

## Ferramenta para análise do deslocamento de pacientes em tratamento de câncer de mama no SUS a partir do Registro Hospitalar de Câncer

A tool for analysis of patients' travels for treating breast cancer in the Brazilian Public Health System based on the national Hospital Cancer Registry

Herramienta para el análisis del movimiento de pacientes en tratamiento de cáncer de mama en el SUS a partir del Registro Hospitalario de Cáncer

Luis Fernando Reis Tavares Pais (<https://orcid.org/0000-0002-4802-0196>)<sup>1</sup>  
José Luiz Barbosa Bevilacqua (<https://orcid.org/0000-0002-3052-4062>)<sup>2</sup>  
José Maria Parente de Oliveira (<https://orcid.org/0000-0002-7803-1718>)<sup>3</sup>  
Ricardo da Silva Santos (<https://orcid.org/0000-0001-8933-790X>)<sup>3,4</sup>

**Resumo** Foram estimados 704 mil casos de câncer no Brasil no triênio 2023-2025, sendo o câncer de mama o tipo mais incidente, com 10,37% do total. A organização das redes de atenção se baseia no modelo regionalizado. Tal organização, contudo, faz com que os pacientes algumas vezes tenham de percorrer longas distâncias para receber o diagnóstico e o tratamento. O deslocamento de pacientes vem sendo alvo de estudos em todo o mundo. No contexto internacional, o principal foco é analisar a associação entre o deslocamento e o desfecho clínico. No cenário nacional, o foco ainda se restringe à caracterização dos deslocamentos, e para isso usa a base de dados do sistema de informações ambulatoriais (SIA-SUS). Considerando a necessidade de avaliar o uso de uma base que permita análises mais detalhadas sobre o deslocamento de pacientes, o objetivo deste trabalho foi apresentar o desenvolvimento de uma ferramenta computacional para análise de deslocamentos de pacientes em tratamento de câncer de mama no SUS, a partir dos dados do Registro Hospitalar de Câncer (RHC). A ferramenta foi desenvolvida e implementada e, apesar de limitações potenciais, mostrou-se eficiente na produção de informações.

**Palavras-chave** Análise de dados, Mineração de dados, Neoplasias da mama, Regionalização da saúde, Acesso aos serviços de saúde

**Abstract** In Brazil, 704,000 cases of cancer have been estimated for the 2023-2025 triennium, with breast cancer being the most common type, accounting for 10.37% of the total. The organization of care networks is based on the regionalized model. However, this structure means that patients are sometimes obliged to travel long distances to receive the diagnosis and treatment. Patient displacement has been the subject of studies worldwide. In the international context, the focus is on analyzing the association between displacement and clinical outcome. In the national scene, the focus is still restricted to the characterization of displacement, and for this purpose, it uses the database of the outpatient information system (SIA-SUS). In view of the need to evaluate the use of a database that allows more detailed analysis of patient displacement, the scope of this study was to present the development of a computational tool for analyzing the displacement of patients undergoing breast cancer treatment in the SUS, using data from the Hospital Cancer Registry (RHC). The tool was developed and implemented, and, despite potential limitations, it proved to be efficient in generating information.

**Key words** Data analysis, Data mining, Breast cancer, Regionalization of health, Access to health services

**Resumen** En Brasil, se estimaron 704 mil casos en el trienio 2023-2025, siendo el cáncer de mama el tipo más común, representando el 10,37% del total. La organización de las redes de atención se basa en el modelo regionalizado. Sin embargo, esta organización implica que a veces los pacientes tienen que viajar largas distancias para recibir diagnóstico y tratamiento. El movimiento de pacientes ha sido objeto de estudios en todo el mundo. En el contexto internacional, el enfoque principal es analizar la asociación entre el desplazamiento y el resultado clínico. En el escenario nacional, el enfoque aún se restringe a la caracterización de los desplazamientos, y para ello se utiliza la base de datos del sistema de información ambulatoria (SIA-SUS). Considerando la necesidad de evaluar el uso de una base de datos que permita análisis más detallados sobre el movimiento de pacientes, el objetivo de este trabajo fue presentar el desarrollo de una herramienta computacional para analizar el movimiento de pacientes en tratamiento de cáncer de mama en el SUS, con base en datos del Registro Hospitalario de Cáncer (RHC). La herramienta fue desarrollada e implementada y, a pesar de las posibles limitaciones, demostró ser eficiente en la producción de información.

**Palabras clave** Análisis de datos, Minería de datos, Neoplasias de mama, Regionalización de la salud, Acceso a servicios de salud

<sup>1</sup> Universidade de Mogi das Cruzes. Av. Dr. Cândido Xavier de Almeida e Souza 200. 08780-911 Mogi das Cruzes SP Brasil. [luisfernandopais@gmail.com](mailto:luisfernandopais@gmail.com)

<sup>2</sup> Hospital Sírio Libanês. São Paulo SP Brasil.

<sup>3</sup> Instituto Tecnológico de Aeronáutica. São José dos Campos SP Brasil

<sup>4</sup> Compumédica Pesquisa & Inovação. São Paulo SP Brasil.

## Introdução

O câncer vem sendo considerado a principal barreira ao aumento da expectativa de vida em todos os países, com taxas de incidência e de mortalidade em crescimento acelerado<sup>1</sup>. No Brasil, estima-se mais de 704 mil novos casos no triênio 2023-2025, sendo o câncer de mama o tipo mais incidente com 10,37% do total<sup>2</sup>.

Em 2013, foi instituída no Brasil a Política Nacional para a Prevenção e Controle do Câncer, que enfatiza a organização de redes de atenção regionalizadas e descentralizadas<sup>3</sup>. Tal política pressupõe que o acesso aos serviços seja feito em unidades próximas ao local de residência dos pacientes; no entanto, há alguns fatores no Sistema Único de Saúde (SUS) que dificultam a formação de redes regionais homogêneas: a heterogeneidade do território brasileiro, a formalização de responsabilidades institucionais em cada um dos níveis de atenção e a delicada questão político-administrativa da manutenção de autonomia dos governos locais com as regulações federais e estaduais, entre outros<sup>4</sup>.

As diferenças, entretanto, ocorrem até mesmo em um único estado. Alves e Magalhães<sup>5</sup> citam a região de saúde do norte de Minas Gerais como um exemplo de onde a distribuição geográfica não está otimizada, tendo somente um centro municipal com serviços oncológicos para atender toda a região, incluindo cidades a mais de 300 quilômetros de distância.

A consolidação de centros especializados em oncologia, buscando a otimização dos recursos de saúde, é comum em várias partes do mundo. Contudo, isso faz com que os pacientes muitas vezes tenham de percorrer longas distâncias para receberem o diagnóstico e o tratamento adequados.

O deslocamento de pacientes em tratamento de câncer vem sendo alvo de estudos em todo o mundo.

