

O Uso do *Google Trends* para Análise de Interesse por Informações sobre o Câncer no Brasil: Aspectos Teórico-Metodológicos

doi: <https://doi.org/10.32635/2176-9745.RBC.2023v69n2.3847>

The Use of Google Trends for Analysis of Interest for Information about Cancer in Brazil: Theoretical-Methodological Aspects
El Uso de *Google Trends* para el Análisis de Interés en Información sobre el Cáncer en el Brasil: Aspectos Teórico-Metodológicos

Fernando Lopes Tavares de Lima¹

INTRODUÇÃO

No Brasil, estimam-se 704 mil novos casos de câncer para cada ano do triênio 2023-2025¹. Esse cenário reafirma o câncer como um crescente problema de saúde pública no país. A atual Política Nacional para Prevenção e Controle do Câncer (PNPCC)² institui o estímulo à formulação de estratégias de comunicação com a população, visando à disseminação e ampliação do conhecimento sobre o câncer, seus fatores de risco e as diretrizes de prevenção.

Tal estímulo deve acompanhar ações práticas que ampliem o conhecimento da sociedade sobre como prevenir e detectar precocemente o câncer³. Porém, diante das dificuldades de acessar essas ações, a Internet tem sido utilizada e, nesse ambiente, o buscador *Google* é um espaço privilegiado, considerado uma importante fonte de informação em saúde⁴. O uso disseminado dessa ferramenta tem gerado um enorme banco de dados acerca do interesse da população sobre qualquer tema, incluindo o câncer.

Pesquisadores têm utilizado o *Google Trends* (GT) como ferramenta para acessar esses dados e identificar os interesses populacionais por informações em saúde⁵. Assim, o comportamento baseado na Internet é analisado para se examinar o comportamento humano real⁶.

Apesar da importância das pesquisas com o GT, há preocupações com a robustez metodológica dos artigos. Em 2014, identificou-se que apenas 7% dos artigos forneceram as informações necessárias para reprodução da busca e somente 34% documentaram a data em que a busca foi realizada⁷.

Assim, diante da importância dessa ferramenta, o objetivo deste artigo é apresentar uma orientação teórico-metodológica sobre seu uso, bem como suas indicações e limitações no âmbito da prevenção e controle do câncer no Brasil.

DESENVOLVIMENTO

O *GOOGLE TRENDS*

O GT é um portal on-line gratuito criado em 2006 pela *Google Inc.* que apresenta a frequência relativa com que um determinado termo é pesquisado no buscador *Google* em todo o mundo⁸. Assim, com base nas consultas reais dos usuários, o GT fornece informações sobre o que foi e o que é tendência de interesses de internautas⁶.

Em razão do seu processamento rápido, os dados estão disponíveis em tempo real⁵ e sua utilização tem sido associada à utilidade para detectar prontamente um fenômeno, à sua representatividade e ao seu alto potencial de aplicação nas mais diversas áreas⁹.

De forma geral, as pesquisas que utilizam o GT visam a quatro objetivos: medir o interesse sobre um tema; detectar suas variações ou sazonalidades; analisar as relações dessa variação com eventos específicos; e prever eventos relacionados à saúde⁶.

UTILIZANDO O *GOOGLE TRENDS*

Para utilizar o GT, é necessário acessar <https://trends.google.com/> e digitar uma estratégia de busca. A partir da apresentação dos resultados, fica disponível o conjunto de ferramentas para aperfeiçoar essa busca: (1) estratégia de busca; (2) área geográfica; (3) período; (4) categoria; e (5) modalidade de pesquisa. Outra forma de acessar as ferramentas é clicar na aba “Explorar”, na página inicial.

A construção de uma estratégia de busca deve considerar que não há diferenciação entre letras maiúsculas e minúsculas, mas há em relação aos acentos, formas plurais e erros de ortografia⁸. Um exemplo é a busca pelo termo “câncer de próstata” com e sem acento. No Brasil, apesar da forma correta de escrita ser com acento, há um maior volume de buscas pela expressão sem acento (Figura 1A). Isso exemplifica a importância de o pesquisador

¹Instituto Nacional de Câncer (INCA), Coordenação de Prevenção e Vigilância (Conprev), Divisão de Pesquisa Populacional. Rio de Janeiro (RJ), Brasil. E-mail: flima@inca.gov.br. Orcid iD: <https://orcid.org/0000-0002-8618-7608>

Endereço para correspondência: Fernando Lopes Tavares de Lima. Rua Marquês de Pombal, 125, 7º andar – Centro. Rio de Janeiro (RJ), Brasil. CEP 20230-240. E-mail: flima@inca.gov.br



analisar as diferentes possibilidades de busca em sua língua e construir uma estratégia que não perca dados.

