

El Uso de *Google Trends* para el Análisis de Interés en Información sobre el Cáncer en el 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
O Uso do *Google Trends* para Análise de Interesse por Informações sobre o Câncer no Brasil: Aspectos Teórico-Metodológicos

Fernando Lopes Tavares de Lima¹

INTRODUCCIÓN

En el Brasil, se estiman 704 000 nuevos casos de cáncer para cada año del trienio 2023-2025¹. Este escenario reafirma al cáncer como un problema creciente de salud pública en el país. La actual Política Nacional para Prevención y Control del Cáncer (PNPCC)² instituye el estímulo para la formulación de estrategias de comunicación con la población, buscando la disseminación y ampliación del conocimiento sobre el cáncer, sus factores de riesgo y las directrices de prevención.

Dicho estímulo debe venir acompañado de acciones prácticas que amplíen el conocimiento de la sociedad sobre cómo prevenir y detectar precozmente el cáncer³. Aunque, ante las dificultades de acceder a estas acciones, Internet ha sido utilizada y, en ese ambiente, el buscador *Google* es un espacio privilegiado, considerado una importante fuente de información en salud⁴. El uso disseminado de esta herramienta ha generado una enorme base de datos acerca del interés de la población sobre cualquier tema, incluyendo el cáncer.

Los investigadores han venido utilizando *Google Trends* (GT) como una herramienta para acceder a estos datos e identificar los intereses poblacionales por informaciones en salud⁵. Así, el comportamiento basado en Internet es analizado para examinar el comportamiento humano real⁶.

A pesar de la importancia de las investigaciones con GT, hay preocupaciones con la robustez metodológica de los artículos. En 2014, se identificó que solo el 7% de los artículos proporcionaron las informaciones necesarias para reproducción de la búsqueda y solamente el 34% documentaron la fecha en que esta búsqueda fue realizada⁷.

Así, ante la importancia de esta herramienta, el objetivo de este artículo es presentar una orientación teórico-metodológica sobre su uso, así como sus indicaciones y limitaciones en el ámbito de la prevención y control del cáncer en el Brasil.

DESARROLLO

GOOGLE TRENDS

GT es un portal on-line gratuito creado en 2006 por *Google Inc.* que presenta la frecuencia relativa con la que un determinado término es buscado en el buscador *Google* en todo el mundo⁸. Así, con base en las consultas reales de los usuarios, GT brinda informaciones sobre lo que fue y lo que es tendencia de intereses de internautas⁶.

En virtud de su procesamiento rápido, los datos están disponibles en tiempo real⁵ y su utilización ha sido asociada a la utilidad para detectar rápidamente un fenómeno, a su representatividad y a su alto potencial de aplicación en las más diversas áreas⁹.

De forma general, las investigaciones que utilizan GT buscan cuatro objetivos: medir el interés sobre un tema; detectar sus variaciones o estacionalidades; analizar las relaciones de esta variación con eventos específicos; y prever eventos relacionados a la salud⁶.

UTILIZANDO GOOGLE TRENDS

Para utilizar GT, es necesario entrar a <https://trends.google.com/> y digitar una estrategia de búsqueda. A partir de la presentación de los resultados, queda disponible el conjunto de herramientas para perfeccionar esta búsqueda: (1) estrategia de búsqueda; (2) área geográfica; (3) período; (4) categoría; y (5) modalidad de búsqueda. Otra forma de acceder a las herramientas es clicar en la pestaña “Explorar”, en la página inicial.

La construcción de una estrategia de búsqueda debe considerar que no hay diferenciación entre letras mayúsculas y minúsculas, pero la hay en relación a los acentos, formas plurales y errores de ortografía⁸. Un ejemplo es la búsqueda por el término “cáncer de próstata” con y sin acento. En el Brasil, a pesar de que la forma correcta de escritura es con acento, hay un mayor volumen de búsquedas por la expresión sin acento (Figura 1A). Esto

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>

Dirección para correspondencia: 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



ejemplifica la importancia de que el investigador analice las diferentes posibilidades de búsqueda en su lengua y construya una estrategia que no pierda datos.