No contexto internacional, o principal foco é a análise da associação entre o deslocamento dos pacientes e o impacto no estadiamento e no desfecho clínico<sup>6-11</sup>. Tais estudos, com exceção de Fabian *et al.*<sup>8</sup>, mostram associação negativa entre o tempo de viagem e o estágio do câncer no diagnóstico, assim como seu desfecho clínico. Esse estudo comparou a apresentação clínica de pacientes com retinoblastoma em países da Europa e da África. Os resultados mostraram que os pacientes de países africanos viajam distâncias menores, no entanto, apresentam o estágio do câncer no diagnóstico mais avançado.

No cenário brasileiro, o foco se restringe à caracterização dos deslocamentos, sem, entre-

tanto, analisar os impactos associados a esses deslocamentos, e para isso usa a base de dados do sistema de informações ambulatoriais (SIA-SUS)<sup>4,5,12-14</sup>.

Deve-se esperar que a evolução natural dos estudos nacionais seja a análise da associação entre o deslocamento do paciente e o seu desfecho clínico. No entanto, esse próximo estágio requer o uso de uma base de dados com informações mais detalhadas sobre o tratamento do paciente, tal como o Registro Hospitalar de Câncer (RHC)<sup>15,16</sup>.

Considerando a necessidade de avaliar a viabilidade de usar a base de dados do RHC, o objetivo deste trabalho foi apresentar o desenvolvimento de uma ferramenta computacional para análise de deslocamentos de pacientes em tratamento de câncer de mama no SUS, a partir dos dados do RHC.

## Métodos

O processo de desenvolvimento e validação da ferramenta consistiu em cinco grandes etapas: (1) carga inicial de dados do RHC; (2) seleção do subconjunto de dados; (3) pré-processamento dos dados; (4) elaboração do painel de análises descritivas; e (5) validação da ferramenta.

A carga inicial de dados consistiu em coletar os dados públicos do RHC e armazená-los em um banco de dados local. A próxima etapa consistiu em filtrar, do banco local, um subconjunto de dados restrito ao escopo do estudo desejado; em nosso caso, câncer de mama. O pré-processamento corresponde à preparação dos dados, visando deixá-los em um formato adequado à elaboração das análises. A quarta etapa corresponde à elaboração de um painel contendo as análises descritivas, por meio de gráficos e tabelas dinâmicas. Por fim, a quinta etapa consistiu na validação da ferramenta.

### Carga inicial de dados

Os RHC podem ser caracterizados como centros de informação situados em unidades hospitalares (UH) de assistência oncológica, cujas atividades incluem: coleta de dados dos pacientes com diagnóstico confirmado de câncer atendidos nessas unidades; armazenamento e processamento desses dados; e análise e divulgação dessas informações<sup>17</sup>.

O Inca é o órgão responsável pela consolidação e publicação em âmbito nacional das informações geradas pelos RHC de todo o Brasil. Para isso, desenvolveu dois sistemas: SisRHC e

IntegradorRHC. O IntegradorRHC, é um sistema *web* que efetua a consolidação e divulgação dos dados hospitalares incluídos no SisRHC<sup>15,18</sup>.

As UH enviam os dados através da internet e cabe às coordenações estaduais fazerem uma primeira consolidação estadual. Posteriormente, é feita também uma consolidação nacional, para que só então os dados possam ser divulgados no *site* do IntegradorRHC. A principal função dessa consolidação é identificar redundâncias (casos em que o paciente foi registrado em mais de um centro de tratamento) e eliminá-las, evitando assim que o número de ocorrências seja indevidamente superestimado<sup>18</sup>.

Foram carregados os arquivos do RHC disponibilizados no *site* do Inca, contemplando o período de 2000 a 2017. A data da primeira consulta foi o atributo usado para a delimitação do período.

A fim de avaliar a coerência dos dados carregados em cada ano, dois métodos foram adotados: o volume de registros, para identificar se haveria algum período muito discrepante dos demais; e uma consulta comparativa no *site* do DATASUS, para comprovar se a quantidade de casos carregados estaria coerente com os valores apresentados pelo DATASUS. Caso fosse encontrada alguma discrepância, o arquivo deveria ser desconsiderado da carga e recarregado após identificação e correção do problema.

### **Seleção do subconjunto de dados (*dataset*)**

O processo de carga incluiu no banco de dados registros referentes a todos os tipos de câncer. No entanto, cada estudo elaborado pela ferramenta se refere a apenas um tipo de câncer, em nosso caso, o de mama.

A seleção deve ser feita a partir do atributo “CID Primário”. Em nosso estudo foram considerados todos os registros cujos três primeiros dígitos do CID eram “C50”.

Além da seleção pelo atributo “CID Primário”, outros critérios de seleção também foram adotados: atributo “sexo” com o valor igual a “2” (feminino); e atributo “UF residencial” diferente de “99” (este código indica uma unidade da federação desconhecida ou fora do Brasil).

A existência de atributos sem preenchimento na base de dados do RHC não constituiu um critério de perda de registros; uma vez que o propósito primordial da ferramenta é mostrar o deslocamento dos pacientes, os únicos atributos que obrigatoriamente devem estar preenchidos corretamente são “município de residência” do paciente e “município da unidade” de tratamento.

### **Pré-processamento**

A atividade mais importante da etapa de pré-processamento é a inclusão do atributo “distância percorrida”, que deve armazenar a distância percorrida da residência do paciente à unidade hospitalar de tratamento.

O deslocamento do paciente foi calculado a partir da distância entre os municípios de residência do paciente e da unidade de tratamento, uma vez que o endereço exato do paciente não está disponível no arquivo origem. Algumas premissas foram assumidas para esse cálculo.

A primeira delas foi que a distância seria considerada zero (0 km) quando os municípios de residência e da unidade hospitalar fossem os mesmos, independentemente do porte da cidade.

Uma vez que não há como identificar qual foi o transporte utilizado pelo paciente, adotou-se como segunda premissa considerar o meio de transporte rodoviário como a forma de deslocamento.

A última premissa foi que o cálculo da distância adotaria a malha rodoviária existente no momento do cálculo, outubro de 2019, e refletiu, portanto, o caminho disponível nessa ocasião. Se, por acaso, no ano 2000 não existia uma estrada pavimentada entre dois municípios que existia em 2019, o deslocamento talvez pudesse ter sido maior, no entanto, isso não foi considerado no cálculo.

Caso as condições anteriores não pudessem ser atendidas, a distância seria calculada usando a diferença geodésica entre dois pontos, através de sua latitude e longitude. O cálculo geodésico não considera o relevo, as estradas e as barreiras naturais entre municípios, por isso não é de todo correto, mas foi considerado neste estudo como uma aproximação válida. Tal cálculo poderia ser útil, por exemplo, em municípios localizados em ilhas (como Fernando de Noronha) ou cujo acesso é somente via barcos.

Para calcular a distância de deslocamento de cada ocorrência da base, foi feita inicialmente uma consolidação dos dados, a fim de identificar apenas os pares origem-destino unívocos. Por meio de uma planilha auxiliar, os dados da base principal foram agrupados, sendo selecionadas apenas as quatro colunas necessárias para calcular a distância de deslocamento: a UF e o município de residência do paciente, e a UF e o município da unidade hospitalar.

