

# Cálculos de Estimativas de Câncer Regionais pela Razão Incidência/Mortalidade: Tutorial com Apoio de *Scripts* do *Software R*

<https://doi.org/10.32635/2176-9745.RBC.2025v71n1.4836>

*Calculations of Regional Cancer Estimates Using the Incidence/Mortality Ratio: Tutorial with the Support of R Software Scripts*  
Cálculos de Estimaciones Regionales de Cáncer Utilizando la Razón Incidencia/Mortalidad: Tutorial con el Apoyo de *Scripts* del *Software R*

Gil Patrus Pena<sup>1</sup>; Sara Oliveira Ribeiro<sup>2</sup>

## RESUMO

**Introdução:** Um dos principais papéis da vigilância do câncer é fornecer ao gestor um cenário da situação do câncer, projetando expectativas sobre a incidência dos diferentes tipos de câncer nos territórios, possibilitando o planejamento racional das ações de prevenção, detecção precoce e tratamento direcionados aos principais tipos de câncer incidentes naquela Região. A elaboração de estimativas de câncer constitui uma das principais ferramentas nesse sentido, mas sua elaboração detalhada pelos diferentes tipos de câncer e regiões é complexa e trabalhosa. O *software R* é uma ferramenta poderosa de manipulação de dados e seu uso ainda é pouco difundido nas atividades de vigilância em saúde.

**Objetivo:** Descrever a metodologia do cálculo das estimativas de incidência do câncer e disponibilizar o conjunto de *scripts* do *software R* utilizado, com o objetivo de difundir e fomentar o seu uso nas atividades de vigilância do câncer. **Método:** É apresentada a elaboração das Estimativas de Câncer em Minas Gerais, para o ano de 2024, por Macrorregiões de Saúde. Todo o processo de manipulação e cálculos foi estruturado em *scripts* do R, que são disponibilizados como um suplemento. **Resultados:** A elaboração de estimativas é potencializada pelo uso do *software R*, pela facilidade da manipulação da grande quantidade de dados e dos cálculos necessários, além da possibilidade de apresentação dos resultados em gráficos e mapas. **Conclusão:** Com a disponibilização dos *scripts*, espera-se que todo o processo possa ser adaptado por outros Estados, para a elaboração de suas próprias estimativas regionais.

**Palavras-chave:** Vigilância em Saúde Pública; Neoplasias/epidemiologia; Incidência.

## ABSTRACT

**Introduction:** One of the main roles of cancer surveillance is to provide the manager a scenario of the cancer status, predicting expectations about the incidence of different types of cancer in the territories, enabling rational planning of prevention, early detection and treatment actions for the main types of cancer occurring in that region. The preparation of cancer estimates is one of the main tools in this regard, but their detailed elaboration for different types of cancer and regions is complex and laborious. Software R is a powerful data manipulation tool and its use is still not widespread in health surveillance activities. **Objective:** To describe the methodology for calculating cancer incidence estimates and make available the set of software R scripts used, with the objective of disseminating and promoting its use in cancer surveillance activities. **Method:** Description of the preparation of 2024 Cancer Estimates in Minas Gerais by Health Macro-regions. The entire manipulation and calculation process was structured in R scripts, which are available as a supplement. **Results:** The preparation of estimates is enhanced by the use of software R, facilitating the handling of large amount of data and the necessary calculations, in addition to the possibility of presenting results in graphs and maps. **Conclusion:** With the availability of the scripts, it is expected that the entire process is adjusted by other states to prepare their own regional estimates.

**Key words:** Public Health Surveillance; Neoplasms/epidemiology; Incidence.

## RESUMEN

**Introducción:** Una de las funciones principales de la vigilancia del cáncer es brindar al gestor un escenario de la situación del cáncer, proyectando expectativas sobre la incidencia de los distintos tipos de cáncer en los territorios, posibilitando la planificación racional de las acciones de prevención, detección temprana y tratamiento dirigidos a los principales tipos de cáncer que ocurren en esa región. La elaboración de estimaciones de cáncer es una de las principales herramientas en este sentido, pero su elaboración detallada para diferentes tipos de cáncer y regiones es compleja y laboriosa. El *software R* es una poderosa herramienta de manipulación de datos y su uso aún no está extendido en las actividades de vigilancia de la salud. **Objetivo:** Describir la metodología para el cálculo de estimaciones de incidencia de cáncer y poner a disposición el conjunto de *scripts* del *software R* utilizados, con el objetivo de difundir y promover su uso en actividades de vigilancia del cáncer. **Método:** Se presenta la elaboración de las Estimaciones de Cáncer en Minas Gerais, para el año 2024, por macrorregiones de salud. Todo el proceso de manipulación y cálculo se estructuró en *scripts* R, que están disponibles como suplemento. **Resultados:** La elaboración de estimaciones se ve favorecida por el uso del *software R*, dada la facilidad para manejar la gran cantidad de datos y los cálculos necesarios, además de la posibilidad de presentar los resultados en gráficos y mapas. **Conclusión:** Con la disponibilidad de los *scripts*, se espera que otros Estados puedan adaptar todo el proceso para preparar sus propias estimaciones regionales.

**Palabras clave:** Vigilancia en Salud Pública; Neoplasias/epidemiología; Incidencia.

<sup>1</sup>Registro de Câncer de Base Populacional de Belo Horizonte, Coordenação de Vigilância das Doenças e Agravos Crônicos não Transmissíveis e Câncer. Superintendência de Vigilância Epidemiológica, Subsecretaria de Vigilância em Saúde, Secretaria de Estado da Saúde de Minas Gerais. Belo Horizonte (MG), Brasil. E-mail: gil.pena@saude.mg.gov.br. Orcid iD: <https://orcid.org/0000-0002-9395-6372>

<sup>2</sup>Superintendência de Vigilância Epidemiológica, Subsecretaria de Vigilância em Saúde, Secretaria de Estado da Saúde de Minas Gerais. Belo Horizonte (MG), Brasil. E-mail: sara.oribeiro@yahoo.com.br. Orcid iD: <https://orcid.org/0009-0008-1433-3056>

**Endereço para correspondência:** Gil Patrus Pena. Secretaria de Estado da Saúde. Rodovia João Paulo II, 4143, Edifício Minas, 13º andar – Serra Verde. Belo Horizonte (MG), Brasil. CEP 31630-900. E-mail: gil.pena@saude.mg.gov.br



## INTRODUÇÃO

O câncer é uma doença complexa que engloba numerosos tipos histológicos e diferentes topografias. As ações de vigilância da doença envolvem o monitoramento da incidência e da mortalidade. Nas Regiões não cobertas por Registros de Câncer de Base Populacional (RCBP), a incidência pode ser estimada por modelagem a partir das informações de mortalidade, usando razões de incidência/mortalidade derivadas de dados de RCBP específicos da Região.

A elaboração de estimativas de câncer, como as publicadas periodicamente pelo Instituto Nacional de Câncer (INCA), para as Unidades da Federação (UF) e capitais do Brasil, para tipos selecionados de câncer e por sexo, é trabalhosa. Para cada tipo de câncer, sexo e Região, são necessários cálculos relativos à população, ao número de óbitos, obtendo-se taxas de mortalidade bruta e ajustada. No cálculo da mortalidade ajustada, é necessário calcular as taxas de mortalidade específicas por idade e aplicar essas taxas a uma população padrão. É necessário ainda obter as taxas de incidência por tipo de câncer e sexo nos RCBP selecionados<sup>1</sup>.