A estratégia de busca pode ser composta por termos e expressões. O uso de aspas em expressões limita os resultados a essa forma exata de escrita⁷. A Figura 1B exemplifica uma ligeira variação para uma busca por “câncer de próstata” com e sem aspas.

Ao digitar alguns termos no GT, abrem-se automaticamente opções de busca por tópicos específicos. Os tópicos são um grupo de termos sinônimos já categorizados em qualquer idioma, o que aumenta o poder da busca⁸. Um exemplo é apresentado na Figura 1C, que compara a busca por “câncer de próstata” como termo de pesquisa e como tópico.

Outra estratégia importante é a pesquisa agregada de termos ou expressões. Para tanto, utiliza-se o recurso “+”. Isso agrega diferentes termos dentro do mesmo grupo de estratégia de busca (limitado a cinco grupos simultâneos de pesquisa com 25 termos cada). Assim, podem ser incluídas as variações do termo, bem como erros ortográficos comuns⁵.

Deve-se delimitar uma Região Geográfica de abrangência. Pode-se selecionar o mundo todo ou um país específico. Além disso, alguns países possuem dados por sub-regiões^{5,7}, como as Unidades Federativas no Brasil.

A Figura 1D exemplifica a comparação de duas regiões, São Paulo e Minas Gerais, para o tópico “Câncer de próstata” no último ano. O valor maior em Minas Gerais não significa um valor absoluto maior, pois a medida é relativa⁸.

O GT oferece duas modalidades de dados. A primeira é classificada como “tempo real” e corresponde aos últimos sete dias. A segunda é uma amostra separada da primeira que vai desde 2004 a até 72 horas antes da pesquisa⁸.

O pesquisador pode escolher por um período já configurado ou personalizá-lo. Independente da escolha, é importante notar que os dados são normalizados pelo período selecionado. Assim, a busca deve ser composta exatamente pelo período de interesse⁷.

A busca pode ser realizada com base em uma categoria. Esse recurso é importante para eliminar ruídos de dados, principalmente nos casos em que a palavra possui diferentes significados. A palavra “câncer”, por exemplo, pode significar uma doença ou o signo do zodíaco. A seleção da categoria “saúde” pode delimitar os resultados (efeito parecido com a escolha de busca por tópico). Destaca-se que isso afeta todos os grupos de termos de pesquisa.

Como modalidade de pesquisa, o GT oferece as opções: “Pesquisa na Web”, “Pesquisa de Imagens”, “Pesquisa de