O cálculo da distância em metros foi feito por meio de uma interface (Application Programming Interface, ou API), disponibilizada

pelo Google em sua plataforma relacionada a mapas.

Após o cálculo para todos os pares origem-destino da tabela auxiliar, cada registro do *dataset* foi atualizado com a distância correspondente ao deslocamento do paciente.

Também foi necessário incluir um atributo sequencial numérico que permite a identificação unívoca de cada registro, algo essencial para execução de algoritmos de mineração de dados.

Outra alteração necessária no *dataset* foi o desmembramento do atributo “tipo tratamento”. No arquivo original do RHC, o valor desse atributo é a concatenação dos códigos dos diversos tipos de tratamento. Assim, para se obter uma listagem completa de todos os tipos, deve-se usar uma função do banco de dados que separa esses códigos, cada um em uma linha, mantendo os demais atributos. Por exemplo, uma ocorrência que originalmente tinha esse formato:

UF	Município	Tipo Tratamento
SP	Suzano	Cir Qui Rad

Após a execução da função é desmembrada em três linhas:

UF	Município	Tipo Tratamento
SP	Suzano	Cir
SP	Suzano	Qui
SP	Suzano	Rad

Outros atributos incluídos no subconjunto de dados foram as esferas administrativas existentes entre a UF e o município. Isso possibilita a elaboração de análises a partir de diferentes níveis hierárquicos, desde o município, passando pelas regiões e macrorregiões de saúde e pelo estado, até a região geográfica.

Essa inclusão foi feita tanto para o município residencial quanto para o município da unidade hospitalar. Com esses dados, é possível identificar se determinado tratamento ocorreu dentro ou fora da esfera em análise.

### Elaboração do painel de análises descritivas

As análises descritivas foram separadas em quatro grupos: tipo de tratamento, perfil do paciente, perfil do tumor e deslocamento; que foram em seguida divididos em subgrupos. Esses agrupamentos foram feitos para segregar as análises com características semelhantes, permitindo um estudo exploratório das características por região e possibilitando melhor compreensão do deslocamento nas regiões envolvidas.

Para cada grupo e subgrupo foram criadas páginas *web* com botões e atalhos para auxiliar na navegação. As análises dos três primeiros

grupos foram desenvolvidas com recursos de tabelas dinâmicas e gráficos. Em cada subgrupo é possível filtrar os dados de um determinado estado e compará-los com os resultados do Brasil. Também é possível filtrar o período de análise, que pode incluir um ou mais anos.

As variáveis utilizadas nesses três primeiros grupos incluem: “tipo de tratamento”, “localização primária”, “lateralidade”, “estadiamento”, “faixa etária”, “grau instrução”, “raça”, “histórico familiar”, “tabagismo”, “alcoolismo”, “mais de um tumor”, “origem encaminhamento”.

Algumas dessas variáveis apresentam alto percentual de não preenchimento ou são preenchidas com o valor “sem informação”, no entanto elas foram mantidas nas análises. Um dos benefícios em mantê-las é a possibilidade de elaborar relatórios que mostrem o percentual de preenchimento dessas variáveis, o que pode subsidiar futuras ações de melhoria da qualidade dessa base de dados.

As análises do quarto grupo, destinado à descrição dos deslocamentos, foram elaboradas no software Microsoft Power BI e inclui painéis interativos com recursos de navegação *web* e opções de segmentação e detalhamento (*drill-down*) dos dados.

Além das análises descritivas, a ferramenta oferece suporte para o desenvolvimento de modelos de mineração de dados. Tais modelos são implementados internamente na ferramenta, não estando disponíveis de maneira direta ao usuário final, e podem ser usados em duas fases distintas: no pré-processamento dos dados e no apoio à interpretação dos resultados das análises descritivas.

Na fase de pré-processamento foi usada a técnica de agrupamento (*clustering*), a fim de discretizar os valores contínuos e substituí-los por faixas de valores, o que facilita a análise estatística.

Para a elaboração dos modelos de mineração de dados foi usado o software MS-SQL Analysis Services.

A técnica usada foi o agrupamento (*clustering*), aplicada na discretização dos atributos numéricos do *dataset* com intervalo de valores muito amplo e variado. A discretização por agrupamento foi aplicada aos três atributos numéricos do *dataset*: “distância de deslocamento”, “dias até o tratamento” e “idade”.

### Validação da ferramenta

Zhuang et al.<sup>19</sup> propõem uma estrutura abrangente para avaliar painéis de dados em

saúde, no entanto, considerando o objetivo da ferramenta proposta, foram validados apenas dois aspectos: consistência dos dados e usabilidade.

A validação da consistência dos valores produzidos pela ferramenta foi feita por meio da comparação direta com os valores compilados a partir da ferramenta do DATASUS denominada TABNET. Em muitos casos, devido à limitação do TABNET, foi necessária a compilação de dois ou mais relatórios em uma planilha Excel, e posteriormente a comparação desses valores com os valores apresentados pela ferramenta.

A avaliação da usabilidade foi feita pelo uso direto da ferramenta por dois usuários com experiência em saúde pública. Os usuários atribuíram notas, em uma escala de zero a dez, para os seguintes critérios: desempenho da ferramenta, *layout* da tela, formato de apresentação da informação, funcionalidades de filtros e detalhamentos (*drill-down*).

## Resultados

A base de dados foi carregada inicialmente com os registros do RHC do período de 2000 a 2017. No entanto, foi verificado que o ano de 2017 continha uma quantidade de registros muito inferior aos anos anteriores, apenas 40% dos registros de 2016. Dessa forma, o arquivo referente àquele ano foi desconsiderado. Foram então mantidos os dados de 2000 a 2016, totalizando 3.176.138 casos.

Aplicando os critérios de seleção definidos para a criação do *dataset*, foram selecionados apenas os casos de câncer de mama, com 492.148 casos.

Considerando a falta de preenchimento do atributo “município de residência”, foram descartados apenas 125 registros (0,00002%), totalizando 492.023 casos.

Usando a interface disponibilizada pelo Google Maps (API) foi possível calcular a distância de deslocamento para 99,1% dos registros do *dataset*. O 0,9% restante não pode ser obtido pela API por corresponder a locais em que não havia transporte rodoviário disponível. A maioria dos casos era de cidades da região Norte, como Afuá, no Pará, e Parintins, no Amazonas – cidades acessíveis somente por via fluvial ou aérea – ou, ainda, a ilha de Fernando de Noronha, em Pernambuco. Para esses casos, foi obtida a distância geodésica.

Uma vez concluídos a carga e os pré-processamentos do *dataset*, o painel de análises des-

critivas foi criado e a ferramenta foi implementada, encontrando-se disponível no *site* [www.minersus.com.br](http://www.minersus.com.br).

Foram incluídos na tela inicial de navegação quatro botões que direcionam o usuário a cada um dos grupos de análise proporcionados pela ferramenta: (1) perfil do tratamento, (2) perfil do paciente, (3) perfil do tumor e (4) perfil do deslocamento.

