The Journal of Laryngology & Otology

cambridge.org/jlo

Using Google Trends to understand information-seeking behaviour about throat cancer

M Faoury, T Upile and N Patel

Department of Otolaryngology – Head and Neck Surgery, University Hospital Southampton, UK

Abstract

Objective. Many people seek health information from internet sources. Understanding this behaviour can help inform healthcare delivery. This study aimed to review Google Trends as a method for investigating internet-based information-seeking behaviour related to throat cancer in terms of quantity, content and thematic analysis.

Method. Data was collected using Google Trends. Normalised data was created using the search terms 'throat cancer', 'cancer', 'HPV', 'laryngeal cancer' and 'head and neck cancer'. The search data was used to analyse the temporal and geographical interest pattern of these terms from 2004 to 2015.

Results. Three important peaks in searches for 'throat cancer' were identified. The first and greatest increase in interest was in September 2010, and there were also peaks in June 2013 and in October 2011.

Conclusion. Internet-search analysis can provide an insight into the information-seeking behaviour of the public. Mass media can hugely affect this information-seeking behaviour. Possessing tools to investigate and understand information-seeking behaviour may be used to improve healthcare delivery.

Main Article

Dr M Faoury takes responsibility for the integrity of the content of the paper

Presented at the British Association of Head and Neck Oncologists annual scientific meeting, 12 May 2017, London, UK.

Cite this article: Faoury M, Upile T, Patel N. Using Google Trends to understand information-seeking behaviour about throat cancer. *J Laryngol Otol* 2019;**133**:610–614. https://doi.org/10.1017/S0022215119001348

Accepted: 15 April 2019
First published online: 8 July 2019

Key words:
Pharynx; Papillomavirus Infections; Head And Neck Neoplasms; Correlation Of Data; Data Analysis

Author for correspondence:
Dr Morad Faoury, Department of Otolaryngology – Head and Neck Surgery, University Hospital Southampton, Tremona Road, Southampton SO16 6YD, UK
E-mail: moradfaoury@gmail.com
Fax: 02380 333043

Introduction

Studies have shown that many people seek health information from internet sources.[1–3] Understanding this behaviour can help inform healthcare delivery and may also reduce patients' disease-specific anxiety and uncertainty if reliable sources are available.[1]

Google® is the world's most popular search engine[4–6] and has been shown to be the best search engine for both average precision (70 per cent) and average response time (2 seconds).[7] Seeking health information through a search engine has become a very common occurrence. In 2009, it was estimated that 83 per cent of American adults used the internet, and 61 per cent used the internet to search for online health-related information.[2,3] Google Trends (or Google Insights, the previous and similar Google tool) is an online search analysis tool. It analyses a percentage of Google web searches to determine how many searches have been performed for the entered terms compared to the total number of Google searches performed during a specific period of time. It collects useful information that can be used to monitor health information seeking behaviour trends, epidemiology, aetiology and management of specific health conditions. Scrutinising such information now constitutes a new research discipline termed infodemiology, which is the study of the determinants and distribution of health information.[8] An example of this is the use of Google Trends' real-time surveillance system to understand public health epidemic issues, including disease hotspots, as has been done with influenza outbreaks.[9,10]

Our study aimed to review the use of Google Trends as a new method for evaluating interest in health information on a particular topic. Our primary aim was to investigate internet-based information-seeking behaviour related to the search 'throat cancer' in terms of quantity, content and thematic analysis. Our secondary aim was to evaluate the relationship between 'throat cancer' searches and related mass media stories.

Materials and methods

Google Trends was chosen as the tool to interrogate searches on the internet. Google Trends is a tool that allows users to interact with internet search data. It analyses a sample of the billions of daily Google search results and provides information on geospatial and temporal patterns in search volumes for user-specified terms.[11] Google Trends creates a 'search volume index', which represents the relative search volume for a search term indexed against the overall search volume.[2] Search volume index values were adjusted to a normalised data scale of 0–100.

Google Trends allows the popularity of search terms and trends to be evaluated, with different features and options[12] including: allowing the user to compare the volume of searches between two or more terms; the ability to show information related to the search-term overlaid on a chart (showing how new events affect search popularity); breaking down related searches into topics and queries, which makes it possible to find related topics and specific keywords that users also searched for; and making the data easier to compare and interpret by normalising the results, which avoids the problem of regions with the greatest search volume always ranking highest.

Data that are excluded by Google Trends represent low volume search terms, defined by Google Trends as 'searches made by few people'. Google Trends only analyses data for popular searches and repeated searches from a certain user over a short period of time (defined as 'duplicate searches' by Google Trends).

There are many products and programs that can be described as Google Trends alternatives. Most of these products are listed under the category of search engine optimisation, but only few are able to do what Google Trends offers.

For example, Keyword Planner is a free, open source Google tool that was launched by Google AdWords in 2013. It is a tool developed for pay-per-click marketers, and it collects search term numbers on Google as a volume rather than as an interest ratio. Google Trends, unlike Keyword Planner, offers a line graph timeline of the search keywords. This line graph illustrates the peaks of a search and gives a better understanding of that change over a period of time instead of displaying average search volume ranges for each month. Google Trends can also offer search data by sub-region whereas Keyword Planner does not. Additionally, Keyword Planner has no way to compare different keywords across a timeline during a period of time.

Another product called SemRush is a relatively popular subscription-based search engine optimisation tool that relies on competitive intelligence to refine search engine optimisation. It is smart and highly comparable to Google Trends, but it is not free and therefore not available to everyone. It is more of a business-targeted tool, and it needs to be studied further to be considered a reliable research tool.