Embora os cálculos envolvidos na metodologia sejam baseados na epidemiologia clássica, com metodologia relativamente acessível, a grande quantidade de dados e cálculos envolvidos, multiplicados para cada tipo de câncer, Região e sexo, certamente inibe iniciativas de elaboração de estimativas com melhor detalhamento regional.

Em Minas Gerais, o Programa de Avaliação de Vigilância do Câncer e seus Fatores de Risco (PAV-MG) publicou em 2013 a primeira estimativa para as Macrorregiões de Saúde do Estado<sup>2</sup>. Uma atualização das estimativas foi inteiramente calculada, mas, em virtude de uma alteração da regionalização administrativa da saúde<sup>3</sup>, não chegou a ser publicada. Recentemente, uma nova alteração do Plano Diretor de Regionalização em Saúde foi instituída<sup>4</sup>, exemplificando a necessidade de se ter uma análise estruturada das populações e da mortalidade que possa rapidamente ser adaptada a mudanças na organização administrativa do território.

O conhecimento do cenário do câncer em Minas Gerais, detalhado pelas Macrorregiões em Saúde, é necessário, dada a extensão e a diversidade regional, com marcadas variações ambientais, culturais e econômicas. Essas informações permitem ao gestor definir prioridades e planejar ações de enfrentamento da doença conforme as particularidades regionais.

Além das estimativas anteriores realizadas no âmbito da SES-MG<sup>2</sup>, estudos com estimativas de câncer para Sub-Regiões estaduais, calculadas pelo método da razão

da incidência/mortalidade, foram realizados para a 17 Redes Regionais de Atenção à Saúde (RRAS) do Estado de São Paulo<sup>5</sup>.

As estimativas de incidência podem ser produzidas com base em diferentes métodos, de acordo com a disponibilidade de informações em cada país. Nas estimativas mundiais por país, o *Global Cancer Observatory: Cancer Today*<sup>6</sup> (Globocan) aplica diferentes métodos. O método utilizado prioritariamente envolve a projeção de taxas de incidências de RCBP com grande cobertura nacional ou subnacional. Na indisponibilidade de taxas de incidência com cobertura abrangente, utilizam-se estimativas nacionais de mortalidade por modelagem, usando razões de incidência/mortalidade derivadas de dados de registros de câncer específicos da Região ou Regiões vizinhas.

Além do método direto para o cálculo da razão incidência/mortalidade, métodos estatísticos envolvendo modelos de multiníveis de Poisson para a estimativa da razão da incidência/mortalidade foram empregados por Jardim et al.<sup>7</sup>, permitindo considerar a variação dos efeitos aleatórios dos RCBP que não seriam captados pelos modelos tradicionais de efeitos fixos.

O *software* R<sup>8</sup> é uma ferramenta para a manipulação de dados e análises estatísticas. Um dos melhores recursos que o R disponibiliza é a possibilidade de manter em um *script* todas as manipulações de dados e análises feitas. Mesmo nas análises mais complexas, foi possível recomençar e refazer após correções, com muita rapidez. Um *script* organizado permite ter a rotina do trabalho documentada, a automação dos processos e a reprodutibilidade do trabalho<sup>9</sup>. Além disso, o R é um *software* livre, que pode ser instalado sem o pagamento de licenças. O uso do R é ainda potencializado pelos numerosos pacotes que agregam funções específicas ao programa, ampliando os recursos de manipulação dos dados, análise e visualização (gráficos e publicações).

Neste artigo, foi descrita, em detalhe, a metodologia da elaboração das estimativas de câncer no Estado de Minas Gerais, para as 16 Macrorregiões de Saúde, 21 tipos específicos de câncer e para todas as neoplasias, para os sexos masculino e feminino. Todo o percurso de preparação dos dados e cálculos foi estruturado em *scripts* do R que são disponibilizados como um suplemento, podendo ser adaptados para elaboração de estimativas em outros territórios (Brasil, UF e divisões regionais por UF).

## MÉTODO

Toda a manipulação dos dados foi realizada no *software* R, versão 4.3.2.<sup>8</sup> Foi utilizado um computador com processador Intel® Core™ i5-4210U CPU 1.70GHz

2.40 GHz, com 8 GB de RAM e sistema operacional Windows 10 Pro de 64 bits. Os seguintes pacotes com as suas respectivas dependências foram instalados: *dplyr*<sup>10</sup>, *tidyr*<sup>11</sup>, *readxl*<sup>12</sup>, *ggplot2*<sup>13</sup>, *read.dbc*<sup>14</sup>, *car*<sup>15</sup>, *gmodels*<sup>16</sup>, *lmtest*<sup>17</sup>, *rmarkdown*<sup>18</sup>, *knitr*<sup>19</sup> e *sf*<sup>0</sup>. Com exceção dos *scripts* em *rmarkdown* (.Rmd), as análises foram realizadas na interface *Rgui.exe*, um programa executável disponibilizado na instalação do R base que abre um console em uma janela do *Windows*. Os *scripts* do *rmarkdown* (.rmd) foram rodados na interface *Rstudio*<sup>21</sup>. Todos os *scripts* foram inteiramente preparados para a realização deste estudo, com exceção do *script* para as regressões lineares das taxas de mortalidade que foi adaptado de uma versão prévia utilizada na preparação das estimativas pelo INCA<sup>1</sup>. Uma breve descrição das bases utilizadas é apresentada no Material Suplementar.

Os bancos referentes à população foram baixados do Departamento de Informática do Sistema Único de Saúde (DATASUS)<sup>22</sup>. Entre 1980 e 2012, foram usadas informações dos Censos (1980, 1991, 2000 e 2010), contagem (1996) e projeções intercensitárias (1981 a 2012), elaboradas pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e disponibilizadas pelo DATASUS, segundo faixa etária, sexo e situação de domicílio. Foram baixadas as bases de 1980 a 2012 referentes a todo o Brasil, pela página do Tabwin<sup>22</sup>. Para os anos de 2013 a 2021, foram utilizadas as estimativas preliminares elaboradas pelo Ministério da Saúde/Secretaria de Vigilância em Saúde e Ambiente/Departamento de Análise Epidemiológica e Vigilância de Doenças não Transmissíveis/Coordenação-Geral de Informações e Análises Epidemiológicas (MS/SVSA/DAENT/CGIAE)<sup>23</sup>. As informações foram tabuladas ano a ano, para cada sexo, selecionando-se o Estado de Minas Gerais e configurando-se municípios na linha e faixa etária na coluna. A informação referente a 2022 corresponde à população do Censo e foi obtida do IBGE<sup>24</sup>. Com a assistência do *script* do R, as informações da população organizadas pelos 853 municípios de Minas Gerais foram agrupadas pelas Macrorregiões de Saúde, conforme a planilha do Plano Diretor de Regionalização de Minas Gerais (ver mapa na Figura 1 do Material Suplementar). Além das informações por Macrorregião de Saúde, foram preparadas as informações da população de Belo Horizonte e de Poços de Caldas, que têm cobertura por RCBP, com a finalidade de realizar os cálculos de incidência do câncer. Foram empregados grupos de faixa etária de cinco em cinco anos de 0 a 79 anos e maiores de 80 anos. Foi criado um *data.frame* longo com as colunas Macrorregião, ano, sexo, faixa etária e população da macro (disponível como suplemento). Foram elaborados gráficos de pirâmides das populações para cada ano e Macrorregião também com a assistência dos *scripts* e do *software* R.

