UNIVERSIDADE FEDERAL DE CIÊNCIAS DA SAÚDE DE PORTO ALEGRE
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E GESTÃO EM SAÚDE

Felipe de Assis Ribeiro

Visualização de Dados aplicada à Oncologia Pediátrica: Auxiliando a Descobrir Padrões e Tendências

Porto Alegre

2025

Felipe de Assis Ribeiro

Visualização de Dados aplicada à Oncologia Pediátrica para Descoberta de Padrões e Tendências

Dissertação no Programa de Mestrado Acadêmico em Tecnologias da Informação e Gestão em Saúde da Universidade Federal de Ciências da Saúde de Porto Alegre.

Orientadora: Profa. Dra. Isabel Cristina Siqueira da Silva

Co-Orientadora: Profa. Dra. Marialva Sinigaglia

Porto Alegre

2025

Catalogação na Publicação

De Assis Ribeiro, Felipe Visualização de Dados aplicada à Oncologia Pediátrica: Auxiliando a Descobrir Padrões e Tendências / Felipe De Assis Ribeiro. -- 2025.
91 p. : 30 cm.

Dissertação (mestrado) -- Universidade Federal de
Ciências da Saúde de Porto Alegre, Programa de Pós-Graduação em Tecnologias da Informação e Gestão em Saúde, 2025.
Orientador(a): Isabel Cristina Siqueira da Silva; coorientador(a): Marialva Sinigaglia. 1. Análise de Dados. 2. Oncologia Pediátrica. 3. Visualização de Dados. 4. ETL Automatizado. I. Título.

Sistema de Geração de Ficha Catalográfica da UFCSPA com os dados fornecidos pelo(a) autor(a).

AGRADECIMENTOS

Gostaria de agradecer a todos que, de forma direta ou indireta, contribuíram para a realização do meu mestrado.

Começo pelos meus pais, Maria Celina de Assis e Josemar Vasconcelos Ribeiro, que sempre me apoiaram e continuam me apoiando em tudo que faço na vida, e ao meu irmão, Vinicius de Assis Ribeiro, que esteve ao meu lado em diferentes momentos, oferecendo apoio e auxílio.

Agradeço, também, à minha namorada, Julia Gonçalves Kuchle, amor da minha vida, que me apoiou e suportou todos os dias, especialmente durante todo o processo do mestrado. Igualmente, agradeço à minha sogra, Patrícia Gonçalves, e à minha cunhada, Laura Gonçalves Kuchle, que sempre estiveram presentes.

Na sequência, agradeço aos meus amigos, Thawan, Kemilly, Tamara, Thalita, Kaysa, Leonardo, Lara, Alice e Ytalo que, mesmo distante, em Roraima, sempre torceram por mim. Crescemos juntos, desde a primeira série, e é muito gratificante ver essa trajetória de crescimento mútuo. Além dos meus amigos de Roraima, agradeço aos meus amigos de Campina Grande - Pedro, André e Igor, que ainda mantêm contato comigo e sempre torceram por mim.

Gostaria de agradecer, ainda, aos meus tios, tias, primos e primas que me ajudaram durante esses anos longe de Boa Vista, Roraima. Em especial, minha madrinha Maria Assis (Betinha), que sempre me apoiou tirando dúvidas sobre problemas de saúde, tirando dúvidas do dia a dia profissional e sobre casos.

Dedico um agradecimento especial às minhas orientadoras, Isabel Cristina Siqueira da Silva e Marialva Sinigaglia, que foram fundamentais para o desenvolvimento deste mestrado, oferecendo uma orientação dedicada e de alta qualidade além de incentivo às melhores práticas de pesquisa. Agradeço, também, aos meus amigos Alexandro Ávila, José Burmann e Vinicius Vicente, que me auxiliaram no desenvolvimento do produto construído nesse mestrado.

Destaco, ainda, um agradecimento à Virgínia de Nóbrega e outros profissionais do Instituto do Câncer Infantil (ICI), os quais me auxiliaram a compreender melhor o universo do câncer infantojuvenil e o Sistema Único de Saúde (SUS); e ao Eliel Pereira, que me recebeu de braços abertos no Posto de Atendimento Cruzeiro do Sul e foi um grande parceiro na luta pela obtenção dos dados junto à Companhia de Processamento de Dados de Porto Alegre (PROCEMPA).

Por fim, agradeço à Carla Maria Dal Sasso Freitas e à Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) pela oportunidade de participar do projeto CIARS - Inteligência Artificial aplicada à Saúde e à bolsa de estudos recebida junto a este; e finalizo, agradecendo a todos os professores e professoras que marcaram minha trajetória acadêmica, despertando em mim o desejo contínuo de aprendizado e aprimoramento.

A todos vocês, o meu muito obrigado!

RESUMO

A análise de dados aplicada à saúde pode contribuir, de forma significativa, para o processo de gestão da jornada do paciente, desde a realização de um diagnóstico até o período de pós-tratamento. No entanto, a utilização eficaz dos dados enfrenta desafios substanciais devido à heterogeneidade e à diversidade destes, além da falta de estruturação dos registros de saúde. Questões como ausência de padronização, variabilidade na qualidade dos dados e preocupações com a privacidade surgem como obstáculos críticos a serem superados. Neste sentido, este estudo teve, por objetivo, investigar padrões epidemiológicos, tanto em bases de dados públicas como em privadas, a partir da extração e do processamento de dados relacionados à oncologia pediátrica. A partir dos dados coletados, tratados e integrados, foi desenvolvido um sistema *web* composto por quatro *dashboards* interativos, os quais apresentam diferentes tipos de representações visuais interativas de dados de oncologia pediátrica. Procurou-se, através da integração e da visualização de dados, auxiliar a tomada de decisão por parte de profissionais e gestores da área da saúde, que atuam na prevenção e tratamento e combate ao câncer infantojuvenil. A usabilidade do sistema proposto foi avaliada por seis especialistas, que atuam em oncologia pediátrica, e constatou-se a satisfação destes com os *dashboards* interativos em termos de exibição de conexões entre dados que não eram possíveis a partir das bases de dados disponíveis. O sistema, já com informações anonimizadas, está disponível para acesso público, visando auxiliar, principalmente, na tomada de decisão estratégica relacionada ao tratamento e combate do câncer infantojuvenil.

Palavras-Chave: Oncologia Pediátrica, Análise de Dados, Visualização de Dados, ETL Automatizado, Usabilidade.

ABSTRACT

Data analysis applied to healthcare can significantly contribute to the management of the patient journey throughout the entire process, from diagnosis to post-treatment. However, the effective use of this data faces substantial challenges due to the heterogeneity and diversity of these data, in addition to the lack of structure in health records. Issues such as lack of standardization, variability in data quality and privacy concerns emerge as critical obstacles to be overcome. This study aimed to investigate epidemiological patterns, both in public and private databases, based on the extraction and processing of data related to pediatric oncology. Based on the collected, processed and integrated data, a web system was developed consisting of four interactive dashboards, which present different types of interactive visual representations of pediatric oncology data. The aim was to use data visualization to assist decision-making by healthcare professionals and managers who work in the prevention and treatment of childhood cancer. The usability of the proposed system was evaluated by six specialists who work in pediatric oncology, and it was found that they were satisfied with the interactive dashboards in terms of displaying connections between data that were not possible using the available databases. The system is available for public access, since the information collected from the aforementioned databases is already anonymized, and aims to assist, mainly, in strategic decision-making related to the treatment and prevention of childhood cancer.

Keywords: Pediatric Oncology; Data Analysis; Data Visualization; Automated ETL; Usability.

LISTA DE TABELAS

Tabela 1 - Resumo dos trabalhos relacionados 28
Tabela 2 - Problemas encontrados em cada base de dados 41

LISTA DE FIGURAS

Figura 1- Painel de dados sobre oncologia do Estado de Goiás.....23
Figura 2 - Painel de dados sobre a COVID-19.....23
Figura 3 - Etapas de Execução do Projeto.....33
Figura 4 - Mapa de Evidências em Oncologia.....34
Figura 5 - Bases de dados de oncologia pediátrica empregadas neste estudo.....36
Figura 6 - Integração de bases de dados de oncologia pediátrica em quatro dashboards.....40
Figura 7- Pipeline da extração, processamento, integração e visualização dos dados de oncologia pediátrica.....44
Figura 8 - Parte inicial do dashboard exibindo dados do painel de oncologia do DATASUS (SIA, SIH e SISCAN) integrados ao CNES do DATASUS..... 46
Figura 9 - Parte inicial do dashboard exibindo dados do RHC integrados ao CNES do DATASUS.....46
Figura 10 - Parte inicial do dashboard exibindo dados do SIM do DATASUS integrados ao CNES do DATASUS.....47
Figura 11 - Parte inicial do dashboard exibindo dados da SMSPOA..... 47
Figura 12 - Gráfico Sunburst..... 48
Figura 13 - Gráfico Sankey.....48
Figura 14 - Tempo de atuação na Oncologia Pediátrica..... 50
Figura 15 - Distribuição de Profissões/Áreas de Atuação..... 50
Figura 16 - Distribuição das Respostas SUS (Q1–Q10) com Desvio-Padrão.....51
Figura 17 - SUS Escore Final.....52
Figura 18 - Classificações de percentis se associam às pontuações do SUS.....53

LISTA DE ABREVIATURAS E SIGLAS

CCDI	*Childhood Cancer Data Initiative*
CEP	Comitê de Ética em Pesquisa
CIARS	Rede Centro de Inovação em Inteligência Artificial para a Saúde
DATASUS	Departamento de informática do Sistema Único de Saúde do Brasil
FAPERGS	Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul
GERCON	Gerenciamento de Consultas
GERINT	Gerenciamento de Internações Hospitalares
GERPAC	Gerenciamento de Procedimentos Ambulatoriais
GLOBOCAN	Observatório Global do Câncer
IARC	Agência Internacional de Pesquisa sobre o Câncer
ICI	Instituto do Câncer Infantil
INCA	Instituto Nacional do Câncer
OPAS	Organização Pan-Americana da Saúde
PROCEMPA	Empresa de Tecnologia da Informação e Comunicação do Município de Porto Alegre
PICO	População, Intervenção, Comparação e Desfechos
RCBP	Registros de câncer de base populacional
RHC	Registros hospitalares de câncer
RITES	Redes Inovadoras de Tecnologias Estratégicas
SIM	Sistema de Informação de Mortalidade
SIA	Sistemas de Informação Ambulatorial
SIH	Sistema de informação hospitalar
SISCAN	Sistema de informação do câncer
SUS	Sistema Único de Saúde
SUS	*System Usability Scale*
SMSPA	Secretaria Municipal de Saúde de Porto Alegre
UFCSPA	Universidade Federal de Ciências da Saúde de Porto Alegre
VA	Visualização Analítica de Dados

SUMÁRIO

1 INTRODUÇÃO 12
2 OBJETIVOS 16
2.1 Objetivo Geral 16
2.2 Objetivos Específicos 16
3 REFERENCIAL TEÓRICO 17
3.1 O Câncer Infantojuvenil 17
3.2 Ciência de Dados em Saúde 18
3.3 Bancos de Dados Oncológicos 19
3.4 Visualização de Dados em Saúde 21
4 TRABALHOS RELACIONADOS 25
5 MATERIAIS E MÉTODOS 30
5.1 Revisão Bibliográfica 30
5.2 Análise e Seleção de Bases de Dados 30
5.3 Coleta de Requisitos 31
5.4 Pré-processamento e Transformação dos Dados 31
5.5 Análise dos Dados 32
5.6 Visualização de Dados 32
5.7 Avaliação de Usabilidade 32
6. RESULTADOS 34
6.1 Criação do Mapa de Evidências 34
6.2 Definição das Bases de Dados 35
6.3 Coleta de Requisitos do Sistema 37
6.4 Pré-Processamento e Transformação dos Dados 40
6.5 Visualização de Dados 42
6.6 Implementação do Sistema 43
6.7 Avaliação de Usabilidade 49
7. DISCUSSÃO 54
8 CONSIDERAÇÕES ÉTICAS 56
9 PUBLICAÇÕES 57
10 CONCLUSÕES 58
REFERÊNCIAS 60
APÊNDICE A - Documentação PainelOncoped 64
APÊNDICE B – Pareceres dos Comitês de Ética em Pesquisa 72
APÊNDICE C – Formulário SUS 88

1 INTRODUÇÃO

A transformação digital e o avanço acelerado das tecnologias da informação têm impulsionado uma geração massiva de dados em diversos setores da sociedade. Esse fenômeno envolve grandes volumes de dados produzidos de forma contínua e em diferentes formatos, exigindo novas abordagens para seu processamento e análise.

Na área da saúde, essa realidade assume um papel cada vez mais relevante com a aplicação da ciência de dados sobre informações provenientes de serviços, sistemas e práticas de saúde (Marr, 2015). Esse campo abrange a coleta, o armazenamento e a análise de grandes conjuntos de dados com o objetivo de apoiar a organização, o gerenciamento e a recuperação da informação, além de possibilitar a geração de novos conhecimentos (Obermeyer & Emanuel, 2016; Jensen, Jensen & Brunak, 2012).

A importância dessa abordagem se destaca, especialmente, por seu potencial em transformar dados em evidências úteis para a tomada de decisões clínicas e administrativas, contribuindo para a melhoria da qualidade dos serviços de saúde e promovendo discussões sobre o uso ético e estratégico dessas informações (Khan & Qin, 2018; Marr, 2015). A ciência de dados aplicada à saúde facilita a análise desses dados, gerando *insights* e auxiliando na identificação de padrões que podem contribuir para a melhoria na qualidade dos cuidados com a saúde, otimizando o desenvolvimento de medicamentos, personalizando tratamentos e, até mesmo, antecipando surtos de doenças (Li, Lu & Shu, 2018).

No entanto, a natureza heterogênea, diversificada e, muitas vezes, não estruturada dos dados de saúde impõe obstáculos à sua utilização de forma eficaz. Nesse contexto, a falta de padronização, a qualidade variável dos dados e as questões de privacidade emergem como problemas críticos a serem enfrentados (Ienca et al., 2018).

Nota-se, assim, que à medida que os dados de saúde são gerados em ritmo exponencial, as consequências da não aplicação de métodos adequados de tratamento tornam-se ainda mais evidentes. A presença de ruído, valores aberrantes e inconsistências nos dados podem levar a diagnósticos incorretos, tomadas de decisões equivocadas que prejudicam a integridade das pesquisas científicas (Koh et al., 2013). Além disso, questões de privacidade e segurança dos dados ganham

destaque à medida que informações sensíveis dos pacientes são coletadas e compartilhadas em plataformas de análise de dados (Khan & Qin, 2018).

Diante desse cenário, a necessidade de tomar decisões eficazes e precisas torna-se essencial para profissionais de saúde bem como gestores e pesquisadores. Nesse contexto, destaca-se o papel das ferramentas de apoio à tomada de decisão baseadas em dados. A ciência de dados aplicada à saúde possibilita a criação de modelos preditivos e descritivos que auxiliam na previsão de tendências em saúde, identificação precoce de riscos e na personalização dos cuidados de acordo com as necessidades individuais de cada paciente (Batko, 2022). Portanto, a implementação eficaz dessas ferramentas requer uma abordagem multidisciplinar que englobe expertise em saúde, estatística, ciência da computação e ética (Batko, 2022).

Em relação à oncologia, atualmente, existem diferentes bases de dados essenciais para a compreensão da incidência de câncer na população brasileira em geral, tanto públicas quanto privadas. Dentre as públicas, destaca-se o painel oncológico do departamento de informática do Sistema Único de Saúde (DATASUS) (Atty, 2020), que reúne dados de diferentes sistemas, como, por exemplo, o sistema de informação ambulatorial (SIA), o sistema de informação hospitalar (SIH), o sistema de informação do câncer (SISCAN), o sistema de informação de mortalidade (SIM) e o cadastro nacional de estabelecimentos de saúde (CNES). Além do DATASUS, os registros hospitalares de câncer (RHC) e os registros de câncer de base populacional (RCBP) são outras importantes fontes de dados para estudos epidemiológicos da doença.

Tumores infantis diferem dos tumores de adultos quanto ao aspecto morfológico, comportamento clínico e localização primária. Assim, para o entendimento do cenário do câncer infantojuvenil, é necessário extrair, das bases de dados oncológicos, apenas as informações relacionadas à incidência de câncer em crianças e adolescentes a fim de pesquisá-las e investigar padrões epidemiológicos. Porém, dados sobre oncologia pediátrica apresentam obstáculos consideráveis na construção de uma epidemiologia abrangente em função da natureza heterogênea dos registros de dados e das complexidades que envolvem sua coleta.

O registro dos dados, por exemplo, nem sempre é realizado por profissionais que compreendem as especificidades destes e, por vezes, sintomas de câncer infantil podem ser confundidos com outras doenças comuns na infância quando uma investigação mais aprofundada não é realizada. A ausência de *feedbacks* sobre o uso

desses dados na construção de políticas públicas também é um fator que pode comprometer a percepção da importância da coleta e do registro de dados oncológicos, principalmente os pediátricos. Como consequência, os dados podem apresentar inconsistências, as quais, por sua vez, podem levar a diagnósticos incorretos e tomadas de decisões equivocadas, comprometendo a integridade dos dados e tomada de decisão.

Neste contexto, este projeto propôs a análise e a integração dos dados públicos presentes nas bases do DATASUS e do RHC, com foco nos casos de tumores pediátricos, de modo a propor um novo modelo de dados que permitisse o mapeamento e a integração de diferentes bases de conhecimento. Além destas bases públicas, este estudo integrou, também, dados privados oriundos do sistema de regulação da Secretaria Municipal de Saúde de Porto Alegre (SMSPA). Este sistema, abrange uma vasta gama de informações, coletadas através de sistemas de gerenciamento de internações hospitalares do SUS (GERINT), de gerenciamento de consultas especializadas do SUS (GERCON), gerenciamento de procedimentos de alta complexidade (GERPAC) e sistema de informação sobre mortalidade (SIM).

A partir de então, técnicas de visualização analítica de dados foram empregadas através de *dashboards* interativos para auxiliar gestores de saúde na análise dos dados e na tomada de decisões. Ao explorar a riqueza de informações contida em diferentes bancos de dados de saúde, o projeto almejou, além de fornecer *insights* sobre a prevalência e características dos tumores pediátricos, estabelecer um modelo de dados replicável que possa ser estendido para análises de outras enfermidades. Dessa maneira, a proposta não apenas se concentrou na análise de dados, mas também na construção de um mecanismo que possibilitasse uma compreensão mais profunda do panorama de saúde oncopediatria da região da grande Porto Alegre, buscando auxiliar a busca constante da qualidade da assistência médica oferecida à população.

Destaca-se, ainda, que este estudo pretende contribuir para o avanço da área de informática em saúde e de oncologia pediátrica seguindo as diretrizes do Manual de Oslo (OCDE, 2018) para medição e fomento à inovação. Ao propor um modelo de integração de dados replicável, o projeto visa estimular práticas inovadoras de pesquisa, desenvolvimento e difusão de soluções tecnológicas em saúde. Tais questões também indicam o alinhamento do presente estudo Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU), em

especial o ODS 3, referente à saúde e ao bem-estar.

Como pergunta de pesquisa, tem-se a seguinte:

“A visualização de dados em oncologia pediátrica pode auxiliar a investigação de padrões epidemiológicos em bases públicas e privadas?

Por hipótese, esse projeto considera que técnicas de visualização de dados, quando aplicadas à oncologia pediátrica, auxiliam no processo de tomada de decisão por parte de profissionais e gestores que atuam na área da saúde.

Este projeto de mestrado integra a Rede Centro de Inovação em Inteligência Artificial para a Saúde (CIARS) que, por sua vez, faz parte do programa Redes Inovadoras de Tecnologias Estratégicas (RITEs), criado pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS).

2 OBJETIVOS

Neste capítulo, são apresentados o objetivo geral e os específicos desta pesquisa.

2.1 Objetivo Geral

Esta pesquisa teve, por objetivo investigar padrões epidemiológicos, tanto em bases de dados públicas como em privadas, a partir da extração e processamento de dados relacionados à oncologia pediátrica. A partir dos dados coletados, tratados e integrados, foi desenvolvido um sistema *web* composto por quatro *dashboards* interativos, os quais apresentam diferentes tipos de representações visuais interativas de dados de oncologia pediátrica. Procurou-se, através da visualização de dados, auxiliar a tomada de decisão de profissionais e gestores da área da saúde, que atuam na prevenção e tratamento e combate ao câncer infantojuvenil.

2.2 Objetivos Específicos

Por objetivos específicos, este estudo teve:

- Identificar as evidências disponíveis que auxiliam na tomada de decisão por parte de profissionais e gestores da área da saúde ao utilizar visualização analítica de dados;
- Entender as diferentes informações relacionadas à oncologia pediátrica presentes nas bases de dados, anonimizadas, tanto as bases públicas quanto as bases disponibilizadas pela Secretaria Municipal de Saúde de Porto Alegre;
- Identificar técnicas de visualização analítica de dados que possam ser empregadas em um *dashboard* para auxiliar gestores em saúde na análise de informações relacionadas à oncologia pediátrica;
- Projetar, desenvolver e testar um sistema interativo, voltado a profissionais e gestores em saúde, a fim de auxiliar na compreensão dos diferentes aspectos envolvidos no diagnóstico e tratamento relacionados à oncologia pediátrica.

3 REFERENCIAL TEÓRICO

Para a construção do referencial teórico, foram adotados critérios de inclusão que abrangeram a pesquisa de artigos científicos em bases de dados reconhecidas na área da saúde, tais como PubMed, Scopus, Web of Science, SciELO e também fontes nacionais como o Instituto Nacional do Câncer (INCA). A busca considerou publicações no intervalo temporal de 2018 a 2025, visando garantir a atualização e relevância das informações utilizadas.

