

UTILIZAÇÃO DAS FERRAMENTAS DE *BUSINESS INTELLIGENCE* NA ANÁLISE DE DADOS ONCOLÓGICOS

¹ Thyanne Thysysanne de Souza Soares Costa; ² Ariel Stephany Meneses da Silva; ³ Taffarel Melo Torres.

¹ Graduanda em Biotecnologia pela Universidade Federal Rural do Semi-Árido – UFERSA

² Graduada em Biotecnologia pela Universidade Federal Rural do Semi-Árido; ³ Docente da Universidade Federal Rural do Semi-Árido – UFERSA;

Área temática: Biotecnologia e Inovação em Saúde

Modalidade: Comunicação Oral Presencial

E-mail dos autores: oncolthyanne@gmail.com¹; ariel.silva8729@alunos.ufersa.edu.br²
taffareltorres@gmail.com³

RESUMO

INTRODUÇÃO: O câncer é um grupo de mais de 100 doenças caracterizadas pelo crescimento desordenado de células, sendo a segunda principal causa de óbito no Brasil. A estimativa para 2023-2025 projeta mais de 700 mil novos casos devido ao envelhecimento populacional e maior exposição a fatores de risco. A análise de dados de registros como o RHC e o TCGA é essencial para compreender a incidência, prevalência, e fatores de risco, além de auxiliar na personalização dos tratamentos e na formulação de políticas públicas mais eficazes. **OBJETIVO:** Avaliar dados disponíveis no Registro Hospitalar de Câncer do INCA aplicando técnicas de Business Intelligence. **MÉTODOS:** O Registro Hospitalar de Câncer (RHC) do INCA contém informações clínicas detalhadas e desidentificadas de pacientes com câncer de todos os estados brasileiros, coletadas entre 1988 e 2019. Esses dados foram analisados com ferramentas de Business Intelligence, como o Power BI. **RESULTADOS:** Esses dashboards facilitam a investigação de estatísticas regionais e possíveis relações de causa e efeito. **DISCUSSÃO:** A criação de dashboards interativos permite visualizar e interpretar grandes volumes de dados de maneira eficiente, proporcionando a compreensão das tendências e padrões da doença. **CONCLUSÃO:** Essas análises são cruciais para identificar fatores de risco como histórico familiar, consumo de tabaco e álcool, e a relação entre ocupação e tipos de câncer.

Palavras-chave: (Banco de dados), (Oncologia), (Epidemiologia).

1 INTRODUÇÃO

O Câncer é o nome de um conjunto de mais de 100 doenças que têm em comum o crescimento desordenado de células que podem ter como características invadir células e tecidos vizinhos. Segundo o Ministério da Saúde, no Brasil, o câncer representa a segunda principal causa de óbito na população adulta, ficando atrás apenas das doenças cardiovasculares (Santos, 2023).

No triênio de 2023-2025 projeta-se que mais de 700 mil novos casos de câncer serão diagnosticados no país. Esse aumento significativo reflete tanto o envelhecimento da população quanto a maior exposição a fatores de risco, como tabagismo, sedentarismo, alimentação inadequada e exposição a agentes cancerígenos. A detecção precoce e o acesso a tratamentos eficazes são cruciais para melhorar as taxas de sobrevivência e reduzir a mortalidade associada à doença (Instituto Nacional do Câncer, 2023).

O crescimento da incidência e da mortalidade do câncer é cada vez mais evidente, derivado das transições demográficas, dos fatores de risco e da epidemiologia local (Sung et al., 2021). O envelhecimento populacional e a urbanização acelerada contribuem para a maior exposição a agentes cancerígenos. Além disso, mudanças nos hábitos de vida, como dieta inadequada e sedentarismo, têm exacerbado esse cenário preocupante.

Esse problema de saúde pública gera um grande volume de dados, abrangendo informações clínicas até dados genômicos. Assim, é necessário desenvolver e aplicar técnicas de análise de dados, com os bancos de dados sendo ferramentas essenciais nesse processo. Existem diversos bancos de dados públicos que contêm informações valiosas para os avanços da pesquisa oncológica, como os Registros Hospitalares de Câncer (RHC) do Instituto Nacional do Câncer, que fornecem dados clínicos, e o The Cancer Genome Atlas (TCGA), que disponibiliza diversos tipos de dados como exemplo, dados genômicos detalhados de vários tipos de câncer (Costa; Bezerra, 2023).