La estrategia de búsqueda puede estar compuesta por términos y expresiones. El uso de comillas en expresiones limita los resultados a esta forma exacta de escritura⁷. La Figura 1B ejemplifica una ligera variación para una búsqueda por “cáncer de próstata” con y sin las comillas.

Al digitar algunos términos en GT, se abren automáticamente opciones de búsqueda por temas específicos. Los temas son un grupo de términos sinónimos ya categorizados en cualquier idioma, lo que aumenta el poder de la búsqueda⁸. Un ejemplo se presenta en la Figura 1C, que compara la búsqueda por “cáncer de próstata” como término de búsqueda y como tema.

Otra estrategia importante es la búsqueda agregada de términos o expresiones. Para esto, se utiliza el recurso “+”. Esto agrega diferentes términos dentro del mismo grupo de estrategia de búsqueda (limitado a cinco grupos simultáneos de búsqueda con 25 términos cada uno). Así, pueden incluirse las variaciones del término, así como errores ortográficos comunes⁵.

Se debe delimitar una Región Geográfica de cobertura. Se puede seleccionar todo el mundo o un país específico.

Además, algunos países poseen datos por subregiones^{5,7}, como las Unidades Federales en el Brasil.

La Figura 1D ejemplifica la comparación de dos regiones, São Paulo y Minas Gerais, para el tema “Cáncer de próstata” en el último año. El valor mayor en Minas Gerais no significa un valor absoluto mayor, pues la medida es relativa⁸.

GT ofrece dos modalidades de datos. La primera es clasificada como “tiempo real” y corresponde a los últimos siete días. La segunda es una muestra separada de la primera que va desde 2004 hasta 72 horas antes de la búsqueda⁸.

El investigador puede elegir un período ya configurado o personalizarlo. Independientemente de la elección, es importante notar que los datos son normalizados para el período seleccionado. Así, la búsqueda debe estar compuesta exactamente por el período de interés⁷.

La búsqueda puede realizarse con base en una categoría. Este recurso es importante para eliminar ruidos de datos, principalmente en los casos en los que la palabra tiene diferentes significados. La palabra “cáncer”, por ejemplo, puede significar una enfermedad o el signo del zodiaco. La selección de la categoría “salud” puede delimitar los resultados (efecto parecido con la elección de búsqueda por tema). Hay que destacar que esto afecta a todos los grupos de términos de búsqueda.