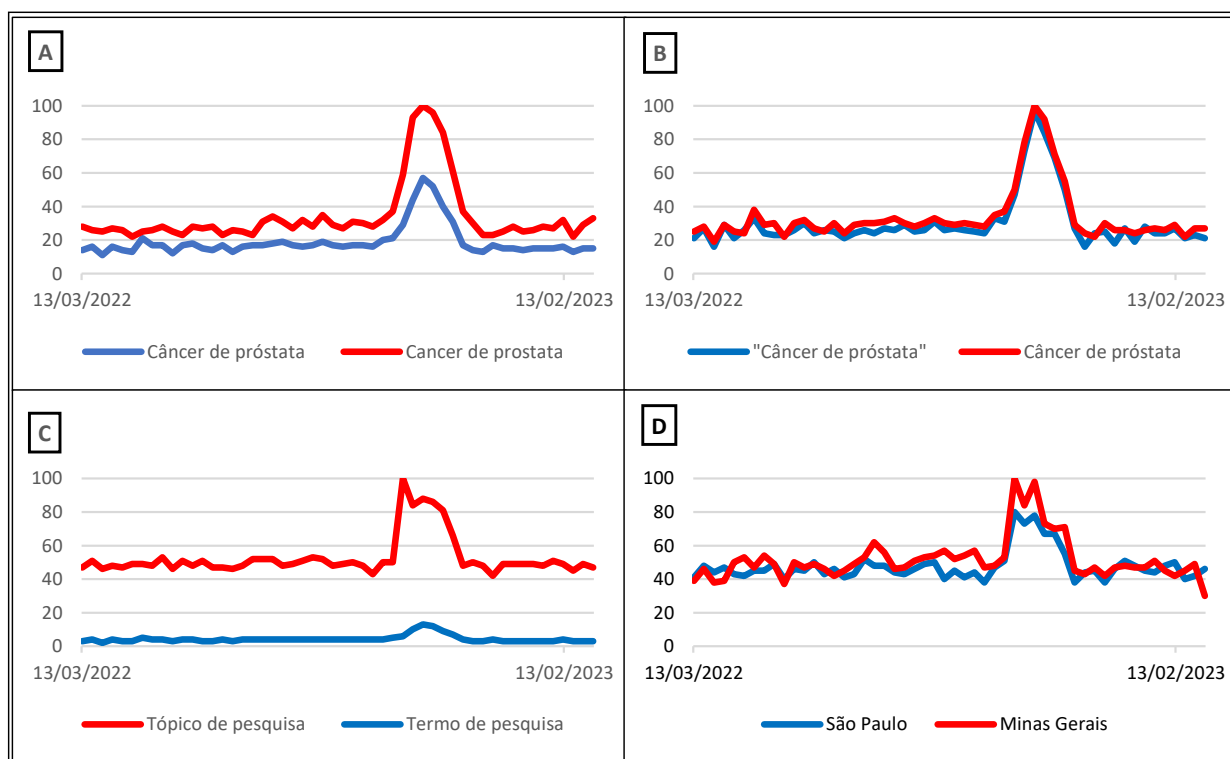


Figura 1. Exemplos do resultado do uso de diferentes ferramentas para construção da estratégia de busca no Google Trends. Imagem A: diferença do uso do acento na expressão “câncer de próstata” no Brasil. Imagem B: diferença do uso de aspas na expressão “câncer de próstata” no Brasil. Imagem C: diferença da seleção de tópico ou termo de pesquisa para a busca da expressão “câncer de próstata”. Imagem D: diferença do volume de busca por “câncer de próstata” em duas Unidades da Federação no Brasil

Fonte: Google Trends⁸, acesso em 3/4/2023.

Notícias”, “*Google Shopping*” e “Pesquisa do *YouTube*”. A opção “Pesquisa na Web” se refere ao buscador *Google* e é o objeto deste artigo. Porém, a depender do objetivo da pesquisa, pode-se fazer outra seleção⁵. A seleção da modalidade de pesquisa afeta todos os grupos de termos de pesquisa.

RESULTADOS DA BUSCA

O resultado da busca no GT é o volume de pesquisa relativo (VPR), dimensionado entre zero a 100. O valor zero indica volumes de buscas muito baixos e o valor de 100 representa o pico de interesse⁸. Esse índice é um valor normalizado para um determinado local e período, permitindo comparações⁸.

Na plataforma on-line do GT, os resultados são apresentados em quatro painéis: *Interesse ao longo do tempo*; *Interesse por sub-região*; *Assuntos relacionados*; e *Pesquisas relacionadas*.

Interesse ao longo do tempo

Esse painel apresenta um gráfico de linhas, em que o eixo vertical é o valor de VPR e o horizontal é o tempo escalonado a depender do período selecionado (pode ser em minutos, horas, dias, semanas e meses).

A Figura 2A reproduz o painel *Interesse ao longo do tempo*, no qual se verifica a diferença entre o interesse de internautas brasileiros e portugueses pelo tópico “Câncer de próstata” no último ano.

Interesse por sub-região

O segundo painel apresenta um gráfico de mapa dividindo a região selecionada em suas sub-regiões, com cores mais escuras onde o VPR é maior. Além do gráfico de mapa, é apresentada uma lista das sub-regiões classificadas de acordo com a popularidade do termo⁸.

A Figura 2B exemplifica o painel *Interesse por sub-região* para uma busca pelo interesse de brasileiros pelo tópico “Câncer cervical” no último ano.