A análise de dados na área da saúde permite a compreensão do problema, proporcionando estimativas mais precisas sobre a incidência e prevalência das doenças. Isso facilita a identificação de padrões e tendências, auxiliando na compreensão das causas e fatores de risco associados ao câncer. Além disso, a análise de dados é crucial para a busca por melhores tratamentos, ajudando a identificar novas terapias e a personalizar os cuidados com base nas características individuais dos pacientes. A utilização eficaz dos dados também pode melhorar significativamente as políticas públicas de saúde. Dessa forma, a análise de dados não só contribui para avanços científicos, mas também para a promoção de um sistema de saúde eficaz (Shikida; Monasterio; Nery, 2021).

2 MÉTODO

O Registro Hospitalar de Câncer (RHC) é um banco de dados brasileiro pertencente ao Instituto Nacional do Câncer (INCA). Esse banco de dados contém informações clínicas detalhadas

de pacientes com câncer, protegendo a identidade dos mesmos para garantir a confidencialidade. Os dados incluem idade, sexo, tratamentos realizados (como quimioterapia e radioterapia), estadiamento do tumor (TNM), histórico de consumo de tabaco e álcool, histórico familiar de câncer, ocupação, hospital de origem, entre outras informações relevantes.

Figura 01: Exemplo de algumas variáveis encontradas no Dicionário das variáveis da base de dados.

Nº do campo na ficha de cadastro	Variável	Descrição	Domínio
44	ALCOOLIS	Histórico de consumo de bebida alcoólica	1.Nunca; 2.Ex-consumidor; 3.Sim; 4.Não avaliado; 8.Não se aplica; 9.Sem informação
22	ANOPRIDI	Ano do diagnóstico	aaaa
42	ANTRI	Ano da triagem	dd/mm/aaaa
base de dados SP	BASDIAGSP	Base mais importante para o diagnóstico do tumor	1.Exame clínico 2.Recursos auxiliares não microscópicos 3.Confirmação microscópica 4.Sem informação
24	BASMAIMP	Base mais importante para o diagnóstico do tumor	1.Clínica; 2.Pesquisa clínica; 3.Exame por imagem; 4.Marcadores tumorais; 5.Citologia; 6.Histologia da metástase; 7.Histologia do tumor primário; 9. Sem informação
47	CLIATEN	Clínicas do primeiro atendimento - entrada do paciente	Codificação segundo Tabela de Clínicas do SisRH
31	CLITRAT	Clínica de início do tratamento	Codificação segundo Tabela de Clínicas do SisRH
SIS	CNS	Número do CNS do Hospital	Codificação segundo tabela do Cadastro Nacional de Estab. de Saúde
32	DATAINITRT	Data do início do primeiro tratamento específico para o tumor, no hospital	dd/mm/aaaa

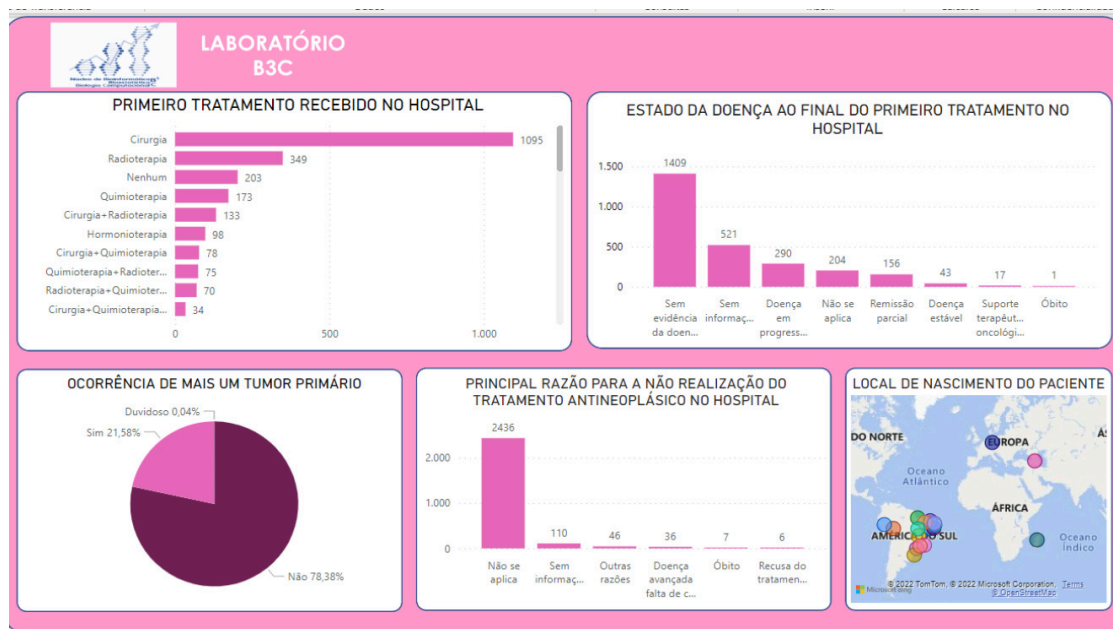
Fonte: Autoria própria adaptado do Sistema de Registro Hospitalar de Câncer (Instituto Nacional do Câncer)

