RESEARCH ARTICLE

Using Google Trends Data to Study Public Interest in Breast Cancer Screening in Malaysia

Mazlyfarina Mohamad*, Hui Sin Kok

Abstract

Objective: This study aims to investigate the public pattern in seeking breast cancer screening information in Malaysia using Google Trends. **Methods:** The Google Trends database was evaluated for the relative Internet search popularity of breast cancer and screening-related search terms from 2007 to 2018. **Results:** Result showed downward trends in breast cancer search, whereas mammogram and tomosynthesis search fluctuated consistently. A significant increment was found during Pink October month. Breast cancer search term achieved the highest popularity in the east coast of Malaysia with [x^2 (5, N=661) = 110.93, P<0.05], whereas mammogram attained the highest search volume in central Malaysia [x^2 (4, N=67) = 18.90, P<0.05]. The cross-correlation for breast cancer was moderate among northern Malaysia, Sabah, and Sarawak ($0.3 \leq r_s \leq 0.7$). **Conclusion:** Public interest trend in breast cancer screening is strongly correlated with the breast cancer awareness campaign, Pink October. Breast cancer screening should be promoted in the rural areas in Malaysia.

Keywords: Breast cancer- mammogram- tomosynthesis- breast cancer screening- early detection of cancer

Introduction

Breast cancer is the most prevalent cancer that affects females worldwide and in Malaysia. It comprises approximately 30.4% of all cancers in Malaysia. It is the leading cause of cancer death among women from all ethnic and age groups in Malaysia, accounting for approximately 11% of all medically certified deaths (Hadi et al., 2010). The Malaysian National Cancer Registry (Azizah et al., 2016) reported that the number of breast cancer incidents has increased since 2009 to 2011, and the number of cases is expected to increase in the coming years. Studies have shown that the majority of breast cancer patients in Malaysia exhibit symptoms for the first time at a late stage (Ibrahim et al., 2012; Cheng et al., 2015; Saxena et al., 2012), thereby reflecting the necessity for increasing awareness and early detection of the disease among Malaysians. One of the challenges faced by the Ministry of Health of Malaysia is promoting interest in breast cancer screening, particularly among high-risk women.

Mammography, clinical breast examination, and breast self-examination are considered effective methods for the early detection of breast cancer (Akhtari–Zavare et al., 2015; Hasan et al., 2015). Mammography imaging of breasts has played a vital role in breast cancer detection. The early detection of breast cancer increases the survival rate of a patient because this disease can be treated easily and effectively during its early stages. Women diagnosed with breast cancer at an early stage have double the five-year survival rate of patients diagnosed at a late stage (Oussama and Khatib, 2006). Early detection not only increases the survival rate but is also considerably more cost-effective than treating the disease in its late stages. However, breast screening behavior among Asian women is low compared with their Western counterparts due to social and cultural variations, such as lifestyle, perceptions of health and illness, and health-seeking behavior (Adeeb et al., 2008; Al–Naggar et al., 2009; Rahmah et al., 2013; Norsa`adah et al., 2012; Hassan et al., 2017).

In this new era of science and technology, the growing use of electronic devices has led to increasing data production on the Internet. The Internet has become an important part of people's lives. At present, seeking health information on the Internet has become a common practice among the people of Malaysia. Individuals seek information, such as symptoms and diseases, healthcare tips, and treatment methods. Such behavior seeks to fill in the gap between the information that they already have and what they still need to know (Ahadzadeh et al., 2017). The percentage of Internet users in Malaysia in 2014 and 2016 was 66.6% and 76.9%, respectively, according to a survey on Internet users in Malaysia (Malaysia Communications and Multimedia Commission, 2017). Meanwhile, the percentage of Internet users searching for health-related information online was 77.2%. However, this survey was conducted by interviewing 3,469 respondents, and it does not reflect the trends in Internet search activities

Centre for Health and Applied Sciences, Faculty of Health Science, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, 50300 Kuala Lumpur, Malaysia. *For Correspondence: mazlyfarina@ukm.edu.my

in Malaysia. Therefore, Google Trends is used because it can accommodate a massive number of queries to create traffic data that can be analyzed to show hidden interest cycles and their seasonality.