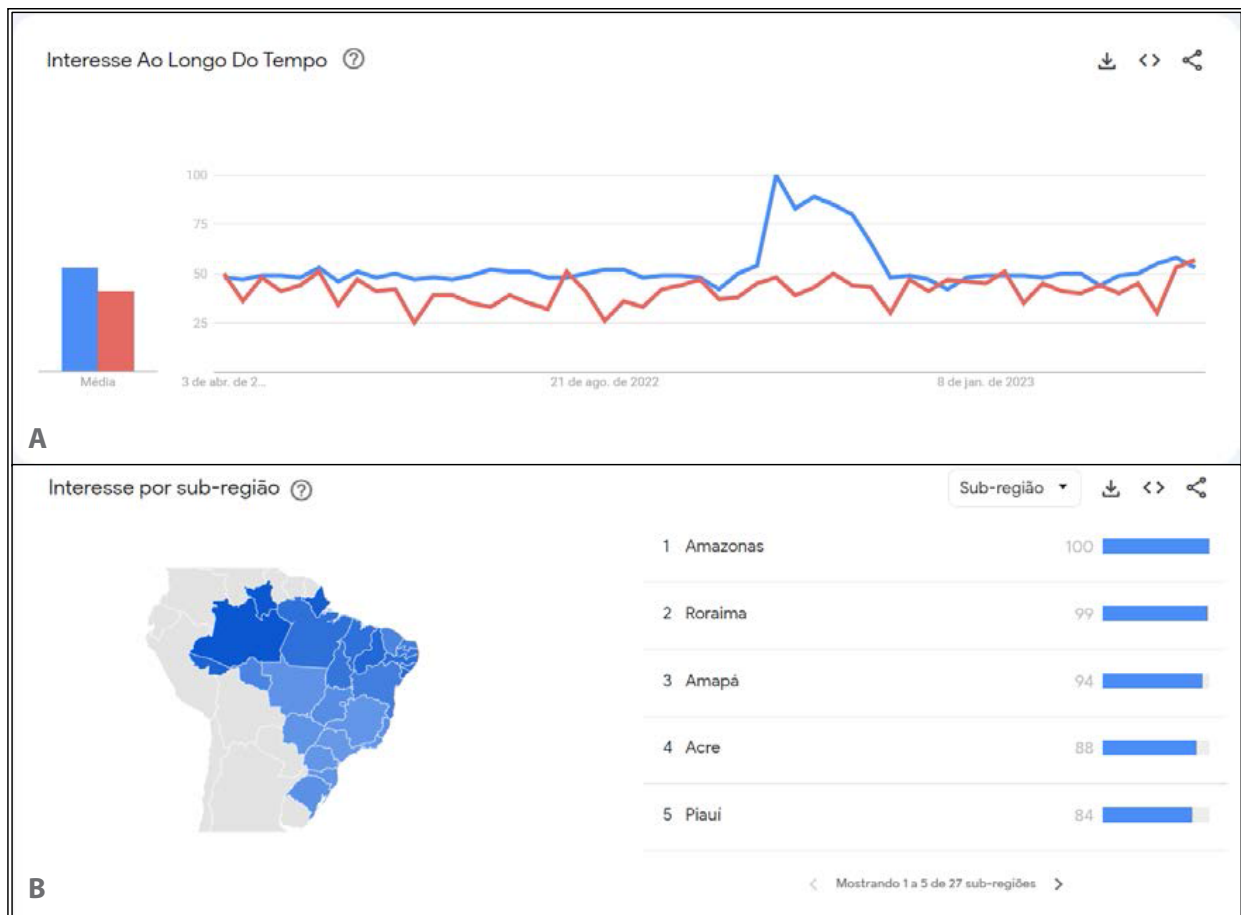


Figura 2. Reprodução dos painéis de *Interesse ao longo do tempo* e *Interesse por sub-região* do Google Trends. Imagem **A**: painel de *Interesse ao longo do tempo* de internautas brasileiros (tom mais escuro) e portugueses (tom mais claro) pelo tópico “câncer de próstata” no último ano. Imagem **B**: painel de *Interesse por sub-região* de internautas brasileiros pelo tópico “câncer cervical” no último ano, no qual as imagens mais escuras representam valores mais altos

Fonte: Google Trends⁸, acesso em 3/4/2023.

Assuntos relacionados

Nesse painel, são apresentados tópicos que também foram pesquisados pelos usuários que buscaram o termo de interesse. É uma importante ferramenta para *insights*. Essa apresentação é feita em duas métricas: principais e em ascensão. Os principais são os assuntos relacionados mais pesquisados para toda série histórica. Os assuntos em ascensão são os que tiveram aumento repentino na frequência. A Figura 3A e a Figura 3B apresentam as duas opções de visualização do painel, respectivamente.

Pesquisas relacionadas

O último painel se assemelha ao anterior na apresentação e função, mas apresenta os termos que também foram pesquisados pelos usuários que buscaram pelo tema de interesse. A Figura 3C (principais) e a Figura 3D (em ascensão) exemplificam as pesquisas relacionadas à busca de brasileiros pelo tópico “Câncer de próstata” nos últimos cinco anos.