O primeiro grupo de análise (perfil do tratamento) mostra, de forma gráfica, os percentuais anuais do Brasil e das unidades da federação, referentes à quantidade de casos em que houve um determinado tipo de tratamento em relação à quantidade total de ocorrências naquele mesmo período. Considerando o exemplo da Figura 1(A), observamos que o valor percentual referente ao tratamento de quimioterapia no país em 2010 foi de 50%. Tal percentual corresponde à ocorrência de 18.588 tratamentos para um total de 37.441 casos.

Conforme podemos observar na Figura 1, a ferramenta permite que se comparem os tipos de tratamento usados em um ou mais estados ou regiões.

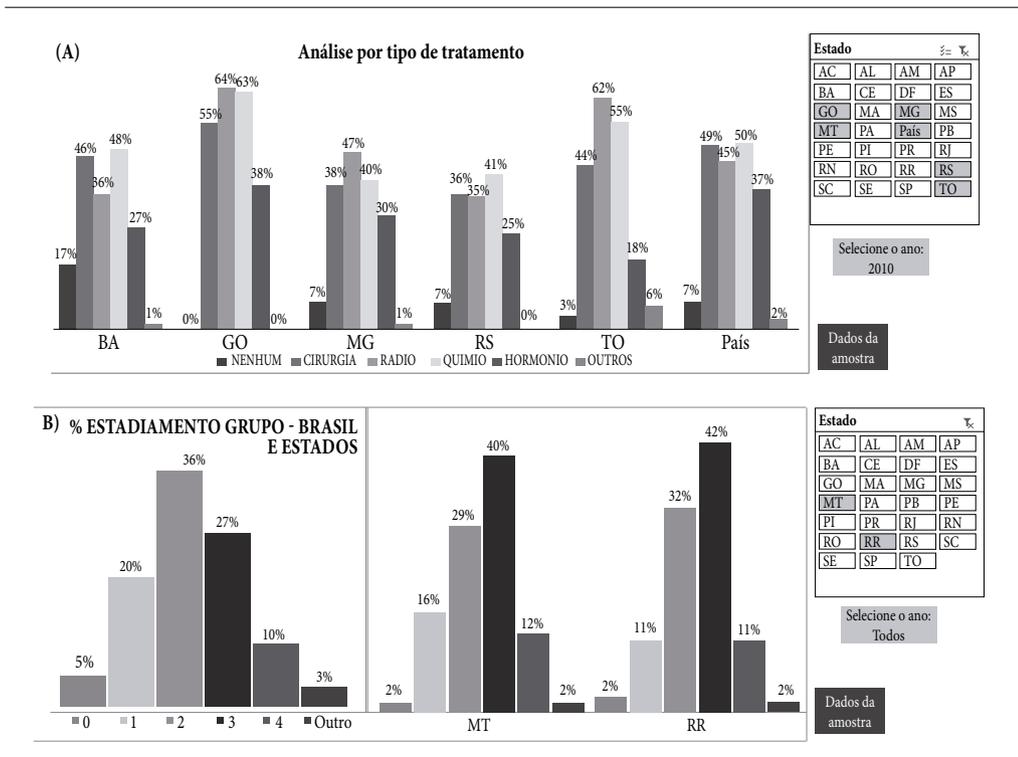
Ainda nesse grupo de análise, várias outras opções de gráficos podem ser criadas, inclusive aqueles que mostram a evolução temporal dos tratamentos adotados pelos estados/regiões.

O segundo grupo de análise (perfil do paciente) permite a elaboração de gráficos que descrevem o perfil dos pacientes. Foram implementados gráficos para oito variáveis: “faixa etária”; “grau de instrução”; “tabagismo”; “alcoolismo”; “histórico familiar”; “raça”; “mais de um tumor” e “origem de encaminhamento”.

Com exceção da “faixa etária”, que apresenta 100% de preenchimento, as demais variáveis relacionadas ao perfil do paciente têm elevado percentual de ausência de informação. “Grau de instrução”, “raça”, “origem de encaminhamento” e “mais de um tumor” apresentam um percentual superior a 30% sem preenchimento. “Histórico familiar”, “tabagismo” e “alcoolismo” estão sem informação para mais de 50% dos casos.

O terceiro grupo de análise (perfil do tumor) permite a elaboração de gráficos que descrevem, quantitativamente, o perfil do tumor dos pacientes em estudo. As variáveis relacionadas ao perfil do tumor incluídas nesse grupo de análise são: “estadiamento”, “localização primária” e “lateralidade”.

Em relação ao percentual de preenchimento dessas variáveis, “lateralidade” apresenta 11% de valores não preenchidos, “estadiamento” contém 27% e “localização primária” 55% dos casos.



**Figura 1.** Análises por gráficos (perfil de tratamento e perfil do tumor).

Fonte: Autores.

Padronizou-se o formato dos gráficos do segundo e terceiro grupos de análise, mostrando os dados gerais do Brasil no lado esquerdo da tela, os dados dos estados selecionados na parte central e os filtros na parte direita (Figura 1).

Em todas as páginas de análises descritivas da ferramenta está disponível o botão “dados da amostra”, que leva a outra página com os dados exibidos em formato tabular (Figura 2).

Na tabela “dados completos” é possível identificar todos os valores possíveis para a variável em análise e suas respectivas ocorrências na amostra selecionada. Os valores selecionados na tela anterior para estados e ano podem ser alterados manualmente também nessa página. A tabela permite filtros por outros valores do atributo e seu conteúdo pode ser copiado para uma planilha.

O quarto grupo de análise (deslocamento) contém um menu de navegação dividido em duas grandes visões: “distâncias” e “esferas de tratamento”. A visão “distâncias” exibe na tela principal uma tabela com duas colunas. A primeira coluna mostra a esfera selecionada (inicialmente agrupada por estados), e a segunda,

o valor da distância média de deslocamento, em quilômetros, para cada esfera em questão (Figura 3). Por exemplo, a distância média de deslocamento dos pacientes em tratamento de câncer de mama em Minas Gerais é de 86,96 km, enquanto o deslocamento médio no estado do Maranhão é de 228,23 (considerando o período de 2000 a 2016).

Os valores são exibidos com uma coloração graduada do branco ao vermelho, indicando, respectivamente, das menores para as maiores distâncias (quanto mais forte o tom de vermelho, maior a distância média de deslocamento).

Ao clicar no ícone “mais” (+) ao lado de cada estado, pode-se abrir a hierarquia seguinte (macrorregião de saúde); da mesma forma, uma ação semelhante abre o próximo nível (região de saúde) e, em seguida, a hierarquia mais baixa (município). Ao abrir o nível municipal, são exibidas as unidades hospitalares nas quais os pacientes daquela cidade foram se tratar – para facilitar a identificação, aparecem também a UF e a cidade de destino.

