



INTRODUÇÃO / OBJETIVO

A pandemia de COVID-19 começou em dezembro de 2019, quando foram identificados casos de pneumonia em Wuhan, na China. Em janeiro de 2020, um novo coronavírus, o SARS-CoV-2, foi confirmado como o agente causador da doença. A rápida disseminação levou a OMS a declarar a COVID-19 como uma pandemia em março de 2020. Medidas de distanciamento social foram adotadas em todo o mundo para conter o vírus, afetando a rotina das pessoas e evidenciando questões como a fragilidade dos sistemas de saúde e as desigualdades sociais. De modo a auxiliar no estudo e desenvolvimento de estratégias de mitigação do impacto de epidemias e pandemias, foi disponibilizado o Banco de Dados da COVID-19, desenvolvido a partir de informações disponibilizadas pelo DATASUS. Nesse repositório de origem, os dados contêm informações clínicas dos pacientes, bem como a evolução dos casos. No entanto, trata-se de um repositório com lacunas e inconsistências nas informações. Assim, o objetivo deste trabalho foi desenvolver um repositório com dados verificados em relação à consistência e completude.

MATERIAL E MÉTODOS

Seleção dos Dados

Para a construção do Banco de Dados da Covid-19, foram utilizados recursos públicos do DATASUS. Os quais, foram armazenados no formato CSV (*Comma-separated values*), juntamente com um dicionário de dados para o mapeamento das informações e as variáveis armazenadas no arquivo CSV. Segundo o portal do DATASUS, os dados passam por atualizações quinzenais, especificamente nas quartas-feiras. O arquivo de dados disponibilizado pelo portal, contém aproximadamente 170 variáveis, incluindo informações completas sobre os pacientes, como local de atendimento, evolução do quadro, entre outras.

Processamento de Dados

O pré-processamento dos dados foi realizado utilizando a linguagem de programação R (R Core Team, 2023) por meio da IDE (*Integrated Development Environment*) RStudio [Posit team, 2023]. O pré-processamento dos dados se deu a partir da importação dos dados para o ambiente do RStudio. Dentro desta etapa, foram selecionadas as variáveis de fatores de risco, a evolução e resultado do teste de COVID pela técnica de biologia molecular PCR, UF (unidade federativa), classificação final do exame de diagnóstico e idade do paciente. As variáveis utilizadas para a realização deste estudo foram: Unidade Federativa; Evolução do caso; Resultado do teste RT-PCR; Diagnóstico final do caso; Data do 1º sintoma do caso, comorbidades que o paciente possuía: Cardiopatia, Hematologia, Doença Hepática, Asma, Doença Renal, Doença Neurológica, Diabetes, Obesidade e idade. Após a seleção das variáveis, foram realizadas a inclusão de variáveis derivadas, para realizar uma análise por região do Brasil. Para que fosse possível realizar a seleção dos dados por ano de ocorrência, foi adicionada a variável ano, baseando-se na variável da data dos primeiros sintomas. Por fim, foi incluída a variável de faixa etária.

Em relação aos filtros de inclusão dos pacientes, foram selecionados apenas os dados de pacientes internados e com resultado do teste de PCR positivo para COVID-19. Adicionalmente, para os fatores de risco e evolução clínica, foram excluídos pacientes com registros com valores ignorados, vazios ou nulos. A fim de facilitar a aplicação das análises, após a seleção dos dados de interesse, foram mesclados os dados de todos os anos (2020,2021,2022,2023), uma vez que são disponibilizados separadamente