3.1 O Câncer Infantojuvenil

A sobrevida em 5 anos de crianças e adolescentes com diagnóstico de câncer é de aproximadamente 80% nos países de alta renda, entretanto essas estimativas são reduzidas em países de baixa e média renda, onde vivem mais de 80% das crianças com câncer. Devido à melhora no tratamento de doenças infecciosas e não transmissíveis nesses países, o tratamento do câncer tornou-se o foco principal, sendo uma das principais causas de morte em crianças e adolescentes (Lupo, 2022).

No Brasil, o câncer infantil é a primeira causa de morte por doença em crianças de 1 a 19 anos, superando as doenças infecciosas. O Instituto Nacional do Câncer (INCA) estimou 704 mil novos casos de câncer pediátrico no Brasil para cada ano do triênio 2023-2025, com estimativa de sobrevida de 64%, aliada a uma enorme disparidade entre as regiões (Inca, 2023). Os tumores infantis diferem dos de adultos quanto aos seus aspectos morfológicos, comportamento clínico e localizações primárias e, por isso, devem ser estudados e tratados separadamente.

Crianças tratadas em centro de oncologia pediátrica possuem 80% de chances de cura, contudo, quando são tratadas em protocolos e centros de adultos, a chance de cura cai para 40%. As chances de cura efetiva no câncer infantil dependem, ainda do tipo de câncer e estão intimamente relacionadas com a capacidade de diagnosticar precocemente e com precisão, com o tratamento dentro de protocolos clínicos e equipe multidisciplinar especializada em oncologia pediátrica e com a eficácia em detectar doença metastática ou recidivada (Inca, 2022).

Além destes, existem outros fatores que também desempenham um papel crítico na qualidade e cuidado do tratamento desses pacientes. Como exemplo, pode-se citar o acesso precário ao atendimento integral ao paciente, o diagnóstico tardio,

os programas educacionais inadequados, a falta de decisões baseadas em protocolo e especialização do centro, a indisponibilidade de equipes multidisciplinares e a escassez de recursos públicos e políticas governamentais específicas (Lupo, 2022).

Na área da oncologia pediátrica, existem obstáculos consideráveis na construção de uma epidemiologia abrangente e precisa. Isso se deve, em grande parte, à natureza incongruente dos registros de dados e às complexidades que envolvem a coleta destes (Teixeira, 2019). As etapas cruciais de registro e coleta de informações frequentemente se deparam com desafios, pois os responsáveis nem sempre compreendem plenamente a importância de cada informação coletada.

A ausência de *feedbacks* sobre o uso desses dados na construção de políticas públicas pode levar a uma lacuna dessa compreensão. No entanto, é vital reconhecer que os dados constituem o ponto de partida para a elaboração de qualquer política, estratégia ou programa, independentemente da área de atuação, incluindo a saúde.

3.2 Ciência de Dados em Saúde

A análise de dados na área da saúde tem sido fundamental na detecção precoce de doenças, permitindo a identificação de padrões e sinais precoces que podem indicar condições médicas (Wang, 2012). Algoritmos de aprendizado de máquina têm sido aplicados com sucesso na interpretação de imagens médicas, como radiografias e ressonâncias magnéticas, para auxiliar no diagnóstico de doenças como câncer e doenças cardiovasculares (Esteva, 2017).

A ciência de dados permite a personalização dos cuidados de saúde com base nas características individuais de cada paciente (Obermeyer & Emanuel, 2016). Modelos preditivos podem ser desenvolvidos para prever a resposta de pacientes a determinados tratamentos, considerando fatores como histórico médico, perfil genético e estilo de vida (Rajkomar, 2018). Isso possibilita a oferta de cuidados mais eficazes e direcionados, melhorando os resultados para os pacientes.

A análise de dados, em larga escala, tem sido importante para a compreensão de tendências epidemiológicas e para a identificação de fatores de risco para doenças (Hripcsak & Albers, 2013). Por meio da mineração de dados de saúde pública, os pesquisadores podem identificar padrões de propagação de doenças, prever surtos e implementar medidas preventivas de forma proativa (Brownstein, 2010). Essas análises fornecem importantes informações para aprimorar os cuidados paliativos e o

suporte aos pacientes, abordando questões como efeitos tardios do tratamento e necessidades psicossociais (LeCun, 2015).

Na gestão eficiente de recursos e na otimização dos serviços de saúde, a análise de dados também exerce um papel fundamental (Obermeyer & Emanuel, 2016). Ao se analisar dados de utilização de serviços de saúde, é possível identificar áreas de melhoria, reduzir custos desnecessários e garantir que os recursos sejam alocados de forma eficaz para atender às necessidades da população (Wang et al., 2018).

Apesar dos benefícios significativos, a utilização da ciência de dados em saúde também apresenta desafios importantes, como questões de privacidade, segurança e interpretação responsável dos resultados (Beam & Kohane, 2018). É essencial garantir que os dados sejam utilizados de forma ética e transparente, respeitando os direitos e a dignidade dos pacientes.

3.3 Bancos de Dados Oncológicos

Atualmente, existe um banco de dados global sobre câncer denominado Observatório Global do Câncer (GLOBOCAN), desenvolvido pela Agência Internacional de Pesquisa sobre o Câncer (IARC). Este observatório abrange dados de 185 países, oferecendo informações estatísticas detalhadas sobre a incidência, mortalidade e prevalência de diversos tipos de câncer, tanto em adultos quanto em crianças. Essas informações são essenciais para compreender o impacto global da doença, permitindo análises comparativas entre diferentes países e regiões e auxiliando na definição de estratégias internacionais de combate ao câncer.

No que diz respeito aos tumores pediátricos, diferentes iniciativas têm sido desenvolvidas ao redor do mundo visando a criação de bancos de dados dedicados e especializados. Nos Estados Unidos, destaca-se a *Childhood Cancer Data Initiative* (CCDI), implementada pelo Instituto Nacional do Câncer. A CCDI tem por objetivo centralizar e disponibilizar dados abrangentes sobre câncer infantil, impulsionando pesquisas que buscam acelerar a descoberta de novos tratamentos, melhorar as estratégias terapêuticas existentes e promover resultados mais positivos para crianças e adolescentes com câncer, além de oferecer suporte a sobreviventes e suas famílias (Flores-Toro, 2023).

Na Alemanha, existe o Registro Alemão de Câncer Infantil, hospedado pelo

Instituto de Estatística Médica, Epidemiologia e Informática do Centro Médico Universitário da Universidade Johannes Gutenberg, em Mainz. Este registro cobre aproximadamente 95% do território alemão, coletando informações detalhadas desde 2009. Este registro é amplamente reconhecido por sua abrangência e precisão, permitindo a realização de estudos epidemiológicos aprofundados e o acompanhamento da qualidade do atendimento e a eficácia dos tratamentos disponibilizados aos pacientes pediátricos no país (Kaatsch, 2010).

No Brasil, o Instituto Nacional do Câncer (INCA) administra dois importantes bancos de dados relacionados à oncologia: o Registro de Câncer de Base Populacional (RCBP) e o Registro Hospitalar de Câncer (RHC).

O RCBP é composto por mais de 20 registros ativos espalhados pelo país, com concentração especial na região sudeste do país. Seu principal objetivo é fornecer uma visão abrangente sobre a incidência, distribuição geográfica e tendências temporais dos diversos tipos de câncer na população brasileira, contemplando dados provenientes tanto da rede pública quanto da rede privada. As informações do RCBP subsidiam importantes estudos epidemiológicos, ajudam a identificar populações mais vulneráveis e servem como base para avaliar e aprimorar programas de prevenção e controle do câncer em nível nacional (INCA, 2022).

Já o RHC tem um enfoque clínico especificamente voltado ao monitoramento e avaliação dos procedimentos realizados em hospitais da rede pública. O RHC coleta dados detalhados sobre os tratamentos administrados, a qualidade assistencial oferecida aos pacientes e os resultados clínicos alcançados. Este registro permite aos gestores hospitalares e pesquisadores realizarem análises críticas sobre a eficácia e eficiência dos tratamentos, identificando pontos de melhoria na prestação dos serviços oncológicos (INCA, 2022).

Outra fonte que se destaca é o Painel Oncológico (PO), gerenciado pelo Departamento de Informática do Sistema Único de Saúde (DATASUS). Este painel surgiu como resposta direta à Lei nº 12.732/2012, popularmente conhecida como "Lei dos 60 dias", que assegura aos pacientes diagnosticados com câncer o direito de iniciar seu tratamento em até 60 dias após o diagnóstico confirmado. O PO integra diferentes bases de dados existentes no DATASUS, permitindo o acompanhamento em tempo real do intervalo entre o diagnóstico e o início do tratamento. Além disso, PO é o único sistema brasileiro que proporciona uma visão integrada dos dados oriundos de diferentes fontes dentro do sistema público de saúde, sendo uma

ferramenta estratégica para a gestão pública e para o controle social da qualidade e eficiência da assistência oncológica prestada no país (Atty, 2020).

Por fim, destaca-se, também, as bases de dados privadas provenientes da Prefeitura de Porto Alegre, especificamente do complexo regulador formado pelos sistemas Gerenciamento de Internações Hospitalares (GERINT), Gerenciamento de Consultas (GERCON) e Gerenciamento de Procedimentos Ambulatoriais (GERPAC), que constituem importantes fontes de dados relacionados à oncologia. Embora o objetivo principal desses sistemas seja realizar o gerenciamento e a regulação da fila de espera do Sistema Único de Saúde (SUS), eles acabam funcionando também como repositórios valiosos de informações clínicas e administrativas dos pacientes oncológicos. Entre os dados disponíveis, estão registros referentes à primeira consulta do paciente, datas de início do tratamento, quantidades e tipos de procedimentos realizados, além de indicadores essenciais para o acompanhamento assistencial e epidemiológico.

Essas informações, inicialmente coletadas para fins administrativos e regulatórios, proporcionam um panorama abrangente sobre a jornada do paciente oncológico dentro do sistema de saúde municipal. Dessa forma, tornam-se especialmente relevantes para estudos epidemiológicos, análises de eficiência e eficácia dos tratamentos, bem como para a formulação de políticas públicas e estratégias para melhoria do atendimento ao paciente com câncer.

3.4 Visualização de Dados em Saúde

A visualização de dados tem se tornado cada vez mais relevante na área da saúde, especialmente diante do crescimento exponencial dos dados clínicos e epidemiológicos. Em cenários como o monitoramento de surtos, a avaliação de políticas públicas ou a gestão hospitalar, o uso dessas técnicas visuais permite interpretar informações complexas com muito mais agilidade. Representações visuais interativas podem auxiliar profissionais da saúde a identificar padrões, tendências e *outliers* nos dados de maneira rápida e intuitiva (Milani et al., 2020).

O maior potencial da visualização está justamente em combinar a percepção e o raciocínio humano com o poder gráfico computacional. Como destaca Freitas et al. (2001), essa sinergia entre humanos e máquinas permite extrair conhecimento de forma mais eficaz, principalmente quando lidamos com grandes volumes de dados

não estruturados ou heterogêneos, o que é bastante comum na saúde. Com uma boa interface visual, dados que antes pareciam caóticos passam a fazer sentido e a orientar decisões mais embasadas.

Na prática, a visualização de dados tem sido aplicada com sucesso em contextos como o acompanhamento de leitos hospitalares, a análise da cobertura vacinal, a comparação de taxas de mortalidade entre regiões e a avaliação de desfechos clínicos. Quando diferentes representações visuais são integradas em um *dashboard,* o qual constitui um painel interativo, é possível visualizar dados relacionados a comorbidades, à faixa etária e ao tempo médio de internação, a fim de apoiar decisões de gestores sobre alocação de recursos ou priorização de atendimentos (Milani et al., 2020).

Outro ponto importante é que a visualização não deve ser aplicada apenas ao final da análise, como forma de apresentação. Seu uso já nas fases iniciais, especialmente no pré-processamento dos dados, pode facilitar a identificação de problemas como valores faltantes, dados duplicados ou inconsistência de variáveis. Representações visuais como *diagramas de caixas*, gráficos de densidade ou, mesmo, gráficos de valores ausentes, ajudam a detectar esses problemas de forma visual, acelerando o processo de limpeza e preparação dos dados (Milani et al., 2020).

Além disso, ao integrar dados de múltiplas fontes, como sistemas hospitalares, registros eletrônicos de saúde (EHRs), bases públicas e sensores ambientais, a visualização torna-se ainda mais poderosa. Representações como *glyphs* e diagramas ampliam a capacidade analítica dos profissionais, permitindo explorar relações não óbvias e descobrir padrões escondidos nos dados (Freitas et al., 2001). Essa integração pode ser observada em exemplos práticos, como:

- Painel de dados sobre oncologia do Estado de Goiás[1] , que reúne informações sobre incidência, mortalidade, tipos de tumor e distribuição geográfica dos casos, permitindo análises estratégicas para planejamento da atenção oncológica (Figura 1);

- Painel de dados sobre a COVID-19[2], que apresenta em tempo real dados sobre

[1] https://indicadores.saude.go.gov.br/public/oncologia.html

[2] https://coronavirus.jhu.edu/map.html

casos confirmados, óbitos, vacinação e ocupação de leitos, sendo fundamental no enfrentamento da pandemia e na tomada de decisões rápidas por parte das autoridades de saúde (Figura 2).

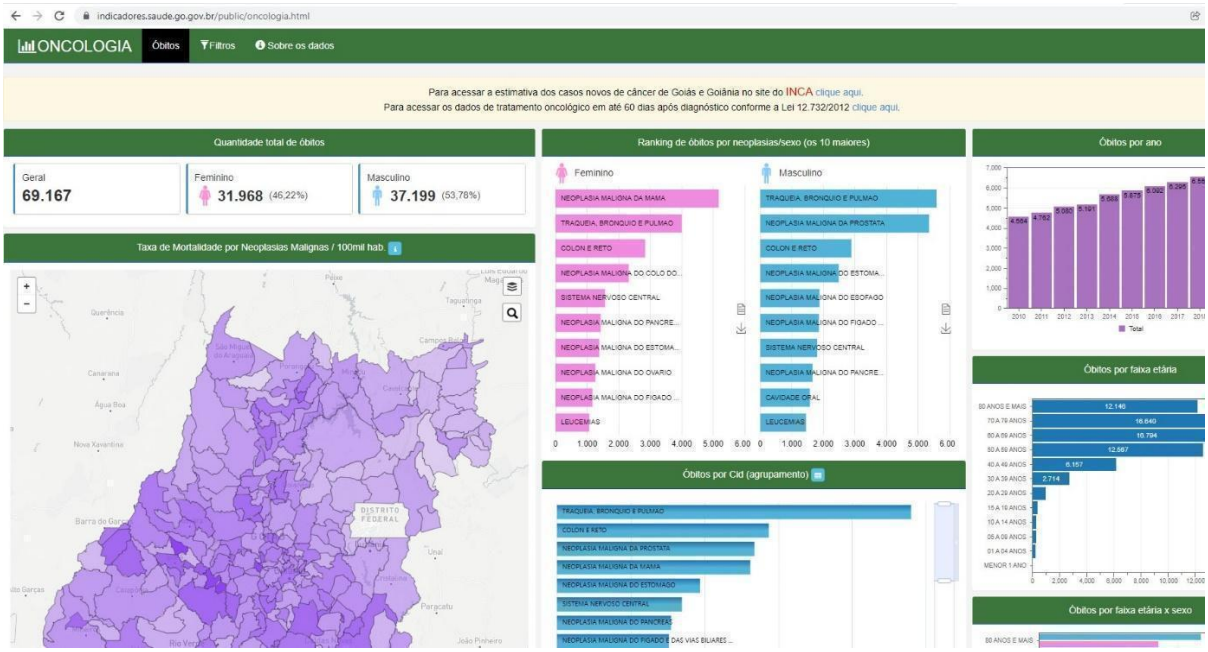


***Figura 1** - Painel de dados sobre oncologia do estado de Goiás*

Fonte: Secretaria de Estado da Saúde de Goiás

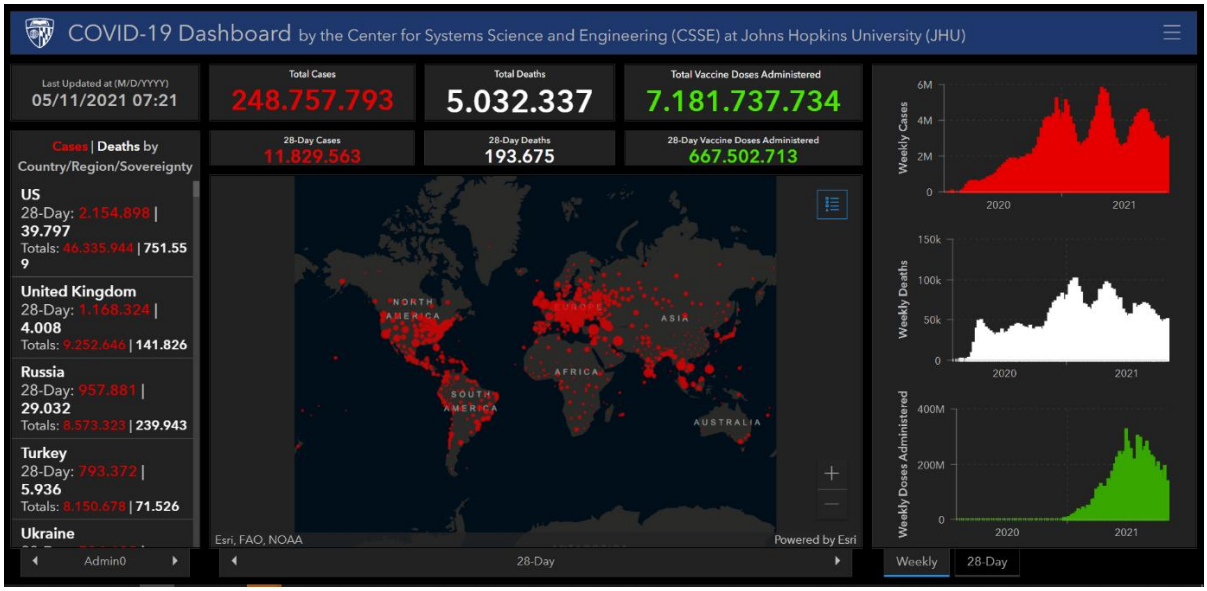


***Figura 2** - Painel de dados sobre a COVID-19*

Fonte: Universidade Johns Hopkins (JHU)

A partir do exposto, nota-se que a visualização de dados aplicada à saúde constitui uma ferramenta estratégica, a qual pode contribuir para a realização de diagnósticos mais rápidos, para a tomada de decisões clínicas mais assertivas e para

a proposição de políticas públicas mais eficazes. Ao tornar os dados mais acessíveis e compreensíveis, a visualização de dados fortalece iniciativas de cuidado centrado no paciente, aprimorando a gestão dos serviços de saúde.

4 TRABALHOS RELACIONADOS

Este capítulo discute-se alguns dos estudos que embasaram esta proposta de projeto. Estes foram elencados a partir de critérios de inclusão que abrangeram a busca por artigos científicos em bases de dados reconhecidas na área da saúde, como PubMed, Scopus, Web of Science, SciELO, e fontes nacionais como o Instituto Nacional do Câncer (INCA). A pesquisa considerou publicações no período de 2018 a 2025, com o objetivo de garantir a atualidade e relevância das informações utilizadas em temas como oncologia pediátrica, epidemiologia, coleta e análise de dados em saúde, e visualização de dados em saúde.

O estudo apresentado por (Buhlinger et al., 2021) teve como objetivo realizar uma análise pré-implementação do registro nacional de câncer pediátrico da Etiópia, visando melhorar a coleta e análise de dados sobre câncer em crianças no país, tendo sido delineado em torno de três objetivos específicos. Primeiramente, buscou-se caracterizar a ferramenta atual baseada em planilhas utilizada nos locais participantes, incluindo a análise das variáveis coletadas em comparação com os padrões estabelecidos pela Organização Mundial da Saúde (OMS), além de descrever a inserção e a completude dos dados. Em segundo lugar, avaliou-se as percepções da equipe hospitalar em relação a um registro aprimorado. Por fim, foram identificadas lacunas no fluxo de trabalho relacionadas à documentação.

A equipe e a liderança hospitalar manifestaram percepções positivas em relação a um registro aprimorado de câncer pediátrico, o que foi reforçado por interações adicionais. A avaliação do fluxo de trabalho revelou diversas lacunas, as quais foram abordadas sistematicamente por meio de uma abordagem científica de implementação em três fases. Além disso, a avaliação evidenciou que a ferramenta existente baseada em planilhas não incluía todas as variáveis recomendadas pela OMS e apresentava inconsistências no preenchimento devido a lacunas no fluxo de trabalho.

Já o estudo de (Jeanquartier et al., 2019) teve por objetivo investigar a disponibilidade de recursos na web relacionados ao câncer pediátrico. Para tanto, foi conduzida uma análise comparativa entre os recursos gerais e aqueles especificamente direcionados ao estudo dessa área. Onze recursos específicos do câncer pediátrico, incluindo ferramentas e bases de dados, foram identificados e comparados com outras fontes abertas de informações biomédicas. A metodologia

envolveu a avaliação de cinco casos de uso, centrados no exemplo da pesquisa de tumores cerebrais, abrangendo cenários definidos pelo usuário, bem como tarefas de mineração de dados, além de examinar recursos de visualização analítica interativa. Os resultados revelaram uma disparidade significativa na quantidade e na apresentação de informações entre os recursos analisados.

Os casos de uso, de tal estudo, foram baseados nos sistemas *Pediatric Cancer Data Commons* (Pedican), *Pediatric Cancer* (PeCan), *Pediatric Cancer BioPortal* (PedcBioPortal), *University of California Santa Cruz Xena Browser* e *International Cancer Genome Consortium* (ICGC). Enquanto o Pedican oferecia uma quantidade significativa de entradas, mas com poucos recursos de seleção, outros, como PeCan e PedcBioPortal incluíam ferramentas de análise visual, sendo que o PedcBioPortal integrava dados publicados e novos provenientes de consórcios. O UCSC se destacou pela análise aprofundada de dados genômicos, enquanto o portal de dados do ICGC oferecia várias opções para análise de dados, com foco em projetos de câncer em adultos.