A Figura 3 mostra um exemplo desses recursos. O estado de Minas Gerais foi destaca-

Dados Completos	Localização Primária	Ocorrências	% Ocorrência
	Lesão sobreposta	30.659	6%
	Mamilo	9.027	2%
	Porção central	17.367	4%
	Prolongamento axilar	6.092	1%
	Quadrante inferior interno	14.485	3%
	Quadrante inferior externo	15.639	3%
	Quadrante superior externo	105.324	21%
	Quadrante superior interno	23.886	5%
	(vazio)	269.548	55%
<b>Total Geral</b>	<b>492.027</b>	<b>100%</b>	

Estado	Lesao sobreposta	Mamilo	Porcao Central	Prolongamento axilar	Quadrante inferior interno	Quadrante inferior externo	Quadrante superior externo	Quadrante superior interno	Total
ES	12%	4%	4%	0%	2%	2%	75%	1%	100%
MG	17%	2%	4%	1%	7%	7%	53%	8%	100%
RJ	4%	1%	18%	0%	7%	10%	48%	13%	100%
SP	29%	3%	7%	2%	6%	7%	38%	9%	100%

Figura 2. Dados da amostra para localização primária da região Sudeste.

Fonte: Autores.



Ano da Consulta



Registros

68K

Faixas de deslocamento



### Distância média da residência ao hospital

Estado	Dist. Média (km)
<input checked="" type="checkbox"/> GO	287.61
<input checked="" type="checkbox"/> MA	228.23
<input checked="" type="checkbox"/> MG	86.96
<input checked="" type="checkbox"/> CENTRO	
<input checked="" type="checkbox"/> CENTRO SUL	149.01
<input checked="" type="checkbox"/> BARBACENA	
<input checked="" type="checkbox"/> ALFREDO VASCONCELOS	46.02
<input checked="" type="checkbox"/> ALTO RIO DOCE	
<input checked="" type="checkbox"/> ANTONIO CARLOS	59.95
<input checked="" type="checkbox"/> BARBACENA	37.72
BA-VITÓRIA DA CONQUISTA - HOSPITAL GERAL DE VITÓRIA DA CONQUISTA	1,057.73
MG-BARBACENA - HOSPITAL IBIAPABA CEBAMS	0.00
MG-BELO HORIZONTE -	171.72
MG-BELO HORIZONTE -ASSOCIAÇÃO MARIO PENNA	171.72
MG-BELO HORIZONTE - HOSPITAL ALBERTO CAVALCANTI	171.72
MG-BELO HORIZONTE - HOSPITAL DA BALEIA	171.72
MG-BELO HORIZONTE - HOSPITAL FELICIO ROCHO	171.72
MG-BELO HORIZONTE - SANTA CASA DE BELO HORIZONTE	171.72
MG-JUIZ DE FORA - HOSPITAL DR JOÃO FELICIO	102.77
MG-JUIZ DE FORA - HOSPITAL MARIA JOSÉ BAETA REIS ASCOMCER	102.77
MG-JUIZ DE FORA - ONCOLÓGICO	102.77
MG-SÃO JOÃO DEL REI - SANTA CASA DE MISERICÓRDIA DE SÃO JOÃO DEL REI	60.55
RJ-RIO DE JANEIRO - MS INCA HOSPITAL DO CÂNCER III	278.27
SP-BARRETOS - FUNDAÇÃO PIO XII BARRETOS	615.60
SP-SÃO PAULO - A C CAMARGO CANCER CENTER	527.31
<input checked="" type="checkbox"/> CAPELA NOVA	126.40
<input checked="" type="checkbox"/> CARANDAÍ	95.81

Figura 3. Visão Distância – distância média da residência ao hospital.

Fonte: Autores.

do, selecionando em seguida a macrorregião de saúde Centro-Sul, depois a região de saúde Barbacena, e por fim o município de Barbacena. Mesmo com alguns deslocamentos distantes para Bahia, São Paulo e Rio de Janeiro, a distância média é de 37,72 km. Subindo um nível, é possível ver que a média daquela região de saú-

de é de 53,97 km, e esse número aumenta para 149,01 km ao se considerar a macrorregião de saúde Centro-Sul.

A visão “esferas de tratamento” foi construída com a ideia de que se possa navegar por todas as divisões administrativas, numa sequência hierárquica, do nível maior para o menor. Em

cada um dos níveis são exibidos os percentuais de tratamentos feitos dentro (D) ou fora (F) daquela esfera administrativa.

O primeiro nível começa com os estados e os percentuais D/F de cada um, em formato tabular. Pode-se clicar no estado desejado e navegar (usando a opção “detalhar”) para a esfera seguinte. Aparecem então todas as macrorregiões de saúde para o estado pré-selecionado, com seus percentuais de tratamento D/F. Usando o mesmo conceito, pode-se navegar para o detalhamento seguinte (região de saúde) e, mais uma vez, até a lista de municípios que compõem aquela divisão hierárquica. Ao chegar a esse último nível (esfera municipal), pode-se clicar no ícone “mais” (+) ao lado do município para exibir as unidades hospitalares em que os pacientes foram se tratar.

A Figura 4 mostra um exemplo desse tipo de análise. É possível verificar que o contador à esquerda apresenta o número arredondado de 10 mil registros, que é a soma dos totais de ocorrências exibidos na tabela (dentro: 4.067, fora: 6.284).

Pela imagem, pode-se ver que apenas cinco das doze regiões de saúde exibidas devem ter unidades hospitalares de tratamento oncológico (“Alfenas Machado”, “Passos Piumhi”, “Poços de Caldas”, “Pouso Alegre” e “Varginha”), pois nas outras o percentual de tratamento externo é de 100% dos casos. É visível também que as regiões com maior predominância de atendimento interno são as de Varginha e Poços de Caldas, com respectivamente 95% e 91% das ocorrências nessa situação.

Em relação aos modelos de mineração de dados usados para discretização de atributos, foram elaborados sete modelos diferentes para o atributo distância, executados em sequência e variando os parâmetros do algoritmo, a fim de identificar os melhores intervalos possíveis. As faixas (arredondadas) obtidas pelo modelo vencedor estão indicadas na Figura 5(A).

Para a discretização do atributo “dias até o início do tratamento” foram elaborados cinco modelos, de modo a identificar os melhores intervalos possíveis. As faixas (arredondadas) obtidas pelo modelo vencedor estão indicadas na Figura 5(B).

Para o terceiro atributo (“idade”) foram elaborados dez modelos diferentes, de forma análoga aos anteriores. Entretanto, os grupos identificados não se mostraram relevantes para as análises, por isso foram descartados. Para as faixas etárias, optou-se por utilizar a classificação indicada por especialista da saúde pública.

Em relação à validação da ferramenta, no aspecto de consistência dos dados não foi detectada nenhuma divergência entre os valores produzidos pela ferramenta e aqueles compilados a partir do TABNET. Sobre a avaliação da usabilidade, todos os critérios avaliados receberam notas acima de 8. O critério que recebeu menor nota (8.1) foi o desempenho, mas isso se justifica pelo fato de a ferramenta ter sido avaliada em um computador com processadores e memória aquém do recomendado.