As informações de mortalidade de 1979 a 2022, provenientes do Sistema de Informação sobre Mortalidade (SIM), foram baixadas da página do Tabwin<sup>22</sup>. Entre 1996 e 2022, as causas básicas estão codificadas pela décima revisão da Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde (CID-10)<sup>25</sup> e, entre 1979 e 1995, pela nona revisão (CID-9)<sup>26</sup>. Não foram incluídos na análise óbitos sem informação de sexo (n=373; 0,06%), idade (n=834; 0,14%) e residência (n=546; 0,09%). Os óbitos por câncer foram agrupados de acordo com a Tabela 1.

Os dados dos casos classificados como útero sem outra especificação (SOE) (C55) foram incorporados na análise, na tentativa de produzir taxas mais fidedignas para as localizações colo do útero (C53) e corpo do útero (C54). Essa realocação é recomendada no sentido de produzir taxas que permitam comparações significativas entre populações. Os métodos de realocação foram baseados na idade e nas distribuições específicas dos casos<sup>27</sup>. Na série completa, entre 1979 e 2022, a proporção de casos “útero SOE (C55)” correspondeu a 35,5% do conjunto de casos “12-colo do útero (C53)”, “13-corpo do útero (C54)” e “útero SOE (C55)”. Quando a idade foi menor do que 50 anos, todos os casos “útero SOE (C55)” foram realocados para “12-colo do útero (C53)”. Se a idade era maior ou igual a 50 anos, os casos “útero SOE (C55)” foram realocados para “12-colo do útero (C53)” na proporção de casos:

$$\frac{\text{"12 – Colo do útero (C53)"}}{\text{"12 – Colo do útero (C53)" + "13 – Corpo do útero (C54)"}}$$

para cada Macrorregião e ano;

e para “13-corpo do útero (C54)”, na proporção de:

$$1 - \frac{\text{"12 – colo do útero (C53)"}}{\text{"12 – colo do útero (C53)" + "13 – corpo do útero (C54)"}}$$

para cada Macrorregião e ano.

Com as informações de população e do número de óbitos, foram calculadas as taxas de mortalidade específicas por neoplasia, idade e sexo, as taxas brutas e as taxas ajustadas pelo método direto, de acordo com a população mundial (Tabela 1 do Material Suplementar).

A taxa de mortalidade específica por faixa etária (i) foi calculada para cada Macrorregião (m), neoplasia (n), sexo (s) e ano (a).

$$\text{Taxa de mortalidade específica}_{\text{imnsa}} = \frac{\text{número de óbitos}_{\text{imnsa}}}{\text{população}_{\text{imnsa}}}$$

**Legendas:** i = faixa etária; m = Macrorregião, n = grupo de neoplasia; s = sexo; a = ano.



Tabela 1. Grupos de neoplasia conforme a codificação pela CID-9 (1979-1995) e CID-10 (1996-2022)

Grupos de neoplasia	Códigos CID-9 (1979-1995)	Códigos CID-10 (1996-2022) Três primeiros dígitos
01-Cavidade oral (C00-C10)	140, 141, 142, 144, 145, 146	C00, C01, C02, C03, C04, C05, C06, C07, C08, C09, C10
02-Esôfago (C15)	150	C15
03-Estômago (C16)	151	C16
04-Cólon e reto (C18-C21)	153, 154	C18, C19, C20, C21
05-Fígado e vias biliares intra-hepáticas (C22)	155	C22
06-Pâncreas (C25)	157	C25
07-Laringe (C32)	161	C32
08-Traqueia, brônquio e pulmão (C33-C34)	162	C33, C34
09-Melanoma maligno da pele (C43)	172	C43
10-Outras neoplasias malignas da pele (C44)	173	C44
11-Mama feminina (C50)	174	C50 e sexo feminino
12-Colo do útero (C53)	180	C53
13-Corpo do útero (C54)	182	C54
Útero SOE (C55)*	179	C55
14-Ovário (C56)	183	C56
15-Próstata (C61)	185	C61
16-Bexiga (C67)	188	C67
17-Sistema nervoso central (C70-C72)	191, 192	C70, C71, C72
18-Glândula tireoide (C73)	193	C73
19-Linfoma de Hodgkin (C81)	201	C81
20-Linfoma não Hodgkin (C82-C85; C96)	200, 202	C82, C83, C84, C85, C96
21-Leucemias (C91-C95)	204, 205, 206, 207	C91, C92, C93, C94, C95
22-Outras neoplasias (C00-C97; D46)	147, 148, 149, 152, 156, 158, 159, 160, 163, 164, 165, 170, 171, 175, 176, 177, 178, 181, 184, 186, 187, 189, 190, 194, 195, 196, 197, 198, 203, 238	C11, C12, C13, C14, C17, C23, C24, C26, C27, C28, C29, C30, C31, C35, C36, C37, C38, C39, C40, C41, C45, C46, C47, C48, C49, C51, C52, C57, C58, C59, C60, C62, C63, C64, C65, C66, C68, C69, C74, C75, C76, C77, C78, C79, C80, C86, C88, C90, C97, D46

**Legenda:** (\*) a localização em útero sem outra especificação (SOE) foi realocada para as localizações em colo uterino e corpo uterino<sup>27</sup>.

Essa taxa foi aplicada à população padrão, resultando no número de óbitos esperados em cada faixa etária. O somatório do número de óbitos esperados foi então dividido pelo total da população padrão, para obtenção da taxa ajustada por idade.

$$\text{Taxa de mortalidade ajustada}_{\text{mnsa}} = \frac{\sum (\text{Taxa de mortalidade específica}_{\text{imnsa}} \times \text{População padrão}_i)}{\sum \text{População padrão}_i}$$

A taxa de mortalidade bruta foi calculada para cada Macrorregião (m), tipo de câncer (n), sexo (s) e ano (a):

$$\text{Taxa de mortalidade bruta}_{\text{mnsa}} = \frac{\sum \text{número de óbitos}_{\text{imnsa}}}{\sum \text{população}_{\text{imnsa}}}$$

As taxas de mortalidade bruta e ajustada, para cada tipo de câncer, sexo e Macrorregião, entre 1979 e 2022, foram submetidas a uma regressão linear, colocando-se o ano como variável numérica independente. O *script* para a realização das regressões foi desenvolvido em *rmarkdown*, adaptado de um *script* previamente utilizado na realização das estimativas pelo INCA. Os resultados das regressões,

com os respectivos coeficientes de determinação ( $r^2$ , uma medida de ajuste do modelo), foram analisados em um documento *word* (.docx), juntamente com gráficos de dispersão da taxa bruta de mortalidade *versus* ano do óbito, da distribuição dos resíduos em relação aos valores ajustados e à taxa bruta, e um gráfico QQ normal (para cada neoplasia, sexo e Macrorregião). Se a regressão foi considerada adequada (valor de  $r^2 > 0,7$  e distribuição aleatória dos resíduos), utilizaram-se as taxas calculadas a partir da equação obtida com a regressão. Nos outros casos, optou-se pela taxa média (bruta ou ajustada) dos últimos cinco anos (2018-2022).