A comparação das informações sobre mutações proeminentes no glioma destacou a diversidade na alocação de dados e a presença de biomarcadores diversos. Como conclusão, observa-se que os recursos disponíveis na internet para o estudo de cânceres pediátricos são menos abundantes e menos conhecidos em comparação àqueles voltados para o câncer em adultos.

O estudo realizado por (Krejčí et al., 2021), por sua vez, teve como objetivo desenvolver um portal destinado a fornecer informações sobre a epidemiologia do câncer infantil na República Tcheca. Os dados sobre câncer infantil foram obtidos do registro nacional de câncer do país, validados em bases de dados clínicos e combinados com dados de pacientes hospitalizados e extraídos de certidões de óbito. Esses dados foram, então, utilizados para determinar taxas de incidência e sobrevivência de pacientes com câncer infantil diagnosticados entre 1994 e 2016. Esse portal está disponível desde novembro de 2019, oferecendo informações básicas sobre grupos de diagnóstico de câncer infantil por meio de relatórios analíticos interativos. Os resultados publicados no portal desenvolvido foram verificados através da validação teórica e teste do usuário, resultando em uma ferramenta que permite análises automatizadas mesmo para leigos.

Destaca-se, ainda, o estudo realizado pelos autores Ansari e Martin (2022), os quais desenvolveram um estudo com o objetivo de criar uma lista de verificação

(*checklist*) de usabilidade voltada, especificamente, para *dashboards* de saúde pública. A proposta surgiu diante da constatação de que muitos painéis de dados em saúde apresentam problemas de usabilidade que dificultam seu uso por diferentes públicos, como profissionais da saúde pública, pesquisadores acadêmicos e o público geral. Para isso, os autores realizaram uma análise sistemática de todos os dashboards públicos disponíveis em sites de departamentos estaduais de saúde dos Estados Unidos que apresentavam dados sobre infecções sexualmente transmissíveis, totalizando treze painéis avaliados.

A metodologia adotada baseou-se na aplicação de um conjunto de onze princípios de visualização de informação amplamente reconhecidos na literatura, incluindo organização espacial, codificação das informações, consistência, remoção de elementos desnecessários, reconhecimento em vez de lembrança, ações mínimas, redução do conjunto de dados, flexibilidade para diferentes tipos de usuários, compreensibilidade dos conteúdos, integridade científica e legibilidade. Seis avaliadores, com conhecimentos em usabilidade e experiência em saúde pública, foram responsáveis por aplicar esses critérios em cada *dashboard*, atribuindo notas e fornecendo comentários qualitativos.

Os resultados mostraram que os *dashboards* apresentavam níveis variados de complexidade, sendo que os problemas de usabilidade mais críticos estavam concentrados em três áreas principais: flexibilidade da experiência do usuário, compreensão do conteúdo (especialmente para usuários não especialistas) e integridade científica (voltada aos especialistas). Em contrapartida, os princípios com menor incidência de problemas foram a remoção de elementos desnecessários, a minimização de ações necessárias para navegação e a consistência visual e funcional entre seções dos painéis.

A partir dessas análises, os autores desenvolveram um *checklist* de usabilidade, com recomendações práticas para cada princípio avaliado, objetivando orientar futuros desenvolvedores de *dashboards* em saúde pública. O estudo destaca a importância de equilibrar complexidade e usabilidade para atender adequadamente às necessidades de diferentes grupos de usuários.

Por fim, Almasi et al. (2023) realizaram uma revisão sistemática com o objetivo de identificar os principais questionários utilizados na avaliação de usabilidade de dashboards em saúde e propor critérios específicos para essa finalidade. O estudo analisou vinte e nove artigos extraídos das bases PubMed, Scopus e Web of Science,

sem restrição de data, focando exclusivamente em trabalhos que utilizaram questionários para avaliar dashboards em ambientes clínicos e administrativos. Os resultados mostraram que a maioria dos estudos empregava instrumentos como o *System Usability Scale* (SUS), *Technology Acceptance Model* (TAM), Teoria Unificada da Aceitação e Uso da Tecnologia (UTAUT) e outros, que não foram originalmente desenvolvidos para avaliar *dashboards*.

A partir dessa análise, os autores propuseram um conjunto de critérios específicos para a avaliação da usabilidade de *dashboards*, incluindo utilidade, operabilidade, facilidade de aprendizado, facilidade de uso, adequação às tarefas, interface do usuário, conteúdo e capacidades do sistema. O estudo foi concluído com a apresentação desses critérios como uma base para futuras avaliações mais precisas e contextualizadas.

A Tabela 1 apresenta um resumo dos trabalhos discutidos nesta seção.

Tabela 1 - Resumo dos trabalhos relacionados

Estudo	Objetivo	Resultados
Buhlinger (2021)	Analisar a pré-implementação do Registro Nacional de Câncer Pediátrico da Etiópia	Foi constatado que a ferramenta se baseia em planilhas e não inclui todas as variáveis recomendadas pela OMS e apresenta inconsistências no preenchimento devido a lacunas no fluxo de trabalho.
Jeanquartier (2019)	Investigar a disponibilidade de recursos na web sobre câncer pediátrico, comparando-os com recursos gerais	Foi observado que os recursos disponíveis para o estudo de cânceres pediátricos, na internet, são menos abundantes e menos conhecidos em comparação com aqueles voltados para o câncer em adultos.
Krejčí (2021)	Desenvolver o Sistema Tcheco de Informações sobre o Câncer Infantil	Foi desenvolvido o portal proposto, o qual oferece informações básicas sobre grupos de diagnóstico de câncer infantil por meio de relatórios analíticos interativos.
Almasi et al. (2023)	Investigar os questionários utilizados na avaliação de usabilidade de dashboards em saúde e propor critérios específicos para essa avaliação.	Foi identificado que a maioria dos estudos utilizou questionários de usabilidade bem difundidos na literatura. Como resultado, os autores propuseram critérios específicos para avaliar *dashboards*, incluindo: utilidade, operacionalidade, facilidade de uso, interface, conteúdo, entre outros.

Os estudos, discutidos neste capítulo, oferecem uma visão do cenário atual da pesquisa em sistemas de coleta e visualização de dados relacionados ao câncer

pediátrico, destacando tanto os avanços quanto as lacunas existentes. No entanto, apesar dos esforços apresentados, ainda há uma escassez de trabalhos publicados nesta área. Esta lacuna ressalta a necessidade premente de mais investimentos e colaborações na pesquisa sobre câncer pediátrico, a fim de melhorar a compreensão, o tratamento e os resultados para as crianças afetadas por esta doença.

5 MATERIAIS E MÉTODOS

Esta seção apresenta as diferentes etapas envolvidas na realização deste estudo. A execução destas e os resultados obtidos são apresentados na sequência, no Capítulo 6.

5.1 Revisão Bibliográfica

Esta pesquisa começou com a realização de uma revisão bibliográfica, a qual embasou a fundamentação teórica deste trabalho (Capítulo 3). Na sequência, foi realizado um levantamento de trabalhos relacionados ao tema proposto (Capítulo 4), visando entender o impacto da análise e da visualização analítica de dados na tomada de decisão por profissionais e gestores da área da oncologia.

A partir dos trabalhos relacionados pesquisados, empregou-se uma metodologia de revisão que se baseia na construção de um mapa de evidência, que apresenta informações sobre diferentes intervenções relacionadas ao uso de dashboards e ferramentas interativas, agrupando-as conforme as categorias estabelecidas pela metodologia PICO (população, intervenção, comparação e desfechos). Para tanto, o autor deste estudo realizou um curso de vinte horas, oferecido pela Organização Pan-Americana da Saúde (OPAS), que forneceu as bases de conceito e práticas necessárias empregadas na criação de mapas de evidência.

Para o desenvolvimento do mapa, utilizou-se a linguagem de programação Python.

5.2 Análise e Seleção de Bases de Dados

Na sequência, foi realizado o levantamento das bases de dados existentes, com foco naquelas que contêm informações oncológicas relevantes. Após o levantamento, foi realizada uma triagem das bases de dados para selecionar aquelas que ofereciam dados específicos sobre o câncer infantil, com a intenção de fornecer uma visão clara e precisa da situação da oncologia pediátrica no Brasil e no Rio Grande do Sul. Decidiu-se, então, pelo uso das bases públicas DATASUS e RHC-INCA e das bases privadas de dados fornecidas pela Secretaria Municipal de Saúde de Porto Alegre (SMSPA). A escolha dessas bases considerou, além da qualidade e

especificidade dos dados, a abrangência e confiabilidade das fontes.

Para acessar os dados públicos, foi utilizado o diretório de arquivos do DATASUS disponível em: ftp://ftp.datasus.gov.br/dissemin/publicos/ e o sistema IntegradorRHC do INCA. Já os dados provenientes dos sistemas da SMSPA foram acessados diretamente na rede interna da Empresa de Tecnologia da Informação e Comunicação do Município de Porto Alegre (PROCEMPA).

5.3 Coleta de Requisitos

A partir de então, foram coletados requisitos que embasaram a forma como os dados das bases selecionadas seriam preparados, integrados e analisados e as formas de representações visuais necessárias. Tais requisitos estão ligados ao entendimento de aspectos relacionados à oncologia infantil, como o tempo de atendimento, o local de atendimento, remissão, óbito entre outras.

Para tanto, este estudo seguiu a abordagem do *design* centrado no usuário (DCU), a qual priorizou as necessidades, os desejos e as limitações dos usuários durante todo o processo de criação do painel oncológico. O DCU foi aplicado com a participação de uma equipe multidisciplinar de profissionais, com interesse em investigar padrões epidemiológicos sobre o câncer infantojuvenil, que colaborou com esta pesquisa.

5.4 Pré-processamento e Transformação dos Dados

Com as bases de dados definidas, a próxima etapa deste projeto foi realizar o pré-processamento destes. A preparação dos dados começou com o estudo das documentações disponíveis para cada base de dados. A seguir, estes passaram pelas etapas de pré-processamento e transformação para garantir sua integridade e qualidade. Para tanto, foi realizada a verificação da integridade dos dados, sua normalização e padronização.

Em relação às tecnologias empregadas para o pré-processamento, utilizou-se a biblioteca *Ydata-profiling*[3], que permite gerar relatórios detalhados sobre a qualidade dos dados, identificando rapidamente padrões, valores ausentes e inconsistências.

[3] https://docs.profiling.ydata.ai/latest/

Dado o volume e a complexidade das bases, após a definição do modelo de de dados, foi implementado um processo de integração de dados automatizado conhecido por ETL (*extract, transform, load)*, esse processo é responsável por extrair, transformar e carregar os dados provenientes de diferentes fontes, garantindo sua padronização e integração. O ETL foi implementado na linguagem de programação Python.

5.5 Análise dos Dados

Com os dados devidamente preparados, a equipe realizou análises estatísticas detalhadas com o objetivo de identificar padrões, tendências e *insights* relevantes. A análise dos dados foi conduzida utilizando tecnologias baseadas em ciência de dados, com foco principal na linguagem de programação Python. Bibliotecas como *Pandas*[4] e *NumPy*[5] foram empregadas para a manipulação e o processamento eficaz dos dados. Essas ferramentas garantiram a otimização das operações, estabelecendo uma base sólida para análises mais avançadas.

5.6 Visualização de Dados

Após a etapa de análise de dados, as representações visuais e os *dashboards* foram desenvolvidos utilizando a linguagem de programação Python aliada às bibliotecas *Dash*[6], Matplotlib[7] e Plotly[8]. Para garantir responsividade e acessibilidade, adotou-se o *framework* Bootstrap[9], que habilitou recursos como o modo para daltonismo e permitiu o uso de paletas de cores integradas.

5.7 Avaliação de Usabilidade

A última etapa foi a realização da avaliação de usabilidade dos *dashboards*, visando analisar a facilidade de uso destes e como usuários em potencial interagiam

[4] https://pandas.pydata.org/
[5] https://numpy.org/
[6] https://plotly.com/dash/
[7] https://matplotlib.org/
[8] https://plotly.com/
[9] https://getbootstrap.com/

com o mesmo para realizar tarefas. O objetivo foi identificar possíveis problemas e pontos de melhoria. Para tanto, empregou-se a metodologia *System Usability Scale* (SUS). Criada por John Brooke em 1986, o questionário SUS avalia a efetividade, eficiência e satisfação dos usuários em relação a um sistema interativo.

A Figura 3 apresenta o fluxograma referente às etapas de realização deste projeto, juntamente com os respectivos responsáveis pela execução de cada uma delas.

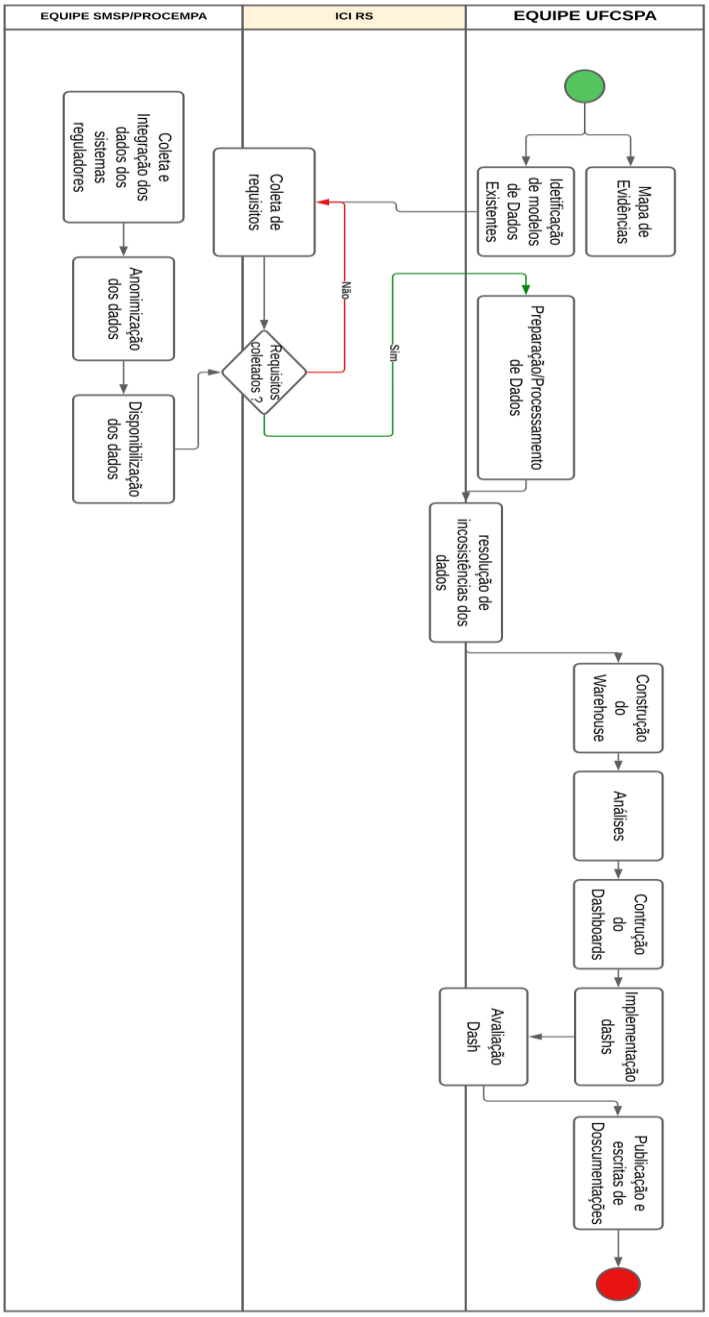


Figura 3 *- Etapas de execução do projeto*

Fonte: Autor

6. RESULTADOS

Este capítulo apresenta o detalhamento da execução das etapas de desenvolvimento do painel OncoPed, listadas no Capítulo 5, bem como os resultados de cada uma.

6.1 Criação do Mapa de Evidências

A partir do levantamento de trabalhos relacionados, apresentados no Capítulo 4, foi desenvolvido um mapa de evidências sobre o uso de *dashboards* e técnicas de visualização analítica de dados aplicados à oncologia. A Figura 4 traz a representação gráfica do mapa de evidências criado, onde pode-se observar informações sobre diferentes intervenções relacionadas ao uso de *dashboards* e ferramentas interativas, agrupando-as conforme as categorias estabelecidas pela metodologia PICO (população, intervenção, comparação e desfechos).

A análise visual proporcionada pelo mapa de evidências permitiu identificar rapidamente quais intervenções possuem maior respaldo na literatura e em quais áreas ainda existem lacunas significativas de pesquisa. Por exemplo, observou-se uma concentração de estudos sobre dashboards clínicos integrados a prontuários eletrônicos, com efeitos positivos reportados na tomada de decisão e na comunicação com pacientes, enquanto intervenções voltadas especificamente para tipos raros de câncer ou para contextos de saúde pública permanecem pouco exploradas. Essas informações auxiliaram nas etapas seguintes do estudo, destacando a importância de incorporar bases de dados abrangentes e ligadas à saúde pública, de modo a ampliar o alcance das análises e contemplar cenários ainda pouco investigados.

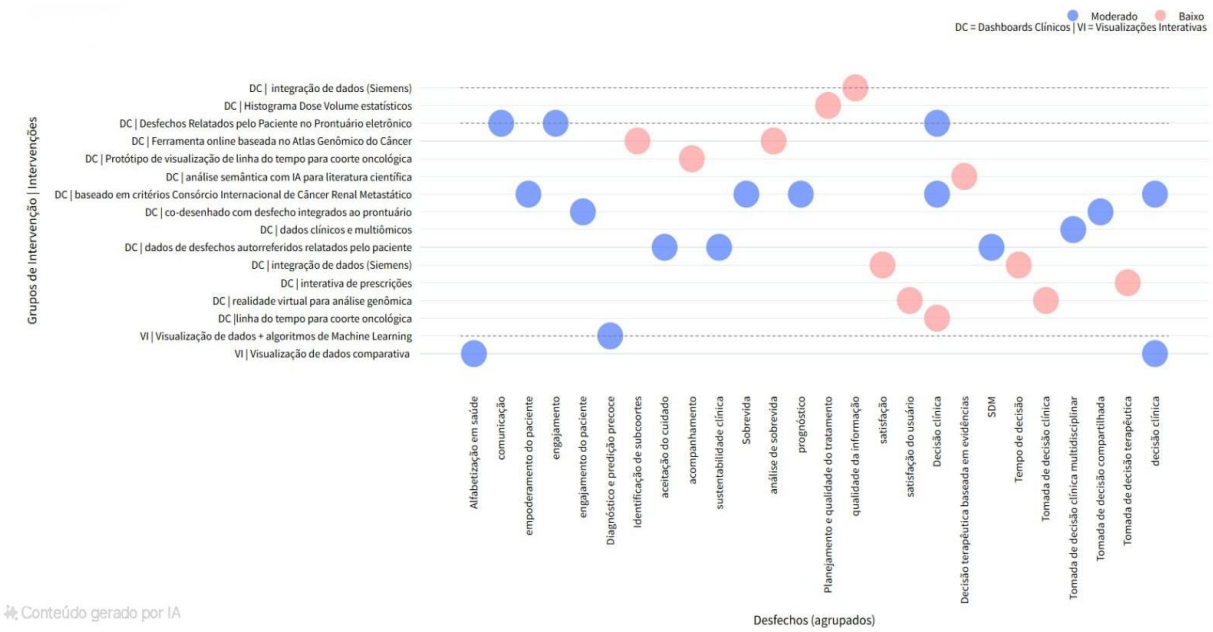


Figura 4 *- Mapa de evidências em oncologia*

Fonte: Autor

6.2 Definição das Bases de Dados

Para o levantamento das bases de dados existentes, das quais se pudesse extrair informações oncológicas pediátricas para este estudo, foi realizada uma triagem das bases para selecionar aquelas que possuíam dados específicos sobre o câncer infantil. Dentre diferentes possibilidades, foram selecionadas as bases de dados públicas do DATASUS e do RHC-INCA bem como as bases privadas fornecidas pela Secretaria Municipal de Saúde de Porto Alegre (SMSPA).

A Figura 5 ilustra as bases de dados selecionadas para este estudo, as quais estão detalhadas a seguir.

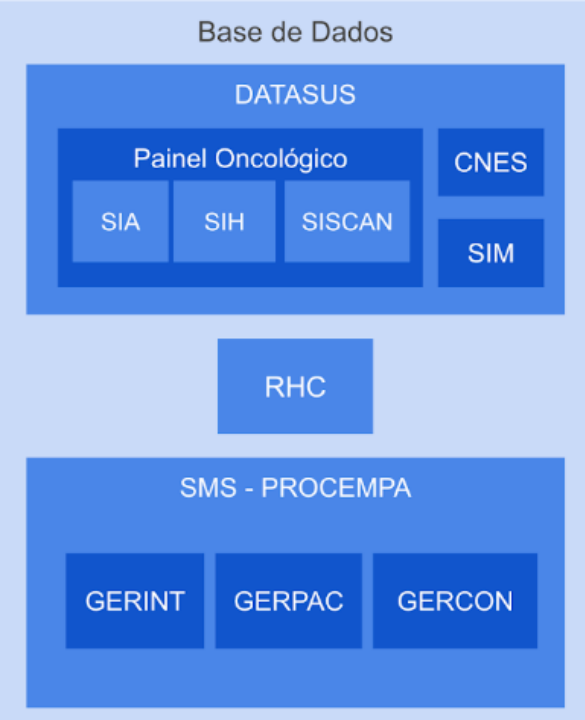


Figura 5 *- Bases de dados de oncologia pediátrica empregadas neste estudo.*

Fonte: Autor

Em relação ao DATASUS, foram selecionadas três bases principais: o Painel Oncológico, o Sistema de Informação sobre Mortalidade (SIM) e o Cadastro Nacional de Estabelecimentos de Saúde (CNES).