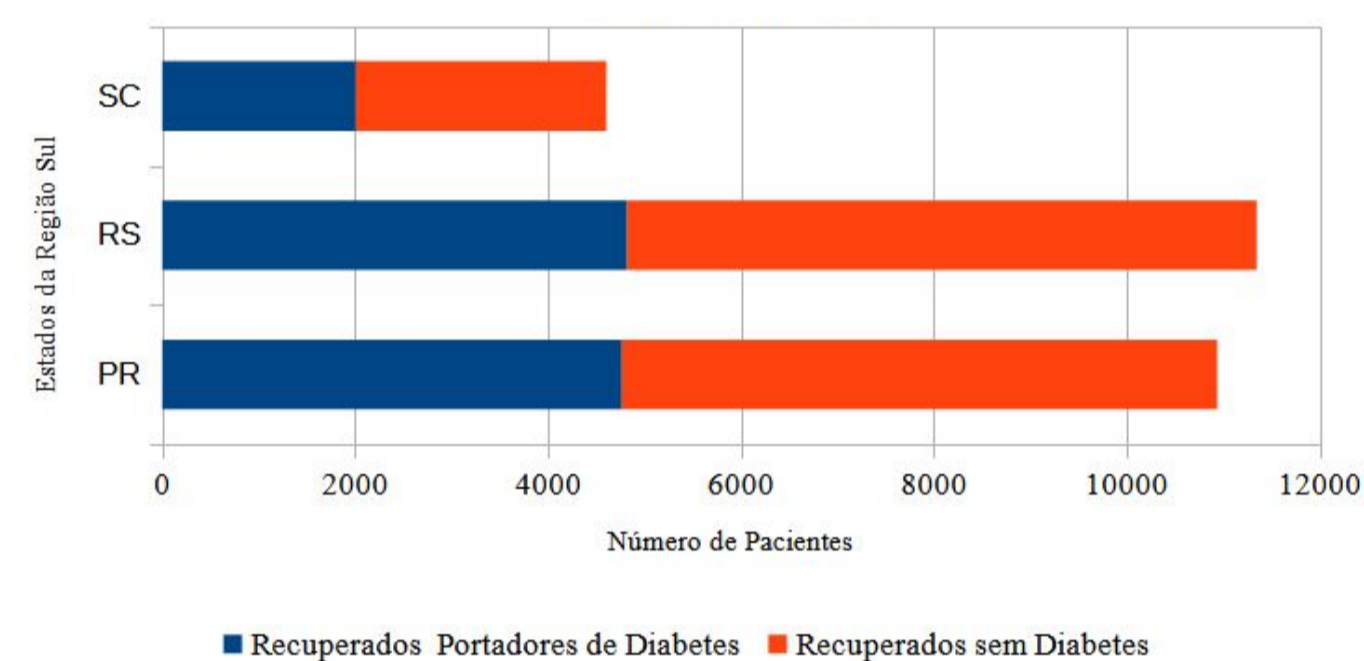
RESULTADOS

Produção Científica a partir de Dados da Covid 19

Segundo Bufrem (2020, p. 103) “[...] a produção e a comunicação científicas são fatores cruciais para compreender a doença e seus efeitos, na busca de soluções possíveis, se não efetivas ao menos redutoras de riscos”. Dessa forma, vê-se que produção e a comunicação científica têm um papel importante no desenvolvimento de melhores estratégias de compreensão da Covid-19.

