

## A caracterização clínica e laboratorial de pacientes com neutropenia febril a partir da base de dados MIMIC-III

### **Pesquisadores responsáveis:**

- Miguel Pisani Leme Ferreira
  - Aluno do curso de graduação em Medicina da FCM-Unicamp
- Erich Vinicius de Paula
  - Professor MS 3.1 do Departamento de Clínica Médica da FCM – Unicamp
  - Pesquisador do CNPq (Bolsa PQ nível 2)

### **RESUMO**

A neutropenia febril (NF) pós quimioterapia é uma condição grave, associada a alto risco de sepse e de suas complicações. Embora estudos de grandes coortes tenham descrito modelos de estratificação de risco para a sepse na NF, a descrição da evolução clínica de pacientes com NF que necessitam de terapia intensiva é um tema menos abordado, em geral descrito em estudos unicêntricos com número limitado de pacientes e de variáveis. Atualmente o papel dos neutrófilos é compreendido como fundamental na fisiopatologia da sepse, tanto para a erradicação de patógenos quanto no próprio mecanismo de disfunção orgânica, através da geração das redes extracelulares de neutrófilos (NETs). Neste contexto é razoável supor que a fisiopatologia da sepse e da disfunção orgânica a ela associada seja distinta em pacientes com NF. Recentemente foram tornados públicos dados clínicos e laboratoriais detalhados de mais de 40.000 pacientes internados em uma unidade de terapia intensiva nos EUA através da base de dados MIMIC-III (*Medical Information Mart for Intensive Care III*), a partir da qual já foram gerados mais de 80 trabalhos nos últimos 2 anos. O objetivo deste projeto é extrair desta base dados de pacientes com NF e sepse para (i) descrever de forma sistematizada as características clínicas das disfunções orgânicas observadas nestes pacientes, e (ii) avaliar o impacto da neutropenia e da plaquetopenia (bem como da relação entre estas duas variáveis) em desfechos clínicos relevantes da sepse neste grupo.

### **A pandemia da COVID-19 e os impactos no projeto**

Com a chegada da pandemia da COVID-19, surgiu a possibilidade de utilizar a estrutura e a metodologia já desenvolvidas neste projeto, para elaborar uma pesquisa científica de alta relevância e impacto social, relacionada à fisiopatologia da COVID-19. Utilizando a base de dados MIMIC-III como fonte dos dados, foram traçados os seguintes objetivos para esta pesquisa correlacionada: (i) obter informações de pacientes internados em unidades de cuidado intensivo devido a quadros respiratórios de etiologia viral e bacteriana, a fim de identificar parâmetros clínicos definidores de gravidade; (ii) identificar parâmetros hematológicos, através da análise de exames

laboratoriais da base, relacionados aos quadros graves, bem como a ocorrência de fenômenos tromboembólicos, de pacientes com diagnóstico de pneumonia de etiologia viral e bacteriana; (iii) obter dados suficientes para correlacionar, posteriormente, a evolução dos quadros de pneumonias de etiologia viral e bacteriana com as evidências científicas disponíveis acerca da fisiopatologia da pneumonia associada à COVID-19.

### A base de dados MIMIC-III

A MIMIC-III (*Medical Information Mart for Intensive Care III*) é uma extensa base de dados, aberta e gratuita, que disponibiliza informações clínicas de mais de 40.000 pacientes admitidos em unidades de cuidados intensivos<sup>15</sup>. A base é composta por dados de um único hospital terciário, o *Beth Israel Deaconess Medical Center*, em Boston, Massachusetts, EUA – hospital afiliado à Universidade de Harvard, e que foram colhidos de internações entre os anos de 2001 e 2012. O direito ao acesso à base é obtido após a realização de um cadastro digital, a finalização de um curso online e a aprovação bilateral de um termo de uso. Após a aprovação, o usuário pesquisador pode fazer a análise dos dados de modo irrestrito em estudos analíticos de quaisquer áreas que desejar. A base permite a análise de informações como dados vitais, anotações de prontuário registradas pelos profissionais da saúde, diagnósticos, resultados de exames laboratoriais, prescrições, administração de fluidos, informações sobre procedimentos, laudos de exames de imagem, informações de internação e dados de sobrevida e mortalidade dos pacientes. Grande parte das informações possui registro temporal, o que possibilita estudar a evolução dos quadros clínicos dos pacientes através da análise de múltiplos parâmetros, como dados vitais e resultados de exames laboratoriais<sup>16</sup>. A grande quantidade de dados disponíveis, o caráter aberto da base, a presença de registros temporais e a elevada confiabilidade das informações faz com que a MIMIC-III seja um ambiente ideal para a reprodução de resultados e para a validação de hipóteses em estudos científicos.