Do Painel Oncológico, por sua vez, foram utilizadas três bases de dados: Sistema de Informação Ambulatorial (SIA), Sistema de Informação Hospitalar (SIH) e Sistema de Informação de Câncer (SISCAN). Estas estão detalhadas a seguir:

- SIA: contém registros detalhados sobre atendimentos ambulatoriais relacionados à oncologia, fornecendo dados sobre consultas e tratamentos realizados fora do ambiente hospitalar, de locais que atuam no Sistema Único de Saúde;
- SIH: apresenta informações sobre internações hospitalares dos hospitais públicos do Brasil, incluindo dados sobre procedimentos oncológicos realizados durante a internação dos pacientes;
- Sistema de Informação de Câncer (SISCAN): reúne dados sobre o rastreamento e o diagnóstico precoce do câncer, sendo fonte essencial para analisar a detecção precoce da doença;

A base SIM, do DATASUS, foi utilizada na análise das taxas de óbito e de sobrevida de pacientes pediátricos com câncer, fornecendo dados fundamentais para o entendimento do impacto da doença. Já a base CNES, também do DATASUS, contém informações sobre hospitais e centros de saúde especializados no tratamento oncológico, permitindo mapear os locais de tratamento disponíveis;

Além das bases públicas do DATASUS, optou-se por utilizar, também, a base do Registro Hospitalar de Câncer (RHC), a qual é mantida pelo Instituto Nacional de Câncer (INCA). Esta oferece registros detalhados sobre os pacientes atendidos em unidades de oncologia hospitalar, os quais são essenciais para a avaliação da qualidade do atendimento e da evolução do tratamento oncológico.

Em termos de bases privadas, foram utilizados três provenientes dos sistemas do complexo regulador da Secretaria Municipal de Saúde de Porto Alegre (SMSPA), desenvolvidos pela Empresa Pública de Tecnologia da Informação e Comunicação (PROCEMPA), os quais regulam o fluxo de pacientes no SUS:

- Gerenciador de consultas do SUS (GERCON): Reúne dados das consultas e exames solicitados;
- Gerenciador de internações do SUS (GERINT): Reúne dados das internações e cirurgias solicitadas;
- Gerenciamento de procedimentos ambulatoriais (GERPAC): Reúne informações dos procedimentos solicitados.

A escolha dessas bases considerou, além da qualidade e especificidade dos dados, a abrangência e confiabilidade das fontes. A partir de então, os requisitos do sistema puderam ser definidos, os quais delinearam a proposta de integração dos dados e as técnicas de visualização que seriam utilizadas. As próximas seções abordam estas definições.

6.3 Coleta de Requisitos do Sistema

Este estudo seguiu a abordagem do *design* centrado no usuário para o projeto e desenvolvimento de sistemas e, para tanto, contou com a participação de uma equipe multidisciplinar de profissionais com interesse em investigar padrões epidemiológicos sobre o câncer infantojuvenil.

Esta equipe foi formada por profissionais que atuam no Instituto do Câncer Infantil (ICI) de Porto Alegre, sendo eles:

- Uma médica hematologista pediátrica;
- Um médico oncologista pediátrico e superintendente de uma instituição focada na detecção precoce e no tratamento do câncer infantojuvenil;
- Um médico pediatra, que atua em uma secretaria municipal de saúde;
- Uma bióloga e líder de projetos na área da bioinformática; e
- Um informata biomédico.

Tais profissionais atuaram como *stakeholders* durante todo o desenvolvimento do projeto.

Durante esta análise, definiu-se perguntas que as representações visuais precisariam auxiliar a responder, as quais foram categorizadas em tempo de atendimento, adequação do local de atendimento, deslocamento do paciente e informações sobre óbito:

a) Tempo de Atendimento

- Qual é o tempo médio entre o diagnóstico e o início do tratamento?
- Como os tempos entre o diagnóstico e o início do tratamento (ou tempo de espera) variam entre unidades habilitadas e não habilitadas?

b) Adequação do Local de Atendimento

- O hospital de atendimento é habilitado para oncologia pediátrica?
- Qual é o número de pacientes atendidos em locais não habilitados para oncologia pediátrica?
- Como os tempos de espera e tratamento variam entre unidades habilitadas e não habilitadas?

c) Deslocamento do Paciente

- Qual é a média de distância percorrida pelos pacientes entre sua residência e o local de atendimento?
- Quais são as regiões com maiores índices de deslocamento para tratamento oncológico?

d) Informações sobre Óbito

- Qual é a data do óbito e a causa (CID) para pacientes oncológicos pediátricos?
- Qual é a taxa de mortalidade de acordo com o tipo de câncer, idade e local de atendimento?
- Qual a sobrevida dos diferentes tumores pediátricos?

A partir de tais perguntas, definiu-se os requisitos que precisariam ser atendidos para respondê-las:

- Preparar, padronizar e integrar dados de oncologia pediátrica;
- Automatizar a coleta e a atualização dos dados de diferentes bases;
- Implementar quatro *dashboards*, os quais apresentariam representações visuais relacionadas à integração entre dados do painel oncológico e do CNES, dados do RHC e do CNES, dados do SIM e do CNES e, ainda, dados do GERINT, do GERPAC e do GERCON.
- Definir as representações visuais e as formas de interação presente nos *dashboards*.

A Figura 6 apresenta o planejamento dos *dashboards* presentes neste estudo.

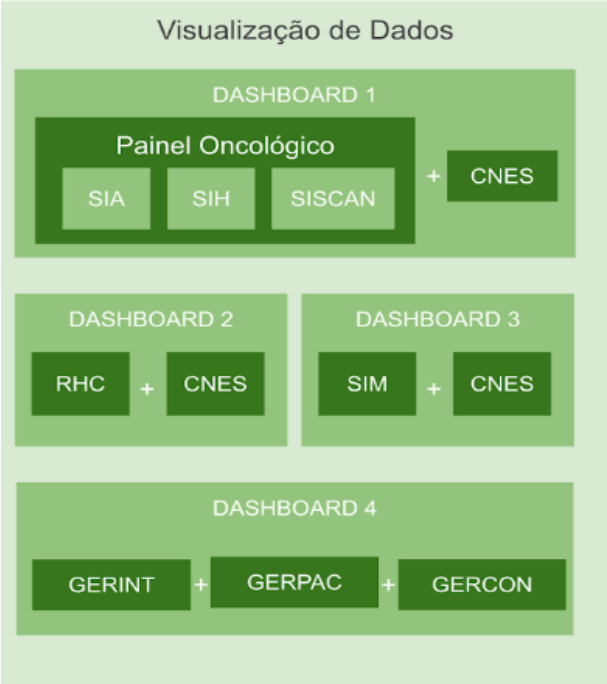


Figura 6 *- Integração de bases de dados de oncologia pediátrica em quatro dashboards.*

Fonte: Autor

6.4 Pré-Processamento e Transformação dos Dados

As bases públicas fornecidas pelo DATASUS contavam com documentação completa, incluindo dicionários de variáveis e portarias relacionadas aos dados utilizados. Em relação à base RHC do INCA, estava disponível apenas o dicionário de variáveis.

Já os dados provenientes da PROCEMPA, por sua vez, não possuíam documentação prévia, exigindo a realização de reuniões com a equipe de dados do complexo regulador. Essas reuniões objetivaram compreender a finalidade e as variáveis dos sistemas que alimentam essas bases. A partir dessas discussões, foram elaboradas tanto a documentação quanto o dicionário de variáveis específicas para as bases da PROCEMPA.

As equipes da UFCSPA e do Instituto do Câncer Infantil (ICI) realizaram uma avaliação detalhada da integridade dos dados provenientes dessas fontes, buscando identificar problemas como *outliers*, dados faltantes e registros duplicados nas variáveis consideradas obrigatórias. Na Tabela 2, é possível visualizar os problemas encontrados em cada base de dados.

Tabela 2 - Problemas encontrados em cada base de dados

Base de dados	Problemas
Painel Oncológico (Datasus)	1. Variável de estadiamento do tumor utilizado não é adequado para tumores pediátricos e *missing data* 2. Variável data de Início de tratamento, mais da metade da base de dados sem a informação 3. Local de tratamento e Informação sobre o primeiro tratamento sem Informação
RHC (tabulador INCA)	1. Data de Óbitos preenchidas, com outliers(ex 99/99/9999) 2. Estadiamento dos tumores não é adequado para tumores pediátricos (utilizar TNM) 3. Tempo de atualização de 2 anos
SIM (Datasus)	1. Layouts diferentes entre estados, dificulta integração em uma única base. 2. Alguns casos com a data de óbito inválidas(ex: 99/99/999) 3. Causa básica do óbito em alguns caso utilizar CIDs genéricos por exemplo causas naturais inespecíficas 4. Duplicidades de Registros 5. Dois anos para atualização dos dados.
SMSPA	1. Falta de documentação sobre os sistemas e variáveis 2. Apenas quatro dos seis Hospitais habilitados em oncologia pediátrica estão na base de dados 3. Demora na implementação desses sistemas em outros municípios fora de Porto Alegre.

Após a coleta dos dados das bases selecionadas, estes precisaram passar por uma etapa de pré-processamento para garantir sua integridade e qualidade, o qual foi realizado através da verificação da integridade dos dados e da normalização e padronização dos dados.

A verificação da integridade dos dados, realizada em três etapas principais:

- Detecção de valores ausentes ou inconsistentes, garantindo que os registros possuíam informações completas sempre que possível. Nos casos em que a ausência de dados compromete a análise, optou-se pela remoção desses registros;
- Identificação e eliminação de registros duplicados, evitando que informações fossem contabilizadas mais de uma vez, o que poderia distorcer os resultados;

- Análise detalhada de cada base de dados — devido à falta de identificador único comum entre elas, não foi possível realizar uma integração automática. Como alternativa, cada base foi analisada separadamente, permitindo a correção de erros internos e a padronização de informações dentro de cada conjunto de dados;

Já a normalização e a padronização dos dados, mesmo sem integração automatizada, incluíram:

- Padronização de nomenclaturas para evitar variações nos nomes de categorias, diagnósticos e outras informações relevantes;
- Organização e reestruturação dos códigos de identificação de doenças conforme a Classificação Internacional de Doenças (CID-10) e Classificação Internacional de Doenças para Oncologia (CID-O), garantindo que todos os registros seguissem um mesmo padrão;
- Ajuste de formatos de datas, convertendo-as para um padrão unificado, o que facilitou a análise temporal dentro de cada base;
- Tratamento de caracteres especiais e remoção de inconsistências textuais, tornando os registros mais homogêneos e evitando problemas durante a análise.

Esse processo foi essencial para corrigir inconsistências, remover duplicidades e preparar os dados para análises futuras.

6.5 Visualização de Dados

As técnicas de visualização empregadas nos *dashboards* foram definidas a partir das perguntas apresentadas na seção 6.3, de modo a mapear a forma de representação visual mais adequada aos tipos de dados a serem analisados.

Assim, as representações visuais selecionadas foram:

- Gráficos do tipo rosca (*donut*): Indicados para mostrar a relação entre porcentagens e proporções de dados categóricos, pois é um tipo de visualização que auxilia a comparar categorias e como elas se relacionam com

o todo;

- Gráficos de barra: Indicados para representar a comparação de valores entre diferentes categorias ou para exibir distribuições;
- Gráficos *sunburst*: Indicados para representar dados hierárquicos, principalmente quando é preciso representar a dimensionalidade dos dados;
- Gráficos coropléticos: Indicados para representar dados espaciais e quantitativos em um mapa;
- Gráfico sankey: Indicado para representar fluxos, movimentos ou mudanças entre diferentes entidades ou estados;
- Gráfico de coordenadas paralelas: Indicado para representar graficamente dados multivariados, onde cada observação ou ponto de dados é representado por uma linha que intersecta uma série de eixos paralelos.

Estas seis formas de representação visual são utilizadas nos quatro *dashboards* citados na Figura 6.

6.6 Implementação do Sistema

Após as definições das bases e do modelo de dados, dos requisitos do sistema e das representações visuais, deu-se início à implementação do sistema proposto. Visando a automatização e a eficiência na extração de dados, foi necessário, inicialmente, desenvolver um módulo específico para coletar, padronizar e integrar os dados extraídos das bases de dados DATASUS, RHC e SMS PROCEMPA. A Figura 7 apresenta o pipeline que resume o fluxo de entrada e saída de dados do sistema desenvolvido.

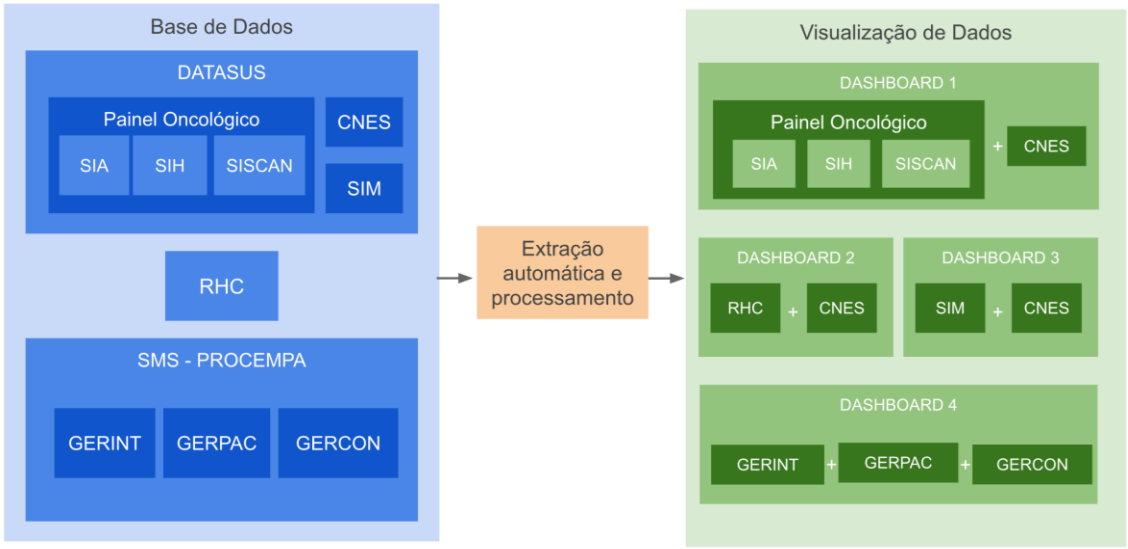


Figura 7- Pipeline da extração, processamento, integração e visualização dos dados de oncologia pediátrica

Fonte: Autor

O módulo desenvolvido monitora continuamente os servidores FTP e realiza, de forma automatizada, o *download* das bases de dados sempre que novos arquivos são disponibilizados ou quando há alterações nos dados existentes. Dessa maneira, garante-se a atualização ininterrupta das informações para análise. Além disso, a solução automatiza a descompactação dos arquivos e a conversão para os formatos necessários, eliminando a necessidade de qualquer intervenção manual. No caso da base de dados da Prefeitura de Porto Alegre, os arquivos são compartilhados por meio do Drive Procempa.

Após a análise das características das bases de dados selecionadas para este estudo e das definições dos *dashboards* que iriam apresentá-las, deu-se início à implementação do sistema *web* para visualização dos dados oncológicos extraídos. Os painéis construídos têm o propósito de oferecer a profissionais de saúde, gestores públicos e pesquisadores uma visualização clara e detalhada dos indicadores mais relevantes sobre o tratamento oncológico pediátrico no Brasil e no Rio Grande do Sul. As representações gráficas disponíveis nos dashboards auxiliam na interpretação das seguintes informações:

- A quantidade total de diagnósticos registrados em oncologia pediátrica, permitindo a análise de tendências e padrões ao longo do tempo;
- O número de pacientes que iniciaram tratamento oncológico, fornecendo uma

visão mais precisa sobre o início do tratamento e suas possíveis lacunas;

- O tempo médio entre o diagnóstico e o início do tratamento, o que é fundamental para avaliar a agilidade no atendimento e o impacto do tempo no prognóstico dos pacientes;
- O diagnóstico mais frequente em pacientes pediátricos, com destaque para os tipos de câncer mais comuns em crianças e adolescentes;
- Os tipos mais comuns de tratamentos iniciais administrados aos pacientes, permitindo uma análise sobre os procedimentos mais frequentemente adotadas para cada tumor;
- Os locais de saúde habilitados para tratamento de oncologia pediátrica, fornecendo uma visão sobre a distribuição geográfica e a capacidade de atendimento;
- Dados sobre mortalidade infantil relacionada à oncologia pediátrica, o que permite avaliar o impacto do câncer em crianças e adolescentes e as possíveis lacunas no tratamento.

As Figuras 8, 9, 10 e 11 apresentam uma parte de cada um dos *dashboards* desenvolvidos. O conjunto destes painéis foi nomeado como “Painel Oncológico Pediátrico” ou “Painel OncoPed”.

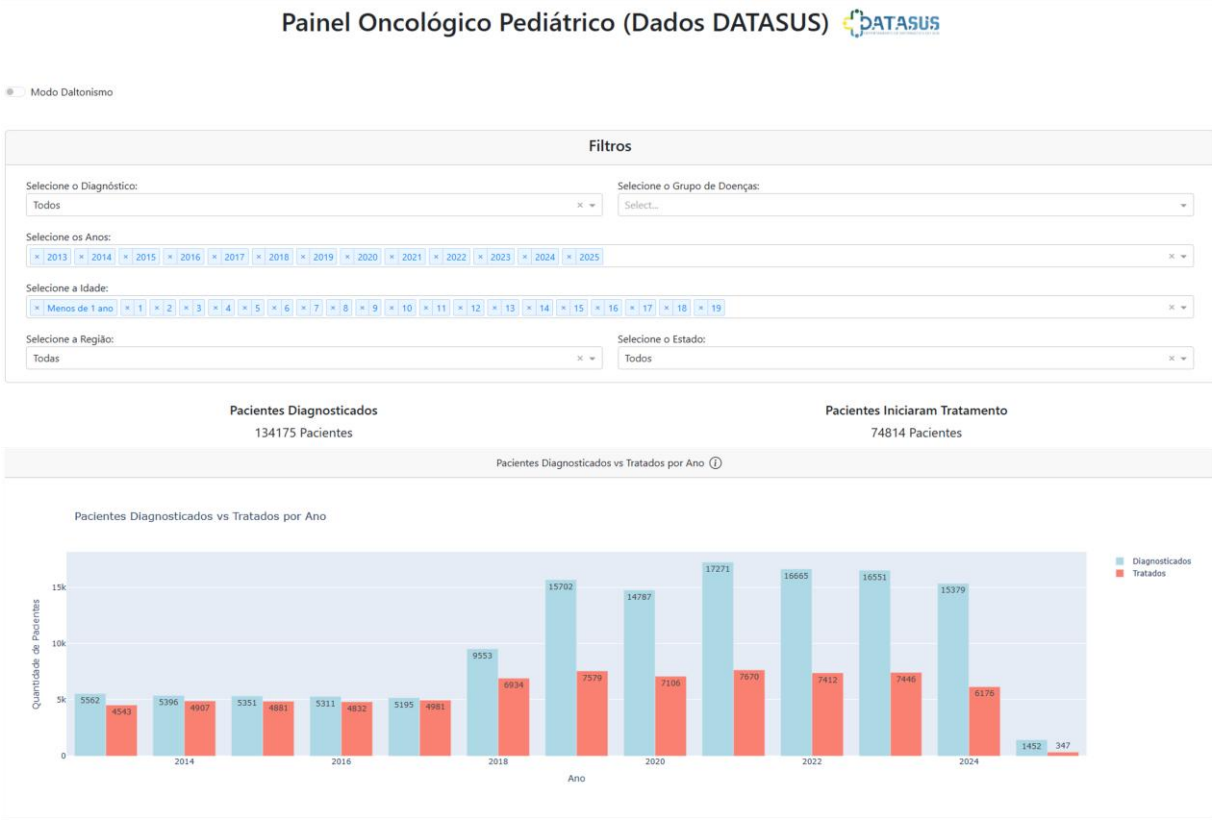


Figura 8 *- Parte inicial do dashboard exibindo dados do painel de oncologia do DATASUS (SIA, SIH e SISCAN) integrados ao CNES do DATASUS*