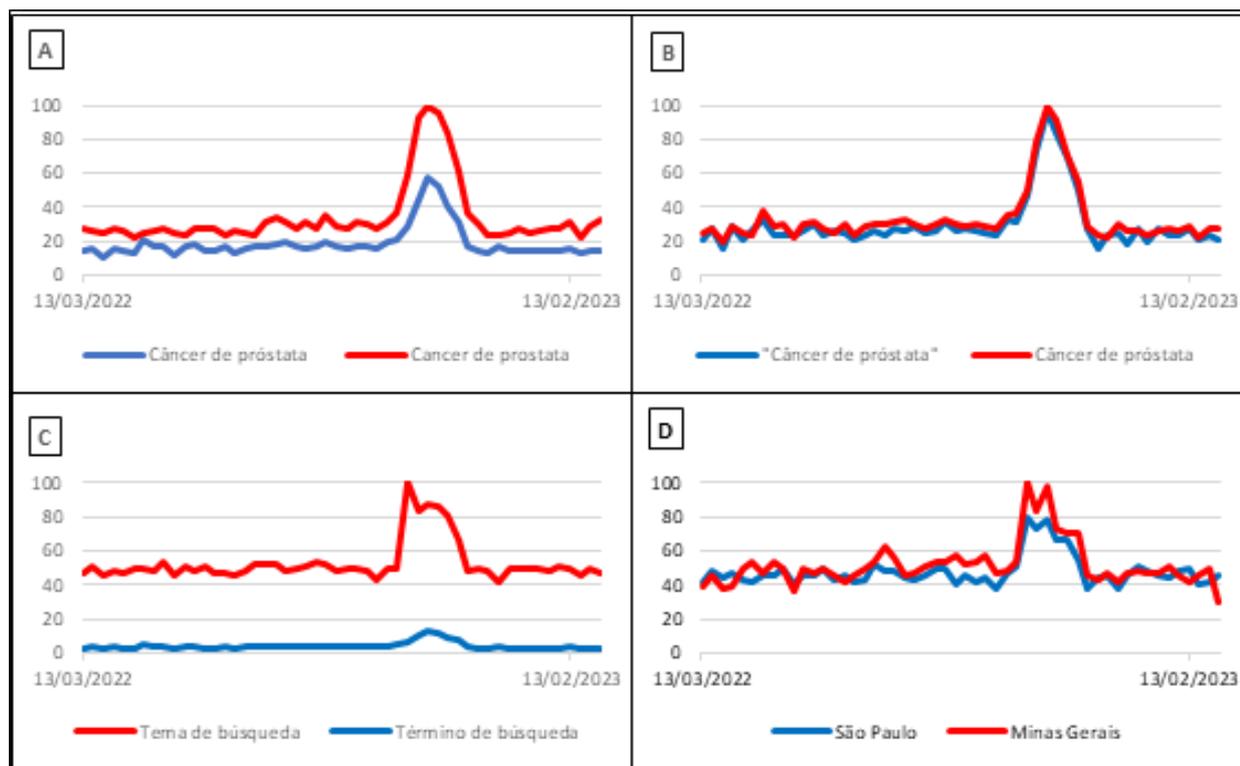


Figura 1. Ejemplos del resultado del uso de diferentes herramientas para construcción de la estrategia de búsqueda en Google Trends. Imagen A: diferencia del uso del acento en la expresión “cáncer de próstata” en el Brasil. Imagen B: diferencia del uso de comillas en la expresión “cáncer de próstata” en el Brasil. Imagen C: diferencia de la selección de tema o término de investigación para la búsqueda de la expresión “cáncer de próstata”. Imagen D: diferencia del volumen de búsqueda por “cáncer de próstata” en dos Unidades de la Federación en el Brasil

Fuente: Google Trends⁸, visitado en 3/4/2023.

Como modalidad de búsqueda, GT ofrece las opciones: “Búsqueda Web”, “Búsqueda de Imágenes”, “Búsqueda de Noticias”, “Google Shopping” y “Búsqueda de YouTube”. La opción “Búsqueda Web” se refiere al buscador *Google* y es el objeto de este artículo. Aunque, dependiendo del objetivo de la búsqueda, se puede hacer otra selección⁵. La selección de la modalidad de búsqueda afecta a todos los grupos de términos de búsqueda.

RESULTADOS DE LA BÚSQUEDA

El resultado de la búsqueda en GT es el volumen de búsqueda relativo (VBR), dimensionado desde cero a cien. El valor cero indica volúmenes de búsquedas muy bajos y el valor cien representa el pico de interés⁸. Este índice es un valor normalizado para un determinado lugar y período, permitiendo comparaciones⁸.

En la plataforma on-line de GT, los resultados se presentan en cuatro paneles: *Interés a lo largo del tiempo*; *Interés por subregión*; *Temas relacionados*; y *Búsquedas relacionadas*.

Interés a lo largo del tiempo

Este panel presenta un gráfico de líneas, en el que el eje vertical es el valor de VBR y el horizontal es el tiempo escalonado dependiendo del período seleccionado (puede ser en minutos, horas, días, semanas y meses).

La Figura 2A reproduce el panel *Interés a lo largo del tiempo*, en el cual se verifica la diferencia entre el interés de los internautas brasileños y portugueses por el tema “Cáncer de próstata” en el último año.

Interés por subregión

El segundo panel presenta un gráfico de mapa dividiendo la región seleccionada en sus subregiones, con colores más oscuros donde el VBR es mayor. Además del gráfico de mapa, se presenta una lista de las subregiones clasificadas de acuerdo con la popularidad del término⁸.

La Figura 2B ejemplifica el panel *Interés por subregión* para una búsqueda por el interés de brasileños por el tema “Cáncer cervical” en el último año.

