

# Relationship between searching for breast cancer on the web and number of cases of breast cancer

Relação entre pesquisa de câncer de mama na web e o número de casos de câncer de mama: o papel da idade no paciente oncológico

Caroline Tanure Castro<sup>1</sup>, Jessica Barbosa Santos<sup>1</sup>, Davi Neri Araujo<sup>2</sup>, Fernanda Freitas Lemos Lopes<sup>1,3</sup>, Vinicius Lima<sup>1</sup>, Tiago Cavalcante<sup>1</sup>, Maria Laura da-Silva<sup>1,3</sup>, Luiz Fernando Quintanilha<sup>1</sup>, Artur Trancoso Lopo Queiroz<sup>3</sup>, Kiyoshi Ferreira Fukutani<sup>1,3</sup>

## ABSTRACT

**Introduction:** Breast cancer is the second most frequent cancer in the world and the first among women. Online searches are used to acquire information about symptoms and treatment. Therefore, such searches can be used in trend analysis, helping to predict outbreaks. **Objectives:** Describe/compare the epidemiologic patients' profile with benign and malignant breast cancer and investigate the relationship between the number of cases of benign and malignant breast cancer with internet searches for "breast cancer", from 2009 to 2013, in different age range. **Methods:** This epidemiological study used Google Trends and DATASUS (SISMAMA) data. The data were standardized in *Microsoft Excel* and for statistical analysis, we used the chi-square test and Pearson's correlation test. **Results:** Our sample is characterized by the predominance of patients in the age group between 35 and 54 years (73.91% and 51.55% for the benign and malignant group, respectively), with an increase in the frequency of older patients in the malignant group ( $p < 0.0001$ ). No statistically significant variation related to patients' education and race was observed. A negative correlation ( $r = -0.94$ ,  $p = 0.014$ ) was observed between the number of cases of malignant neoplasm and the search for breast cancer on the web, and a positive correlation involving benign neoplasia and in different age groups. **Conclusion:** The age proportion are different between patients with malignant and benign cancer. A negative correlation between patients with malignant breast cancer and searches on *Google Trends*, and a positive correlation in different age groups in benign cancer.

**Keywords:** Breast neoplasms; Internet; Epidemiology.

1. Centro Universitario UniFTEC, Faculdade de Medicina - Salvador - Bahia - Brazil
2. Universidade Federal da Bahia, Faculdade de Medicina - Salvador - Bahia - Brazil.
3. Fundação Oswaldo Cruz, Instituto Gonçalo Moniz - Salvador - Bahia - Brazil.

**Funding source:** None.

**Conflicts of interest:** The authors declare no conflict of interest relevant to this manuscript.

**Clinical Trials:** Not Applicable

**Correspondence author:** Kiyoshi Ferreira Fukutani. Rua Waldemar Falcão, 121, Candeal - Salvador/BA, CEP: 40296-710 Tel. (71) 3176-2234  
E-mail: ferreirafk@gmail.com

Received on: December, 2020 | Accepted on: February 2, 2021  
DOI: <https://doi.org/10.5935/2526-8732.20210011>



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. (<http://creativecommons.org/licenses/by-nc-sa/4.0/>).

## RESUMO

**Introdução:** A neoplasia de mama é o segundo câncer mais frequente no mundo e o primeiro entre as mulheres. As buscas online são utilizadas para adquirir informações acerca de sintomas e tratamento. Tais buscas podem ser utilizadas em análises de tendências, auxiliando na predição de surtos. **Objetivos:** Descrever/comparar o perfil epidemiológico e investigar a correlação entre os indivíduos com câncer de mama benigno e maligno com as pesquisas na internet por “câncer de mama”, no período de 2009 a 2013, em diferentes faixas etárias. **Métodos:** Trata-se de um estudo epidemiológico que utilizou dados secundários do DATASUS (SISMAMA) e do *Google Trends*. Os dados foram padronizados em *Microsoft Excel* e para a análise estatística utilizamos o teste qui-quadrado e o teste de correlação de Pearson. **Resultados:** A nossa amostragem é caracterizada pelo predomínio de indivíduos na faixa de idades entre 35 a 54 anos (73.91% e 51.55% para o grupo benigno e maligno, respectivamente), com aumento da frequência de indivíduos mais velhos no grupo maligno ( $p < 0.0001$ ). Não encontramos variações estatisticamente significantes quanto à escolaridade e quanto à raça dos indivíduos. Houve correlação negativa ( $r = -0.94$ ,  $p = 0.014$ ) quando relacionamos o número de casos de neoplasia maligna com a busca por câncer de mama na web e positiva envolvendo a neoplasia benigna e em diferentes faixas etárias. **Conclusão:** As proporções das faixas etárias são diferentes entre indivíduos com câncer maligno e benigno e existe uma correlação negativa entre os indivíduos com câncer de mama maligno e as buscas no *Google Trends*, enquanto há uma correlação positiva nas diferentes faixas etárias no câncer benigno.