Para calcular a população projetada para as Macrorregiões, foram utilizadas as projeções de população para o Estado de Minas Gerais como um todo, por faixa etária e sexo (projeção da população do Brasil e UF por sexo e idade para o período 2010-2060, edição 2018, produzidas pelo IBGE e disponibilizadas no DATASUS)<sup>23</sup>. O cálculo da população para cada Macrorregião foi feito conforme a metodologia do IBGE<sup>28</sup>, considerando a tendência de crescimento da população entre os censos de 2010 e 2022<sup>28</sup>:

$$\begin{aligned} \text{População}_m &= a_i \times \text{População}_{MG} + b_i \\ a_i &= \frac{\text{População}_{m2022} - \text{População}_{m2010}}{\text{População}_{MG2022} - \text{População}_{MG2010}} \\ b_i &= \text{População}_{m2010} - a_i \times \text{População}_{MG2010} \end{aligned}$$

Onde:

População<sub>m</sub> = População projetada da Macrorregião  
 População<sub>MG</sub> = População estimada para Minas Gerais  
 População<sub>m2022</sub> = População da macro no censo de 2022  
 População<sub>MG2022</sub> = População de MG no censo de 2022  
 População<sub>m2010</sub> = População da macro no censo de 2010  
 População<sub>MG2010</sub> = População de MG no censo de 2010

No tabulador do INCA<sup>29</sup>, foi tabulado o número de casos de câncer por faixa etária, segundo a CID-10, para cada sexo, nos RCBP de Belo Horizonte (informações de 2015 a 2019) e de Poços de Caldas (informações de 2010 a 2014). As populações cobertas pelos RCBP de Belo Horizonte (2.315.560 habitantes) e de Poços de Caldas (163.742 habitantes) representam 12,1% da população do Estado (20.539.989 habitantes), considerando os resultados do Censo de 2022. Os casos notificados como C55 (útero SOE) foram realocados nos códigos C53 e C54, seguindo a mesma metodologia utilizada para os óbitos, justificando a obtenção das informações das faixas etárias. A razão da incidência/mortalidade foi calculada para Belo Horizonte e Poços de Caldas de acordo com metodologia usada pelo INCA<sup>1</sup> para cada tipo de neoplasia (subscrito *n* nas equações) e sexo (subscrito *s* nas equações). Não foram previstos ajustes para os casos em que a razão da incidência/mortalidade resultasse valor menor do que 1.

$$\text{Razão} \frac{\text{incidência}}{\text{mortalidade}}_{ns} = \frac{\frac{\text{Número de casos novos}_{ns}}{\sqrt{\text{População}}}}{\frac{\text{Número de óbitos}_{ns}}{\sqrt{\text{População}}}}$$

As estimativas de incidência foram realizadas utilizando-se a média da razão incidência/mortalidade calculada para os registros de Poços de Caldas e de Belo Horizonte. Essa razão foi aplicada às taxas brutas e ajustadas de mortalidade estimadas por regressão para o ano de 2024, para cada sexo (indicado pelo subscrito *s*), Macrorregião (subscrito *m*) e neoplasia (subscrito *n*). Quando a regressão não se mostrou adequada, foram utilizadas as taxas de mortalidade médias (bruta e ajustada) entre 2018 e 2022 (cinco anos).

$$\text{Taxa incidência estimada}_{mns2024} = \text{Taxa mortalidade}_{mns2024} \times \text{Razão} \frac{\text{incidência}}{\text{mortalidade}}_{ns}$$

Os valores absolutos para o número de casos novos foram calculados aplicando-se a taxa bruta estimada à população projetada para 2024, para cada Macrorregião e sexo.

$$\text{Número de casos novos}_{mns2024} = \frac{\text{Taxa incidência estimada}_{mns2024} \times \text{População}_{mns2024}}{100 \text{ mil}}$$

Seguindo a metodologia das estimativas do INCA<sup>1</sup>, somente são apresentados os resultados do número de casos novos se ele for maior ou igual a 20. Os valores absolutos foram arredondados para múltiplos de 10.

As taxas ajustadas de incidência estimadas foram utilizadas para a representação espacial com base nas distribuições das taxas por quartil.

Como toda a metodologia de cálculo foi estruturada em *scripts*, foi possível replicá-la para ambos os sexos, para todas as neoplasias (incluindo ou não pele não melanoma) e o Estado como um todo. Desse modo, o número de casos estimados para ambos os sexos não necessariamente corresponde exatamente à soma aritmética do número de casos estimados para os sexos masculino e feminino. Da mesma maneira, os números de casos estimados para todas as neoplasias e todas as neoplasias exceto pele não melanoma, e para todo o Estado não necessariamente correspondem às somas aritméticas do número de casos estimados para cada neoplasia (incluindo ou não pele não melanoma) ou para cada Macrorregião, respectivamente.

A consolidação dos resultados para publicação foi feita em um *script* do *rmarkdown*, sendo apresentados em texto os valores dos casos estimados para Minas Gerais como um todo e das cinco Macrorregiões com as maiores taxas brutas, para cada sexo. Os resultados das taxas estimadas e o número de casos novos para cada uma das Macrorregiões do Estado são apresentados em uma



tabela (ilustrada no Quadro 2 do Material Suplementar). Foram ainda elaborados mapas com as taxas ajustadas de incidência por sexo, com base na distribuição por quartil. Neste artigo, apenas a título de demonstração da metodologia utilizada, apresentam-se as informações referentes a todas as neoplasias, exceto pele não melanoma, também considerando a importância dessa informação na programação da atenção na alta complexidade estadual.

## RESULTADOS

As pirâmides populacionais para as Macrorregiões do Estado demonstram uma variação grande no contingente populacional entre as Macrorregiões de saúde (Figura 2 do Material Suplementar). Nas diferentes Macrorregiões do Estado, nota-se um envelhecimento da população, com a mudança na forma do gráfico entre 1980 e 2022.

Para ilustrar o funcionamento do *script*, apresentam-se os resultados obtidos para todas as neoplasias, exceto pele não melanoma, para o Estado de Minas Gerais (sexo feminino, masculino e ambos os sexos). Na Figura 1, encontram-se as taxas de mortalidade brutas e ajustadas, entre 1979 e 2022, para os sexos masculino e feminino e ambos os sexos.

Os resultados das taxas projetadas pela regressão linear, os valores do  $r^2$  para cada regressão e a taxa média os últimos cinco anos são apresentados na Tabela 2.

Para os cálculos de incidência e da razão incidência/mortalidade, foram trabalhadas as informações dos últimos cinco anos disponíveis para os RCBP de Poços de Caldas (2010-2014) e de Belo Horizonte (2015-2019). Os resultados dos cálculos para todas as neoplasias, exceto pele não melanoma, são apresentados na Tabela 3.

A razão da incidência/mortalidade calculada para as populações de Belo Horizonte e de Poços de Caldas foi aplicada sobre as taxas de mortalidade projetadas para 2024, permitindo estimar as taxas brutas e ajustadas de incidência para os demais territórios. As taxas brutas estimadas, por sua vez, aplicadas sobre a população do território resultam no número de casos novos estimados para aquela população. Um exemplo desses cálculos é apresentado na Tabela 4.