Fonte: Autor

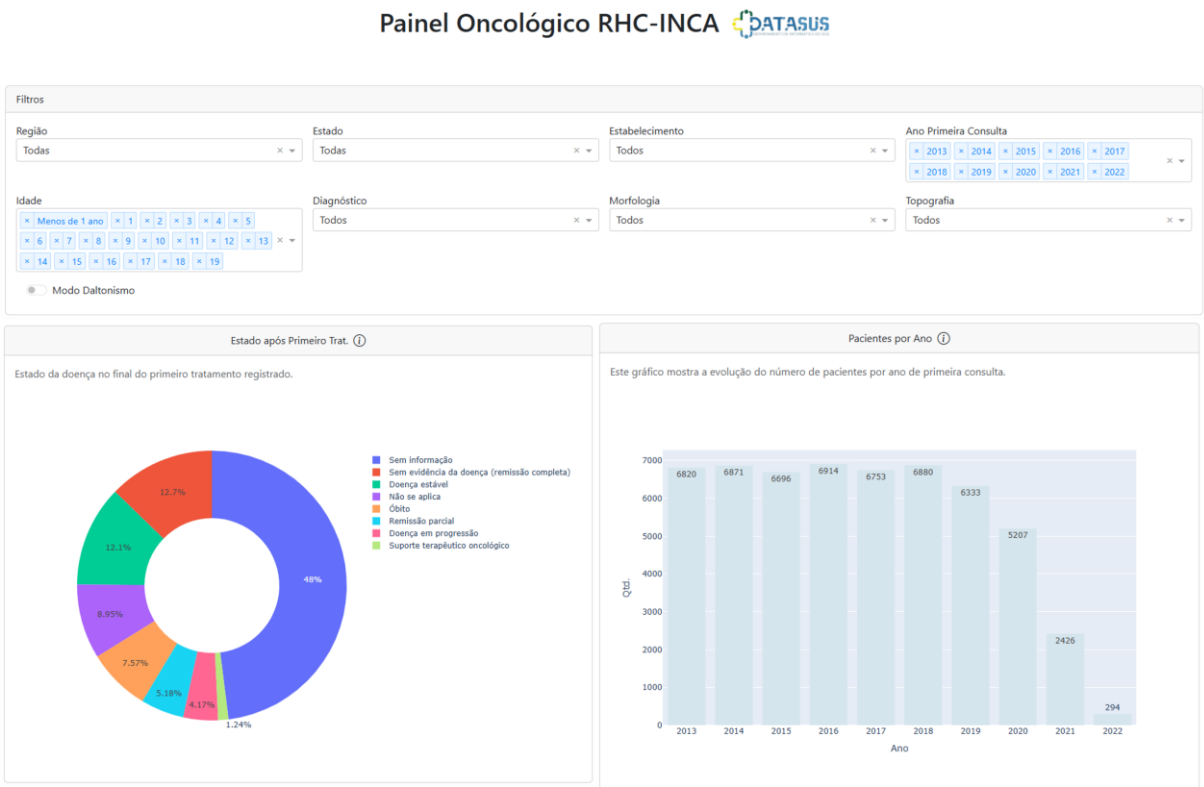


Figura 9 *- Parte inicial do dashboard exibindo dados do RHC integrados ao CNES do DATASUS*

Fonte: Autor



Figura 10 *- Parte inicial do dashboard exibindo dados do SIM do DATASUS integrados ao CNES do DATASUS*

Fonte: Autor

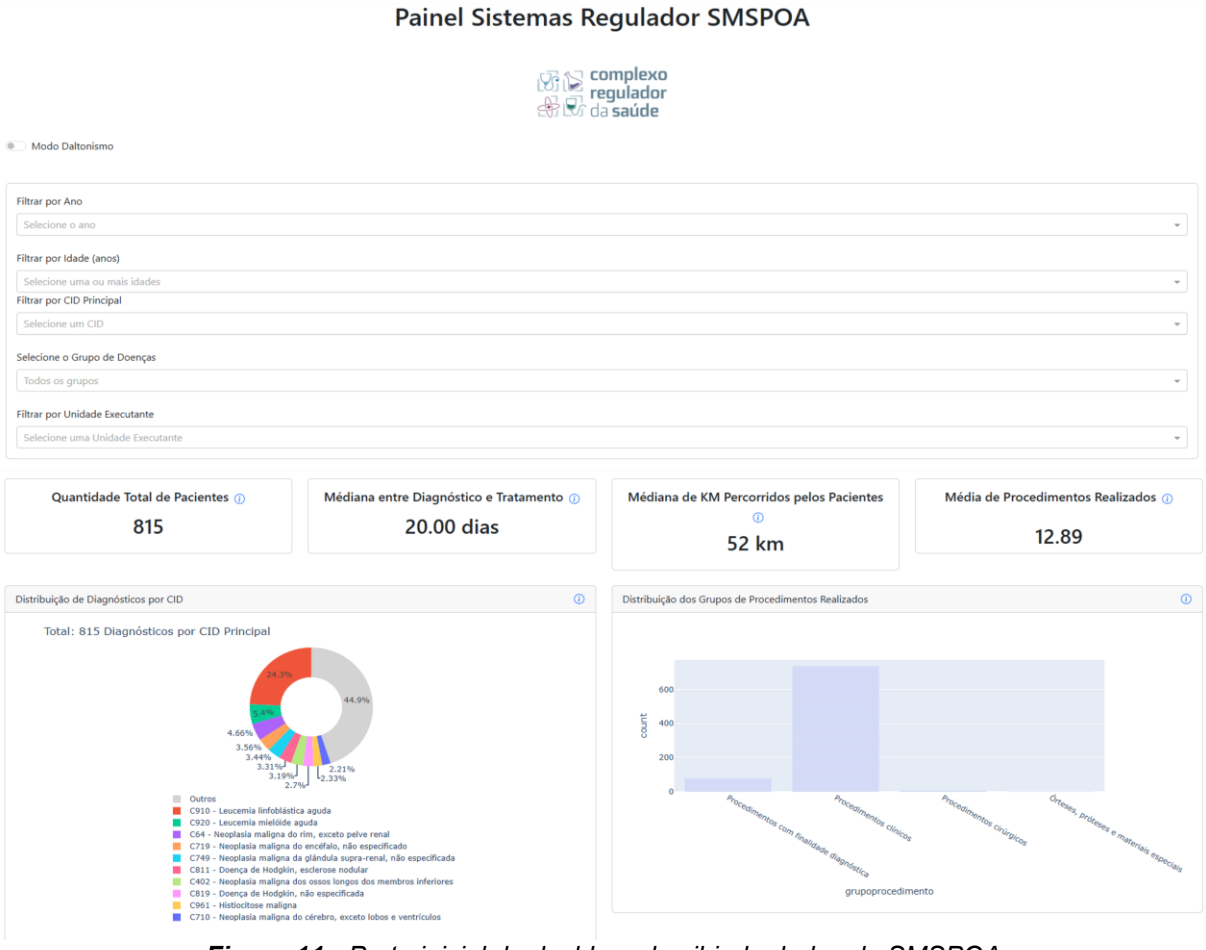


Figura 11 *- Parte inicial do dashboard exibindo dados da SMSPOA*

Fonte: Autor

As Figuras 12 e 13 apresentam exemplos de outros tipos de visualização que também estão presentes nos *dashboards* (representações *sunburst* e *sankey*). Todas estas permitem interação com o usuário, destacando as informações dos gráficos de acordo com a necessidade do usuário.

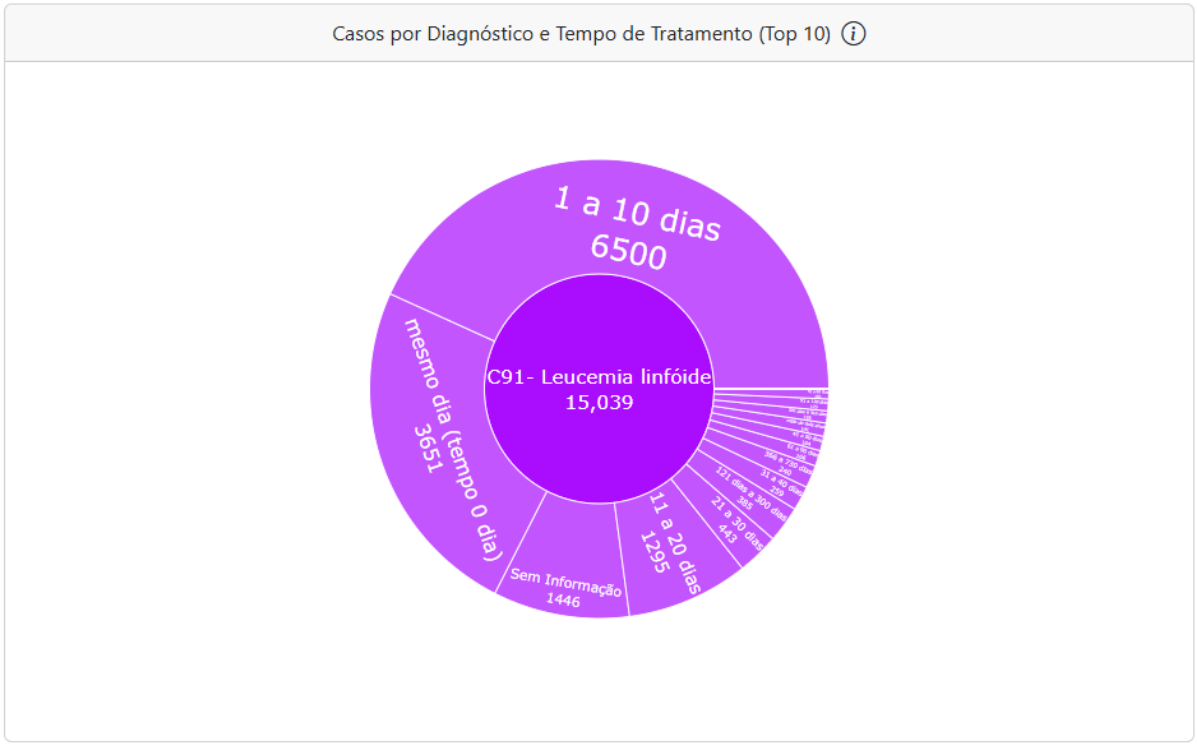


Figura 12 *- Gráfico sunburst*

Fonte: Autor

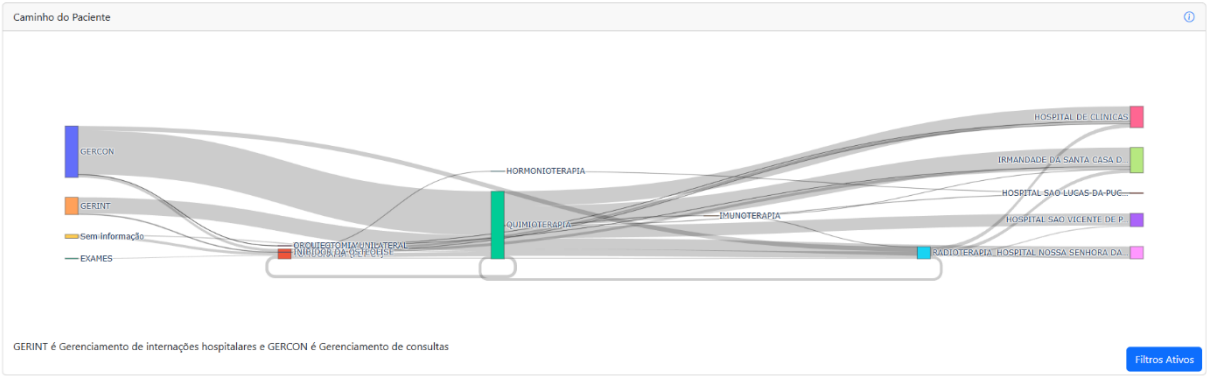


Figura 13 *- Gráfico sankey*

Fonte: Autor

Além das representações visuais em forma de diferentes gráficos interativos, foi implementado um sistema de filtros, permitindo que os usuários explorem os dados de forma personalizada. Os filtros possibilitam ajustes conforme variáveis como idade,

tipo de câncer, estado e hospital responsável pelo atendimento, garantindo que diferentes perfis de usuários possam extrair informações relevantes para suas necessidades específicas.

Após a finalização do desenvolvimento, a aplicação foi hospedada na plataforma *Railway*[10], escolhida por oferecer uma solução gratuita, segura e compatível com as necessidades do projeto. Dessa forma é possível acessar os quatro *dashboards* desenvolvidos de forma remota à ferramenta, permitindo que profissionais de saúde, gestores e pesquisadores pudessem acessar os dados de forma prática, auxiliando na tomada de decisões baseadas em dados do mundo real.

O sistema pode ser acessado através do *link* https://dashpaineloncoped.web.app.

6.7 Avaliação de Usabilidade

Como mencionado na seção 5.8, para avaliação de usabilidade dos quatro dashboards, foi desenvolvido um questionário, composto por dez afirmações balanceadas entre aspectos positivos e negativos, com base no *System Usability Scale* (SUS). No SUS, cada avaliação das afirmativas segue a escala LIkert, uma ferramenta de pesquisa e avaliação que mede atitudes, opiniões e comportamentos de indivíduos, utilizando análise de afirmativas com opções de resposta graduadas. Nesta escala, que varia de 1 a 5, os significados das respostas são:

- 1: discordo totalmente;
- 2: discordo;
- 3: neutro, não concordo, nem discordo;
- 4: concordo; e
- 5: concordo totalmente.

O formulário com as afirmativas SUS foi disponibilizado de forma *online* conforme apresentado no Apêndice C. Seis profissionais possuem experiência que varia de sete meses a três anos em atuação junto à oncologia pediátrica responderam ao formulário, como demonstrado nas Figuras 14.

[10] https://railway.com/

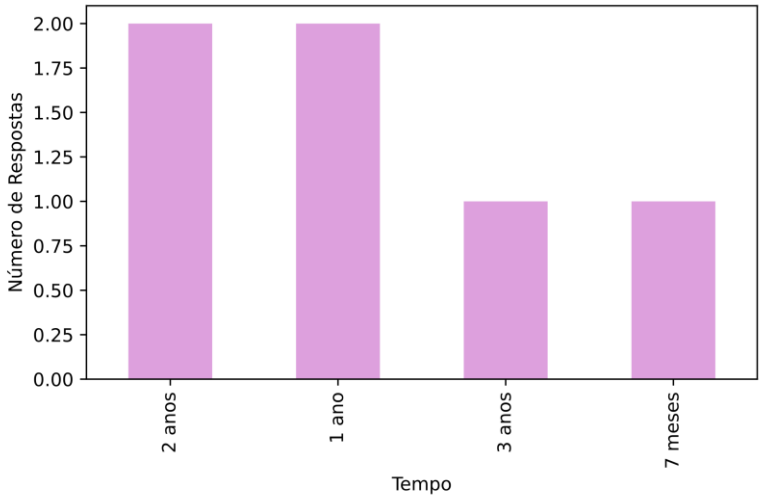


Figura 14 *- Tempo de atuação na oncologia pediátrica*

Fonte: Autor

Já na Figura 15, pode-se observar a área de atuação de cada participante, sendo dois em Administração, um enfermeiro, um oncologista pediatra, um fisioterapeuta e um cirurgião-dentista.

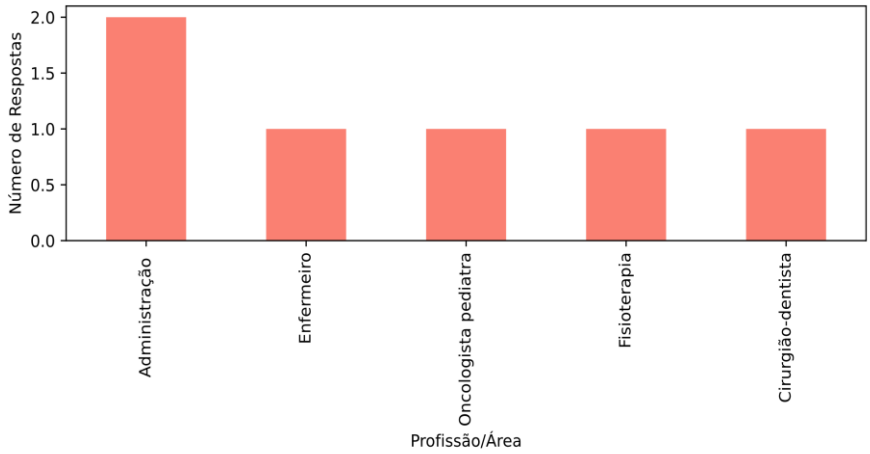


Figura 15 *- Distribuição de profissões/áreas de atuação*

Fonte: Autor

A Figura 16, por sua vez, apresenta o gráfico *boxplot* das distribuições das respostas às dez questões do questionário SUS, exibindo medianas, intervalo interquartil e os desvios-padrão (σ), permitindo avaliar a dispersão e identificar possíveis *outliers* em cada item. É importante ressaltar que, na metodologia SUS, questões de número ímpar (1, 3, 5, 7 e 9) indicam aspectos positivos sobre a usabilidade do sistema, logo quanto mais próximo do número 5 uma resposta estiver,

melhor é a avaliação. Já as questões de números pares (4, 6, 8 e 10) indicam aspectos negativos e, quanto mais próximo de 1 a resposta, melhor avaliada foi a usabilidade do sistema.

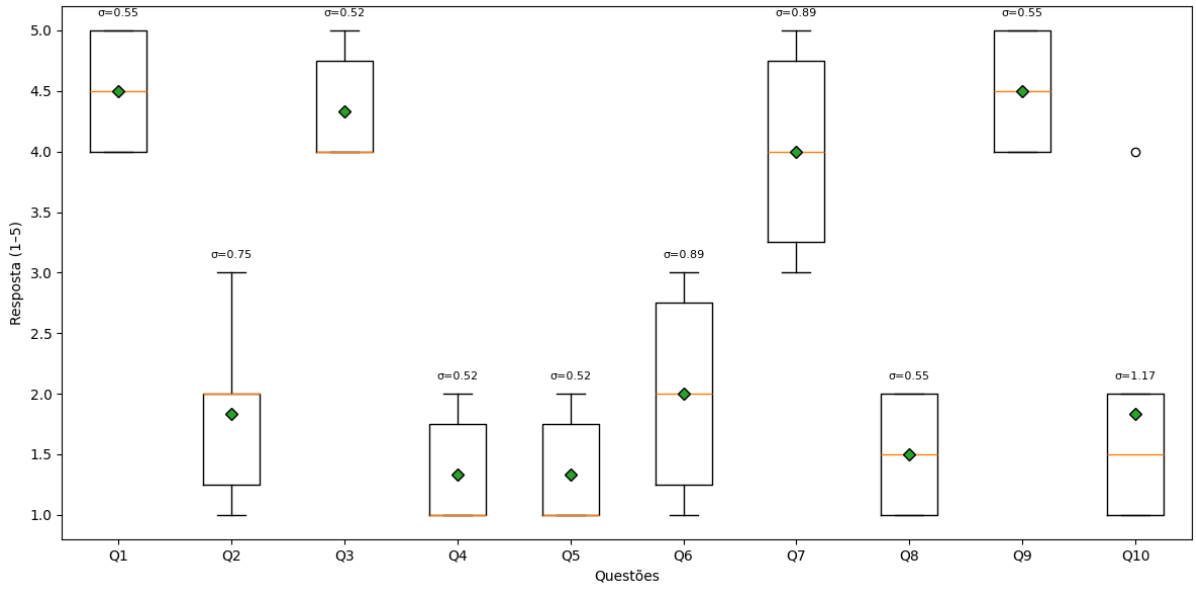


Figura 16 *- Distribuição das respostas SUS (Q1–Q10) com desvio-padrão*

Fonte: Autor

Na análise da Figura 16 verifica-se que a questão 2 apresentou um desvio-padrão moderado, pois as respostas se concentraram majoritariamente no valor 2, com variação entre 1 e 3 e um único registro em 3. Já a questão 6 exibiu amplitude ainda maior, com respostas distribuídas de forma homogênea entre 1, 2 e 3, sugerindo que, aproximadamente, metade dos participantes percebeu pouca inconsistência e a outra metade, alguma inconsistência. Por fim, a questão 7 evidenciou alto grau de dispersão no espectro positivo, apresentando frequências semelhantes para as notas 3, 4 e 5, indicando que alguns usuários assimilaram o uso do painel muito rapidamente (nota 5), enquanto outros demandam um período maior de adaptação (nota 3).

O cálculo do SUS seguiu o método padrão, subtraindo-se 1 ponto de cada resposta ímpar e 5 pontos de cada resposta par. No gráfico da Figura 17, apresenta-se o escore final do questionário SUS por respondente. Somou-se, então, os valores resultantes e multiplicou-se o total por 2,5. Obteve-se, assim, uma pontuação média de 75.42 (DP = 11.55), com valores mínimo e máximo de 65.0 e 90,00, respectivamente.

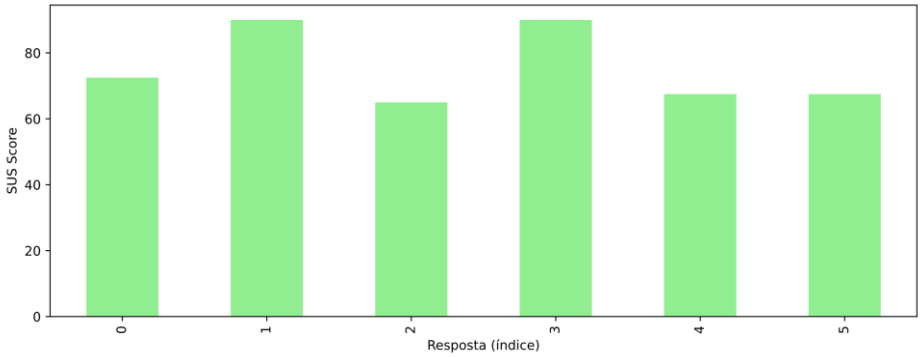


Figura 17 *- SUS escore final*

Fonte: Autor

Para analisar os resultados do escore final e a pontuação média do questionário SUS, observa-se que as pontuações são divididas em cinco faixas, A, B, C, D e F, as quais refletem diferentes níveis de usabilidade (SAURO, 2011). A faixa A, que cobre escores entre 80,0 e 100,0, indica usabilidade excelente, ou seja, os usuários percebem a interface como intuitiva, eficiente e muito satisfatória, exigindo mínimo esforço para aprender e usar o sistema.

A faixa B, de 70,0 a 79,9, corresponde a uma usabilidade boa, sinalizando que o painel funciona bem no geral, mas ainda há pequenos pontos a aperfeiçoar para atingir o nível de excelência. A faixa C, com escores entre 50,0 e 69,9, representa usabilidade aceitável; aqui, o sistema atende às necessidades básicas, mas demanda melhorias mais expressivas em aspectos como clareza de instruções, *feedback* ao usuário e consistência visual.

Já a faixa D, que vai de 25,0 a 49,9, denúncia usabilidade ruim, apontando que os usuários encontram dificuldades significativas de navegação e entendimento, o que pode levar ao abandono ou uso incorreto de funcionalidades. Por fim, a faixa F, abrangendo 0,0 a 24,9, sinaliza usabilidade inaceitável, evidenciando que o sistema falha em atender requisitos fundamentais de interação e precisa de reformulação completa antes de ser disponibilizado aos usuários. A Figura 18 sintetiza o conceito de distribuição das faixas no questionário SUS.