Google Trends is a well-known online search tool that provides near real-time trend data and shows how frequently a keyword has been queried on Google's search engine at a specific time interval. Google Trends provides daily and weekly reports on the volume of queries related to various industries. It may help in predicting present issues. For example, in the field of epidemiology, Polgreen et al., (2008) and Ginsberg et al., (2009) demonstrated that search data could help predict the incidence of influenza-like diseases. Their works were widely publicized and stimulated further findings in epidemiology (Brownstein et al., 2009; Wilson et al., 2009; Corley et al., 2010; Hulth et al., 2009; Pelat et al., 2009; Valdivia and Monge–Corella, 2010). Whenever a search term is entered, Google Trends will present its cycles with respect to the highest number of searches for the term. These findings will have implications in developing programs for the early detection of breast cancer and for understanding geographical variations in breast cancer screening awareness among Malaysians.

The objectives of this study are as follows: (1) to explore public interest trends in breast cancer and its screening-related search terms from 2007 to 2018, (2) to determine the correlation among people's search interests in various regions in Malaysia, and (3) to identify information search patterns related to the breast cancer awareness campaign. We use Google Trends and hypothesize that the trends of these queries may be affected by the implementation of the breast cancer awareness campaign in October.

Materials and Methods

Data Queries

Data from Google Trends were mined and saved in a comma-separated values (CSV) file, which stores tabular data (numbers and text) in plain text. Numerical data is presented as relative search volume (RSV), which is computed as the percentage of queries on a particular term for a given location and time period. The RSV values indicate the ratio of the search volume of specific Google's queries to the search volume of overall Google's queries performed in specific regions and time intervals. Google Trends has normalised data using the highest query share of that term over the time series and presented on a scale of 0 to 100. Each data point is divided by the total searches of the geography and time range it represents to compare relative popularity. Otherwise, places with the most search volume would be ranked highest. Technology evolves and the number of people searching on Google changes with the development of technology, that is, the search volume in earlier years was much smaller than that currently. Hence, raw search numbers were unsuitable for comparing searches then and now. Google Trends incorporated normalisation of data, namely, comparison across dates and countries or cities, which allows research into deeper insights. In this study, Google Trends was

used to explore Internet activity related to breast cancer and its screening-related queries on public interest cycle in Malaysia. The study period spans January 1, 2007 to February 1, 2018. The popularity of search terms was compared according to six regions in Malaysia: northern, central, southern, east coast, Sabah, and Sarawak.

Breast cancer and its screening-related queries were listed in English and Malay languages to select relevant search terms for Malaysia. "Breast cancer," "mammogram," and "tomosynthesis" (kanser payudara, mamogram, and tomosintesis in the Malay language) were used as search terms in this study. These search terms were selected after considering their face validity, excluding their plural forms, which resulted in low weekly RSV values (Vasconcellos–Silva et al., 2017). Related search queries used by Internet users from Malaysia were exported from Google Trends in comma separated value (csv) files and each query was weighted by its respective RSV value.

Statistical Analysis

A chi square test for goodness of fit was used for the analysis of differences in the popularity of breast cancer search terms among distinct regions in Malaysia. Pairwise cross-correlation analysis was performed to examine the consistency of trend data between states in Malaysia to show the direction and degree of search volumes in one state in accordance with the search volumes in another (Foroughi et al., 2016). High cross-correlations between states indicate temporal patterns in information-seeking behavior. The SPSS version 24.0 (IBM Incorporated, New York, USA) and R software (version 3.4.3; https://www.r-project.org) was used for statistical analysis.

Results

The results show the trend patterns for 5 searches over 11 years. The term "breast cancer" has the highest annual mean of searches, followed by the terms "mammogram" and "tomosynthesis" (Figure 1). Overall, "breast cancer" searches exhibits a downward trend. Searches decreased gradually from 2007 to 2009, followed by a steady decline from 2010 to 2018. By contrast, "mammogram" searches presented a slightly upward trend, whereas "tomosynthesis" searches remained nearly constant throughout the study period.