**Descritores:** Neoplasias da mama; Internet; Epidemiologia.

## INTRODUCTION

Breast neoplasm is a disease caused by the disordered multiplication of breast cells and it is the second most common neoplasm in the world among the female population.<sup>[1]</sup> In Brazil, around 66,280 new cases of breast cancer every year in the 2020-2022 triennium<sup>[2]</sup> and the percentage of new cases can reach 29% depending on the Brazilian region.<sup>[3]</sup> Breast neoplasm occupies the first position in frequency in all regions with an estimated risk of 81.06 per 100 thousand in the Southeast Region; 71.16 per 100 thousand in the South Region; 45.24 per 100 thousand in the Midwest Region; 44.29 per 100 thousand in the Northeast Region; and 21.34 per 100 thousand in the North Region.<sup>[2]</sup> However, the population and epidemiological information provided by the Ministries of Health, Science, Technology, Innovations and Communications (MCTIC) are fundamental in the development of public strategies to combat and prevent this disease.<sup>[4]</sup>

The data collection in Brazil started a new era with the SUS Informatics Department (DATASUS), which contains information such as: health indicators, epidemiology, morbidity, health care, demographic, and socioeconomic information that became widely used to support population surveys in Brazil.<sup>[5]</sup> Furthermore, other internet platform has been considered a source of useful information for the scientific community and, in 2012, *Google Launches*, a tool able to analyzes the internet search, that

platform was called *Google Trends*, that is free and accessible.<sup>[6,7]</sup> The *Google Trends* has been used to investigate epidemiological trends worldwide, generating several scientific works that try to relate searches with the occurrences of cases.<sup>[8,9,10,11]</sup>

This research has increasingly and incorporated new technologies able to predict how the internet searches are related to epidemiological data, however, little research has been found involving breast cancer, especially in Brazil. To investigate this, the present work describes and compare the number of cases of benign and malignant breast neoplasm registered in the cancer information system in Brazil and correlate it with the number of breast cancer surveys in the internet from 2009 to 2013.

## METHODS

This is an epidemiological study that used secondary data from DATASUS (<http://www.datasus.gov.br>) and *Google Trends* (<https://trends.google.com/trends>). In both databases, the information collected was restricted to the period from 2009 to 2013, accessed on 1/26/2021. The data obtained from SISMAMA-DATASUS (Cancer Information System and Breast Cancer Information System) were filtered for women who have the positive diagnostic examination of breast neoplasm among all age groups. The information obtained from DATASUS were organized in databases to add the information about the internet searches for “breast cancer” (câncer de

mama - in Portuguese) on Google Trends. The search criteria of Google Trends were: the keyword "breast cancer", being filtered by searches from state of Bahia. All data were hosted and standardized in Microsoft Excel spreadsheets. For the description of the data, descriptive statistics were used (measures of central tendency and measures of dispersion) and for analytical statistics, the chi-square test was used to assess the difference in proportions between the population with malignant and benign breast neoplasm and we also used the Pearson correlation test between the number and number of cases from DATASUS and internet searches for "breast cancer". All tests were considered statistically significant for  $p \leq 5$ . The present study followed the ethical norms of Resolution N. 466/12. Because we used a secondary data in the public domain and without identification, there is no need to apply the informed consent form.