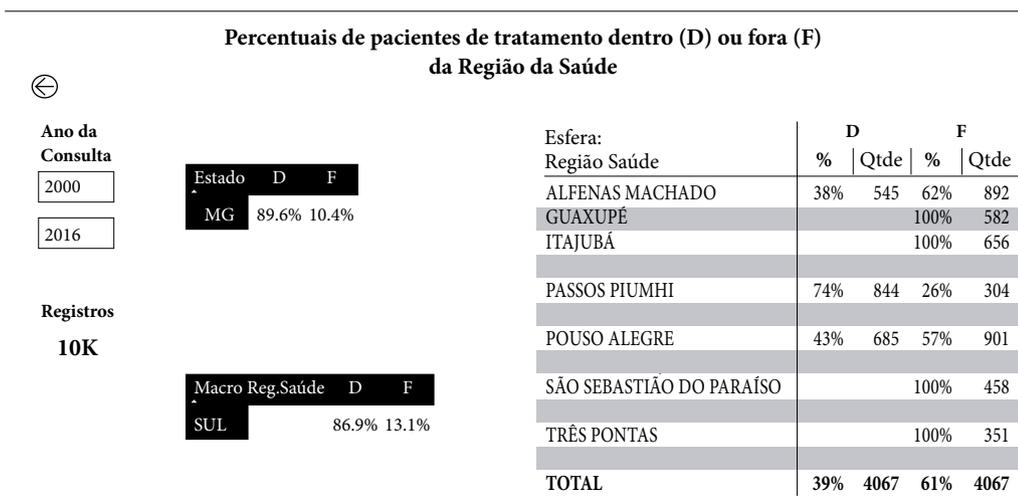
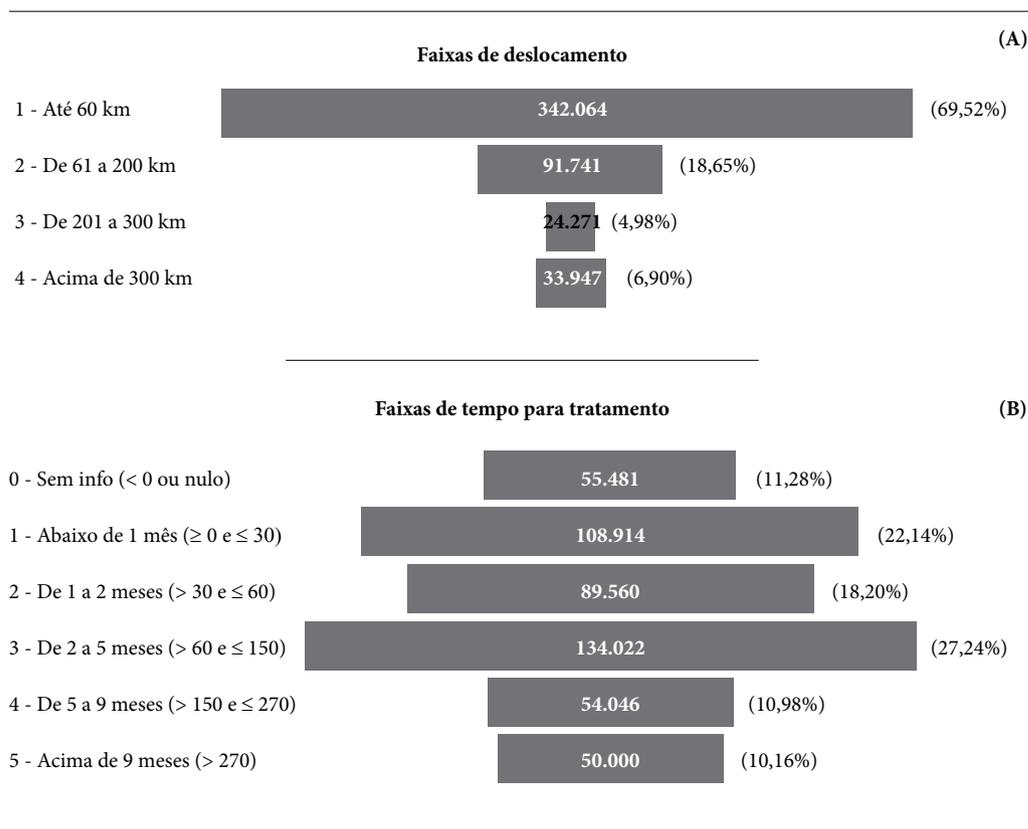


Figura 4. Visão Esferas – macrorregião de saúde Sul de Minas Gerais.



**Figura 5.** Agrupamentos (deslocamentos e tempo de início de tratamento).

Fonte: Autores.

## Discussão

Primeiramente, deve-se ressaltar que as análises sobre câncer de mama apresentadas neste trabalho foram elaboradas com o objetivo de validar as funcionalidades e características da ferramenta implementada. Portanto, tais análises não têm a pretensão de descrever um panorama completo do tratamento de câncer de mama do país, mas sim mostrar o potencial da ferramenta.

As análises criadas no primeiro grupo (perfil do tratamento), embora composta por gráficos simples, permitem visualizar informações importantes para a elaboração de políticas/ações de saúde pública. No exemplo da Figura 1 é

possível identificar que a Bahia é o estado, entre aqueles analisados, que apresenta o maior percentual de pacientes que não fizeram nenhum tratamento (17%), bem superior à média do país (7%). O estado que mais fez “outros” tratamentos foi Tocantins (6%), bem acima dos demais, que têm 1% ou 0%. O percentual de radioterapia é bem alto em Goiás (64%) e no Tocantins (62%). Os valores encontrados para cirurgia também têm algumas variações, com o menor índice no Rio Grande do Sul (36%) e o maior em Goiás (55%), enquanto a média nacional foi de 49% em 2010.

A Figura 1(B) mostra que dois estados da federação destoaram do perfil brasileiro em relação ao estadiamento do tumor. Em todas as

outras unidades da federação o estágio 2 é o que tem os maiores percentuais, enquanto em Mato Grosso e em Roraima o estágio 3 é que fica em primeiro (com, respectivamente, 40% e 42%). Vale ressaltar que a tela apresentada é um recorte de uma seleção que incluía todos os estados.

Apesar do percentual relativamente alto de não preenchimento das variáveis relacionadas ao perfil do paciente, algumas informações interessantes foram obtidas a partir das análises criadas nesse grupo. Por exemplo, um gráfico criado para analisar a classificação etária evidenciou que em Tocantins, no ano de 2010, o percentual de mulheres em tratamento na faixa de 40 a 49 anos (40%) estava bem acima da média brasileira (25%), ocupando a maior parcela entre todas as faixas desse estado. O provável é que tenha sido um fenômeno pontual daquele ano, pois, quando foi selecionado o período acumulado, o número mudou para 27%, ficando abaixo da faixa de maiores de 60 anos (com 31%).

Outro gráfico elaborado no grupo “perfil do paciente” foi sobre o tabagismo, selecionando os estados da região Sudeste no ano de 2014. Esse gráfico mostrou que os pacientes residentes em São Paulo apresentaram um perfil diferente do resto da região Sudeste e do Brasil, com 41% de ex-fumantes. O percentual é bem mais alto que dos outros três estados. O Rio de Janeiro, por outro lado, mostrou o maior percentual de fumantes daquela amostra, com 19% do total de pacientes.