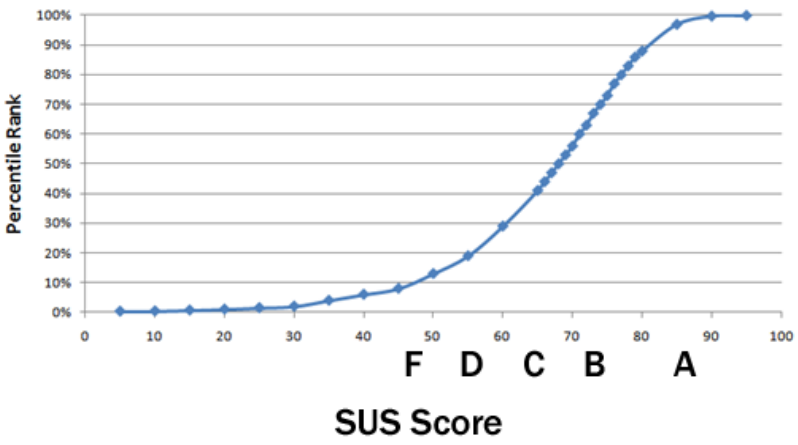


Figura 18 - *Classificações de percentis se associam às pontuações do SUS*

Fonte: SAURO, 2011

A média de 75.42 obtida na avaliação de usabilidade do “PainelOncoPed”, com valores mínimo de 65.0 e máximo de 90,00, o posiciona na faixa B, indicando que sua interface apresenta uma usabilidade boa e bom funcionamento. Nota-se, no entanto, que ainda há melhorias que podem ser realizadas de modo a otimizar a apresentação de dados e, possivelmente, reduzir a sobrecarga cognitiva dos usuários.

Além da avaliação pelo questionário SUS, alguns usuários forneceram *feedbacks,* destacando a satisfação com o recurso de acessibilidade voltado ao daltonismo e, de modo geral, com as funcionalidades oferecidas e o pioneirismo em focar na análise visual de dados oncológicos pediátricos.

7. DISCUSSÃO

A visualização analítica de dados aplicada à saúde pode contribuir, de forma significativa, para o processo de gestão da jornada do paciente, desde a realização de um diagnóstico até o período de pós-tratamento. Para tanto, a qualidade dos dados deve ser avaliada, principalmente, em relação à heterogeneidade e à diversidade destes, além da falta de estruturação dos registros de saúde. Problemas como ausência de padronização, variabilidade na qualidade dos dados e preocupações com a privacidade são comuns de ocorrerem, principalmente quando diferentes bases de dados precisam ser integradas.

Em relação, especificamente, a bases de dados com informações sobre tumores infantis, nota-se uma lacuna a ser preenchida, uma vez que estes possuem particularidades em relação a tumores em adultos quanto ao aspecto morfológico, comportamento clínico e localização primária. Tal fato motivou a proposta e desenvolvimento deste estudo, cujos resultados foram apresentados no Capítulo 6.

Esta pesquisa teve início com a realização de uma revisão da literatura sobre os principais conceitos envolvidos e trabalhos relacionados. A partir de então, foi construído um mapa de evidências, o qual identificou e sintetizou evidências científicas sobre o impacto da visualização analítica de dados na tomada de decisão por profissionais e gestores da área da oncologia.

Essa ferramenta interativa proporcionou uma visão clara e objetiva dos níveis de confiança das evidências disponíveis, explorando as relações entre intervenções específicas e desfechos, como a tomada de decisão clínica, a gestão de pacientes, a satisfação e o engajamento. Além disso, o mapa facilitou a identificação de lacunas nas pesquisas existentes, indicando que essa área necessita de mais estudos aprofundados e proporcionando um panorama geral das evidências atuais sobre o tema.

De posse dessas evidências, reuniu-se uma equipe multidisciplinar que colaborou com esta pesquisa durante toda sua execução a partir da metodologia de *design* centrado no usuário (DCU). O DCU busca entender profundamente o público-alvo para desenvolver soluções que sejam úteis, desejáveis e acessíveis. Neste trabalho, o emprego do DCU foi fundamental para a definição das bases de dados utilizadas, avaliando o quão estes dados eram representativos e relevantes de modo a contribuir para a tomada de decisão por parte de profissionais e gestores da área

da saúde em relação à prevenção e ao tratamento do câncer infantojuvenil. Também para a definição do modelo de dados e dos *dashboards* que exibem estes e constituem o Painel OncoPed.

Para a coleta, pré-processamento, transformação e carregamento dos dados, tanto em bases de dados públicas como em privadas, foi implementado um processo ETL automatizado, o qual possibilitou a atualização contínua dos *dashboards* a partir do monitoramento e da extração automática de novos dados. O Painel OncoPed foi desenvolvido com base em um sistema *web,* composto por quatro *dashboards* interativos, os quais apresentam diferentes tipos de representações visuais interativas de dados de oncologia pediátrica. Procurou-se, através da visualização de dados, auxiliar a investigação de padrões epidemiológicos e, consequentemente, a tomada de decisão de profissionais e gestores da área da saúde, que atuam na prevenção e tratamento e combate ao câncer infantojuvenil.

Além da definição das técnicas de visualização de dados, a forma de interação com os dados também foi definida, incluindo filtros dinâmicos que atualizavam os gráficos de acordo com as escolhas do usuário, permitindo focar e/ou destacar pontos de interesse ao mesmo tempo que relacionava diferentes representações visuais. Buscou-se, assim, disponibilizar estas informações para facilitar o entendimento do quadro de oncologia pediátrica no país.

8 CONSIDERAÇÕES ÉTICAS

A pesquisa em questão foi submetida para análise e provada em três comitês de ética em pesquisa (CEP): CEP da UFCSPA - local de origem da pesquisa, CEP do Hospital de Clínicas de Porto Alegre (HCPA) - local vinculado às pesquisas realizadas pelos membros do ICI, e CEP da SMSPA, com o objetivo de assegurar o acesso aos dados dos sistemas desta. O registro da aprovação nos CEPs foi realizado sob o CAAE 76218523.0.0000.5345 e número do comprovante 140054/2023.

9 PUBLICAÇÕES

Este projeto de mestrado gerou cinco publicações:

- Capítulo de livro (2024): Data Analytics and Visualization in Healthcare Decision-Making: Examples, Benefits & Challenges Related to Epidemiological Scenarios, no livro Applied Interdisciplinary Research: Integrating Information Technologies and Health.
- Mostra de Trabalhos de Ensino, Pesquisa e Extensão da UFCSPA (2023): apresentação oral de trabalho em andamento com o título “Análise de Dados em Oncologia Pediátrica: Investigação de Padrões Epidemiológicos em Banco de Dados Públicos e Privados”;
- XX Congresso Brasileiro de Informática em Saúde (2024): apresentação de pôster de trabalho em andamento com o título “Análise e visualização de dados em oncologia pediátrica: investigação de padrões epidemiológicos em banco de dados”;
- XIX Congresso de Oncologia Pediátrica (2025): apresentação de pôster de trabalho em andamento com o título “Análise e visualização de dados em oncologia pediátrica: investigação de padrões epidemiológicos em banco de dados”;
- Simpósio Brasileiro de Computação Aplicada à Saúde (2025): aceite de artigo (trabalho completo) e apresentação oral com o título “Investigação de Padrões Epidemiológicos em Oncologia Pediátrica a partir de Integração e Visualização de Dados”.

10 CONCLUSÕES

Em um país no qual cada secretaria municipal de saúde decide, de forma autônoma, quais sistemas adotar, desde prontuários eletrônicos até sistemas de regulação de filas e de resultados de exames, a heterogeneidade de plataformas impõe desafios significativos à integração de dados e à geração de *insights* em escala nacional. Nesse contexto, a proposta de construir *dashboards* com intuito de apresentar mostrar dados oncológicos pediátricos, em tempo real, demonstrou-se essencial para harmonizar informações oriundas de DATASUS, RHC-INCA e do complexo regulador da SMSPA. A construção de um *pipeline* de processo ETL (*extract, transform, load*) robusto permitiu não apenas a extração e padronização de dados públicos, mas também enfatizou a importância de parcerias entre universidade, órgãos públicos e iniciativa privada para garantir acesso, qualidade e atualização contínua das bases de dados.

A aplicação de técnicas avançadas de visualização de dados revelou-se capaz de traduzir grandes volumes de dados em informação e *dashboards* interativos voltados a apoiar a tomada de decisão de gestores e profissionais de saúde. Os *dashboards* desenvolvidos facilitam a identificação de locais habilitados e não habilitados que atenderam pacientes oncológicos, distribuição dos diagnósticos e distribuição geográfica, contribuindo para um entendimento mais profundo das disparidades regionais e das tendências temporais.

Entretanto, a execução deste projeto enfrentou desafios expressivos. Em primeiro lugar, constatou-se uma heterogeneidade na qualidade e no detalhamento dos dados: as bases do DATASUS e do RHC-INCA dispõem de documentação abrangente, porém apresentam registros com outliers e valores ausentes. Já os sistemas municipais, embora ofereçam conjuntos mais completos, carecem de dicionários de variáveis consolidados, o que exigiu múltiplas rodadas de alinhamento com as equipes da PROCEMPA. Em segundo lugar, questões éticas e legais relacionadas à privacidade e segurança dos dados sensíveis demandaram a aprovação em três comitês de ética em pesquisa, um processo que, embora imprescindível, prolongou o cronograma e impactou o acesso aos dados da SMSPA.

A avaliação de usabilidade realizada por meio da metodologia *System Usability Scale* (SUS) indicou, em caráter preliminar, que os usuários reconhecem a utilidade e a facilidade de navegação dos painéis, mostrou que o painel desenvolvido tem um

caráter promissor para a utilização no dia a dia pelos profissionais da área.

Por fim, o desenvolvimento do protótipo e a coleta de requisitos com *stakeholders* do ICI e da SMSPA comprovaram o valor do engajamento multidisciplinar e da co-criação, uma vez que a interação direta com profissionais de saúde permitiu priorizar funcionalidades de maior impacto operacional.

Este trabalho demonstrou que a visualização de dados, aliada a um processo de ETL bem estruturado e a parcerias estratégicas, pode transformar dados dispersos sobre oncologia pediátrica em informações acionáveis para gestores e profissionais do Sistema Único de Saúde (SUS).

Os principais resultados incluem:

- Modelo de dados integrado que consolida estruturas heterogêneas de diferentes fontes, facilitando consultas e análises comparativas;
- *Dashboards* interativos capazes de revelar padrões epidemiológicos, fluxos de atendimento e indicadores de deslocamento de forma intuitiva;
- Pipeline de ETL automatizado que garante a atualização periódica das bases, reduzindo o esforço manual e aumentando a confiabilidade dos dados;
- Validação de usabilidade que aponta aceitação dos usuários e oportunidades de refinamento em termos de performance e usabilidade.

Como limitações, destacam-se a dependência de políticas municipais de compartilhamento de dados, documentação incompleta e desatualizada como por exemplo a falta de dicionários consistentes para todos os bancos de dados, falta de preenchimento em campos importantes, dados de mortalidade podem ter omissões ou atrasos de notificação, além disso, há uma grande demora na atualização da base dos dados o que inviabiliza a identificação em tempo real de anormalidades. Para trabalhos futuros, recomenda-se expandir o escopo para outras ontologias de doenças, incorporar análises preditivas baseadas em machine learning e aprofundar a avaliação de impacto dos dashboards em fluxos de trabalho reais, mensurando ganhos em eficiência e qualidade do cuidado oncológico pediátrico.

REFERÊNCIAS

ANDRIENKO, G.; ANDRIENKO, N.; DYKES, J.; FABRIKANT, S.; WACHOWICZ, M. Space, time and visual analytics. International Journal of Geographical Information Science, v. 24, n. 10, p. 1577-1600, 2010. DOI: 10.1080/13658816.2010.508043.

ArcGIS. Dashboard COVID-19. Disponível em: https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6. Acesso em: 22 fev. 2024.

ATTY, B. F. Painel oncológico DATASUS: um novo instrumento para a gestão e avaliação da política de saúde no Brasil. 2020. Tese (Doutorado em Oncologia) - Faculdade de Medicina, Universidade de São Paulo, São Paulo, 2020.

BATKO, K.; ŚLĘZAK, D. The use of Big Data Analytics in healthcare. Journal of Big Data, v. 9, n. 1, p. 3, 2022. DOI: 10.1186/s40537-021-00553-4.

BEAM, A. L.; KOHANE, I. S. Big Data and Machine Learning in Health Care. JAMA, v. 319, n. 13, p. 1317-1318, 2018. DOI: 10.1001/jama.2017.18391.

BROWNSTEIN, J. S.; FREIFELD, C. C.; MADOFF, L. C. Digital disease detection--harnessing the Web for public health surveillance. New England Journal of Medicine, v. 360, n. 21, p. 2153-2155, 2009. DOI: 10.1056/NEJMp0900702.

BUHLINGER, K. et al. Análise pré-implementação do Registro Nacional de Câncer Pediátrico da Etiópia. Journal of Global Oncology, v. 7, n. 1, e2100155, 2021.

CHAN, T.; SEBOK-SYER, S. S.; THOMA, B. Learning analytics in medical education assessment: the past, the present, and the future. AEM Education and Training, v. 2, n. 2, p. 178-187, 2018.

CHUI, K.; WENGER, J.; COHEN, S.; NAUMOVA, E. Visual Analytics for Epidemiologists: Understanding the Interactions Between Age, Time, and Disease with Multi-Panel Graphs. PloS one, v. 6, p. e14683, 2011. DOI: 10.1371/journal.pone.0014683.

COURTIOL, P. et al. Deep learning-based classification of mesothelioma improves prediction of patient outcome. Nature Medicine, v. 25, n. 10, p. 1519-1525, 2019. DOI: 10.1038/s41591-019-0583-3.

ESTEVA, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature, v. 542, p. 115-118, 2017. DOI: 10.1038/nature21056.

FLORES-TORO, J. et al. O papel da iniciativa de dados sobre o câncer infantil na pesquisa oncológica pediátrica. Pediatric Blood & Cancer, v. 70, n. 4, e30132, 2023.

FREITAS, A. D. et al. Uma abordagem baseada em ontologia para o mapeamento de dados. Journal of the Brazilian Computer Society, v. 7, n. 2, p. 1-13, 2001.

HALAMKA, J. D.; MANDL, K. D.; TANG, P. C. Early experiences with personal health

records. Journal of the American Medical Informatics Association, v. 19, n. 1, p. 1-2, 2012. DOI: 10.1136/amiajnl-2011-000576.

HRIPCSAK, G.; ALBERS, D. J. Next-generation phenotyping of electronic health records. Journal of the American Medical Informatics Association, v. 20, n. 1, p. 117-121, 2013. DOI: 10.1136/amiajnl-2012-001145.

IENCA, M. et al. Considerations for ethics review of big data health research: a scoping review. PloS One, v. 13, n. 10, p. e0204937, 2018. DOI: 10.1371/journal.pone.0204937.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Câncer infantojuvenil. Disponível em: https://www.inca.gov.br/tipos-de-cancer/cancer-infantojuvenil. Acesso em: 22 dez. 2024.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Estimativa | 2023 Incidência de Câncer no Brasil. Disponível em: https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/estimativa-2023.pdf. Acesso em: 22 nov. 2024.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Registros de Câncer de Base Populacional. 2022. Disponível em: https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/registros/base-populacional. Acesso em: 22 nov. 2024.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Registros Hospitalares de Câncer. 2022. Disponível em: https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/registros/rhc. Acesso em: 22 nov. 2024.

JEANQUARTIER, F.; JEAN-QUARTIER, C.; HOLZINGER, A. Case-Based Reasoning with Open Data for Pediatric Cancer Research: A Method and Ontology to Reorganize the International Classification of Diseases. BioData Mining, v. 12, n. 2, p. 1-15, 2019. DOI: 10.1186/s13040-018-0190-8.

JENSEN, P. B.; JENSEN, L. J.; BRUNAK, S. Mining electronic health records: towards better research applications and clinical care. Nature Reviews Genetics, v. 13, n. 6, p. 395-405, 2012. DOI: 10.1038/nrg3208.

KAATSCH, P. The German childhood cancer registry. European Journal of Cancer, v. 46, n. 11, p. 1957-1961, 2010.

KHAN, S. S.; QIN, G. A review on healthcare data analytics: Recent advances and future directions. Journal of King Saud University-Computer and Information Sciences, 2018.

KOH, C. et al. Data mining in cancer research: an introduction. Journal of the National Cancer Institute Monographs, v. 2013, n. 46, p. 1-8, 2013.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. Nature, v. 521, n. 7553, p. 436-444, 2015. DOI: 10.1038/nature14539.

LI, X.; LU, Z.; SHU, X. A survey on big data applications and technologies in health informatics. Journal of King Saud University-Computer and Information Sciences, 2018.

LUPO, P. J. O. et al. Pediatric cancer survival and treatment in low- and middle-income countries. Cancer Epidemiology, Biomarkers & Prevention, v. 31, n. 11, p. 1957-1966, 2022.

MANDAWAT, A.; EBERLY, L.; BORDER, W. A Cardio-Oncology Data Commons: Lessons from Pediatric Oncology. Current Cardiology Reports, v. 21, n. 10, p. 128, 2019. DOI: 10.1007/s11886-019-1212-y.

MARR, B. Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results. Chichester: John Wiley & Sons, 2015.

MILANI, S. A. et al. Visualização de dados em saúde: uma revisão da literatura. Jornal de Análise de Dados e Visualização, v. 8, n. 3, p. 112-125, 2020.

NAÇÕES UNIDAS. Escritório das Nações Unidas no Brasil. Objetivo de Desenvolvimento Sustentável 3: Saúde e Bem-estar. Disponível em: https://brasil.un.org/pt-br/sdgs/3. Acesso em: 29 maio 2025.

NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, Haifa. p. 807-814.

OBERMEYER, Z.; EMANUEL, E. J. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. New England Journal of Medicine, v. 375, n. 13, p. 1216-1219, 2016. DOI: 10.1056/NEJMp1606181.

OCDE; EUROSTAT. Manual de Oslo 2018: diretrizes para coleta, relato e uso de dados sobre inovação. 4. ed. Paris: OECD Publishing, 2018. DOI: 10.1787/9789264304604-en.

PRATHER, J. C.; LOBACH, D. F.; GOODWIN, L. K. Medical data mining: knowledge discovery in a clinical data warehouse. In: Proceedings of the AMIA Annual Fall Symposium, 1997, Nashville. p. 101-105.

RAJKOMAR, A. et al. Scalable and accurate deep learning with electronic health records. npj Digital Medicine, v. 1, p. 18, 2018. DOI: 10.1038/s41746-018-0029-1.

SAURO, Jeff. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Denver: MeasuringU LLC, 2011. Disponível em: https://measuringu.com/product/susguide/. Acesso em: 20 maio 2025.

SECRETARIA DE ESTADO DA SAÚDE DE GOIÁS. Indicadores de Saúde - Oncologia. Disponível em: https://indicadores.saude.go.gov.br/public/oncologia.html. Acesso em: 22 fev. 2024.

TEIXEIRA, L. M. D.; ALMEIDA, M. B. Princípios ontológicos no suporte a terminologias

clínicas: método e ontologia para reorganização da Classificação Internacional de Doenças. Ciência Da Informação, v. 48, n. 1, p. 25-39, 2019. DOI: 10.18225/ci.inf.v48i1.4291.

WANG, J. et al. Análise de dados em grande escala para a melhoria do tratamento oncológico. Journal of Healthcare Engineering, v. 2018, Article ID 123456, 2018.

YANG, D. et al. Osteosarcoma Explorer: A Data Commons With Clinical, Genomic, Protein, and Tissue Imaging Data for Osteosarcoma Research. JCO Clinical Cancer Informatics, v. 7, p. e2300104, 2023. DOI: 10.1200/CCI.23.00104.

APÊNDICE A - Documentação PainelOncoped

Visão Geral da Plataforma

O Painel OncoPed é uma solução de visualização interativa de dados focada em oncologia pediátrica. A plataforma integra dados oriundos de diferentes fontes como DATASUS, RHC, SIM e SMS/PROCEMPA, com o objetivo de apoiar o monitoramento, avaliação e tomada de decisão por parte de gestores e profissionais de saúde.

Painel Oncologia (DATASUS)

ETL (Extract, Transform, Load)
Código disponível em:

- Fonte de dados: Pasta FTP: Painel Oncologia DATASUS (ftp://ftp.datasus.gov.br/dissemin/publicos/painel_oncologia/)
Pasta FTP: CNES - Habilitações dos hospitais (ftp://ftp.datasus.gov.br/dissemin/publicos/CNES/200508_/)
- Extração via automação crontab dos arquivos DBC
- Transformação do arquivo em dbc em csv, remoção de duplicatas, ajuste de variáveis (IDADE , ano_diagn e ano_tratam) números com parte fracionária (*float)* em inteiras e junção dos dados de informação de habilitação dos hospitais (CNES), *replaces* dos nomes dos tumores.

Perguntas de Negócio

- Qual é o tempo médio entre o diagnóstico e o início do tratamento?
- Há diferença nos tempos de atendimento dependendo do tipo de câncer ou do local de atendimento?
- O hospital de atendimento é habilitado para oncologia pediátrica?
- Qual é o número de pacientes atendidos em locais não habilitados para oncologia pediátrica?
- Como os tempos entre o diagnóstico e o início do tratamento de espera e tratamento variam entre unidades habilitadas e não habilitadas?

Montagem no Dashboard

- Ferramenta utilizada: Dash com python.
- Gráficos de distribuição por tumor, idade, tempo até início do tratamento.