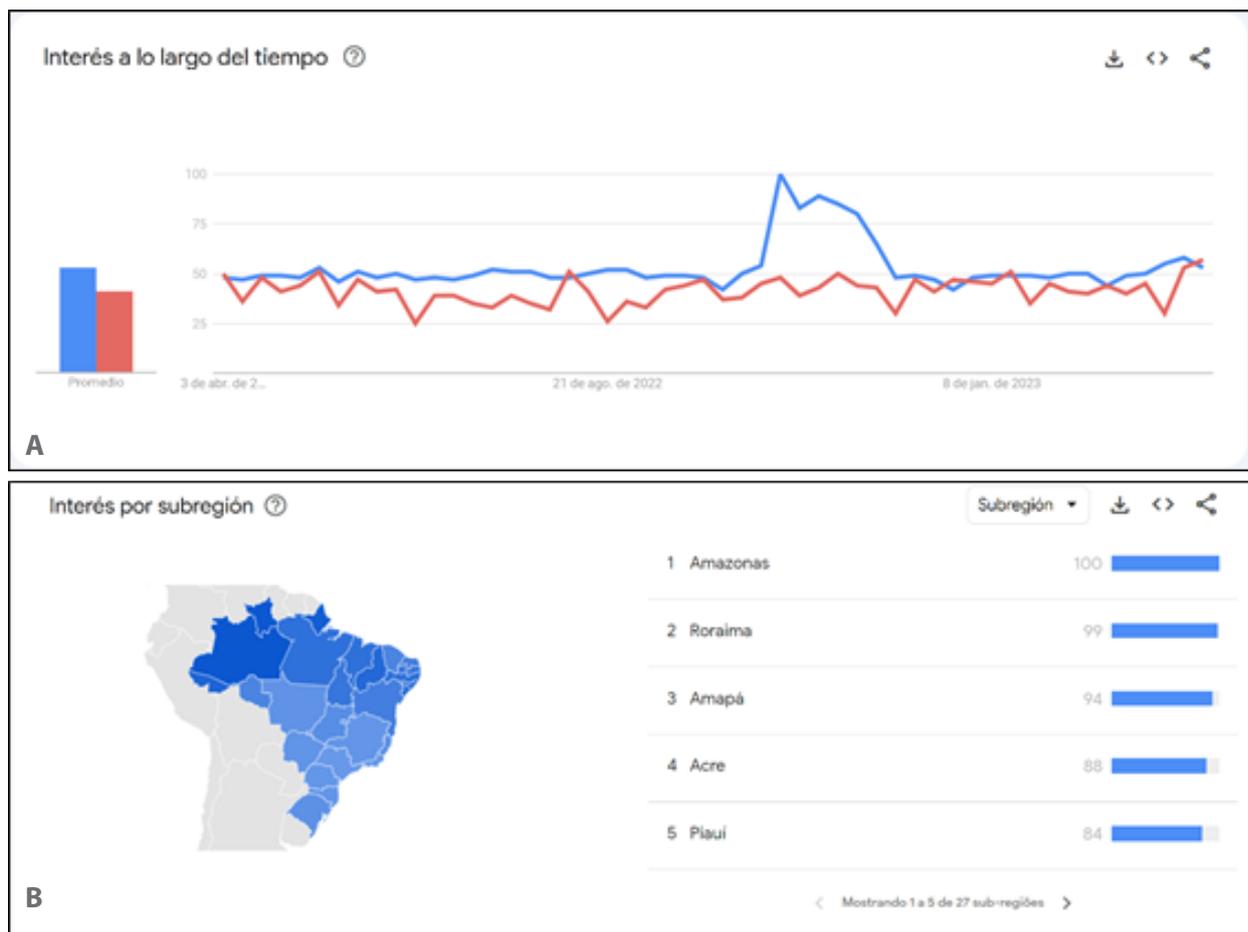


Figura 2. Reproducción de los paneles de *Interés a lo largo del tiempo* e *Interés por subregión* de Google Trends. Imagen **A**: panel de *Interés a lo largo del tiempo* de internautas brasileños (tono más oscuro) y portugueses (tono más claro) por el tema “cáncer de próstata” en el último año. Imagen **B**: panel de *Interés por subregión* de internautas brasileños por el tema “cáncer cervical” en el último año, en el cual las imágenes más oscuras representan valores más altos

Fuente: Google Trends⁸, visitado en 3/4/2023.