## **MATERIAIS E MÉTODOS**

### Fonte dos dados

Os dados foram obtidos da base de dados aberta “MIMIC-III”, hospedada na plataforma PhysioNet<sup>17</sup>. O requisito inicial para a solicitação do acesso à base de dados foi a realização do curso online “*Data or Specimens Only Research*”, no site do programa CITI (<http://www.citiprogram.org/>), com nota mínima para aprovação de 90 em 100 pontos. Após isso, foi submetida uma requisição de acesso à base, na plataforma PhysioNet (<http://mimic.physionet.org/>), onde foi preenchido um termo de utilização dos dados. Depois de aprovada a requisição de acesso, realizou-se o *download* dos arquivos da base de dados, em formato csv (*comma-separated value*), e a instalação dos mesmos em um banco de dados local PostgreSQL versão 11.2. Concluída a



instalação do banco de dados local, a obtenção dos dados foi feita através de consultas (*queries*) ao banco utilizando o administrador pgAdmin versão 4. Os dados obtidos foram, então, transferidos, e serão analisados em programas editores de planilhas e texto.

## RESULTADOS E DISCUSSÃO

A obtenção dos dados foi feita através de consultas (*queries*) ao banco de dados local cuja origem dos dados se deu na base MIMIC-III.

No total, foram obtidos 6.176 resultados de exames que indicaram neutropenia grave (i), referentes a um total de 886 pacientes. Desses, apenas uma parcela dos exames havia sido colhida nas primeiras 120 horas (ii e iii). Finalmente, obteve-se o total de 514 pacientes únicos (iv) que manifestaram neutropenia grave nas primeiras 120 horas de internação.

A maior parte das informações relevantes para o estudo já foram obtidas através de algoritmos desenvolvidos e sequências de consultas ao banco de dados. São estas: relação de pacientes e informações (sexo e idade), duração total de internação em unidade intensiva, ocorrência de óbito durante a internação, além de outras informações que serão descritas a seguir. Obteve-se, também, dados de exames laboratoriais relevantes para correlações clínicas, como a contagem de neutrófilos, hemoglobina e plaquetas.

Além disso, foi feita uma triagem manual de todos os diagnósticos de cada um dos pacientes neutropênicos selecionados. A triagem visava ao agrupamento dos pacientes em diagnósticos de interesse. Para o trabalho em questão, definimos como diagnósticos de interesse os diagnósticos referentes a neoplasias (e, principalmente, neoplasias hematológicas – leucemias, linfomas e mielomas) e que indicassem neutropenias secundárias ao uso de fármacos (quimioterápicos anti-câncer, principalmente). Assim, cada paciente foi atribuído a um grupo diagnóstico a fim de organizar os dados obtidos e possibilitar uma análise estatística posterior. Por fim, agrupamos os dados obtidos em uma tabela parcial, contendo as informações dos pacientes, diagnósticos e dados laboratoriais, para possibilitar o início da análise dos resultados e orientar de forma mais estruturada a obtenção de eventuais dados restantes.

Para a conclusão do trabalho, portanto, ainda resta a fase de análise estatística dos dados obtidos, que será realizada com a colaboração do escritório de estatística da FCM-UNICAMP e utilizando os softwares GraphPad Prism, SPSS e R.

### A pandemia da COVID-19 e os impactos no projeto

Com a chegada da pandemia da COVID-19, surgiu a possibilidade de utilizar a estrutura e a metodologia já desenvolvidas neste projeto, para elaborar uma pesquisa científica de alta relevância e impacto social, relacionada à fisiopatologia da COVID-19. Assim, utilizamos a mesma base de dados, os mesmos programas e a mesma lógica básica para realizar a extração dos dados clínicos essenciais para o desenvolvimento da pesquisa.

Finalizada a obtenção dos dados gerais, iniciamos a extração de dados oriundos de resultados de hemogramas realizados na internação. Foram estes: dosagem de hemoglobina, contagem de plaquetas, leucócitos, linfócitos, neutrófilos, monócitos, basófilos, eosinófilos. A seguir, extraímos os dados necessários para o cálculo do escore da ISTH (*International Society on Thrombosis and Haemostasis*) de risco para CIVD (coagulação intravascular disseminada): Dímero D, contagem de plaquetas, dosagem de fibrinogênio e o Tempo de Protrombina. Para o cálculo do escore de risco, foram consideradas as maiores medidas no intervalo de 48 horas após a internação. Para a disponibilização dos resultados de cada exame individualmente, a fim de realizar outras análises, foi considerado o intervalo de 120h após a internação.