O Registro Hospitalar de Câncer (RHC) abrange informações de todos os estados brasileiros, com dados coletados entre 1988 e 2019. A análise desses dados foi realizada utilizando ferramentas de Business Intelligence, como o Power BI.

3 RESULTADOS

Os dados do Registro Hospitalar de Câncer (RHC) foram extraídos e convertidos em planilhas do Excel, contendo informações desidentificadas dos pacientes. Utilizando o dicionário de variáveis fornecido pelo próprio INCA, esses dados foram tratados e organizados de forma adequada. Posteriormente, os dados tratados foram inseridos no Power BI, permitindo a criação de dashboards interativos.

Figura 02: Gráficos com os dados de 2000 - 2019 de pacientes com câncer (RN)



Fonte: Autoria própria.

Esses dashboards possibilitam realizar diversas análises, proporcionando uma visão abrangente e detalhada das informações sobre o câncer no Brasil. Além disso, permitem a investigação das estatísticas de um determinado tipo de câncer em uma região específica, bem como a análise de possíveis relações de causa e efeito.

4 DISCUSSÃO

A análise dos dados do Registro Hospitalar de Câncer (RHC) utilizando ferramentas de Business Intelligence, como o Power BI, permite contribuições importantes sobre a epidemiologia do câncer no Brasil. A criação de dashboards interativos permite visualizar e interpretar grandes volumes de dados de maneira eficiente, proporcionando a compreensão das tendências e padrões da doença.

Essas análises são relevantes para a identificação de fatores de risco, como histórico familiar, consumo de tabaco e álcool, e a relação entre ocupação e tipos específicos de câncer. Além disso, as ferramentas de Business Intelligence permitem a análise de dados em tempo real, facilitando a identificação de mudanças nas tendências e possibilitando intervenções mais eficazes.

5 CONCLUSÃO

A análise de dados oncológicos utilizando ferramentas como o Power BI representa um avanço significativo na compreensão e no combate ao câncer. Essas ferramentas facilitam a identificação de fatores de risco, a personalização dos tratamentos e a formulação de políticas públicas.

REFERÊNCIAS

Costa, TTSS; Bezerra, LVC. **Bioinformática na luta contra o câncer: os bancos de dados na pesquisa oncológica**. BIOINFO. ISSN: 2764-8273. Vol. 3. p.10 (2023). doi: 10.51780/bioinfo-03-10

SANTOS, Marcell et al. Estimativa de incidência de câncer no Brasil, 2023-2025. **Revista Brasileira de Cancerologia**, v. 69, n. 1, 2023.

Instituto Nacional de Câncer José Alencar Gomes da Silva. ABC do câncer: abordagens básicas para o controle do câncer. 6. ed. rev. atual. Rio de Janeiro: INCA, 2020.

Instituto Nacional de Câncer (Brasil). Estimativa 2023: incidência de câncer no Brasil. Rio de Janeiro: INCA, 2022.

Instituto Nacional de Câncer (Brasil). **Registros Hospitalares de Câncer (RHC)**. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/registros/rhc>. Acesso em: 26 jul. 2024.

SHIKIDA, Claudio D.; MONASTERIO, Leonardo; NERY, Pedro Fernando. Guia brasileiro de análise de dados: armadilhas & soluções. 2021.

SUNG, Hyuna et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: a cancer journal for clinicians**, v. 71, n. 3, p. 209-249, 2021.