O último passo na elaboração das estimativas é a publicação dos resultados. O pacote *rmarkdown* possibilita a conjugação dos resultados das análises com elementos de texto, gráficos, tabelas e mapas, com a elaboração de um documento (no formato de *word*, *pdf* ou *html*). Como exemplo do documento obtido com a execução do *script*, um trecho é apresentado no Quadro 1 do Material Suplementar.

No documento elaborado, ainda foram incluídos tabelas e mapas. Um exemplo de tabela produzida

no *rmarkdown* encontra-se no Quadro 2 do Material Suplementar. A representação espacial das taxas de incidência ajustadas estimadas para as Macrorregiões do Estado, organizadas pela distribuição em quartis, é ilustrada na Figura 2 do Material Suplementar.

## DISCUSSÃO

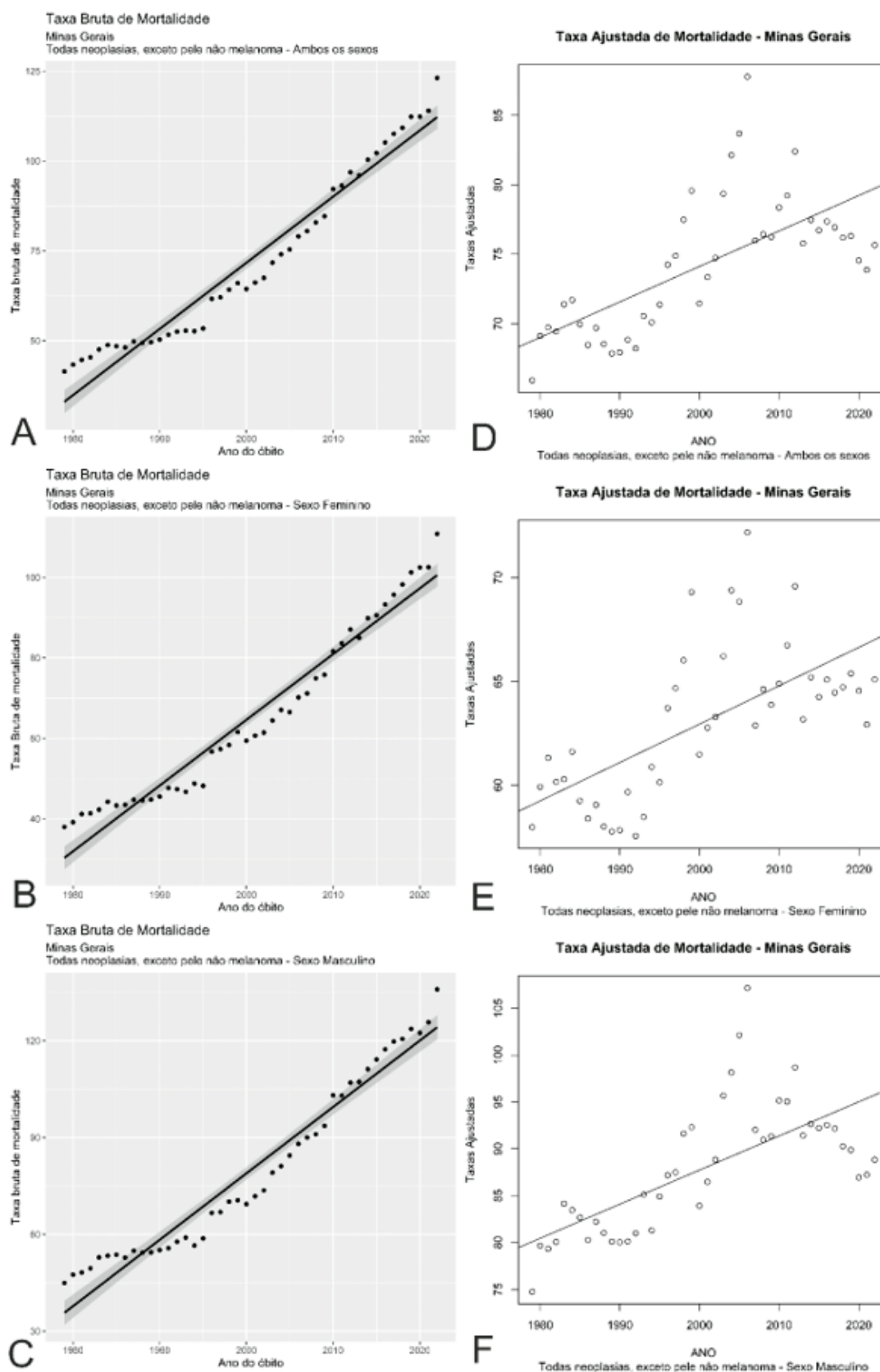
Nas ações da vigilância em saúde, não se pode mais abdicar dos recursos de gerenciamento e análise de informações, como os disponibilizados pelo *software* R e seus pacotes. O domínio dessas tecnologias é essencial para manter a atualidade das análises de situação, incorporar o maior número de informações ao presente estudo e utilizar métodos estatísticos adequados a cada análise. Além disso, o *software* R e seus pacotes contribuem muito na apresentação dos resultados, na elaboração de gráficos, mapas, e até mesmo de texto, como foi feito neste estudo.

Ao contrário de *softwares* com recursos de apontar e clicar (*point and click*), o aprendizado do *software* R não é intuitivo, exigindo estudo e prática. É uma ferramenta gratuita (*software* livre) capaz de lidar com uma quantidade de dados muito além das 1.048.576 linhas disponíveis em uma planilha de *Excel*. Além disso, toda manipulação e análise estruturadas em *scripts* permitem que se refaça todo o processo sempre que necessário, com relativa simplicidade, possibilitando a atualização das informações. Como exemplo, a reestruturação da regionalização pode ser rapidamente incorporada às análises.

O compartilhamento dos *scripts* é outra vantagem, permitindo que a análise desenvolvida para Minas Gerais possa ser adaptada para outros Estados e para o Brasil como um todo. Os *scripts* utilizados neste estudo estão disponibilizados (ver *link* no Material Suplementar), na expectativa que sejam utilizados e aperfeiçoados.

Uma experiência prévia no compartilhamento de *scripts* do R, para a preparação das bases de dados dos RCBP para submissão a um estudo internacional, facilitou a participação de registros brasileiros no estudo. A direção do estudo considerou a estratégia útil e disponibilizará os *scripts* adaptados a registros de outros países<sup>30</sup>. Com base em tais experiências, acredita-se que as ações da vigilância em saúde possam ser potencializadas com o uso do *software* R, muito além da elaboração de estimativas.

Sempre que é instalada uma nova versão atualizada do R, é necessário atualizar também os pacotes. Quando um pacote não é atualizado para a nova versão, ele poderá sair do repositório oficial no *Comprehensive R Archive Network* (CRAN) e não estar mais facilmente acessível aos usuários. Como estratégia para lidar com essa situação, é possível manter as versões antigas, com os pacotes previamente instalados.