Os quatro painéis permitem a exportação dos dados em um arquivo do tipo “csv”, o que possibilita análises

em outros *softwares*. Além disso, o GT também permite a incorporação dos painéis em outros *sites* ou seus compartilhamentos em algumas redes sociais^{7,8}.

Para o monitoramento dos valores do interesse, duas ferramentas podem ser utilizadas. A primeira é pela “inscrição” no próprio GT, selecionando os termos e uma região de interesse. Assim, o usuário recebe atualizações periódicas por e-mail⁵.

A segunda forma é a utilização da *Application Programming Interface* (API) do GT, porém, é necessário ter conhecimento de programação. Um exemplo é a biblioteca *Pytrends*, do *Python*. Apesar de mais complexa, permite maior personalização do monitoramento.

A análise dos dados nas pesquisas com o GT é por meio da abordagem quantitativa. A escolha dos testes estatísticos varia conforme os objetivos do estudo. Em razão das diversas possibilidades e limites de espaço, esse processo não será detalhado, mas sugere-se a revisão realizada por Mavragani et al.⁶.

Quatro limitações para o uso dos dados do GT são comumente debatidas. A primeira é o fato de os

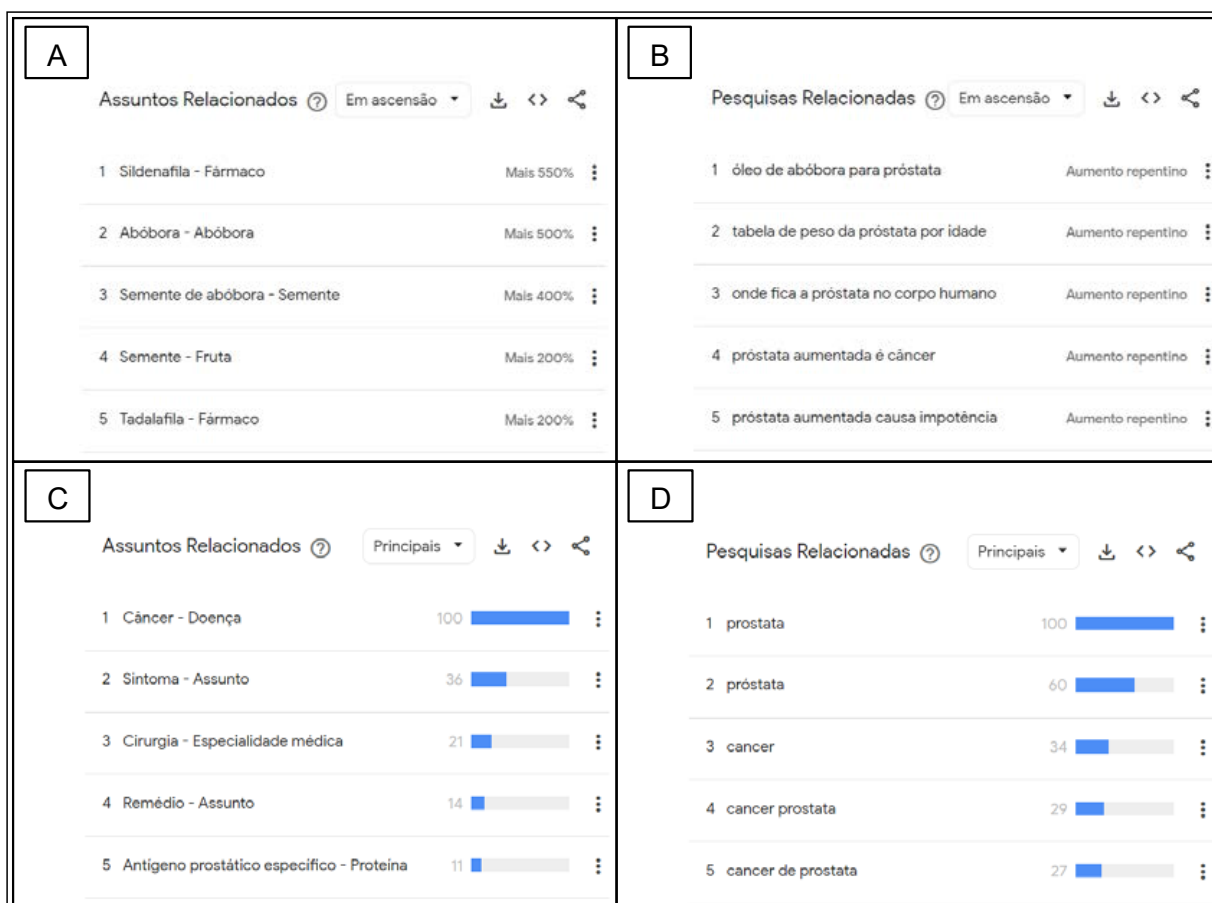


Figura 3. Reprodução dos assuntos e pesquisas relacionados do Google Trends, apresentando a busca pelo interesse de brasileiros pelo tópico “Câncer de próstata” nos últimos cinco anos

Fonte: Google Trends⁸, acesso em 3/4/2023.

resultados apresentados serem uma amostra de tamanho desconhecido. A segunda é que regiões com baixa penetração da Internet ou com menor uso do buscador *Google* podem oferecer dados de interesse limitados. A terceira é a inexistência dos dados absolutos, o que reduz as possibilidades de análise. A quarta ocorre quando há dificuldade de seleção dos termos de busca, o que compromete a exclusão de ruídos^{5,9}.

Apesar desses aspectos, estudos sugerem empiricamente que os resultados são valiosos e válidos na exploração e previsão do comportamento, e estão correlacionados com os dados reais⁵.

CONCLUSÃO

O GT possibilita o acesso de uma amostra das pesquisas realizadas por usuários reais no buscador *Google*. Esses dados fornecem informações valiosas sobre o comportamento relacionado à saúde das populações e as tendências de seus interesses. Os dados brasileiros permitem a comparação entre as diferentes Unidades da Federação e municípios.