Temas relacionados

En este panel, se presentan temas que fueron buscados también por los usuarios que indagaron el término de interés. Es una importante herramienta para *insights*. Esta presentación está compuesta de dos métricas: principales y en aumento. Los principales son los temas relacionados más buscados para toda serie histórica. Los temas en aumento son los que tuvieron una subida repentina en la frecuencia. La Figura 3A y la Figura 3B presentan las dos opciones de visualización del panel, respectivamente.

Búsquedas relacionadas

El último panel se asemeja al anterior en la presentación y función, pero presenta los términos que también fueron buscados por los usuarios que procuraron por el tema de interés. La Figura 3C (principales) y la Figura 3D (en aumento) ejemplifican las búsquedas relacionadas a la indagación de los brasileños por el tema “Cáncer de próstata” en los últimos cinco años.

Los cuatro paneles permiten la exportación de los datos en un archivo del tipo “csv”, lo que posibilita su análisis en otros *softwares*. Además, GT también permite

la incorporación de los paneles en otros *sites* o compartirlos en algunas redes sociales^{7,8}.

Para el monitoreo de los valores de interés, pueden utilizarse dos herramientas. La primera es mediante la “inscripción” en el propio GT, seleccionando los términos y una región de interés. Así, el usuario recibe actualizaciones periódicas por e-mail⁵.

La segunda forma es la utilización de la *Application Programming Interface* (API) de GT, aunque es necesario tener conocimientos de programación. Un ejemplo es la biblioteca *Pytrends*, de *Python*. A pesar de ser más compleja, permite una mayor personalización del monitoreo.

El análisis de los datos en las búsquedas con GT se hace mediante el enfoque cuantitativo. La elección de las pruebas estadísticas varía según los objetivos del estudio. En razón de las diversas posibilidades y límites de espacio, este proceso no se detallará, pero se sugiere la revisión realizada por Mavragani et al.⁶.

Se debaten comúnmente cuatro limitaciones para el uso de los datos de GT. La primera es el hecho de que los resultados presentados sean una muestra de tamaño desconocido. La segunda es que las regiones con baja