## RESULTS

After obtaining the data, individuals with malignant and benign breast cancer present in DATASUS were used for the analysis of the epidemiological profile. Our sample is characterized by the predominance of individuals in the age group that comprises the ages of 35 to 54 years, with the accumulated percentage of 73.91% and 51.55% for the benign and malignant group, respectively. The frequency of malignant manifestation of breast cancer was higher among older women ( $p < 0.0001$ ). In regard of schooling, the highest percentage of females is found in the incomplete elementary school profile, corresponding to 3.81% of cases of malignant neoplasm and 2.33% in cases of benign neoplasm. Interestingly, the high proportion of ignored or blank values, which is equally distributed among the groups ( $p = 0.073$ ). Additionally, concerning to skin color, there was a predominance of brown color with a frequency of 14.18%-11.63% in women with malignant breast neoplasm and women with benign neoplasm, respectively, also with high proportions of missing data ( $p = 0.2291$ ) (Table 1).

After describing the female population with breast neoplasm, we investigated the relationship between Google's search for breast cancer and the number of cases for either malignant or benign neoplasm. Initially, we related the number of cases of malignant breast cancer to the web searches for breast cancer, and found a strong negative correlation ( $r = -0.94$ ;  $p = 0.014$ ) (Figure 1A). When we used cases of benign neoplasm and correlated with breast cancer web searches, we found a positive correlation ( $r = 0.53$ ), but not significant ( $p > 0.05$ ) (Figure 1B). Previously, we analyzed the difference between frequency of breast cancer cases among the different age groups in patients with malignant breast cancer and benign cancer (Table 1:  $p < 0.0001$ ). This evidence made us stratify these women with breast cancer according to age. As a result, the searches for breast cancer on the web were negatively correlated with the number of cases of malignancy in almost all age groups, but

these values were not statistically significant ( $p > 0.05$ ). However, positive correlations were found in women with benign breast cancer in the age groups included in the following class ranges: 25 to 29, 30 to 34, 40 to 44, 45 to 49, and 50 to 54, all significant ( $p < 0.05$ ) (Figure 2).

## DISCUSSION

The women population with breast neoplasm presented in DATASUS revealed that, the cases of malignant neoplasm are more frequent in the age groups between 25 and 49 years compared to cases of benign neoplasm. Other factors such as education and race did not show statistically significant differences in our sample. It is known that breast neoplasms affect young women aged 40 to 50 years old, which corresponds to approximately 34.5% to 29.3% of cases, respectively.<sup>[1]</sup> The early identification of breast cancer can be associated with a worse prognosis, since the diagnosis is made with the disease at a detectable stage.<sup>[12]</sup> Despite this, it is important to recognize that screening for breast neoplasm in Brazil is getting earlier every day and we observe that the age group comprising the ages: 35 to 39, 40 and 59, 60 and 69, and above 70 years old, had respectively 11.75; 41.19; 5.61 and 2.34%, in women with malignant breast neoplasm, while in the United States, the probability of developing breast cancer at older ages is about 12 times higher compared to less advanced age.<sup>[13]</sup>

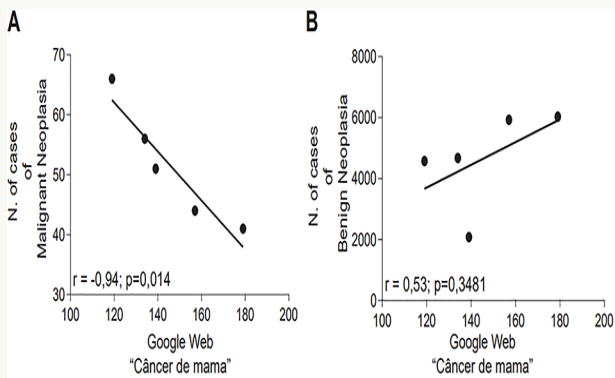
Other studies associate the advanced age and lower education level with an increased risk of cancer.<sup>[14]</sup> We were unable to find an association between educational level and ethnicity in breast cancer cases in Bahia, and it's also in our knowledge that are numerous of other risk factors associated with breast cancer, such as: diet, obesity, physical inactivity, cigarette use and alcohol.<sup>[15]</sup> Such factors were not included in DATASUS, because they were not addressed in this study. Google become a good source of current information, therefore we decided to investigate its role in the casuistry of women with breast cancer in the state of Bahia.