**Figura 1.** Taxas de mortalidade brutas (A, B, C) e ajustadas (D, E, F) para todas as neoplasias, exceto pele não melanoma no Estado de Minas Gerais, para ambos os sexos (A, D), sexo feminino (B, E) e sexo masculino (C, F). As retas indicam a tendência estimada por regressão linear. Eixos y com escalas diferentes entre os gráficos



Tabela 2. Taxas projetadas de mortalidade para o ano de 2024, para todas as neoplasias, exceto pele não melanoma pela equação da reta da regressão linear, com a medida do  $r^2$ . Taxa média dos últimos cinco anos utilizada como alternativa, nos casos em que a regressão não foi considerada satisfatória

Tipo de taxa	Sexo	Taxa projetada pela regressão	$r^2$	Taxa média últimos cinco anos	Taxa escolhida para cálculo
Bruta	Ambos	115,85	0,95	114,12	115,85
Bruta	Feminino	103,82	0,95	103,04	103,82
Bruta	Masculino	128,29	0,95	125,59	128,29
Ajustada	Ambos	80,25	0,45	75,28	75,28
Ajustada	Feminino	67,37	0,42	64,52	64,52
Ajustada	Masculino	96,47	0,47	88,58	88,58

Tabela 3. Cálculo da razão incidência/mortalidade para todas as neoplasias, exceto pele não melanoma

Sexo		Belo Horizonte (2015-2019)			Poços de Caldas (2010-2014)		
		Óbitos	Casos	População	Óbitos	Casos	População
Ambos	Valores absolutos	16.550	36.613	12.461.005	928	2.149	781.295
	Razão I/M	2,39			2,31		
	Mediana	2,35					
Feminino	Valores absolutos	8.417	20.503	6.627.012	410	978	403.729
	Razão I/M	2,43			2,38		
	Mediana	2,41					
Masculino	Valores absolutos	8.133	19.110	5.833.993	518	1.171	377.566
	Razão I/M	2,35			2,26		
	Mediana	2,31					

Legenda: I/M = incidência/mortalidade.

Tabela 4. Cálculo das taxas estimadas brutas e ajustadas para todas as neoplasias, exceto pele não melanoma em Minas Gerais, no ano de 2024

Tipo de taxa	Sexo	Taxa de mortalidade projetada (2024)	Razão incidência/mortalidade	Taxa de Incidência estimada (2024)	População estimada (2024)	Casos novos estimados (2024)	Arredondamento
Taxa bruta	Ambos	115,85	2,35	272,79	21.737.475	59.298,43	59.300
	Feminino	103,82	2,41	250,28	11.037.396	27.624,84	27.620
	Masculino	128,29	2,31	295,72	10.700.079	31.642,01	31.640
Taxa ajustada	Ambos	75,28	2,35	177,26			
	Feminino	64,52	2,41	155,54			
	Masculino	88,58	2,31	204,19			

A elaboração dos *scripts* do R envolve a criação de estratégias para a solução de problemas no gerenciamento de dados e análises, em muitos aspectos limitados pelo próprio conhecimento das múltiplas potencialidades disponibilizadas pela ferramenta. Nesse sentido, muitos aspectos podem ser melhorados e pequenas imperfeições, porventura presentes, podem ser corrigidas.

O Sistema de Vigilância em Saúde tem hoje no DATASUS/Tabwin um grande aliado na disponibilização das informações. Uma melhor coesão das informações,

entretanto, seria desejável. Nas informações de população, por exemplo, foi possível baixar as bases completas no Tabwin de 1980 a 2012, fazer tabulações no DATASUS de estimativas populacionais entre 2013 e 2021, e baixar as informações do Censo de 2022 no *site* do IBGE. Embora os arquivos sejam bem documentados e as fontes de informações preservadas, seria desejável ter todos os dados disponíveis no mesmo formato e em um mesmo diretório. Não há disponível, entretanto, uma retroprojeção para 1979 (ano que inicia a série do SIM),



levando-nos a utilizar a população de 1980 também para 1979. Não há, até o momento, projeções de população a partir de 2023 distribuídas por município e grupos de idade, sendo necessário realizar os cálculos a partir da população projetada para o Estado de Minas Gerais.

As estimativas de incidência devem ser compreendidas como estimativas, não um dado do mundo real. Desse modo, alguma variabilidade nos resultados é esperada, não apenas pela incorporação de novas bases, mas também pelos resultados das análises de regressão e projeção utilizadas. Nas análises realizadas neste estudo, por exemplo, estão incorporadas as informações de mortalidade referentes ao ano de 2022 e os dados populacionais do Censo de 2022, diferentemente das projeções calculadas pelo INCA para 2023<sup>1</sup>, que utilizou dados até 2020. Em comparação com as projeções elaboradas pelo INCA<sup>1</sup>, a presente estimativa projeta maior número de casos, principalmente para o sexo masculino – 52.090 casos na projeção do INCA, 59.300 nesta estimativa (Tabela 2 do Material Suplementar). A maior parte dessa diferença pode ser atribuída ao número de casos estimados para o câncer de próstata – 7.970 casos novos estimados pelo INCA e 11.800 casos novos nesta estimativa. Especificamente neste caso, a metodologia do INCA utiliza a mediana da Região geográfica, de modo a evitar a inflação da taxa de incidência produzida pelo aumento do rastreamento da neoplasia, pelo advento do teste do antígeno específico da próstata (PSA)<sup>1</sup>.

Essa comparação visa apenas demonstrar a consistência e a viabilidade da metodologia proposta no artigo, não estabelecendo uma comparação direta com outras estimativas.

O cálculo das estimativas possui limitações que podem estar relacionadas à qualidade das informações de incidência, da mortalidade e da própria população. No que se refere à incidência, as bases do RCBP de Poços de Caldas foram analisadas e aprovadas para publicação no *Cancer Incidence in Five Continents*, volumes XI e XII, respectivamente. Um critério de qualidade das informações dos registros é a ocorrência de razão incidência/mortalidade menor do que 1, indicando possível subnotificação de casos, se as declarações de óbito são acuradas e a incidência e a sobrevida da neoplasia são constantes. Considerando as razões medianas utilizadas para o cálculo das estimativas, essa situação ocorreu para as neoplasias de fígado e vias biliares intra-hepáticas (sexo masculino, 0,97; sexo feminino, 0,81 e ambos os sexos, 0,93) e pâncreas (sexo feminino, 0,89; e ambos os sexos, 0,94), sabidamente neoplasias de alta letalidade e curta sobrevida. Com a decisão de não propor qualquer ajuste na razão incidência/mortalidade, é possível que alguma subestimativa ocorra para essas neoplasias. Tomando o exemplo extremo da neoplasia de fígado e vias biliares

intra-hepáticas, no sexo feminino, se a razão tivesse sido ajustada para 1, seriam esperados 420 casos novos em Minas Gerais, ao invés de 340 casos novos calculados sem esse ajuste. É possível que parte dos óbitos notificados para essas neoplasias represente casos detectados já em fase terminal, sem tempo hábil para a confirmação do diagnóstico clínico, dificultando a sua inclusão nos Registros de Câncer.

Foi usado o modelo de regressão linear para a projeção das taxas de mortalidade, aplicado a uma longa série temporal (1979-2022). Essa abordagem não capta mudanças na tendência de crescimento ou diminuição das taxas ao longo do tempo. Essa dificuldade foi contornada com o emprego da taxa média dos últimos cinco anos, nos casos em que o ajuste do modelo de regressão não foi considerado adequado.