Related queries refer to related terms that are searched by Internet users using the Google search engine. They can be categorized into "Top" and "Rising" queries. Top searches are terms that are most frequently sought with the term entered in a similar inquiry session within the selected category, country, or region. Rising searches are terms that are searched for with the entered term that have achieved the most significant growth in volume during the requested period. For each rising search term, a percentage of the term's growth is compared with that of the previous period.

Table 1 shows the top four queries for each search term with respect to their RSV values. For the "breast cancer" search term, public queries regarding symptoms have reached peak popularity. By contrast, queries about

Table 1. Top 5 Queries for Each Search Terms in Google and Their Corresponding Search Volumes in Malaysia; January 2007 to February 2018

Search Terms	Top 5 Queries
Breast cancer	breast cancer symptoms
	breast cancer awareness
	symptoms of breast cancer
	breast cancer in malaysia
	breast lump
mammogram	mammogram malaysia
	breast cancer
	mammogram screening
	lppkn
	mammogram test
tomosynthesis	Insufficient data to show

Table 2. Rising 5 Queries for Each Search Terms and Their Corresponding Ggrowth Percentages in Malaysia; January 2007 to February 2018

Search Terms	Rising 5 Queries	Growth Percentage
Breast cancer	breast cancer symptoms	>5,000%
	symptoms of breast cancer	>5,000%
	breast cancer staging	>5,000%
	chemotherapy	>5,000%
	tanda breast cancer	>5,000%
	breast cancer	>5,000%
mammogram	breast cancer	>5,000%
	mammogram screening	>5,000%
	lppkn	>5,000%
	mammogram test	>5,000%
	pap smear	>5,000%
tomosynthesis	Insufficient data to show	-

the search term "mammogram" are highly targeted by Malaysians with an RSV value of 100. Table 2 shows the rising queries for the four search terms with respect to their growth percentage. Related rising queries for all the search terms share the same growth percentage for "Breakout" or more than 5,000%. However, the search term "tomosynthesis" does not have sufficient data to demonstrate public queries.

A chi-square test for goodness of fit (with α=0.05) was performed to assess the differences in the popularity of breast cancer among distinct regions in Malaysia. The chi-square test was statistically significant, x^2 (5, N=661) = 110.93, P<0.05, thereby indicating that the "breast cancer" search term was reported with a significantly greater frequency in the east coast than in other regions of Malaysia (Figure 2).The chi-square test for mammogram was statistically significant, x^2 (4, N=67) = 18.90, P<0.05, thereby indicating that the search term mammogram (mamogram) was reported with a significantly greater frequency in the central region than in other regions in Malaysia, whereas its frequency was lower in Sabah, Sarawak, and the east coast region. For the "breast cancer" search term, pairwise cross-correlation was moderate between Kedah and Pulau Pinang (r = 0.5895, P<0.001) and Negeri Sembilan and Sarawak (r = 0.4667, P<0.001) whereas the rest of the regions exhibited weak correlations (r<0.3). No strong correlation was observed for the "mammogram" search term.