Para melhorar a qualidade e reprodutibilidade dos estudos, é fundamental se atentar aos critérios descritos. E ainda detalhar as estratégias de busca, a região e o período de interesse, além das categorias e locais de busca. A data em que a busca foi realizada e os testes estatísticos utilizados também devem ser apresentados.

Considerando que o GT apresenta resultados em tempo real e possibilita o monitoramento permanente, esse portal se configura como uma importante ferramenta de análise para decisões em saúde pública, incluindo o campo da prevenção e controle do câncer no país.

CONTRIBUIÇÃO

Fernando Lopes Tavares de Lima participou de todas as etapas da construção do artigo, desde a sua concepção até a aprovação da versão final a ser publicada.

DECLARAÇÃO DE CONFLITO DE INTERESSES

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. Ministério da Saúde (BR), Gabinete do Ministro. 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) [Internet]. *Diário Oficial da União, Brasília, DF.* 2013 maio 17 [acesso 2023 mar 10]; Seção 1:129. Disponível em: http://bvsmms.saude.gov.br/bvsm/saudelegis/gm/2013/prt0874_16_05_2013.html
3. Assis M. Comunicação em saúde na prevenção e detecção precoce do câncer: em busca de práticas mais dialógicas e inclusivas. *Rev Bras Cancerol.* 2023;69(1):e-032879. doi: <https://doi.org/10.32635/2176-9745.RBC.2023v69n1.2879>
4. Fahy E, Hardikar R, Fox A, et al. Quality of patient health information on the Internet: reviewing a complex and evolving landscape. *Australas Med J.* 2014;7(1):24-8. doi: <https://doi.org/10.4066/AMJ.2014.1900>
5. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: methodology framework. *JMIR Public Health Surveill.* 2019;5(2):e13439. doi: <https://doi.org/10.2196/13439>
6. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in google trends research: systematic review. *J Med Internet Res.* 2018;20(11):e270. doi: <https://doi.org/10.2196/jmir.9366>
7. Nuti SV, Wayda B, Ranasinghe I, et al. The use of Google Trends in health care research: a systematic review. *PLoS One.* 2014;9(10):e109583. doi: <https://doi.org/10.1371/journal.pone.0109583>
8. Google Trends [Internet]. California: Google; c2023 [acesso 2023 fev 28]. Disponível em: <https://trends.google.com/home>
9. Jun S, Yoo HS, Choi S. Ten years of research change using Google Trends: from the perspective of big data utilizations and applications. *Technol Forecast Soc Change.* 2018;130:69-87. doi: <https://doi.org/10.1016/j.techfore.2017.11.009>

Recebido em 13/3/2023
Aprovado em 22/3/2023