The strong negative correlation found between individuals diagnosed with malignant breast cancer and breast cancer searches on the web, reflects the cancer social stigma which the individuals affected despair, with an imbalance on the emotional and social side, causing loss of their life goals and their social role.<sup>[16]</sup> The decreasing in the desire to use social media, make interactions between people and interactions with the digital media is a behavior associated with depression that can affect about 30% of the population of women with breast cancer around the world.<sup>[17]</sup>

When stratifying the correlation by age, we found a positive correlation between cases of benign breast cancer and web searches in the 25 to 34 age group. This demonstrates that individuals with benign neoplasm at younger ages look for more information

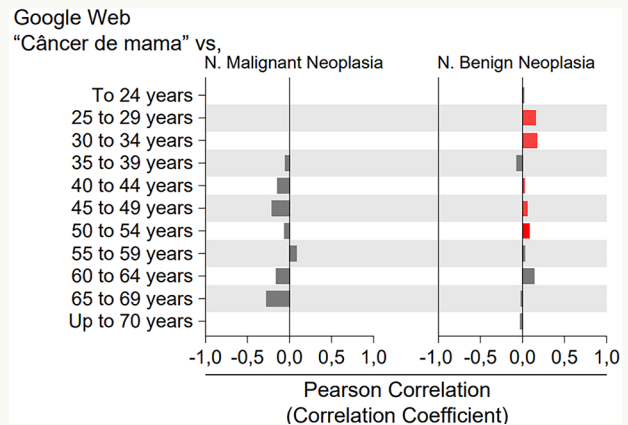
**Table 1.** Characterization and comparison of the female population with breast neoplasm in the years 2009 to 2013.

Age n (%)	Malignant neoplasm	Benign neoplasm	p-value
To 24 years	1354 (5,88%)	0 (0%)	<0,0001
25 to 29 years	1401 (6,08%)	2 (0,78%)	
30 to 34 years	1894 (8,22%)	8 (3,1%)	
35 to 39 years	2708 (11,75%)	20 (7,75%)	
40 to 44 years	4636 (20,12%)	41 (15,89%)	
45 to 49 years	4855 (21,07%)	39 (15,12%)	
50 to 54 years	2937 (12,75%)	33 (12,79%)	
55 to 59 years	1426 (6,19%)	38 (14,73%)	
60 to 64 years	821 (3,56%)	21 (8,14%)	
65 to 69 years	472 (2,05%)	16 (6,2%)	
Upto 70 years	540 (2,34%)	40 (15,5%)	
Education n (%)			
Ignored	20655 (89,63%)	240 (93,02%)	0,073
Illiterate	53 (0,23%)	0 (0%)	
Incomplete Elementary School	877 (3,81%)	6 (2,33%)	
Complete Elementary School	778 (3,38%)	3 (1,16%)	
Complete High School	592 (2,57%)	6 (2,33%)	
Complete College	89 (0,39%)	3 (1,16%)	
Skin color n (%)			
White	126 (0,55%)	4 (1,55%)	0,2991
Black	141 (0,61%)	2 (0,78%)	
Brown	3267 (14,18%)	30 (11,63%)	
Yellow	9 (0,04%)	0 (0%)	
Indigenous	2 (0,01%)	0 (0%)	
Ignored	19499 (84,62%)	222 (86,05%)	



**Figure 1.** Pearson's correlation between. A. number of cases of malignant neoplasms and search for breast cancer on Google; B. Number of cases of benign neoplasms and search for breast cancer on Google. Enter in the years 2009 and 2013, using data from DATASUS and Google Trends.

on the internet regarding their problem. Kübler-Ross (1977),<sup>[6]</sup> defined the five stages of grief as: denial, anger, bargaining, depression, and acceptance; and further studies have shown that the individual's age is directly related to the acceptance level.<sup>[18,19]</sup> The fact is that death is more expected and accepted in the elderly population, while the awareness of their illness being fatal makes young adults look for other methods, manifesting the denial stage.<sup>[18,19,20]</sup> This finding is justified by the access to *Google Trends*.



