, 14 biomarcadores foram desconsiderados. Entre os 32 presentes na análise, somente para a pressão diastólica e a escala de *Glasgow* a suposição de normalidade dos efeitos aleatórios foi satisfeita, além do fato de haver uma conexão importante entre tais biomarcadores e o diagnóstico de PAV, representada pela significância do parâmetro γ e também de apresentarem medidas preditivas satisfatórias. Entretanto, ao se observar os resíduos do submodelo misto para a escala de *Glasgow*, verificou-se que estes se mostraram altamente tendenciosos, e portanto o modelo para tal biomarcador foi descartado, em que pese sua importância biológica neste contexto. Já para o modelo conjunto referente a pressão diastólica, os resíduos do sub-

modelo misto estavam bem comportados, apresentando uma aleatoriedade em torno de zero.

De acordo com o modelo conjunto longitudinal e logístico para a pressão diastólica, tal biomarcador foi influenciado pelo tempo de internação. Em relação ao parâmetro de associação, tem-se que, a cada redução de uma unidade na média da pressão diastólica, ocorre, em média, um aumento de aproximadamente 14% na chance de o paciente desenvolver a PAV. Em termos biológicos, tal interpretação está condizente com a literatura médica, já que a PAV desencadeia um quadro de sepse, que por sua vez leva a uma vasodilatação e, desse modo, provoca uma redução na pressão diastólica.

Por intermédio do gráfico do envelope, desenvolvido com base nos resíduos de *Anscombe* Estudentizados, observou-se que o modelo está adequado, visto que todos os resíduos estavam presentes dentro da região delimitada pelos quantis, tanto para o envelope normal quanto para o semi-normal.

A capacidade preditiva do modelo conjunto para a pressão diastólica se mostrou bastante satisfatória, visto que a sensibilidade, especificidade e área sob a curva resultaram em 75%, 80,6% e 0,815, respectivamente, mostrando que a adição no modelo logístico de medidas para a pressão diastólica ao longo do tempo aumentou de forma significativa o poder de predição, permitindo uma identificando mais sólida de pacientes com ou sem PAV. Todavia, ressalta-se que esses resultados são otimistas, visto que a amostra utilizada para se construir o modelo foi a mesma para se avaliar a capacidade preditiva.

Diante do exposto, é importante esclarecer que este estudo se encontra numa fase preliminar, pois os dados coletados no HUM não foram homogeneizados quanto a doenças crônicas, idade, foco, tempo de internação de outro hospital, entre outros, que possivelmente são fatores de confusão para o diagnóstico da PAV.

Portanto, a modelagem conjunta longitudinal e logística se mostrou uma ferramenta bastante oportuna para se avaliar de que modo a evolução de um biomarcador no tempo pode afetar a chance de o paciente vir a desenvolver PAV.

1.6 Limitações

Como este estudo é tipo "tempo de duração fixado", algumas restrições na interpretação devem ser consideradas para evitar efeitos de confusão, tais como: gravidade dos pacientes ao serem admitidos na UTI, diferentes doenças de base e/ou crônicas, idade, foco infeccioso, diferentes tratamentos, entre outros.

Sob o ponto de vista estatístico, também são necessárias algumas considerações: dificuldades de se obter uma amostra homogênea em virtude do tipo de estudo, consistência insuficiente referente a algumas informações e dados faltantes no instrumento de medida, isto é, os prontuários de cada paciente. Também, ressalta-se o fato desta pesquisa ser resultado de um único centro médico terapêutico intensivo, necessitando-se assim precauções para as inferências.

1.7 Recomendações

Para dar a continuidade nesta pesquisa, parece ser fundamental homogeneizar e aumentar o tamanho da amostra em outros centros médicos terapêuticos intensivos. É importante frisar também que, no presente estudo, considerou-se apenas o diagnóstico pontual do paciente, isto é, se este desenvolveu ou não PAV. Entretanto, para este conjunto de dados, o diagnóstico longitudinal da PAV também se encontra disponível, ou seja, em todos os dias de internação os pacientes foram avaliados quanto ao desenvolvimento de tal complicação. Assim, sugere-se utilizar outras duas metodologias: o modelo conjunto longitudinal para respostas binárias e contínuas (Amini et al. (2018)) e também o modelo conjunto longitudinal e de sobrevivência com fração de cura (Martins, Silva e Andreozzi (2017)) caso haja o interesse em se observar o tempo de ocorrência até o diagnóstico de PAV. Recomenda-se utilizar a fração de cura em virtude de que pode-se identificar claramente um platô na curva de Kaplan-Meier, apresentada na Figura 7. Depois de aproxima-damente 20 dias de internação, nenhum paciente desenvolveu a PAV, caracterizando-se assim o platô.



Figura 7 – Curva de Kaplan-Meier para os dados da PAV

Referências

AGRESTI, A. *An Introduction to Categorical Data Analysis*. Wiley, 2007. (Wiley Series in Probability and Statistics). ISBN 9780470114742. Disponível em: <<u>https://books.google.com.br/books?id=YRtEiDevAi0C></u>. Citado na página 23.