As análises criadas no grupo “deslocamento” evidenciaram alguns aspectos interessantes relacionados ao deslocamento. Um dos resultados encontrados ao se fazer a análise da base completa, sem filtro algum, é que a distância média de deslocamento de pacientes em tratamento de câncer de mama no Brasil foi de 88,4 km. Filtrando os dados somente do primeiro ano do período e em seguida do último foi verificado que o deslocamento médio nacional aumentou, passando de 79,0 km em 2006 para 92,8 km em 2016.

Ainda considerando todo o período, foi observado que 69,5% dos deslocamentos são de até 60 km. Em 2006 esse percentual era de 72,6%, tendo diminuído para 65,8% em 2016. Observou-se também que essa diminuição se refletiu no acréscimo das demais faixas, com destaque para a segunda (entre 61 e 200 km), que cresceu de 16,9% no primeiro ano para 21,4% no último, o que pode justificar o aumento do deslocamento médio relatado no parágrafo anterior.

Os modelos de mineração de dados criados para a discretização de variáveis também evi-

denciaram algumas informações relevantes sobre o deslocamento de pacientes, ainda que esse não fosse o seu propósito. Por exemplo, a discretização do atributo “distância” mostrou que, a despeito das dimensões continentais do Brasil, o tratamento de câncer de mama é concentrado em distâncias curtas (até 60 km) e médias (entre 61 e 200 km), chegando a cerca de 88% das ocorrências. Resultado bem próximo ao do estudo de Vetterlein *et al.*<sup>11</sup>, que englobou todo o território norte-americano e indicou alta concentração nas duas faixas iniciais. Nesse estudo, os autores definiram distâncias curtas como aquelas menores de 20 km, e distância média entre 20 e 80 km, encontrando 87,9%, dos casos nessas faixas.

Outro dado evidenciado foi que 47,1% das ocorrências correspondem a pacientes que se tratam em seu próprio município. Esse resultado é semelhante ao encontrado por Saldanha *et al.*<sup>4</sup>, que analisaram as internações e tratamentos de quimioterapia e radioterapia entre 2014 e 2016, e ao resultado encontrado no estudo de Oliveira *et al.*<sup>12</sup> para radioterapia entre os anos de 2005 e 2006. É, porém, ligeiramente menor do que na pesquisa de Sobreira da Silva *et al.*<sup>14</sup>, que está restrito somente à quimioterapia e ao ano de 2013.

Os modelos para discretização do atributo “dias até o início do tratamento” evidenciaram que as duas faixas iniciais, que indicam tratamentos em até 60 dias, englobam 40,3% dos pacientes. Aqueles que levaram mais de 60 dias para iniciar o tratamento correspondem a cerca de 48% dos casos, estando acima do resultado de 36% encontrado por Renna Junior e Azevedo e Silva<sup>16</sup> – o que pode ser justificado pelos critérios de exclusão utilizados por esses autores diferirem dos nossos.

A elaboração das análises descritivas, bem como dos modelos de discretização, evidenciaram duas potenciais limitações da ferramenta, que são consequências naturais das características da base de dados do RHC.

A primeira delas é a impossibilidade de calcular a distância de deslocamento intramunicipal, o que impede a análise da dinâmica de deslocamentos no município. A segunda é o alto percentual de não preenchimento em algumas variáveis, que pode limitar futuros estudos de correlações entre o deslocamento do paciente e tais variáveis.

Apesar das limitações potenciais, a ferramenta mostrou-se eficiente na produção de informações descritivas sobre os deslocamentos dos pacientes, demonstrando que é possível usar o RHC como base de dados para analisar esses deslocamentos.

## Conclusão

Uma ferramenta computacional que permita, de forma simples, a elaboração de gráficos e relatórios analíticos capazes de descrever os deslocamentos de pacientes em tratamento de câncer no SUS, a partir dos dados do RHC, foi implementada e validada.

Sua implementação se deu por meio de um estudo de caso de câncer de mama em que foram desenvolvidos modelos analíticos que descrevem os deslocamentos de pacientes, considerando o período de 2006 a 2016.

As análises foram agrupadas conforme características relacionadas aos tipos de tratamento, ao tumor, aos pacientes e seus deslocamentos. Esses quatro grupos de análise foram então organizados num painel de navegação interativo, disponibilizado publicamente no *site* [www.minersus.com.br](http://www.minersus.com.br), que possibilita uma ampla gama de estudos para especialistas e gestores em saúde pública.

Assim, uma grande contribuição deste trabalho foi o desenvolvimento de um painel de

análises estatísticas descritivas que permite avaliar características do deslocamento de pacientes, como distância média, percentuais por faixa de deslocamento e percentuais de tratamentos dentro/fora de cada uma das unidades da federação, suas macrorregiões, regiões de saúde, até o município.

As principais limitações da ferramenta são a impossibilidade de calcular a distância de deslocamento intramunicipal e o alto percentual de não preenchimento de algumas variáveis do RHC. No entanto, apesar dessas limitações, a ferramenta mostrou-se eficiente na produção de informações sobre os deslocamentos dos pacientes, demonstrando que é possível usar o RHC como base de dados para analisar esses deslocamentos.

Sugere-se como trabalho futuro a análise descritiva do deslocamento de pacientes para outros tipos de câncer. Também é proposto que seja feita a análise comparativa usando os deslocamentos calculados a partir dos sistemas de informações ambulatoriais do DATASUS.

## Colaboradores

LFRT Pais: execução do trabalho (experimentos, coleta e compilação de dados, análise dos resultados) e redação do artigo. JLB Bevilacqua: revisão crítica dos conceitos relacionados à oncologia e saúde pública, análise dos resultados e revisão da versão final para a publicação. JMP Oliveira: revisão crítica dos conceitos relacionados às técnicas computacionais, análise dos resultados e revisão da versão final para a publicação. RS Santos: concepção da ideia do projeto/trabalho, gerenciamento do projeto e orientação do trabalho, revisão crítica do trabalho, análise dos resultados e revisão da versão final para a publicação.

## Referências

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68(6):394-424.
2. Santos MO, Lima FCS, Martins LFL, Oliveira JFP, Almeida LM, Cancela MC. Estimativa de Incidência de Câncer no Brasil, 2023-2025. *Rev Bras Cancerol* 2023; 69(1):e-213700.
3. Brasil. Ministério da Saúde (MS). Portaria no 874, de 16 de maio de 2013. Institui a Política Nacional para a Prevenção e Controle do Câncer na Rede de Atenção à Saúde das Pessoas com Doenças Crônicas no âmbito do Sistema Único de Saúde (SUS). *Diário Oficial da União* 2013; 17 maio.
4. Saldanha RF, Xavier DR, Carnavalli KM, Lerner K. Estudo de análise de rede do fluxo de pacientes de câncer de mama no Brasil entre 2014 e 2016. *Cad Saude Publica* 2019; 35(7): e00090918.
5. Alves MO, Magalhães SCM. A geografia do câncer de mama no norte de Minas Gerais: do direito ao acesso à saúde. *Hygeia* 2017; 13(26):13-32.
6. Lim J, Ali AS, Prawira A, Sim HW. Impact of travel distance on outcomes for clinical trial patients: the Kinghorn Cancer Centre experience. *Intern Med J* 2023; 53(2):242-249.