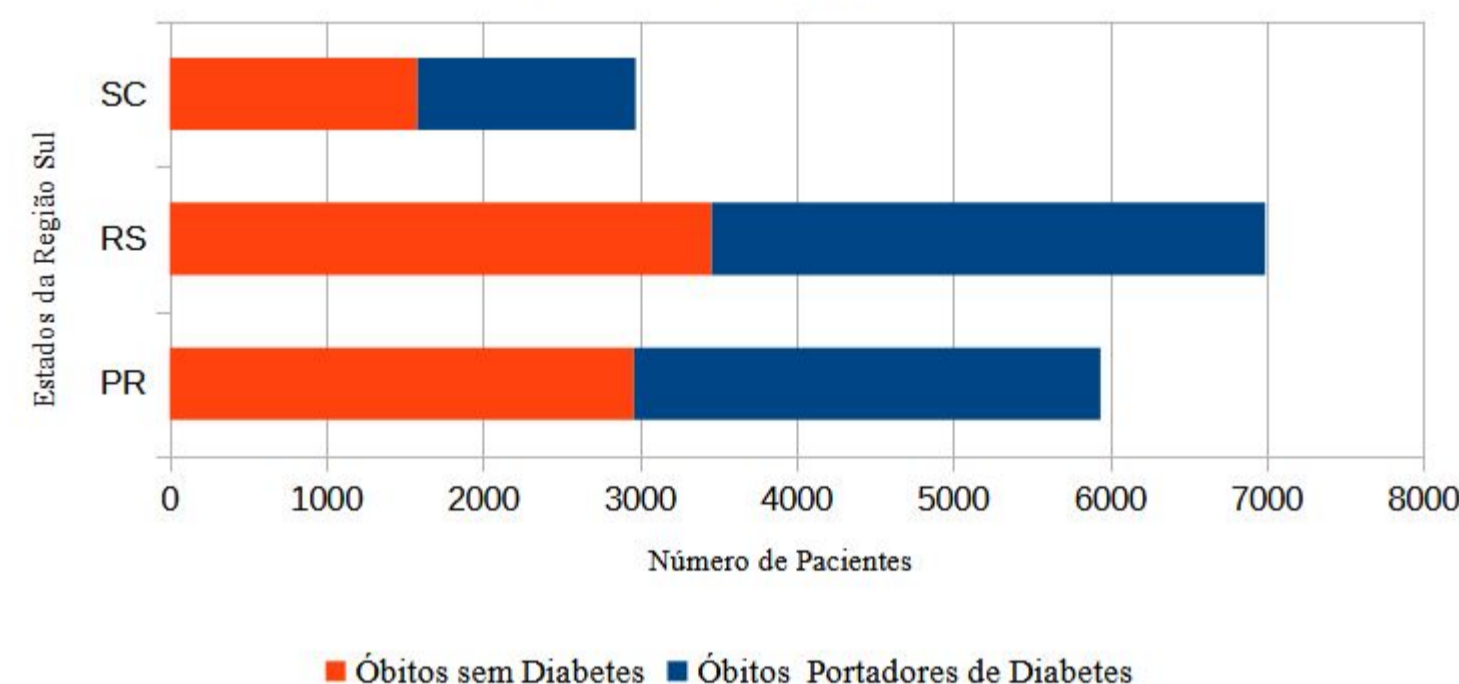
Por meio do acesso às informações, organizadas em um banco de dados, é possível realizar ainda mais análises, gerando mais conhecimento acerca da pandemia. Segundo estudo realizado por Ou *et al.*, 2020, dentre os pacientes COVID-19 positivos, a diabetes é um fator de risco comum, consequentemente, ela foi sugerida como um fator de risco para casos graves e fatais. Este padrão também pôde ser observado por meio da análise dos dados obtidos no DATASUS, com o número de pacientes recuperados da COVID 19 e o número de óbitos, em três unidades federativas da região Sul do Brasil: Paraná, Rio Grande do Sul e Santa Catarina, no ano de 2020. No gráfico 1, temos uma comparação entre o números de pacientes recuperados portadores da Diabetes e os recuperados não portadores da doença. No gráfico 2, é possível visualizar a comparação entre o número de óbitos de pacientes portadores de diabetes e os não portadores da doença.

Gráfico 1. Prevalência da Diabetes em Pacientes Recuperados da Covid-19 no ano de 2020



Fonte: Autora

Gráfico 2. Prevalência da Diabetes em Óbitos por Covid-19 no ano de 2020



Fonte: Autora

CONSIDERAÇÕES FINAIS

Práticas de difusão de informações e dados em acesso aberto e redes de colaboração fazem parte de um modelo de produção científica denominado Ciência Aberta (*Open Science*). Estas iniciativas são importantes para o desenvolvimento e enriquecimento das metodologias de pesquisa. Dessa forma, salienta-se a importância do acesso aberto de dados e informações relevantes, tendo como objetivo o desenvolvimento de metodologias científicas colaborativas, rápidas e efetivas para entender e tratar emergências, como a Covid-19.

REFERÊNCIAS BIBLIOGRÁFICAS

- Bufrem, Leilah Santiago. A pandemia da covid-19 no Brasil: informações e contradições na atual conjuntura. P2P e Inovação, Rio de Janeiro, v. 7, n. 1, p. 101-120, 26 set. 2020. DOI: <https://doi.org/10.21721/p2p.2020v7n1>. Disponível em: <http://revista.ibict.br/p2p/article/view/5461/5073>. Acesso em: 11 maio 2021.
- Croda, J.; Oliveira, W. K.; Frutuoso, R. L.; Mandetta, L. H.; Baia-Dasilva, D. C.; Brito-Souza, J. D.; Monteiro, W. M.; Lacerda, M. V. G. COVID-19 in Brazil: advantages of a socialized unified health system and preparation to contain cases. *Journal of the Brazilian Society of Tropical Medicine* 2020, 53, e20200167.
- DATASUS, disponível em: <https://opendatasus.saude.gov.br/about>
- Posit team (2023). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL <http://www.posit.co/>.
- Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, Guo L, Guo R, Chen T, Hu J, Xiang Z, Mu Z, Chen X, Chen J, Hu K, Jin Q, Wang J, Qian Z. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun.* 2020 Mar 27;11(1):1620. doi: 10.1038/s41467-020-15562-9. Erratum in: *Nat Commun.* 2021 Apr 1;12(1):2144. PMID: 32221306; PMCID: PMC7100515.