Dicionário de Variáveis
Disponível em: https://github.com/felipeasr/Documentacao_PainelOncoped

Painel RHC

ETL

- Registro Hospitalar de Câncer (RHC).
- Automação através do python e selenium para acessar e baixar os dados.
- Extração de arquivos DBF padronizados.
- Transformação de DBF em Csv para análises
- Transformação com padronização de CID-O, unificação de duplicatas.

Perguntas de Negócio

- Incluir apenas os primeiros tratamentos.
- O hospital de atendimento é habilitado para oncologia pediátrica?
- Qual é o número de pacientes atendidos em locais não habilitados para oncologia pediátrica?
- Qual Primeiro tratamento Registrado?
- Qual a causa (CID) do óbito para pacientes oncológicos pediátricos?

Montagem no Dashboard

- Ferramenta utilizada: Dash com python.

Gráficos:

- Pacientes por Ano: evolução do número de pacientes.
- Top 10 Diagnósticos: donut com 10 tumores mais frequentes.
- Base para Diagnóstico: donut mostrando exame confirmatório (“BASMAIMP”).
- Diagnóstico/Trat. Anteriores: donut de registros pré-existentes.
- Histórico Familiar: donut com proporção de histórico familiar.
- Estado após 1º Tratamento: donut do status após primeiro tratamento.
- Registros por Sexo: barras comparando por sexo.

Dicionário de Variáveis

Disponível em: https://github.com/felipeasr/Documentacao_PainelOncoped

Painel SIM

ETL

- Fonte: Sistema de Informações sobre Mortalidade (SIM).
- Extração via automação crontab dos arquivos DBC
- Transformação do arquivo em dbc em csv, remoção de duplicatas, ajuste de variáveis
- Transformação com limpeza, mapeamento CID10 e faixas etárias.
- Foco em mortalidade por câncer em menores de 20 anos.
- Exclusão de causas externas.

Pasta FPT: ftp://ftp.datasus.gov.br/dissemin/úblicos/SIM/CID10/DORES/

Perguntas de Negócio

- Qual é a taxa de mortalidade de acordo com o tipo de câncer, idade e local de atendimento?
- Qual é a data do óbito e a causa (CID) para pacientes oncológicos pediátricos?
- Qual Sobrevida dos tumores Pediátricos

Montagem no Dashboard

- Gráficos:
    - Óbitos por Ano: barras anuais.
    - Top 10 Causas Básicas de Óbito: donut com 10 CIDs.
    - Locais de Ocorrência do Óbito: donut de local.
    - Top 10 Causas por Sexo: barras por sexo.
    - Mapa de Óbitos por Município (RS): coroplético se UF=RS.

Dicionário de Variáveis

Disponível em: https://github.com/felipeasr/Documentacao_PainelOncoped

Painel Sistemas Regulador SMSPOA

ETL

- Fonte: Base da Prefeitura (PROCEMPA/RS).
- Dados Repassados Via Drive Procempa
- Acesso no Pronto atendimento do cruzeiro do Sul
- Transformação dos arquivos xlsx em Csv para os painéis, Integração dos dados dos diferentes sistemas GERINT, GERCON e GERPAC

Perguntas de Negócio

- Qual é o tempo médio entre o diagnóstico e o início do tratamento?

- Como os tempos entre o diagnóstico e o início do tratamento de espera variam entre unidades habilitadas e não habilitadas?
- Qual é a média de distância percorrida pelos pacientes entre sua residência e o local de atendimento?
- Qual é a data do óbito e a causa (CID) para pacientes oncológicos pediátricos?
- Qual é a taxa de mortalidade de acordo com o tipo de câncer, idade e local de atendimento?

Montagem no Dashboard

Python + Dash

Gráfico	**O que mostra**
Distribuição de Diagnósticos por CID	Pizza (donut) com os 10 CIDs mais frequentes e agrupamento “Outros” para o restante.
Distribuição dos Grupos de Procedimentos	Barra mostrando quantos procedimentos de cada grupo (quimioterapia, radioterapia etc.) foram realizados.
Distribuição por Local de Residência	Pizza com os 10 municípios de residência mais comuns entre os pacientes.
Origem do Paciente (Protocolo)	Pizza que mostra a proporção de entradas pelo Gerint, Gercon, Gerapc ou sem informação.
Primeiro Procedimento Registrado (Top 10)	Pizza com os 10 primeiros procedimentos feitos por paciente, agrupando o restante como “Outros”.
Tempo diagnóstico→tratamento por faixa etária	Barras com a mediana de dias entre diagnóstico e tratamento, separadas em faixas etárias (0–4, 5–10 etc.).

Caminho do Paciente (Sankey)	Diagrama de fluxo que ilustra o caminho de cada paciente, desde o protocolo de entrada até a unidade executante.

Dicionário de Variáveis

Disponível em: https://github.com/felipeasr/Documentacao_PainelOncoped

Painel OncoPed - Projeto Arquitetural

User Stories

1. Tempo Médio de Início de Tratamento: Como oncologista pediátrico, quero visualizar o tempo médio entre diagnóstico e início do tratamento para cada tipo de tumor, para que eu possa avaliar atrasos na assistência em diferentes neoplasias.
2. Comparação por Habilitação de Unidade: Como gestor de saúde, quero comparar os tempos de espera entre unidades habilitadas e não habilitadas, para que eu possa identificar disparidades na qualidade do atendimento.
3. Filtragem por Tipo de Unidade: Como analista de dados, quero filtrar pacientes por status de habilitação da unidade, para que eu possa gerar relatórios segmentados e ações direcionadas.
4. Contagem de Pacientes em Unidades Não Habilitadas: Como formulador de políticas públicas, quero saber quantos pacientes foram atendidos em locais não habilitados, para que eu possa priorizar recursos e parcerias.
5. Distância Percorrida pelos Pacientes: Como gestor de saúde, quero calcular a média de distância entre residência e local de atendimento (RS), para que eu possa planejar novas unidades ou transporte assistido.
6. Mapeamento de Deslocamento Regional: Como gestor de saúde, quero identificar regiões com maior índice de deslocamento, para que eu possa sugerir áreas de cobertura prioritária.
7. Dados de Óbito e CID oncologista pediátrico, quero visualizar data de óbito e causa (CID) de pacientes pediátricos (RS e Brasil), para que eu possa monitorar padrões de mortalidade.
8. Curvas de Sobrevida (Kaplan-Meier): Como pesquisador(a), quero gerar curvas de sobrevida por tipo de tumor pediátrico, para que eu possa avaliar prognósticos e planos de estudo.
9. Acesso em Dispositivos: Como usuário, quero que o painel seja responsivo e

otimizado para meu dispositivo.

Requisitos Funcionais

ID	Descrição
FR-1	Calcular e exibir o tempo em mediana (dias) entre diagnóstico e início de tratamento, por tipo de tumor pediátrico.
FR-2	Permitir selecionar um tipo de tumor e visualizar seu tempo médio
FR-3	Comparar tempos de espera entre unidades habilitadas e não habilitadas em gráfico ou tabela comparativa.
FR-4	Atualizar em tempo real todas as visualizações e métricas ao aplicar filtros.
FR-5	Exibir o número total de pacientes atendidos em unidades não habilitadas (indicador numérico e gráfico).
FR-6	Calcular a distância média (km) residência-unidade para pacientes do RS e exibir em indicador e gráfico.
FR-7	Gráfico com da distância percorrida entre os Hospitais e Residências do paciente.
FR-8	Exibir gráfico da causa base do óbito em CID-10 de pacientes pediátricos.
FR-9	Filtrar óbitos por período, CID ou região e apresentar estatísticas agregadas (contagens, taxa de mortalidade).
FR-10	Gerar curvas de sobrevida Kaplan–Meier por tipo de tumor, com seleção de coortes por data de diagnóstico ou faixa etária.
FR-11	Permitir download dos dados brutos em CSV.
FR-12	Garantir que a interface seja responsiva e se adapte a desktop, tablet sem perda de usabilidade.
FR-13	Incluir testes automatizados de responsividade (pontos de quebra) para validar exibição em diferentes larguras de tela.

Requisitos Não Funcionais

ID	Descrição
NFR-1	Tempo de carregamento rápido para páginas e visualizações, com até vários usuários simultâneos.
NFR-2	Disponibilidade mínima de 99,9% mensal.
NFR-3	Realizar backup automático dos dados
NFR-4	Cumprir LGPD, criptografando dados sensíveis

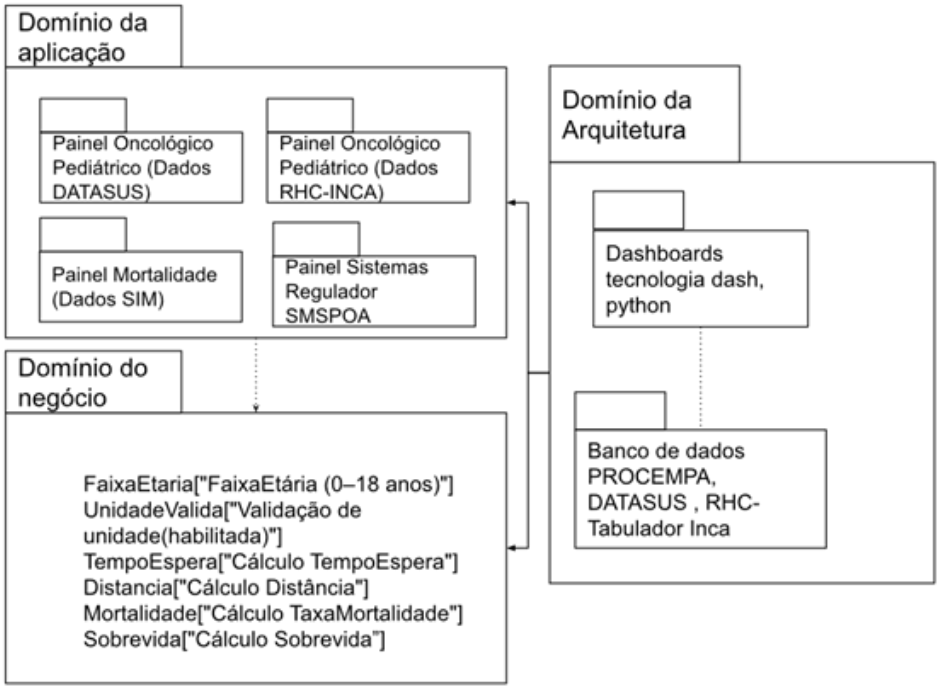
NFR-5	Disponibilizar ajuda contextual (tooltips, guias rápidos) em todos os gráficos.
NFR-6	Manter cobertura de testes unitários ≥ 80% e código modular/reutilizável.
NFR-7	Ter pipeline para testes e deploy automáticos.
NFR-8	Atualizar dados fonte (DATASUS, etc.) mensalmente, com ETL concluído até as 10 h da manhã.

Arquitetura do Projeto

Camadas:

1. Fontes de Dados: DATASUS (SIA, SIH, SISCAN); SMS RS; SIM Brasil.
2. Ingestão (ETL): Scripts Python agendados (cron).
3. Processamento: ydata profiling, pandas e numpy.
4. Persistência: Banco de dados (SQLite).
5. API: Flask/FastAPI expondo endpoints REST.
6. Apresentação: Dash + dash_bootstrap_components, gráficos interativos(plotly).
7. Implantação: railway + Nginx para HTTPS.

Diagrama de Pacotes UML



Personas

Persona 1: Gestor de Saúde

Nome: Carla Silva
Idade: 45 anos
Cargo: Coordenadora de Regulação e Planejamento Estadual
Formação: Administração Pública com MBA em Gestão em Saúde
Local de Trabalho: Secretaria Estadual de Saúde do Rio Grande do Sul
Objetivos no Painel:
- Monitorar indicadores de tempo médio entre diagnóstico e início de tratamento por região
- Identificar unidades não habilitadas com alta demanda para propor habilitação ou readequação de recursos
- Gerar relatórios consolidados para justificar alocação de verbas
Frustrações:
- Dados desatualizados ou inconsistentes
- Dificuldade de comparação entre sistemas
Citação: “Preciso de evidências rápidas e confiáveis para justificar investimentos e capacitar novas unidades.”

Persona 2: Chefe de Oncologia Pediátrica

Nome: Dr. João Pereira
Idade: 50 anos
Cargo: Chefe do Serviço de Oncologia Pediátrica
Formação: Pediatra com residência em Oncologia Pediátrica, Doutorado em Epidemiologia
Local de Trabalho: Hospital de Porto Alegre (unidade habilitada)
Objetivos no Painel:
- Avaliar curvas de sobrevida (Kaplan-Meier) de diferentes tumores
- Comparar tempos de atendimento e mortalidade entre instituições habilitadas
- Mapear causas de óbito para aprimorar diretrizes de cuidado paliativo
Frustrações:

- Dados desatualizados ou inconsistentes
- Não tem BI de Oncologia Pediátrica

Citação: “Quero ter clareza imediata sobre a sobrevida dos meus pacientes e onde podemos melhorar protocolos.”

APÊNDICE B – Pareceres dos Comitês de Ética em Pesquisa

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Análise de Dados em Oncologia Pediátrica: Investigação de Padrões Epidemiológicos em Banco de Dados **Pesquisador:** Isabel Cristina Siqueira da Silva **Área Temática:**
Versão: 2

CAAE: 76218523.0.0000.5345

Instituição Proponente: Universidade Federal de Ciências da Saúde de Porto Alegre

Patrocinador Principal: Universidade Federal de Ciências da Saúde de Porto Alegre

DADOS DO PARECER

Número do Parecer: 6.708.428

Apresentação do Projeto:

As informações elencadas neste campo foram retiradas do arquivo Informações Básicas da Pesquisa (PB_INFORMAÇOES_BÁSICAS_DO_PROJETO_2257419, DE 24/01/2024).
A análise de dados aplicada à saúde pode contribuir, de forma significativa, para o processo de gestão da jornada do paciente em todo o processo, desde a realização de um diagnóstico até o período de póstratamento. No entanto, a utilização eficaz desses dados enfrenta desafios substanciais devido à heterogeneidade e a diversidade destes, além da falta de estruturação dos registros de saúde. Questões como a ausência de padronização,
a variabilidade na qualidade dos dados e preocupações com a privacidade surgem como obstáculos críticos a serem superados.Este projeto tem, como objetivo, mapear as diferentes bases de dados relacionadas à oncologia pediátrica, no município de Porto Alegre/RS e, a partir deste mapeamento, propor um novo modelo de dados bem como elencar técnicas de visualização analítica de dados que auxiliem a tomada de decisão. Para enfrentar desafios como inconsistências nos diagnósticos, a metodologia do projeto inclui a preparação dos dados, a resolução de incongruências nos diagnósticos e a implementação de etapas de execução detalhadas. Os resultados esperados não se limitam à análise de

dados, uma vez que o projeto busca contribuir para a qualidade geral da assistência médica à oncologia pediátrica através da investigação de formas de apoiar decisões através da visualização analítica e interativa de dados.

Objetivo da Pesquisa:

O objetivo primário desse projeto é mapear as diferentes bases de dados relacionadas à oncologia pediátrica, no município de Porto Alegre e, a partir deste mapeamento, propor um novo modelo de dados bem como elencar técnicas de visualização analítica de dados que auxiliem a tomada de decisão por parte de profissionais e gestores da área da saúde.

Avaliação dos Riscos e Benefícios:

Riscos:

Como a modalidade da pesquisa proposta neste projeto é observacional sobre dados secundários e sem identificação, a qual será desenvolvida e testada pela própria equipe proponente, a ocorrência de riscos é mínima e se concentra na possibilidade de desconforto e estresse dos pesquisadores em relação a alguma eventual observação de dados incompletos e/ou não atualizados.

Benefícios:

Por benefícios oriundos desta pesquisa, destaca-se o fornecimento de insights sobre a incidência e características dos tumores pediátricos, estabelecendo um modelo replicável de modelo de dados que possa ser estendido para análises de outras enfermidades no futuro. Dessa maneira, a proposta não apenas se concentra na análise de dados, mas também na construção de um mecanismo que possibilite uma compreensão mais profunda do panorama de saúde da região da grande Porto Alegre, auxiliando na busca constante da qualidade da assistência médica oferecida à população.

Comentários e Considerações sobre a Pesquisa:

Conforme Informações Básicas do projeto na PB, trata-se de de um estudo observacional, envolvendo a análise de dados a serem fornecidos pela Secretaria Municipal de Saúde
de Porto Alegre, com foco nos casos de tumores pediátricos, de modo a propor um novo modelo de dados que permita o mapeamento e a integração de diferentes bases de conhecimento. Caráter acadêmico, realizado para obtenção de título de mestre em Tecnologias de Informação e Gestão em Saúde (TIGSaúde).

Previsão de início da coleta de requisitos, 15 de abril de 2024 e encerramento do estudo, maio de 2025.

Considerações sobre os Termos de apresentação obrigatória:

Os termos obrigatórios apresentados de forma adequada.

Conclusões ou Pendências e Lista de Inadequações:

Sugere-se a aprovação do referido estudo.

Ressalta-se que cabe ao pesquisador responsável encaminhar o(s) relatório(s) - parciais e final - da pesquisa, por meio da Plataforma Brasil, via notificação do tipo ¿relatório¿ para que sejam devidamente apreciadas no CEP, conforme Norma Operacional CNS nº 001/12, item XI.2.d.

Considerações Finais a critério do CEP:

De acordo com o parecer do Relator.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2257419.pdf	24/01/2024 19:25:16		Aceito
Outros	Carta_resposta_ao_CEP_Dados_Oncologicos_Pediatricos_assinado.pdf	24/01/2024 19:24:44	Isabel Cristina Siqueira da Silva	Aceito
Brochura Pesquisa	Analise_dados_oncologia_pediatrica_pos_cep_UFCSPA.pdf	24/01/2024 18:46:38	Isabel Cristina Siqueira da Silva	Aceito
Outros	Formulario_de_Projeto_de_Pesquisa_CEP_SMSPA_assinado.pdf	24/01/2024 17:57:14	Isabel Cristina Siqueira da Silva	Aceito
Outros	Termo_de_compromisso_CEP_SMSPA_assinado.pdf	24/01/2024 17:56:46	Isabel Cristina Siqueira da Silva	Aceito
Outros	TCUD_ASSINADO_POR_TODOS_assinado.pdf	24/01/2024 17:56:24	Isabel Cristina Siqueira da Silva	Aceito
Outros	anexo3_termo_anuencia_SMSP.pdf	24/01/2024 17:32:30	Isabel Cristina Siqueira da Silva	Aceito

Outros	Termo_anuencia_SMSPA.pdf	24/01/2024 17:03:20	Isabel Cristina Siqueira da Silva	Aceito
Projeto Detalhado / Brochura Investigador	Analise_dados_oncologia_pediatrica_cep.docx	30/11/2023 17:15:18	Isabel Cristina Siqueira da Silva	Aceito
Outros	ParecerCPP.pdf	30/11/2023 17:13:58	Isabel Cristina Siqueira da Silva	Aceito
Outros	tcdu.pdf	30/11/2023 17:13:05	Isabel Cristina Siqueira da Silva	Aceito
Outros	termo_compromisso_RELATORIO.pdf	30/11/2023 17:10:05	Isabel Cristina Siqueira da Silva	Aceito
Folha de Rosto	folhaDeRosto_assinada.pdf	30/11/2023 17:08:41	Isabel Cristina Siqueira da Silva	Aceito
Outros	termo_anuencia.pdf	30/11/2023 17:01:57	Isabel Cristina Siqueira da Silva	Aceito

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

PORTO ALEGRE, 18 de Março de 2024

Assinado por:
Fernanda Bordignon Nunes
(Coordenador(a))

PARECER CONSUBSTANCIADO DO CEP

Elaborado pela Instituição Coparticipante

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Análise de Dados em Oncologia Pediátrica: Investigação de Padrões Epidemiológicos em Banco de Dados **Pesquisador:** Isabel Cristina Siqueira da Silva **Área Temática:**
Versão: 1

CAAE: 76218523.0.3002.5327

Instituição Proponente:INSTITUTO DO CANCER INFANTIL DO RIO GRANDE DO SUL **Patrocinador Principal:** Universidade Federal de Ciências da Saúde de Porto Alegre

DADOS DO PARECER

Número do Parecer: 6.731.646

Apresentação do Projeto:

As informações elencadas nos campos "Apresentação do Projeto", "Objetivo da Pesquisa" e "Avaliação dos

Riscos e Benefícios" foram retiradas do arquivo do projeto e das Informações Básicas da Pesquisa PB_INFORMAÇÕES_BÁSICAS_2257419, de 24/01/2024.

Projeto UFCSPA-ICI-SMSPA

A análise de dados aplicada à saúde pode contribuir, de forma significativa, para o processo de gestão da jornada do paciente em todo o processo, desde a realização de um diagnóstico até o período de póstratamento. No entanto, a utilização eficaz desses dados enfrenta desafios substanciais devido à heterogeneidade e a diversidade destes, além da falta de estruturação dos registros de saúde. Questões como a ausência de padronização, a variabilidade na qualidade dos dados e preocupações com a privacidade surgem como obstáculos críticos a serem superados.Este projeto tem, como objetivo, mapear as diferentes bases de dados relacionadas à oncologia pediátrica, no município de Porto Alegre/RS e, a partir deste mapeamento, propor um novo modelo de dados bem como elencar técnicas de visualização analítica de dados que auxiliem a tomada de decisão. Para enfrentar desafios como inconsistências nos diagnósticos, a metodologia do projeto inclui a preparação dos dados, a resolução de incongruências nos diagnósticos e a implementação de etapas de execução detalhadas. Os resultados esperados não se limitam à análise de dados, uma vez que o projeto

busca contribuir para a qualidade geral da assistência médica à oncologia pediátrica através da investigação de formas de apoiar decisões através da visualização analítica e interativa de dados.