A visual inspection of the RSV values was performed 1 month before the campaign, during the campaign, and 1
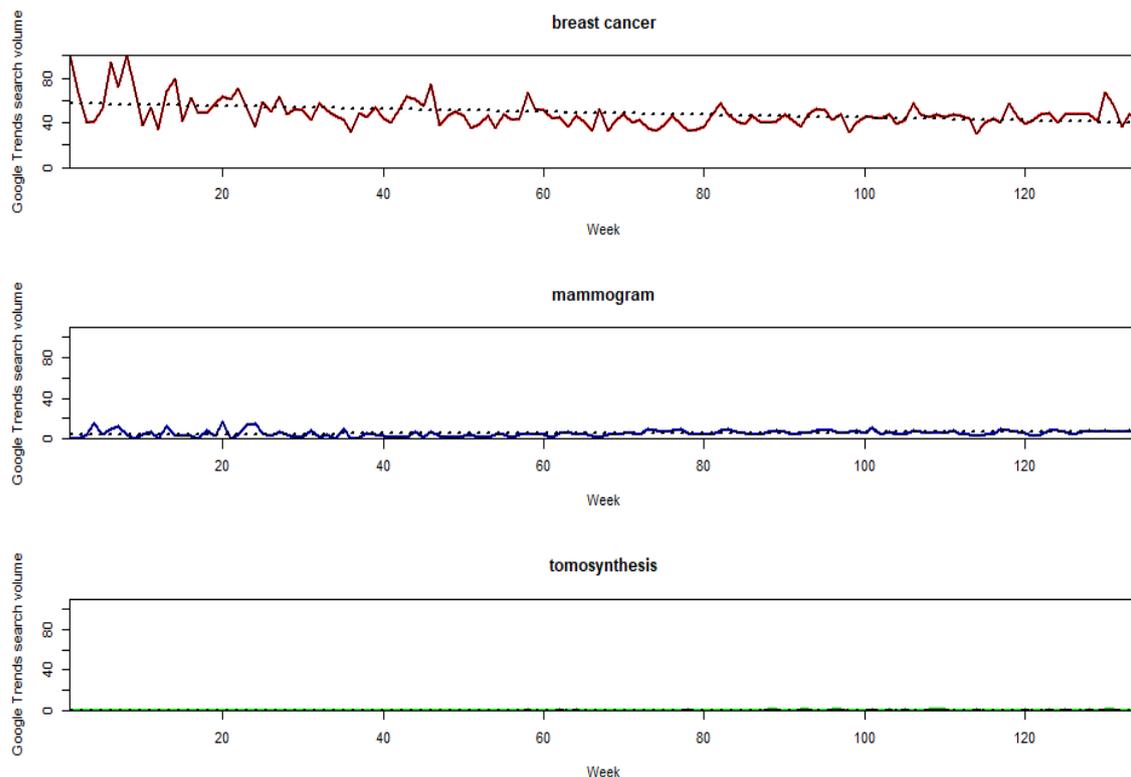

Figure 1. Breast Cancer, Mammogram and Tomosynthesis Relative Search Volume (RSV) for 132 Months
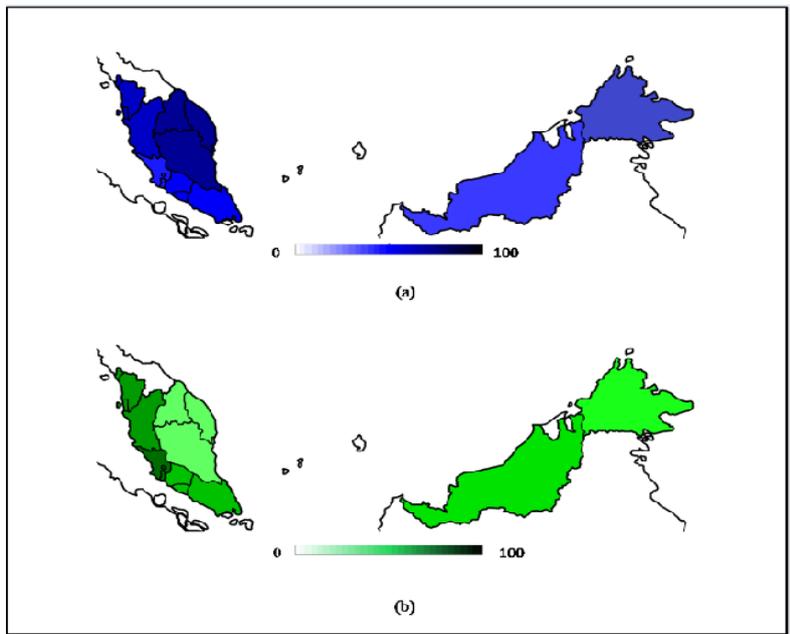
Figure 2. Interest of Search Terms by Regions in Malaysia (a) Breast Cancer (b) Mammogram

month after the campaign to determine whether increases and decreases in search activities were influenced by the breast cancer awareness campaign. As shown in Figure 3, "breast cancer" achieves the greatest increase in search activities during the breast cancer awareness campaign period. Search activities increased as the campaign is introduced and decreased toward the end of the campaign.