**Figure 2.** Pearson correlation in different age between number of cases of malignant and benign neoplasms and breast cancer research on Google Web. Between 2009 and 2013, using data from DATASUS and Google Trends. The red bar represents statistical significance with a value of  $p < 0,05$ , using Pearson's correlation test.

However, one of the limitations of our work, besides the large amount of ignored data, is the restriction of a segment of the population with access to the Internet. After all, the majority of the population with access is composed of young people who have higher socioeconomic status.<sup>[21]</sup> However, with the increasing use of *smartphones* regardless of socioeconomic status, the level of access to

information has become a routine and over the years this device has become the number one option for accessing the internet not only among young people.<sup>[22,23]</sup> Other limitation is the small coverage of DATASUS, since the SISMAMA database does not present updated data, which makes the temporal pattern of our search limited to 5 years (is not possible to investigate the current days); the lack of use of the male population in the present study due to its low casuistry, this made it impossible for us to draw conclusions regarding this specific group. Although, the population with breast cancer is predominantly female and internet searches do not provide the user gender, there is a redundancy of searches for partners and family members who live with such a disease regardless of sex and the total number of searches is corrected by *Google Trends* presenting normalized values.<sup>[7]</sup>

Despite the limitations proposed here and the fragility in some aspects of the study, when relating the number of cases of malignant and benign neoplasms in different age groups, with the searching for information on the internet is capable to identify different behavior pattern that must be decisive to handle the mental of the patients and the artificial intelligence could be an important tool to compose new strategies for Brazilian public health recognizing this nuance.

## CONCLUSION

The DATASUS analysis showed that women with malignant breast neoplasm are older than women with benign breast neoplasm. Factors such as race and education were not a major factor in our sample. When we evaluate DATASUS with *Google Trends*, the cases of individuals with malignant breast neoplasm correlate negatively with the increase in searches on *Google Trends*. In addition, the increase in the number of benign breast neoplasm cases is correlated with the increase in breast cancer web searches in a directly proportional way in women in different age groups. Finally, despite these results, more research is required to better understand the relationship between individuals and the search profile on the web.

## REFERENCES

1. Augusto MN, Nucci LB. O uso de dados públicos de indicadores de saúde em artigos científicos [dissertação]. Campinas (SP): Pontifícia Universidade Católica (PUC); 2015.
2. Bakker KM, Martinez-Bakker ME, Helm B, Stevenson TJ. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. *Proc Natl Acad Sci U S A*. 2016 Jun;113(24):6689-94.
3. Barbosa LNF, Francisco AL. A subjetividade do câncer na cultura: implicações na clínica contemporânea. *Rev SBPH*. 2007;10(1):9-24.
4. Bloom R, Amber KT, Hu S, Kirsner R. Google search trends and skin cancer: evaluating de US population's interest in skin cancer and its association with melanoma outcomes. *JAMA Dermatol*. 2015 Aug;151(8):903-5.
5. Borges ADVS, Silva EF, Mazer SM, Toniollo PB, Valle ERM, Santos MA. Percepção da morte pelo paciente oncológico ao longo do desenvolvimento. *Psicol Estud*. 2006;11(2):361-9.
6. Crippa CG, Hallal ALC, Dellagiustina AR, Traebert EE, Gondin G, Pereira C. Perfil clínico e epidemiológico do câncer de mama em mulheres jovens. *Arq Catarin Med*. 2003 Jul/Sep;32(3):50-8.
7. Fenerick GMP. A utilização de smartphones no acesso à informação científica por jovens estudantes: um estudo de caso [dissertação]. São Carlos (SP): Universidade Federal de São Carlos (UFSCAR); 2017.
8. Ministério da Saúde (BR). Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Síntese de resultados e comentários [Internet]. Rio de Janeiro (RJ): Ministério da Saúde/INCA; 2020; [access in 2021 Jan 20]. Disponível em: <https://www.inca.gov.br/estimativa/sintese-de-resultados-e-comentarios>
9. Ministério da Saúde (BR). Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Câncer de mama: é preciso falar disso. 4a ed. Rio de Janeiro (RJ): Ministério da Saúde/INCA; 2016.
10. Kovacs MJ. Morte e desenvolvimento humano. São Paulo (SP): Casa do Psicólogo; 1992.
11. Kübler-Ross E. Sobre a morte e o morrer. 3a ed. São Paulo (SP): Editora Martins Fontes; 1977.
12. Lima AC, Januário MC, Lima PT, Moura e Silva W. DATASUS: o uso dos sistemas de informação na saúde pública. *REFAS*. 2014 Out;1(3):16-31.
13. Matos JC, Pelloso SM, Carvalho MDB. Prevalence of risk factors for breast neoplasm in the city of Maringá, Paraná state, Brazil. *Rev Latino-Am Enfermagem*. 2010 Jun;18(3):352-9.
14. Melo WA, Souza LAO, Zurita RCM, Carvalho MDB. Fatores associados na mortalidade por câncer de mama no noroeste paranaense. *Rev Eletr Gest Saúde*. 2013;1(1):1809-16.
15. Morsy S, Dang TN, Kamel MG, Zyan AH, Makram OM, Elhady M, et al. Prediction of Zika-confirmed cases in Brazil and Colombia using Google Trends. *Epidemiol Infect*. 2018 Oct;146(13):1625-7.
16. Bottino MB, Fráguas R, Gattaz F. Depressão e câncer. *Rev Psiquiatr Clín*. 2009;36(Suppl 3):109-15.
17. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of Google Trends in health care research: a systematic review. *PLoS One*. 2014 Oct;9(10):e109583.
18. Quintanilha LF, Souza LN, Sanches D, Demarco RS, Fukutani KF. The impact of cancer campaigns in Brazil: a google trends analysis. *Ecancermedicalscience*. 2019;13:963.