As informações da população precisam ser também qualificadas; os resultados do Censo de 2022 revelaram uma divergência entre as projeções populacionais anteriores e a efetiva contagem da população, indicando a necessidade de conciliação dessas informações. Para este estudo, especificamente, não foi realizado qualquer ajuste nas informações da população disponibilizadas pelo IBGE (Censos de 1980, 1991, 2000, 2010 e 2022; contagem de 1996; e projeções intercensitárias de 1981 a 2012) ou pelo Ministério da Saúde (estimativas preliminares de 2013 a 2021). Um possível efeito de subenumeração da população do Censo de 2022 seria um aumento da taxa de mortalidade para esse ano. As projeções de população usadas para 2024, entretanto, são baseadas nas informações até 2018 e não incorporam dados do censo de 2022.

Nas informações sobre mortalidade, fez-se a realocação dos casos notificados com útero SOE, seguindo a metodologia adotada nas estimativas elaboradas pelo INCA. Diferentemente do realizado no estudo de Jardim et al.<sup>7</sup>, não foi feita a realocação de outros códigos de neoplasia maligna mal definidos (C26.0, C26.8, C26.9, C57.8, C57.9, C76.0, C76.2, C76.3, C78-C79 e C80) que, no presente estudo, foram incluídos no grupo outras neoplasias. Considerando apenas a mortalidade entre 1996 e 2022, o grupo de neoplasias mal definidas do trato digestivo (C26.0, C26.8 e C26.9) representou 6,9% do conjunto de óbitos por neoplasias C15-C26, podendo acarretar uma pequena subestimativa de casos atribuídos a esses grupos-alvo específicos.

Pelo planejamento estadual, as Macrorregiões de Saúde constituem a base territorial de planejamento da atenção terciária à saúde que engloba Microrregiões de Saúde com população em torno de 700 mil habitantes e que oferta à sua população serviços de saúde hospitalares de maior densidade tecnológica<sup>4</sup>. Ainda assim, existem



Macrorregiões com população relativamente pequena, como a Macrorregião do Jequitinhonha, com 385.593 habitantes em 2022, levando à maior flutuação das taxas, principalmente para os cânceres de menor frequência. Trata-se de uma limitação para o cálculo das estimativas de neoplasias com menor carga de doença. Não há, entretanto, outra forma de divisão em menor número de agrupamentos. Houve uma tentativa de contornar essa limitação com a decisão de elaborar estimativas para ambos os sexos, o que pode ser adequado no caso de tumores sem nítido predomínio em um dos sexos. Também foi omitido o número de casos novos, quando a estimativa resultou em um número menor do que 20. Na medida em que a atenção oncológica tende cada vez mais à especialização, talvez seja adequado agrupar neoplasias correlatas de acordo com a tendência na especialização em oncologia (tumores hematológicos, tumores do trato digestivo, do sistema genital feminino etc.) como subsídio para o planejamento da organização da rede de atenção. Por outro lado, neoplasias com especificidades próprias no que se refere à prevenção ou detecção precoce (por exemplo, câncer do colo uterino, câncer de mama, câncer colorretal) devem ser avaliadas individualmente.

O Brasil tem uma longa tradição na elaboração de estimativas de câncer, publicadas regularmente desde meados da década de 1990. Dadas as dimensões continentais do país, com apenas parte da população coberta pelos RCBP<sup>31</sup>, as estimativas assumem um papel fundamental no dimensionamento da magnitude e do impacto do câncer nas diferentes Regiões do país. A qualidade das estimativas é altamente dependente da qualidade das informações utilizadas no seu cálculo, incluindo as informações de incidência obtidas dos RCBP, da mortalidade (pelo SIM) e da própria população. O desafio de proporcionar informações da incidência de câncer em Sub-Regiões estaduais pode fornecer ao gestor subsídios para planejar, monitorar e avaliar as ações de controle de câncer e auxiliar na gestão das redes estaduais de atenção oncológica tanto no aspecto da prevenção como no diagnóstico precoce e tratamento como preconizado pela Política Nacional de Prevenção e Controle do Câncer no âmbito do Sistema Único de Saúde (SUS) e pelo Programa Nacional de Navegação da Pessoa com Diagnóstico de Câncer<sup>32</sup>.

## CONCLUSÃO

Neste estudo, foi apresentado todo o percurso metodológico para a elaboração de estimativas de câncer para o Estado de Minas Gerais e suas Macrorregiões de Saúde, e foram disponibilizados como suplemento todos os *scripts* utilizados na preparação das informações,

cálculos necessários e apresentação dos resultados. A expectativa é que este artigo possa ser utilizado como um tutorial, juntamente com os *scripts* disponibilizados, para a elaboração de estimativas em outros territórios e populações, permitindo detalhar a distribuição dos diferentes tipos de câncer em nível intraestadual, onde estejam disponíveis informações adequadas de incidência, ou seja, onde possam ser usadas informações de RCBP existentes, com qualidade de informação e cobertura satisfatórias, específicos da Região ou de Regiões vizinhas.

## AGRADECIMENTOS

À equipe da Divisão de Vigilância e Análise de Situação, da Coordenação de Prevenção e Vigilância (Conprev) do INCA, Rio de Janeiro (RJ), Brasil, na pessoa da saudosa Marcell de Oliveira Santos (*in memoriam*), que muito contribuiu para a realização dessas estimativas regionais. À Fernanda Cristina da Silva de Lima, que contribuiu com uma versão de *script* do *rmarkdown* para a regressão das taxas.

## CONTRIBUIÇÕES

Gil Patrus Pena e Sara Oliveira Ribeiro contribuíram substancialmente na concepção e no planejamento do estudo; na obtenção, análise e interpretação dos dados; na redação e revisão crítica. Gil Patrus Pena elaborou os *scripts* do R. Ambos os autores aprovaram a versão final a ser publicada.

## DECLARAÇÃO DE CONFLITOS DE INTERESSE

Nada a declarar.

## FONTES DE FINANCIAMENTO

Não há.

## REFERÊNCIAS

1. Santos MO, Lima FCS, Martins LFL, et al. Estimativa de incidência de câncer no Brasil, 2023-2025. Rev Bras Cancerol. 2023;69(1):e-213700. doi: <https://doi.org/10.32635/2176-9745.RBC.2023v69n1.3700>
2. Secretaria de Estado da Saúde (MG). Programa de Avaliação e Vigilância do Câncer e seus fatores de risco (PAV-MG). Situação do câncer em Minas Gerais e suas macrorregiões de saúde: estimativas de incidência e mortalidade para o ano 2013, válidas para 2014: perfil da mortalidade: perfil da assistência na alta complexidade. Belo Horizonte: SES-MG; 2013. v. 1.