Discussion

The results demonstrated a downward trend for "breast cancer" in Malaysia. People apparently have low interest in seeking information about breast cancer, particularly in the past few years. Several factors may contribute to this temporal trend searching behavior. First, despite greater focus on colorectal cancer at present and a preoccupation

with breast cancer awareness, gross neglect is observed in most other types of cancer based on the review of Loh et al., (2017) of cancer awareness in Malaysia. Second, although government agencies and nongovernmental organization have been promoting the breast cancer awareness program, the target group does not belong to the most active Internet users in Malaysia, who are below 34 years old (Malaysia Communications and Multimedia Commission 2017). In addition, decreased incidence and lack of resources (e.g., road tour programs, exhibitions) are suggested as factors that contribute to the decline of search activities.

"Breast cancer" was the highest search activity in the east coast region. The Department of Statistics in Malaysia reports that the Malay population is considerably higher in the east coast, where the Malay language is commonly
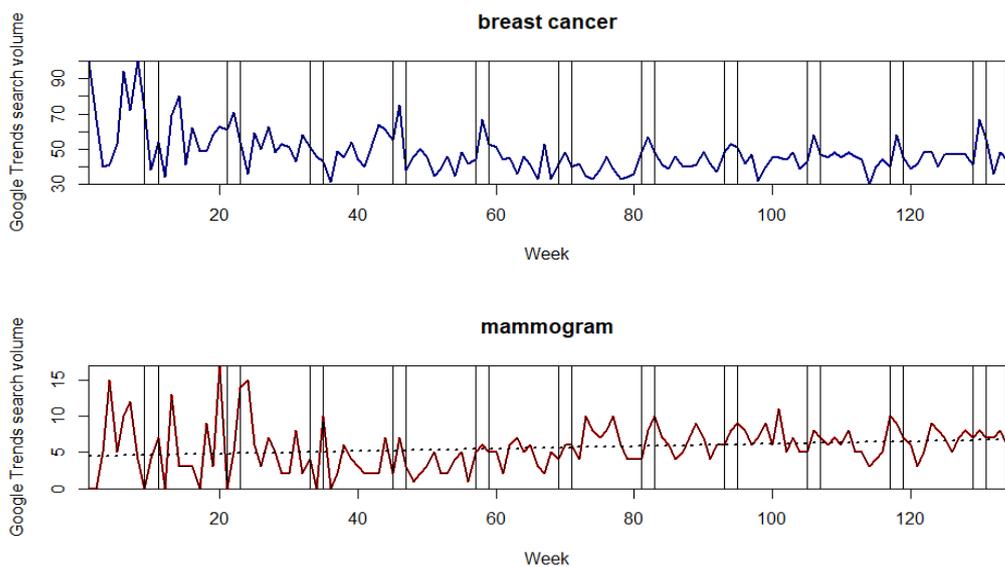

Figure 3. Monthly Relative Search Volume (RSV) for Breast Cancer and Mammogram before, During, and after the Campaign Period from 2012 to 2017

and routinely spoken as the mother tongue in daily life. The increment in breast cancer campaign organized in the Malay language may lead to increasing Malay search terms over time. In addition, Malays have been found to have the highest breast cancer risk among other ethnicities in Malaysia (Hisham and Yip, 2003). The moderate cross correlation among breast cancer trends in Kedah, Pulau Pinang, Negeri Sembilan, and Sarawak indicates that these regions may participate in breast awareness campaigns because of similar information-seeking behavior.