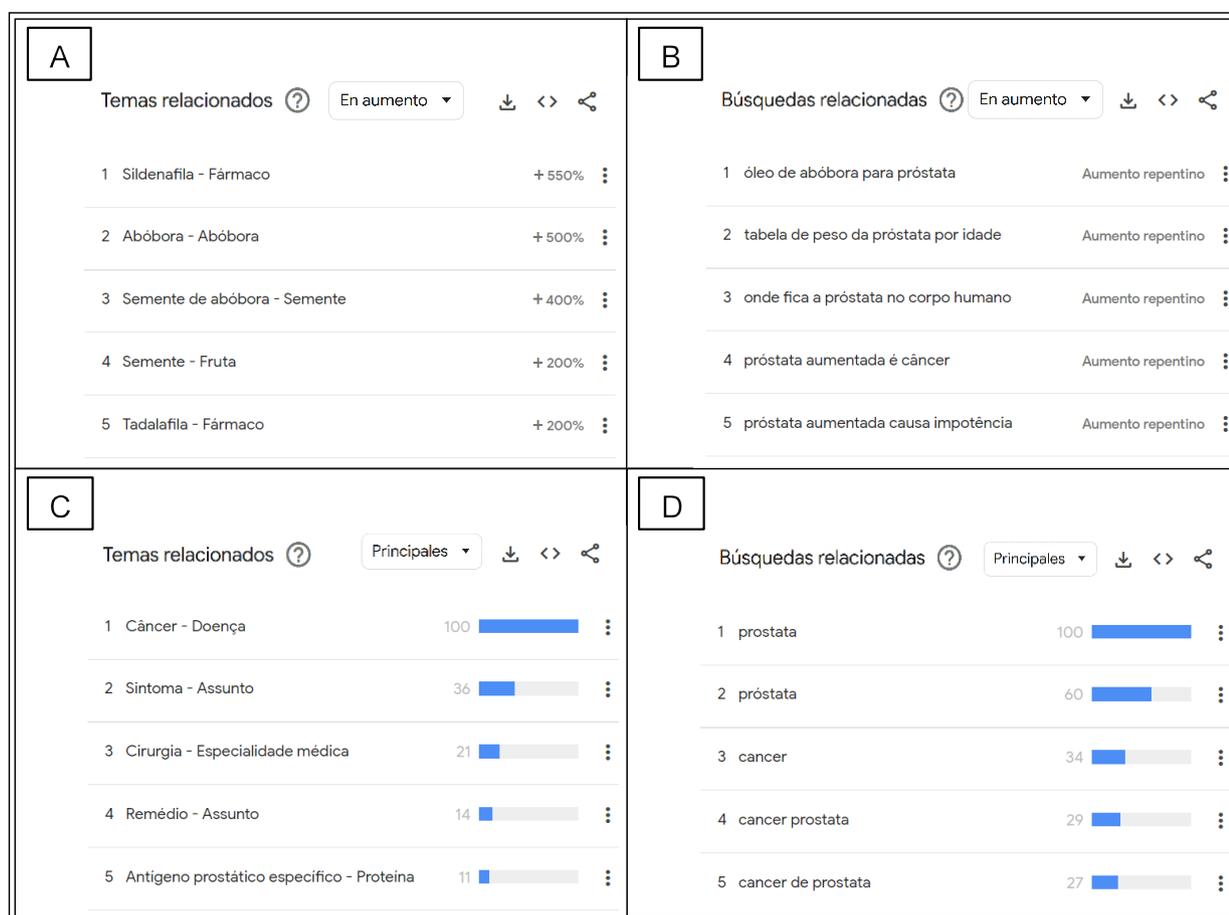


Figura 3. Reproducción de los temas y búsquedas relacionados de Google Trends, presentando la búsqueda por el interés de brasileños por el tema “Cáncer de próstata” en los últimos cinco años

Fuente: Google Trends®, visitado en 3/4/2023.

penetración de Internet o con menor uso del buscador *Google* pueden ofrecer datos limitados de interés. La tercera es la inexistencia de los datos absolutos, lo que reduce las posibilidades de análisis. La cuarta sucede cuando existe dificultad para la selección de los términos de búsqueda, lo que compromete la exclusión de ruidos^{5,9}.

A pesar de estos aspectos, los estudios sugieren empíricamente que los resultados son valiosos y válidos en la exploración y previsión del comportamiento, y están correlacionados con los datos reales⁵.

CONCLUSIÓN

GT posibilita el acceso de una muestra de las búsquedas realizadas por usuarios reales en el buscador *Google*. Estos datos proporcionan informaciones valiosas sobre el comportamiento relacionado a la salud de las poblaciones y las tendencias de sus intereses. Los datos brasileños permiten la comparación entre las diferentes Unidades de la Federación y municipios.

Para mejorar la calidad y reproducibilidad de los estudios, es fundamental atenerse a los criterios descritos. Y además detallar las estrategias de búsqueda, la región y el período de interés, así como las categorías y lugares de búsqueda. La fecha en que la búsqueda fue realizada y las pruebas estadísticas utilizadas también deben presentarse.

Considerando que GT presenta resultados en tiempo real y posibilita el monitoreo permanente, ese portal se configura como una importante herramienta de análisis para tomar decisiones en salud pública, incluyendo el campo de la prevención y control del cáncer en el país.

APORTES

Fernando Lopes Tavares de Lima participó de todas las etapas de la construcción del artículo, desde su concepción hasta la aprobación de la versión final a publicarse.

DECLARACIÓN DE CONFLICTO DE INTERESES

Nada a declarar.

FUENTES DE FINANCIAMIENTO

No hay.

REFERENCIAS

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