3. Secretaria de Estado de Saúde (MG). Subsecretaria de Gestão Regional. Ajuste do Plano Diretor de Regionalização de Saúde de Minas Gerais (PDR/MG) [Internet]. 1. ed. Belo Horizonte: SES-MG; 2020. [Acesso 2024 jul 7]. Disponível em: [https://www.saude.mg.gov.br/images/1\\_noticias/06\\_2023/2-jul-ago-set/regionalizacao/1-PDR%202020.pdf](https://www.saude.mg.gov.br/images/1_noticias/06_2023/2-jul-ago-set/regionalizacao/1-PDR%202020.pdf)
4. Secretaria de Estado de Saúde (MG) [Internet]. Belo Horizonte: SEM-MG; 2023. Plano Diretor de Regionalização, 2023 ago 17. [Acesso 2024 jul 7]. Disponível em: <https://www.saude.mg.gov.br/gestor/regionalizacao>
5. Luizaga CTM, Buchalla CM. Estimativa da incidência de câncer no Estado de São Paulo, Brasil, a partir de dados reais. *Cad Saúde Pública*, 2023;39(2):e00134222. doi: <https://doi.org/10.1590/0102-311XPT134222>
6. Ferlay J, Ervik M, Lam F, et al. Global Cancer Observatory: cancer today [Internet]. Lyon: International Agency for Research on Cancer; 2024 [acesso 2024 out 9]. Disponível em: <https://gco.iarc.who.int/today/en/data-sources-methods>
7. Jardim BC, Junger WL, Daumas RP, et al. Estimativa de incidência de câncer no Brasil e regiões em 2018: aspectos metodológicos. *Cad Saúde Pública*. 2024;40(6):e00131623. doi: <https://doi.org/10.1590/0102-311XPT131623>
8. R: The R Project for Statistical Computing [Internet]. Versão 4.3.2 Viena: The R foundation. 2021 nov 2 - [acesso 2022 set 6]. Disponível em: <https://www.r-project.org>
9. Lemes V, Baccaro FB. Introdução ao uso do programa R [Internet]. Manaus: Instituto Nacional de Pesquisas da Amazônia, 2011. [acesso 2024 out 9]. Disponível em: [https://www.researchgate.net/profile/Victor-Landeiro/publication/275035302\\_Introducao\\_ao\\_uso\\_do\\_programa\\_R/links/553041070cf20ea0a06f67ca/Introducao-ao-uso-do-programa-R.pdf](https://www.researchgate.net/profile/Victor-Landeiro/publication/275035302_Introducao_ao_uso_do_programa_R/links/553041070cf20ea0a06f67ca/Introducao-ao-uso-do-programa-R.pdf)
10. Wickham H, François R, Henry L, et al. dplyr: a grammar of data manipulation [Internet]. Versão 1.1.3. [sem local]: The R foundation. 2023 nov 17. [acesso 2024 out 9]. Disponível em: <https://CRAN.R-project.org/package=dplyr>
11. Wickham H, Vaughan D, Girlich M. Tidy: Tidy Messy Data [Internet]. Versão 1.3.1, [sem local]: The R foundation. 2024 jan 1. [acesso 2024 out 9]. Disponível em: <https://CRAN.R-project.org/package=tidy>
12. Wickham H, Bryan J. readxl: Read Excel Files [Internet]. Versão 1.4.3, [sem local]: The R foundation. 2023 nov 17. [acesso 2024 out 9]. Disponível em: <https://CRAN.R-project.org/package=readxl>
13. Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag; 2016.
14. Petruzalek D. read.dbc: Read Data Stored in DBC (Compressed DBF) Files [Internet]. Versão 1.0.6. [sem local]: The R foundation. 2024 jul 6. [acesso 2024 out 9]. Disponível em: <https://CRAN.R-project.org/package=read.dbc>
15. Fox J, Weisberg S. An R companion to applied regression [Internet]. 3 ed. Thousand Oaks: Sage; 2023. [acesso 2024 out 9]. Disponível em: <https://www.john-fox.ca/Companion/>
16. Warnes GR, Bolker B, Lumley TL, et al. Package 'gmodels', 2024. Versão 1.0.6. [sem local]: The R foundation. 2024. doi: <https://doi.org/10.32614/CRAN.package.gmodels>
17. Zeileis A, Hothorn T (2002). "Diagnostic Checking in Regression Relationships" [Internet]. *R News* [Internet]. 2002 [acesso 2024 jul 17];2(3):7-10. Disponível em [https://cran.r-project.org/doc/Rnews/Rnews\\_2002-3.pdf](https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf)
18. Allaire J, Xie Y, Dervieux C, et al. rmarkdown: Dynamic Documents for R [Internet]. Versão 2.25. São Francisco: GitHub, Inc.; 2023. [Acesso 2024 jul 17]. Disponível em <https://github.com/rstudio/rmarkdown>
19. Xie Y. knitr: A General-Purpose Package for Dynamic Report Generation in R. R package [Internet]. version 1.44. [Acesso 2024 jul 17]. Disponível em <https://yihui.org/knitr/>
20. Pebesma E. "Simple features for R: standardized support for spatial vector data." *R Journal*. 2018;10(1):439-46. doi: <https://doi.org/10.32614/RJ-2018-009>
21. RStudio [Internet]. Versão 2023.12.1+402. Boston: Posit Software, PBC. 2024 abr 1 - [acesso 2024 mar 1]. Disponível em: <http://www.rstudio.com/ide>
22. TABWIN [Internet]. Brasília (DF): DATASUS. c2008 - [acesso 2024 jul 28]. Disponível em: <https://datasus.saude.gov.br/transferecia-de-arquivos/>
23. TABNET [Internet]. Brasília (DF): DATASUS. c2008. [acesso 2024 jul 28]. Disponível em: <http://tabnet.datasus.gov.br/cgi/defthotm.exe?ibge/cnv/popsvsbr.def>
24. SIDRA: Banco de Tabelas Estatísticas [Internet]. Rio de Janeiro: IBGE; [sem data]. [acesso 2024 jul 28]. Disponível em: <https://sidra.ibge.gov.br/tabela/9514>
25. Organização Mundial da Saúde. CID-10: Classificação Estatística Internacional de Doenças e problemas relacionados à saúde. São Paulo: Edusp; 2008.
26. Organização Mundial da Saúde. CID-9: Classificação Estatística Internacional de Doenças: manual de lesões e causas de óbito. São Paulo: Centro Brasileiro para Classificação de Doenças em Português; 1979.
27. Loos AH, Bray F, McCarron P, et al. Sheep and goats: separating cervix and corpus uteri from imprecisely coded uterine cancer deaths, for studies of geographical and temporal variations in mortality. *Eur J Cancer*. 2004;40(18):2794-803. doi: <https://doi.org/10.1016/j.ejca.2004.09.007>
28. Instituto Brasileiro de Geografia e Estatística. Estimativas da população 2021 nota metodológica nº 1. Rio de Janeiro: IBGE; 2021



29. Registro de Câncer de Base Populacional [Internet]. Rio de Janeiro: INCA. [2012] – [acesso 2024 jul 20]. Disponível em: <https://www.inca.gov.br/BasePopIncidencias/Home.action>
30. Pena Gil P, Pongnikorn D, Khan Baloch F, et al. Sharing R scripts to re-code cancer registry data for international research. In: 12º Congresso Brasileiro de Epidemiologia; 2024 nov 23-27; Rio de Janeiro. Rio de Janeiro: Abrasco; 2024.
31. Pena GPM, Santos ATC, Pezzotti CM, et al. Registros de câncer de base populacional no Brasil: relevância, desafios e oportunidades. *Rev Bras Cancerol.* 2025;71(1):e-104878. doi: <https://doi.org/10.32635/2176-9745.RBC.2025v71n1.4878>
32. Presidência da República (BR). Lei nº. 14.758 de 19 de dezembro de 2023. Institui a Política Nacional de Prevenção e Controle do Câncer no âmbito do Sistema Único de Saúde (SUS) e o Programa Nacional de Navegação da Pessoa com Diagnóstico de Câncer; e altera a Lei nº 8.080, de 19 de setembro de 1990 (Lei Orgânica da Saúde). *Diário Oficial da União*, Brasília, DF. 2023 dez 20; Seção 1.

Recebido em 29/7/2024  
Aprovado em 11/11/2024