The "mammogram" search term obtained the highest search activity in the central area of Malaysia but received lower popularity in Sabah, Sarawak, and the east coast region. In Malaysia, many subsidized screening programs that cater to various groups of women are available. At least six major subsidized mammogram screening programs were available in Malaysia in 2015. The Ministry of Health and National Population and Family Development Board caters to the general population, the Social Security Organization caters to women in the formal employment sector, the National Cancer Center caters to rural women, and state governments cater to their constituents. However, the maldistribution of mammogram facilities was observed. Most facilities with mammogram equipment are mainly located in the central and west coast regions of Peninsular Malaysia, where major cities are located. The ratio of mammogram facilities to the target population is satisfactory in the central region, particularly in Kuala Lumpur, where the ratio was 1:10,000 (assuming that one facility with mammogram service had one mammogram machine). However the ratio was poor in other parts of the country and ranged from 1:20,000 to 1:80,000 (Mahmud and Aljunid, 2018). The reason for the low search volume of "mammogram" in certain regions can be the low accessibility of mammography service. In Sabah, for example, mammography is available only in five private hospitals, one private specialist clinic, and three government hospitals, namely, Hospital Queen Elizabeth, Hospital Queen Elizabeth 2, and Sabah Women and Children's Hospital.

The "tomosynthesis" search term presented the least popularity among the five search terms regardless of the related queries, raising queries, and geographical study. Tomosynthesis is a newly developed technology that can improve the detection and characterization of breast lesions (Helvie, 2010). In 2011, Hologic, Inc. received the approval of the Food and Drug Administration (FDA) for the Selenia Dimensions 3D System, a 3D system that is currently the only FDA-approved breast tomosynthesis system. Data are insufficient for demonstration because this technology was introduced only recently, and only a few people are aware of its existence.

One of the interesting findings of this study is that the highest peak search is always observed every October throughout the study period. We strongly believe that the reason for this scenario is the effectiveness of the "Pink October" campaign. In conjunction with National Breast Cancer Awareness Month, Pink October is an annual international health campaign organized by major breast cancer charities to increase awareness of the disease and to raise funds for research on its cause, prevention,

diagnosis, treatment, and cure. The campaign also offers information and support to breast cancer patients and to interested members of the public. Hence, people are exposed to breast cancer information to a greater extent during October because of the massive publicity launched by organizations. Thus, increased search volume for the search terms is observed in Google Trends.

Limitations

This study has limitations. First, web access is still concentrated in (but not limited to) metropolitan areas, which will limit the use of Google Trends in rural areas or regions with a low search volume. Specific subpopulations and their cultural disparities may not be reached by RSV algorithms (Vasconcellos–Silva et al., 2017). Second, Google is not the only search tool available although it has been the most dominant player in the search engine market. People may still use other search engines, such as Yahoo! or Bing. However, other search engines do not have proprietary availability in their web services (Foroughi et al., 2016).

In conclusion, our study shows that the downward trend in "breast cancer" search has increased gradually, whereas "mammogram" and "tomosynthesis" have consistently fluctuated. A significant increment is observed during Pink October month. We conclude that the public interest trend in breast cancer screening is strongly correlated with the breast cancer awareness campaign, Pink October. Therefore, the detection of breast cancer and breast cancer screening should be promoted across Malaysia, particularly in the east and west coast regions.

References

Adeeb N, Nur-Azurah AG, Ong FB, et al (2008). Utility of Pap Smears and mammograms in midlife Urban Malaysian women. *Med Health*, **3**, 59-68.

Ahadzadeh AS, Shari AP (2017). Online health information seeking among Malaysia Women: Technology acceptace model perspective. *Search*, **9**, 47-70.

Akhtari-Zavare M, Juni MH, Ismail IZ, Said SM, Latiff LA (2015). Barriers to breast self examination practice among Malaysian female students: A cross sectional study. *SpringerPlus*, **4**, 692.

Al-Naggar RAM, Isa ZM, Shah SA, et al (2009). Eight year survival among breast cancer Malaysian women from University Kebangsaan Malaysia Medical Centre. *Asian Pac J Cancer Prev*, **10**, 1075–8.

Azizah AM, Nor Saleha IT, Noor Hashimah A, Asmah ZA, Mastulu W (2016). Malaysian National Cancer Registry Report 2007-2011. Ministry of Health Malaysia.

Brownstein JS, Freifeld CC, Madoff LC (2009). Digital disease detection--harnessing the web for public health surveillance. *N Engl J Med*, **360**, 2153-5.

Cheng ML, Ling DY, Nanu PKP, Nording H, Lim CH (2015).