Objetivo da Pesquisa:

Objetivo Primário: mapear as diferentes bases de dados relacionadas à oncologia pediátrica, no município de Porto Alegre e, a partir deste mapeamento, propor um novo modelo de dados bem como elencar técnicas de visualização analítica de dados que auxiliem a tomada de decisão por parte de profissionais e gestores da área da saúde.
Objetivo Secundário:

- Entender as diferentes informações relacionadas à oncologia pediátrica presentes nas bases de dados,anonimizadas, a serem disponibilizadas pela Secretaria Municipal de Saúde de Porto Alegre;
- Identificar modelos de dados, já existentes, relacionados à integração de dados de oncologia pediátrica;- Identificar requisitos, junto a profissionais e gestores da área da saúde, relacionados ao monitoramento dos casos de oncologia pediátrica em Porto Alegre, de modo a auxiliar na análise destes bem como a tomada de decisão;
- Elencar técnicas de modelagem, análise e visualização de dados que auxiliem na identificação de padrõese tendências relevantes nos dados de oncologia pediátrica;
- Projetar, desenvolver e testar um sistema interativo, baseado em técnicas de visualização analítica dedados, voltado a profissionais e gestores em saúde a fim de auxiliar na compreensão dos diferentes aspectos envolvidos no diagnóstico e tratamento relacionados à oncologia pediátrica.

Avaliação dos Riscos e Benefícios:

Riscos: Como a modalidade da pesquisa proposta neste projeto é observacional sobre dados secundários e sem identificação, a qual será desenvolvida e testada pela própria equipe proponente, a ocorrência de riscos é mínima e se concentra na possibilidade de desconforto e estresse dos pesquisadores em relação a alguma eventual observação de dados incompletos e/ou não atualizados.
Benefícios: Por benefícios oriundos desta pesquisa, destaca-se o fornecimento de insights sobre a incidência e características dos tumores pediátricos, estabelecendo um modelo replicável de modelo de dados que possa ser estendido para análises de outras enfermidades no futuro. Dessa maneira, a proposta não apenas se concentra na análise de dados, mas também na

construção de um mecanismo que possibilite uma compreensão mais profunda do panorama de saúde da região da grande Porto Alegre, auxiliando na busca constante da qualidade da assistência médica oferecida à população.

Comentários e Considerações sobre a Pesquisa:

Conforme Informações Básicas do projeto na PB, trata-se de de um estudo observacional, envolvendo a análise de dados a serem fornecidos pela Secretaria Municipal de Saúde de Porto Alegre, com foco nos casos de tumores pediátricos, de modo a propor um novo modelo de dados que permita o mapeamento e a integração de diferentes bases de conhecimento.
Caráter acadêmico, realizado para obtenção de título de mestre em Tecnologias de Informação e Gestão em Saúde (TIGSaúde).
Previsão de início da coleta de requisitos, 15 de abril de 2024 e encerramento do estudo, maio de 2025.

Considerações sobre os Termos de apresentação obrigatória:

Não apresenta TCLE, pois os dados serão anonimizados pelo controlador.

Conclusões ou Pendências e Lista de Inadequações:

O projeto não apresenta pendências e está em condições de aprovação.

Considerações Finais a critério do CEP:

- Diante do exposto, o Comitê de Ética em Pesquisa do Hospital de Clínicas de Porto Alegre, de acordo comas atribuições definidas na Resolução CNS N.º 466/2012 e na Norma Operacional CNS/Conep N.º 001/2013, manifesta-se pela aprovação do projeto de pesquisa proposto.

- Eventos adversos deverão ser comunicados de acordo com as orientações da Comissão Nacional de Éticaem Pesquisa - Conep (Carta Circular N.º 13/2020-CONEP/SECNS/MS). Os desvios de protocolo também deverão ser comunicados em relatórios consolidados, por meio de Notificação.

- Deverão ser apresentados relatórios semestrais e um relatório final. Estes relatórios deverão sersubmetidos acompanhados do Relatório Consolidado de Eventos Adversos Graves (EAGs), conforme preconiza a Carta Circular nº 13/2020-CONEP/SECNS/MS.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Outros	Carta_resposta_ao_CEP_Dados_Oncologicos_Pediatricos_assinado.pdf	24/01/2024 19:24:44	Isabel Cristina Siqueira da Silva	Aceito
Brochura Pesquisa	Analise_dados_oncologia_pediatrica_pos_cep_UFCSPA.pdf	24/01/2024 18:46:38	Isabel Cristina Siqueira da Silva	Aceito
Outros	Formulario_de_Projeto_de_Pesquisa_CEP_SMSPA_assinado.pdf	24/01/2024 17:57:14	Isabel Cristina Siqueira da Silva	Aceito
Outros	Termo_de_compromisso_CEP_SMSPA_assinado.pdf	24/01/2024 17:56:46	Isabel Cristina Siqueira da Silva	Aceito
Outros	TCUD_ASSINADO_POR_TODOS_assinado.pdf	24/01/2024 17:56:24	Isabel Cristina Siqueira da Silva	Aceito
Outros	anexo3_termo_anuencia_SMSP.pdf	24/01/2024 17:32:30	Isabel Cristina Siqueira da Silva	Aceito
Outros	Termo_anuencia_SMSPA.pdf	24/01/2024 17:03:20	Isabel Cristina Siqueira da Silva	Aceito
Projeto Detalhado / Brochura Investigador	Analise_dados_oncologia_pediatrica_cep.docx	30/11/2023 17:15:18	Isabel Cristina Siqueira da Silva	Aceito
Outros	ParecerCPP.pdf	30/11/2023 17:13:58	Isabel Cristina Siqueira da Silva	Aceito
Outros	tcdu.pdf	30/11/2023 17:13:05	Isabel Cristina Siqueira da Silva	Aceito
Outros	termo_compromisso_RELATORIO.pdf	30/11/2023 17:10:05	Isabel Cristina Siqueira da Silva	Aceito
Outros	termo_anuencia.pdf	30/11/2023 17:01:57	Isabel Cristina Siqueira da Silva	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:
Não

PORTO ALEGRE, 28 de Março de 2024

Assinado por:
Daisy Crispim Moreira
(Coordenador(a))

PARECER CONSUBSTANCIADO DO CEP

Elaborado pela Instituição Coparticipante

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Análise de Dados em Oncologia Pediátrica: Investigação de Padrões Epidemiológicos em Banco de Dados **Pesquisador:** Isabel Cristina Siqueira da Silva **Área Temática:**
Versão: 2
CAAE: 76218523.0.3001.5338
Instituição Proponente: Secretaria Municipal de Saúde de Porto Alegre
Patrocinador Principal: Universidade Federal de Ciências da Saúde de Porto Alegre

DADOS DO PARECER

Número do Parecer: 7.217.042

Apresentação do Projeto:
TRATA-SE DE RETORNO DE PENDÊNCIAS - PARECER CONSUBSTANCIADO 6.768.819 DE
16/04/2024.

A presente proposta de pesquisa terá caráter observacional, envolvendo a análise de dados a serem fornecidos pela Secretaria Municipal de Saúde de Porto Alegre, com foco nos casos de tumores pediátricos, de modo a propor um novo modelo de dados que permita o mapeamento e a integração de diferentes bases de conhecimento. A partir de então, técnicas de visualização analítica de dados deverão ser empregadas na construção de dashboards interativos visando auxiliar gestores de saúde na tomada de decisões. Ao explorar a riqueza de informações contida em diferentes bancos de dados de saúde, o projeto almeja, além de fornecer insights sobre a incidência e características dos tumores pediátricos, estabelecer um modelo replicável de modelo de dados que possa ser estendido para análises de outras enfermidades no futuro. Dessa maneira, a proposta não apenas se concentra na análise de dados, mas também na construção de um mecanismo que possibilite uma compreensão mais profunda do panorama de saúde da região da grande Porto Alegre, auxiliando na busca constante da qualidade da assistência médica oferecida à população. Para tanto, esta pesquisa contará com a participação de profissionais do Instituto do Câncer Infantil (ICI) e da Secretaria Municipal de Saúde de Porto Alegre (SMSPA) além de pesquisadores da Universidade Federal de Ciências da

Saúde de Porto Alegre (UFCSPA). Os profissionais do ICI e da SMSPA atuarão como stakeholders e especialistas, participando ativamente do desenvolvimento desta pesquisa junto à equipe da UFCSPA através dos conceitos da metodologia de design centrada no usuário. O projeto proposto terá as principais etapas: - Preparação dos dados, os quais já estarão anonimizados ao chegar para a equipe que atuará nesta pesquisa; - Resolução de inconsistências nos diagnósticos através da identificação padrões de erro nas incongruências e análise de causas subjacentes; - Modelagem dos dados, os quais serão centralizados em um repositório central; - Análise dos dados, visando identificar padrões, tendências e insights relevantes; - Implementação de um dashboard interativo, a partir do emprego de técnicas de visualização analítica de dado; - Avaliação do dashboard pela equipe envolvida no projeto.

Objetivo da Pesquisa:

Objetivo Primário: O objetivo primário desse projeto é mapear as diferentes bases de dados relacionadas à oncologia pediátrica, no município de Porto Alegre e, a partir deste mapeamento, propor um novo modelo de dados bem como elencar técnicas de visualização analítica de dados que auxiliem a tomada de decisão por parte de profissionais e gestores da área da saúde.

Objetivo Secundário: O objetivo secundário desdobra-se em: - Entender as diferentes informações relacionadas à oncologia pediátrica presentes nas bases de dados, anonimizadas, a serem disponibilizadas pela Secretaria Municipal de Saúde de Porto Alegre; - Identificar modelos de dados, já existentes, relacionados à integração de dados de oncologia pediátrica; - Identificar requisitos, junto a profissionais e gestores da área da saúde, relacionados ao monitoramento dos casos de oncologia pediátrica em Porto Alegre, de modo a auxiliar na análise destes bem como a tomada de decisão; - Elencar técnicas de modelagem, análise e visualização de dados que auxiliem na identificação de padrões e tendências relevantes nos dados de oncologia pediátrica; - Projetar, desenvolver e testar um sistema interativo, baseado em técnicas de visualização analítica de dados, voltado a profissionais e gestores em saúde a fim de auxiliar na compreensão dos diferentes aspectos envolvidos no diagnóstico e tratamento relacionados à oncologia pediátrica.

Avaliação dos Riscos e Benefícios:

Riscos: Como a modalidade da pesquisa proposta neste projeto é observacional sobre dados secundários e sem identificação, a qual será desenvolvida e testada pela própria equipe proponente, a ocorrência de riscos é mínima e se concentra na possibilidade de desconforto e estresse dos pesquisadores em relação a alguma eventual observação de dados incompletos

e/ou não atualizados.
Benefícios: Por benefícios oriundos desta pesquisa, destaca-se o fornecimento de insights sobre a incidência e características dos tumores pediátricos, estabelecendo um modelo replicável de modelo de dados que possa ser estendido para análises de outras enfermidades no futuro. Dessa maneira, a proposta não apenas se concentra na análise de dados, mas também na construção de um mecanismo que possibilite uma compreensão mais profunda do panorama de saúde da região da grande Porto Alegre, auxiliando na busca constante da qualidade da assistência médica oferecida à população.

Comentários e Considerações sobre a Pesquisa:

1) Título: Análise de Dados em Oncologia Pediátrica: Investigação de Padrões Epidemiológicos em Bancode Dados
2) Pesquisadora responsável: Isabel Cristina Siqueira da Silva
3) Assistentes de pesquisa: Algemir Lunardi Brunetto, Jorge Luiz Silveira Osorio, Felipe de Assis Ribeiro,

Virginia Tafas da Nóbrega, Marialva Sinigaglia

4) Nível da pesquisa: Mestrado
5) Instituição: Universidade Federal de Ciências da Saúde de Porto Alegre
6) Curso: Programa de Pós Graduação em Tecnologias da Informação e Gestão em Saúde
7) Local de realização do estudo: Secretaria Municipal de Saúde de Porto Alegre - dados de gerenciamentode internações hospitalares (GERINT), de gerenciamento de consultas (GERCON), de gerenciamento de procedimentos ambulatoriais (GERPAC) e de informação sobre mortalidade (SIM).
8) Duração do estudo: 12 meses

9) Número de sujeitos da pesquisa: 10.000
10) Data prevista para conclusão do estudo: maio de 2025

Considerações sobre os Termos de apresentação obrigatória:

1) Termo de Consentimento Livre e Esclarecido: Não apresentado (não se aplica);
2) Termo de uso de imagem e som: Não apresentado (não se aplica);
3) Formulário de projeto de pesquisa: Apresentado;
4) Termo de Anuência Institucional: Apresentado, assinado por Denise Tessler Soltof, da Diretoria-Geral deRegulação (DR), da Secretaria Municipal de Saúde (SMS), da Prefeitura Municipal de Porto Alegre, em 12 de janeiro de 2024;
5) Termo de compromisso de utilização e divulgação dos dados: Apresentado, assinado pela pesquisadoraresponsável, Isabel Cristina Siqueira da Silva, em 24 de janeiro de 2024.

Recomendações:

Não há.

Conclusões ou Pendências e Lista de Inadequações:

ANÁLISE DAS RESPOSTAS ÀS PENDÊNCIAS MENCIONADAS NO PARECER CONSUBSTANCIADO
6.768.819 DE 16/04/2024.

Pendência 1/3 - Considerando que o risco descrito é o de ¿possibilidade de desconforto e estresse dos pesquisadores¿ e que os riscos envolvidos devem dizer respeito aos participantes e não aos pesquisadores, solicita-se que sejam explicitados os riscos para os participantes da pesquisa (entende-se como participantes as pessoas cujos dados estão nas bases de dados que serão utilizadas para fins de pesquisa). Ou seja, mesmo que a pesquisa apresente riscos mínimos aos participantes, os riscos devem ser considerados e devem ser explicitados, bem como as medidas adotadas para minimizar;

Resposta da pesquisadora responsável: ¿O trecho, no texto da pesquisa, que fala sobre os riscos (página 15, item12) foi alterado a fim de complementar a informação: `(...) Como a modalidade da pesquisa proposta neste projeto é observacional sobre dados secundários e sem identificação, a qual será desenvolvida e testada pela

própria equipe proponente, a ocorrência de riscos é mínima e se concentra na possibilidade de desconforto e estresse dos pesquisadores em relação a alguma eventual observação de dados incompletos e/ou não atualizados. Já os riscos relacionados aos participantes (entende-se, como participantes, as pessoas cujos dados estão nas bases de dados que serão utilizadas para fins de pesquisa) envolvem a quebra de confidencialidade dos - minimizado através da assinatura do termo de confidencialidade pelos pesquisadores envolvidos, garantindo o sigilo das informações coletadas. Além deste termo, destaca-se que o acesso aos dados, por parte da equipe, será realizado, exclusivamente, em um computador específico disponibilizado no Posto de Saúde Postão da Cruzeiro, localizado na Rua Prof. Manoel Lobato, 151, Bairro Santa Tereza, na cidade de Porto Alegre. Tal acesso se dará de forma presencial, seguindo todas as normas de segurança estabelecidas pela SMSPA e pela PROCEMPA, garantindo, assim, o uso de uma rede de computadores devidamente configurada para a proteção contínua das informações ao longo de todo o processo (...).¿¿
Situação: pendência atendida.

Pendência 2/3 - Considerando a informação de que ¿dados [serão] fornecidos pela Secretaria

Municipal de Saúde de Porto Alegre¿ e de que se trata de dados pessoais, solicita-se que seja explicitado com quais ¿login¿ e ¿senha¿ os dados serão acessados e quem será o responsável por repassar tais dados à equipe de pesquisa;
Resposta da pesquisadora responsável: ¿O fornecimento dos dados ocorrerá via acesso ao software Power BI, utilizando as credenciais fornecidas pelos servidores da Prefeitura de Porto Alegre para a pesquisadora responsável por esta proposta de pesquisa, como pode ser observado a seguir: Logon de rede: icsilva; Senha: vbqeqpqtfw. Como informado na resposta da pendência 1, o acesso aos dados será realizado em um computador disponibilizado em um endereço específico, sob a

supervisão do servidor Eliel Pereira¿.
Situação: pendência atendida.

Pendência 3/3 - Considerando que o presente protocolo está tramitando neste Comitê de Ética em Pesquisa SMSPA no mês de abril de 2024 e que, no termo de compromisso de uso de dados, o período informado para a coleta teria iniciado no mês de fevereiro, solicita-se que o cronograma de coleta de dados seja ajustado.
Resposta da pesquisadora responsável: ¿Como sugerido, o cronograma foi ajustado e pode ser observado na página 15, item 6, do texto da pesquisa.¿ Situação: pendência atendida.

Considerações Finais a critério do CEP:
O parecer de aprovação do CEP SMSPA deverá ser apresentado à Coordenação responsável, a fim de organizar a inserção da pesquisa no serviço, antes de seu início. Os relatórios semestrais devem ser apresentados ao CEP SMSPA, através de submissão na Plataforma Brasil, como "Notificação".

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_2306026.pdf	31/10/2024 23:44:48		Aceito
Outros	Carta_resposta_CEP_SMSPA_assinado.pdf	31/10/2024 23:40:31	Isabel Cristina Siqueira da Silva	Aceito
Brochura Pesquisa	Analise_dados_oncologia_pediatrica_cep.pdf	31/10/2024 23:39:12	Isabel Cristina Siqueira da Silva	Aceito
Outros	Carta_resposta_ao_CEP_Dados_Oncologicos_Pediatricos_assinado.pdf	24/01/2024 19:24:44	Isabel Cristina Siqueira da Silva	Aceito

Brochura Pesquisa	Analise_dados_oncologia_pediatrica_pos_cep_UFCSPA.pdf	24/01/2024 18:46:38	Isabel Cristina Siqueira da Silva	Aceito

Outros	Formulario_de_Projeto_de_Pesquisa_CEP_SMSPA_assinado.pdf	24/01/2024 17:57:14	Isabel Cristina Siqueira da Silva	Aceito
Outros	Termo_de_compromisso_CEP_SMSPA_assinado.pdf	24/01/2024 17:56:46	Isabel Cristina Siqueira da Silva	Aceito
Outros	TCUD_ASSINADO_POR_TODOS_assinado.pdf	24/01/2024 17:56:24	Isabel Cristina Siqueira da Silva	Aceito
Outros	anexo3_termo_anuencia_SMSP.pdf	24/01/2024 17:32:30	Isabel Cristina Siqueira da Silva	Aceito
Outros	Termo_anuencia_SMSPA.pdf	24/01/2024 17:03:20	Isabel Cristina Siqueira da Silva	Aceito
Projeto Detalhado / Brochura Investigador	Analise_dados_oncologia_pediatrica_cep.docx	30/11/2023 17:15:18	Isabel Cristina Siqueira da Silva	Aceito
Outros	ParecerCPP.pdf	30/11/2023 17:13:58	Isabel Cristina Siqueira da Silva	Aceito
Outros	tcdu.pdf	30/11/2023 17:13:05	Isabel Cristina Siqueira da Silva	Aceito
Outros	termo_compromisso_RELATORIO.pdf	30/11/2023 17:10:05	Isabel Cristina Siqueira da Silva	Aceito
Outros	termo_anuencia.pdf	30/11/2023 17:01:57	Isabel Cristina Siqueira da Silva	Aceito

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

PORTO ALEGRE, 10 de Novembro de 2024

Assinado por:
Alexandre Luis da Silva Ritter
(Coordenador(a))

APÊNDICE C – Formulário SUS

Essa seção está baseada no método de avaliação de usabilidade chamado System Usability Scale (SUS). Por favor, analise cada afirmação de acordo com o seu grau de concordância/discordância, assinalando uma das opções: discordo totalmente, discordo, neutro, concordo, concordo totalmente.

1) Eu gostaria de usar o painel OncoPed sempre que precisasse de informações relacionadas a padrões e tendências na oncologia pediátrica. *

- ○ Discordo totalmente
- ○ Discordo
- ○ Neutro
- ○ Concordo
- ○ Concordo totalmente

2) Eu considero o painel OncoPed desnecessariamente complexo. *

- ○ Discordo totalmente
- ○ Discordo
- ○ Neutro
- ○ Concordo
- ○ Concordo totalmente

3) Eu achei o painel OncoPed fácil de usar. *

- ○ Discordo totalmente
- ○ Discordo
- ○ Neutro
- ○ Concordo
- ○ Concordo totalmente

4) Eu acredito que precisaria de ajuda de uma pessoa com conhecimentos técnicos para usar painel OncoPed. *

- ○ Discordo totalmente
- ○ Discordo
- ○ Neutro
- ○ Concordo
- ○ Concordo totalmente

5) Eu considero que as funcionalidades do painel OncoPed estão bem integradas. *

- ○ Discordo totalmente
- ○ Discordo
- ○ Neutro
- ○ Concordo
- ○ Concordo totalmente

6) Eu considero que o painel OncoPed apresenta muita inconsistência. *

- ○ Discordo totalmente
- ○ Discordo
- ○ Neutro
- ○ Concordo
- ○ Concordo totalmente

7) Eu acredito que as pessoas aprenderão a usar o painel OncoPed rapidamente. *

- ○ Discordo totalmente
- ○ Discordo
- ○ Neutro
- ○ Concordo
- ○ Concordo totalmente

8) Eu achei o painel OncoPed complicado de usar. *

- ○ Discordo totalmente
- ○ Discordo
- ○ Neutro
- ○ Concordo
- ○ Concordo totalmente

9) Eu me senti confiante ao usar o painel OncoPed, entendendo os dados apresentados nos gráficos. *

- ○ Discordo totalmente
- ○ Discordo
- ○ Neutro
- ○ Concordo
- ○ Concordo totalmente

10) Eu precisei aprender várias coisas novas antes de conseguir usar o painel OncoPed. *

- ○ Discordo totalmente
- ○ Discordo
- ○ Neutro
- ○ Concordo
- ○ Concordo totalmente