7. Ambroggi M, Biasini C, Del Giovane C, Fornari F, Cavanna L. Distance as a Barrier to Cancer Diagnosis and Treatment: Review of Literature. *Oncologist* 2015; 20(12):1378-1385.
8. Fabian ID, Stacey AW, Foster A, Kivelä TT, Munier FL, Keren-Froim N, Gomel N, Cassoux N, Sagoo MS, Reddy MA, Harby LA, Zondervan M, Bascaran C, Abdallah E, Abdullahi SU, Boubacar SA, Ademola-Popoola DS, Adio A, Aghaji AE, Portabella SA, Alfa Bio AI, Ali AM, Alia DB, All-Eriksson C, Almeida AA, Alsawidi KM, Antonino R, Astbury NJ, Atsiaya R, Balaguer J, Balwierz W, Barranco H, Popovic MB, Benmiloud S, Guebessi NB, Berete RC, Biddulph SJ, Biewald EM, Blum S, Bobrova N, Boehme M, Bornfeld N, Bouda GC, Bouguila H, Boumedane A, Brichard BG, L MC, Castela G, Català-Mora J, Chantada GL, Chernodrinska VS, Chiwanga FS, Cieslik K, Comsa C, Correa Llano MG, Csóka M, Da Gama IV, Davidson A, Potter P, Desjardins L, Dragomir MD, Bruyn MD, Kettani AE, Elbahi AM, Elgalaly D, Elhaddad AM, Ali Elhassan MM, Elzembely MM, Essuman VA, Evina TGA, Fasina O, Fernández-Teijeiro A, Gandiwa M, Aldana DG, Geel JA, Gizachew Z, Gregersen PA, Guedenon KM, Hadjistilianou T, Hassan S, Hederova S, Hessissen L, Hordofa DF, Hummlen M, Husakova K, Ida R, Ilic VR, Jenkinson H, Amani Kabesha TB, Kabore RL, Kalinaki A, Kapelushnik N, Kardava T, Kemilev PK, Kepak T, Khotenashvili Z, Klett A, Kosh Komba Palet JE, Krivaitiene D, Kruger M, Kyara A, Lachmann ES, Latinović S, Lecuona K, Lukamba RM, Lumbroso L, Lysytsia L, Maka E, Makan M, Manda C, Begue NM, Matende IO, Matua M, Mayet I, Mbumba FB, Mengesha AA, Midená E, Mndeme FG, Mohamedani AA, Moll AC, Moreira C, Msina MS, Msukwa G, Muma KI, Murgoi G, Musa KO, Mustak H, Muyen OM, Naidu G, Naumenko L, Ndoye Roth PA, Nerov V, Nikitovic M, Nkanga ED, Nkumbe H, Nyaywa M, Obono-Obiang G, Oguego NC, Olechowski A, Oscar AH, Osei-Bonsu P, Painter SL, Paintsil V, Paiva L, Pappan R, Parrozzani R, Parulekar M, Pawinska-Wasikowska K, Perić S, Philbert R, Pochop P, Polyakov VG, Pompe MT, Pons JJ, Raobela L, Renner LA, Reynders D, Ribadu D, Riheia MM, Ritter-Sovinz P, Saakyan S, Said AM, Román Pacheco SS, Scanlan TA, Schoeman J, Seregard S, Sherief ST, Cheikh SS, Silva S, Sorochynska T, Ssali G, Stathopoulos C, Kranjc BS, Stones DK, Svojgr K, Sylla F, Tamamyán G, Tandili A, Tateshi B, Theophile T, Traoré F, Tyau-Tyau H, Umar AB, Urbak SF, Ushakova TL, Valeina S, Hoefen Wijsard MV, Veleva-Krasteva NV, Viksnins M, Wackernagel W, Waddell K, Wade PD, Wali Nigeria AH, Wime AD, Dod CW, Yanga JM, Yarovaya VA, Yarovoy AA, Zein E, Sharabi S, Zhilyaeva K, Ziko OA, Bowman R. Travel burden and clinical presentation of retinoblastoma: analysis of 1024 patients from 43 African countries and 518 patients from 40 European countries. *Br J Ophthalmol* 2021; 105(10):1435-1443.
9. Obrochta CA, Parada H, Murphy JD, Nara A, Trinidad D, Araneta MR, Thompson C A. The impact of patient travel time on disparities in treatment for early stage lung cancer in California. *PLoS One* 2022; 17(10):e0272076.
10. Jones AP, Haynes R, Sauerzapf V, Crawford SM, Zhao H, Forman D. Travel times to health care and survival from cancers in Northern England. *Eur J Cancer* 2008; 44(2):269-274.
11. Vetterlein MW, Löttenberg B, Karabon P, Dalela D, Jindal T, Sood A, Chun FK, Trinh QD, Menon M, Abdollah F. Impact of travel distance to the treatment facility on overall mortality in US patients with prostate cancer. *Cancer* 2017; 123(17):3241-3252.
12. Oliveira EXG, Melo ECP, Pinheiro RS, Noronha CP, Carvalho, MS. Acesso à assistência oncológica: mapeamento dos fluxos origem-destino das internações e dos atendimentos ambulatoriais. O caso do câncer de mama. *Cad Saude Publica* 2011; 27(2):317-326.
13. Fonseca BP, Albuquerque PC, Saldanha RF, Zicker F. Geographic accessibility to cancer treatment in Brazil: A network analysis. *Lancet Reg Health Am* 2021; 7:100153.
14. Silva MJS, Melo E, Osorio-de-Castro CGS. Fluxos origem-destino para quimioterapia para o câncer de mama no Brasil: implicações para a assistência farmacêutica. *Cien Saude Colet* 2019; 24(3):1153-1164.
15. Instituto Nacional de Câncer (Inca). *Registros hospitalares de câncer: planejamento e gestão*. Rio de Janeiro: Inca; 2010.
16. Renna Junior NL, Silva GAE. Diagnóstico de câncer de mama em estado avançado no Brasil: análise de dados dos registros hospitalares de câncer (2000-2012). *Rev Bras Ginecol Obstet* 2018; 40(3):127-136.
17. Pinto IV, Ramos DN, Costa MCE, Ferreira CBT, Rebelo MS. Completude e consistência dos dados dos registros hospitalares de câncer no Brasil. *Cad Saude Colet* 2012; 20(1):113-120.
18. Instituto Nacional de Câncer (Inca). *IntegradorRHC: Ferramenta para a Vigilância Hospitalar de Câncer no Brasil*. Rio de Janeiro: Inca; 2011.
19. Zhuang M, Concannon D, Manley E. A Framework for Evaluating Dashboards in Healthcare. *IEEE Trans Vis Comput Graph* 2022; 28(4):1715-1731.

---

Artigo apresentado em 04/07/2023

Aprovado em 21/12/2023

Versão final apresentada em 23/12/2023

---

Editores-chefes: Maria Cecília de Souza Minayo, Romeu Gomes, Antônio Augusto Moura da Silva