Após o término da obtenção dos dados hematológicos, evoluímos para a etapa final: parâmetros clínicos de gravidade e cálculo de escores de gravidade. Foi calculado o tempo total de ventilação mecânica para cada internação, quando presente, sendo a média do tempo sob ventilação mecânica (em horas), de 124,3 horas. Foi realizado o cálculo do SOFA, levando em consideração os parâmetros clínicos mais alterados, das primeiras 24 horas de internação para cada paciente. Para este cálculo, utilizamos como base um script já desenvolvido para esta finalidade, com algumas modificações.

Por fim, organizamos todos os dados obtidos em tabelas gerais, e iniciamos a análise dos dados utilizando o software SPSS.

		Age	SOFA score (24h)	ISTH score	Hb	Platelets
N	Valid	6842,0	6810,0	323,0	6800,0	6801,0
	Missing	0,0	32,0	6519,0	42,0	41,0
Mean		64,7	1,6	4,0	11,4	260550,2
Std. Deviation		15,9	1,9	1,8	2,3	145974,2
Percentiles	25	54,3	0,0	3,0	9,8	169000,0
	50	67,0	1,0	4,0	11,4	237000,0
	75	77,5	2,0	5,0	12,9	324000,0

Exemplo 1 de análise de tabela geral no SPSS

	n	%
Ausente	6.620	96,80%
Presente	222	3,20%

TEP

	n	%
Ausente	6.522	95,30%
Presente	320	4,70%

TVP

	n	%
Ausente	6.356	92,90%
Presente	486	7,10%

Eventos tromboembólicos



	n	%
<b>Viral</b>	156	2,30%
<b>Bacteriana</b>	2.232	32,60%
<b>Não especificada</b>	4.454	65,10%

*Casos de pneumonia*

	n	%
<b>UTI Coronária</b>	31	0,50%
<b>UTI Cirurgia cardíaca</b>	874	12,80%
<b>UTI Cirúrgica</b>	440	6,40%
<b>UTI Médica</b>	3.926	57,40%
<b>UTI Trauma</b>	896	13,10%

*Tipo da UTI*

	n	%
<b>Mulher</b>	2.946	43,10%
<b>Homem</b>	3.896	56,90%

*Distribuição por gênero*

	n	%
<b>Sobrevida</b>	5.521	80,70%
<b>Óbito</b>	1.321	19,30%

*Sobrevida na internação*

## CONSIDERAÇÕES FINAIS

A elaboração deste projeto de pesquisa se mostrou extremamente proveitosa, tanto pela obtenção de dados bastante ricos, quanto pelo desenvolvimento de metodologias e ferramentas para extração de dados de forma rápida e eficiente. É importante, entretanto, pontuar que houve uma mudança considerável no escopo do projeto devido à pandemia da COVID-19. A parte inicial da pesquisa, acerca de neutropenia febril, foi pausada após a fase de obtenção dos dados. Assim, permaneceu pendente a fase de análise estatística, que será realizada futuramente. Logo após essa etapa, assumimos o desafio de realizar uma pesquisa científica praticamente inteira, relacionada à fisiopatologia da COVID-19, em poucos meses, utilizando a metodologia e as ferramentas já desenvolvidas no início do projeto. A mudança no escopo do projeto se deu, principalmente, pela alta relevância científica e social do tema que foi abordado, bem como pela necessidade de se produzir conteúdo científico rapidamente e contribuir para o conhecimento global da doença. Fomos capazes de obter dados históricos relevantes sobre a evolução clínica das pneumonias de diversas etiologias. Agora, utilizaremos estes dados para a realização de análises estatísticas que permitirão comparar e correlacionar a fisiopatologia e a evolução clínica da pneumonia na COVID-19 às pneumonias de outras etiologias. Embora, claro, tenhamos desenvolvido duas pesquisas ainda incompletas neste período, obtivemos dados valiosos que, agora, serão analisados estatisticamente de forma extensa e que gerarão, no futuro próximo, duas pesquisas com produção científica relevante, que serão submetidas à publicação e que terão considerável impacto social.