19. Seifter A, Schwarzwald A, Geis K, Aucott J. The utility of "Google Trends" for epidemiological research: Lyme disease as an example. *Geospatial Health*. 2010 May;4(2):135-7.
20. Singh R, Hellman S, Heimann R. The natural history of breast carcinoma in the elderly. *Cancer*. 2004 May;100(9):1807-13.
21. Tomazelli JG, Silva GA. Rastreamento do câncer de mama no Brasil: uma avaliação da oferta e utilização da rede assistencial do Sistema Único de Saúde no período 2010-2012. *Epidemiol Serv Saúde*. 2017;26(4):713-24.
22. Tondo R. Smartphones e pobreza digital: o consumo de telefones celulares e internet por jovens de camada popular. In: 3º Congresso Internacional de Direito e Contemporaneidade: Mídias e Direitos da Sociedade em Rede. 2015 may 27-29; Santa Maria, Rio Grande do Sul. Santa Maria (RS): Rede CIIDI; 2015.
23. Wang J, Zhang T, Lu Y, Zhou GY, Chen Q, Niu B. Vesicular stomatitis forecasting based on Google Trends. *PLoS One*. 2018 Jan;13(1):e0192141

**Supplementary Table 1.** Task force for implementation the PGx testing project at INCA (alphabetical order).

Maria Thereza Accioly, PhD – INCA National Tumor Bank  
Anna Beatriz Ribeiro Elias, MSc – Division of Clinical Research and Technol. Development  
Mariana Emerenciano, PhD – Div. Clinical Research and Technological Development  
Marcos Ferreira - Technology Information Service  
Antonio Augusto Gonçalves, PhD – Technology Information Service  
Giovana Kovaleski, BSc – Division of Clinical Research and Technological Development  
Marcela B Mansur, PhD – Division of Clinical Research and Technological Development  
Andréia Cristina de Melo, MD, PhD – Div. Clinical Research and Technol. Development  
Vera Lucia Alves da Mota, BSc – Div. Clinical Research and Technological Development  
Alexandre de Mendonça Palladino, MD – Clinical Oncology Service, Hospital do Cancer I  
Guilherme Suarez-Kurtz, MD, PhD – Div. Clinical Research and Technol. Development  
Karolyne Wolch, Biol. student – Div. Clinical Research and Technological Development