Factors influencing late stage of breast cancer at presentation in a District Hospital - Segamat Hospital, Johor. *Med J Malaysia*, **70**, 148-52.

Corley CD, Cook DJ, Mikler AR, Singh KP (2010). Using web and social media for influenza surveillance. *Adv Exp Med Biol*, **680**, 559-64.

Foroughi F, Lam AKY, Ahmadvand A (2016). "googling" for cancer: An infodemiological assessment of online search interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. *JMIR Cancer*, **2**, e5.

Ginsberg J, Mohebbi MH, Patel RS, et al (2009). Detecting influenza epidemics using search engine query data. *Nature*, **457**, 1012-4.

Hadi MA, Hassali MA, Shafie AA, Awaisu A (2010). Evaluation of breast cancer awareness among female university students in Malaysia. *Int J Pharm Pract*, **8**, 29-34.

Hasan TN, Shah SA, Hassan MR, et al (2015). Poor knowledge and practice towards breast cancer among women in Baghdad city, Iraq. *Asian Pac J Cancer Prev*, **16**, 6669-72.

Hassan MR, Ghazi HF, Mohamed AS, Jasmin SJ (2017). Knowledge and practice of breast self-examination among female non-medical students in Universiti Kebangasaan Malaysia (UKM) in Bangi. *Malaysian J Public Health Med*, **17**, 51-8.

Helvie MA (2010). Digital mammography imaging: Breast tomosynthesis and advanced applications. *Radiol Clin North Am*, **48**, 917-29.

Hisham AN, Yip CH (2003). Spectrum of breast cancer in Malaysian women: overview. *World J Surg*, **27**, 921-3.

Hulth A, Rydevik G, Linde A (2009). Web queries as a source for syndromic surveillance. *PLoS One*, **4**, e4378.

Ibrahim NI, Dahlui M, Aina EN, Al-Sadat N (2012). Who are the breast cancer survivors in Malaysia?. *Asian Pac J Cancer Prev*, **13**, 2213-8.

Loh SY, Sunthari S, Tin TS (2017). A review of cancer awareness in Malaysia – what's next?. *J Cancer Oncol*, **1**, 000105.

Mahmud A, Aljunid SM (2018). Availability and accessibility of subsidized mammogram screening program in Peninsular Malaysia: A preliminary study using travel impedance approach. *PLoS One*, **13**, e0191764.

Malaysia Communications and Multimedia Commision 2017. Internet Users Survey 2017. ISSN 1823-2523.

Norsa`adah B, Rahmah MA, Rampal KG, Knight A (2012). Understanding barriers to Malaysian women with breast cancer seeking help. *Asian Pac J Cancer Prev*, **13**, 3723-30.

Oussama MN, Khatib AM (2006). Guidelines for the Early Detection and Screening of Breast Cancer. World Health Organization.

Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron, A-J (2009). More Diseases Tracked by Using Google Trends. Emerging Infectious Diseases, **15**, 1327-8.

Polgreen PM, Chen Y, Pennock DM, Nelson FD (2008). Using Internet Searches for Influenza Surveillance. Clin Infect Dis, 47, 1443-8.

Rahmah MA, Aniza I, Che Engku Nor Bahiyah CEM (2013). Do Elderly women in Malaysia go for mammogram screening?. *Malaysian J Public Health Med*, **13**, 20-6.

Saxena N, Hartman M, Bhoo-Pathy N, et al (2012). Breast cancer in South East Asia: Comparison of presentation and outcome between a middle income and a high income country. *World J Surg*, **36**, 2838-246.

Valdivia A, Monge-Corella S (2010). Diseases tracked by using Google Trends, Spain. *Emerg Infect Dis*, **16**, 168.

Vasconcellos-Silva PR, Carvalho DBF, Trajano V, et al (2017). Using Google Trends data to study public interest in breast cancer screening in Brazil: Why Not a Pink February?. *JMIR Public Health Surveill*, **3**, e17.

Wilson N, Mason K, Tobias M, et al (2009). Interpreting "Google Flu Trends" data for pandemic H1N1 Influenza: The New Zealand Experience. *Eurosurveillance*, **14**, 19386.