AMINI, P.; VERBEKE, G.; ZAYERI, F.; MAHJUB, H.; MAROUFIZADEH, S.; MOGHIMBEIGI, A. Longitudinal joint modelling of binary and continuous outcomes: A comparison of bridge and normal distributions. *Epidemiology, Biostatistics and Public Health*, v. 15, n. 1, 2018. Citado na página 34.

ATKINSON, A. C. Plots, transformations and regression; an introduction to graphical methods of diagnostic regression analysis. [S.I.], 1985. Citado na página 30.

BALTHAZAR, A.; NOWAKONSKI, A. V.; CAPITANI, E. D.; BOTTINI, P.; TERZI, R.; ARAUJO, S. Diagnostic investigation of ventilator-associated pneumonia using bronchoalveolar lavage: comparative study with a postmortem lung biopsy. *Brazilian Journal of Medical and Biological Research*, SciELO Brasil, v. 34, n. 8, p. 993–1001, 2001. Citado na página 17.

CARVALHO, C. Roberto Ribeiro de. Pneumonia associada à ventilação mecânica - editorial. *Jornal Brasileiro de Pneumologia*, SciELO, v. 32, 2006. Citado na página 16.

CHASTRE, J.; FAGON, J.-Y. Ventilator-associated pneumonia. *American journal of respiratory and critical care medicine*, American Thoracic Society New York, NY, v. 165, n. 7, p. 867–903, 2002. Citado na página 16.

DALMORA, C. H.; DEUTSCHENDORF, C.; NAGEL, F. M.; SANTOS, R. P. d.; LISBOA, T. C. Definindo pneumonia associada à ventilação mecânica: um conceito em (des) construção. *Revista brasileira de terapia intensiva. Vol. 25, no. 2 (abr./jun. 2013), p. 81-86*, 2013. Citado na página 16.

GROUP, B. D. W.; JR, A. J. A.; COLBURN, W. A.; DEGRUTTOLA, V. G.; DEMETS, D. L.; DOWNING, G. J.; HOTH, D. F.; OATES, J. A.; PECK, C. C.; SCHOOLEY, R. T. et al. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical pharmacology & therapeutics*, Wiley Online Library, v. 69, n. 3, p. 89–95, 2001. Citado na página 17.

HORROCKS, J.; HEUVEL, M. J. van D. et al. Prediction of pregnancy: a joint model for longitudinal and binary data. *Bayesian Analysis*, International Society for Bayesian Analysis, v. 4, n. 3, p. 523–538, 2009. Citado 2 vezes nas páginas 17 e 19.

KALANURIA, A. A.; ZAI, W.; MIRSKI, M. Ventilator-associated pneumonia in the icu. *Critical care*, BioMed Central, v. 18, n. 2, p. 208, 2014. Citado na página 16.

KLOMPAS, M. Does this patient have ventilator-associated pneumonia? *Jama*, American Medical Association, v. 297, n. 14, p. 1583–1593, 2007. Citado na página 16.

MARTINS, R.; SILVA, G. L.; ANDREOZZI, V. Joint analysis of longitudinal and survival aids data with a spatial fraction of long-term survivors: A bayesian approach. *Biometrical Journal*, Wiley Online Library, v. 59, n. 6, p. 1166–1183, 2017. Citado na página 34.

MILLER, F. Pneumonia associada à ventilação mecânica. 2018. Citado na página 28.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018. Disponível em: https://www.R-project.org/. Citado 3 vezes nas páginas 21, 22 e 31.

RIZOPOULOS, D. Joint models for longitudinal and time-to-event data: With applications in *R*. [S.I.]: Chapman and Hall/CRC, 2012. Citado 2 vezes nas páginas 20 e 21.

SAS Institute Inc. *SAS/IML* [®]. Cary, North Carolina, USA, 2015. Disponível em: http://support.sas.com/documentation/cdl/en/imlug/68150/PDF/default/imlug.pdf>. Citado 2 vezes nas páginas 21 e 31.

SAS Institute Inc. *SAS/STAT* [®]. Cary, North Carolina, USA, 2015. Disponível em: https://support.sas.com/documentation/onlinedoc/stat/141/nlmixed.pdf>. Citado 2 vezes nas páginas 21 e 27.

Sociedade Paulista de Infectologia. Diretrizes sobre pneumonia associada a ventilação mecânica (pav). *São Paulo: Office Editora e Publicidade Ltda*, p. 1–19, 2006. Citado na página 16.

WANG, C.; WANG, N.; WANG, S. Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics*, Wiley Online Library, v. 56, n. 2, p. 487–495, 2000. Citado na página 17.

ZHANG, N.; CHEN, H.; ZOU, Y. A joint model of binary and longitudinal data with nonignorable missingness, with application to marital stress and late-life major depression in women. *Journal of Applied Statistics*, Taylor & Francis, v. 41, n. 5, p. 1028–1039, 2014. Citado na página 17.