Our study data were collected using Google Trends on a Windows® 7 laptop computer. AVG® TuneUp 2015, a computer optimisation program, was used to clean up the Google Chrome browser by removing cookies before data collection was started.

A search of Google Trends' normalised data (using normalised data prevented data corruption) was performed using the search terms: 'throat cancer', 'cancer', 'HPV' (for human papilloma virus), 'laryngeal cancer', 'head and neck cancer', 'tonsil cancer' and 'oral cancer'. The terms were chosen based on a combination of technical terms and layman's terms. The latter were not necessarily medically accurate but were typical of layman's terms used to describe the upper aerodigestive tract. The term 'HPV' was used because of the importance of the infection in many upper aerodigestive tract cancers.

The search data were further analysed by subject item, using a temporal and a geographical representation of the interest pattern from 2004 to 2015. These modified analyses were performed using Google Trends and Microsoft Excel® (2007) spread sheet software. Filters were included to relate to UK searches.

The data that were excluded by Google Trends included low volume search terms and repeated searches from the same user over a short period of time (defined as duplicate searches by Google Trends). This stringency enabled a temporal and a geographical representation of the interest pattern. The geographical representation included options for a worldwide or regional view. In our study, the regional option was used to obtain the interest pattern within the UK. The option of 'news headlines' added the impact of media to the temporal interest pattern.

The interest over time is represented by numbers on the graph that reflect how many searches have been carried out for a search term, relative to the total number of searches done on Google over time. To explain further, those numbers do not represent the absolute search volume, because they are normalised data and presented on a scale from 0–100 in order to reduce data redundancy and improve data integrity. Each point on the graph is divided by the highest point and multiplied by 100. When there is not enough data, 0 (normalised data) is shown. In regard to regional interest, the numbers represent the search volume relative to the highest point on the map, which is always 100 (normalised data).

Google Trends can also identify searches that are related to a particular search term and these related terms are divided into 'topics' and 'queries'. For each sub-category in the related searches, there is the option to look at 'top' searches and 'rising' searches. The top searches option shows popular search terms that are similar to the term entered, whereas the rising search option shows searches that have grown significantly in popularity over a given time period when compared with a preceding time period.

Google Trends data are computed using a sampling method, and the results therefore change by a few per cent from day to day. Because of privacy considerations, only searches with meaningful values (a high search volume) are followed and calculated. There is a considerable amount of online help available through links on the site, which give details of how the data is collected.

Results

By running a Google Trends comparative analysis, we identified three important peaks in throat cancer interest over time. The highest peak in interest was in September 2010 (normalised data = 100), the second highest peak was in June 2013 (normalised data = 84), and the third highest peak was in October 2011 (normalised data = 58) (Figure 1).

The highest recorded peak from September 2010 (based on the searches for throat cancer and HPV (normalised data = 24 and 62, respectively)) was most likely a result of a famous actor's throat cancer diagnosis (date of diagnosis August 2010). This is an increase of roughly 2.2 times the throat cancer interest in August 2010 (throat cancer, actor and HPV (normalised data = 11, 36 and 28, respectively; $p < 0.01$).

June 2013 had the second highest peak of interest in throat cancer (throat cancer and HPV (normalised data = 20 and 92, respectively), an approximate increase of 2.5 times from the previous month (throat cancer; and HPV (normalised data = 8 and 56, respectively; $p < 0.01$)). This was driven by news stories on the relation between HPV and throat cancer and on speculations about the aetiology of a renowned actor's disease (Figure 2).

The third peak in October 2011 correlates to news items regarding the possible relationship between HPV and throat cancer. It shows a 1.55 increase in throat cancer interest from the previous month (Figure 3).

https://doi.org/10.1017/S0022215119001348 Published online by Cambridge University Press

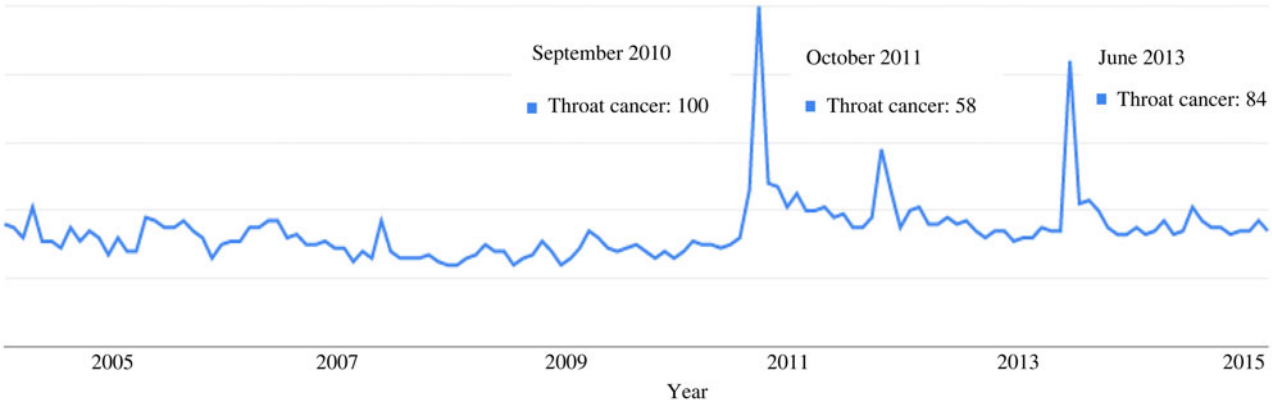

Fig. 1. Graph showing peaks in 'throat cancer' searches on Google over time.

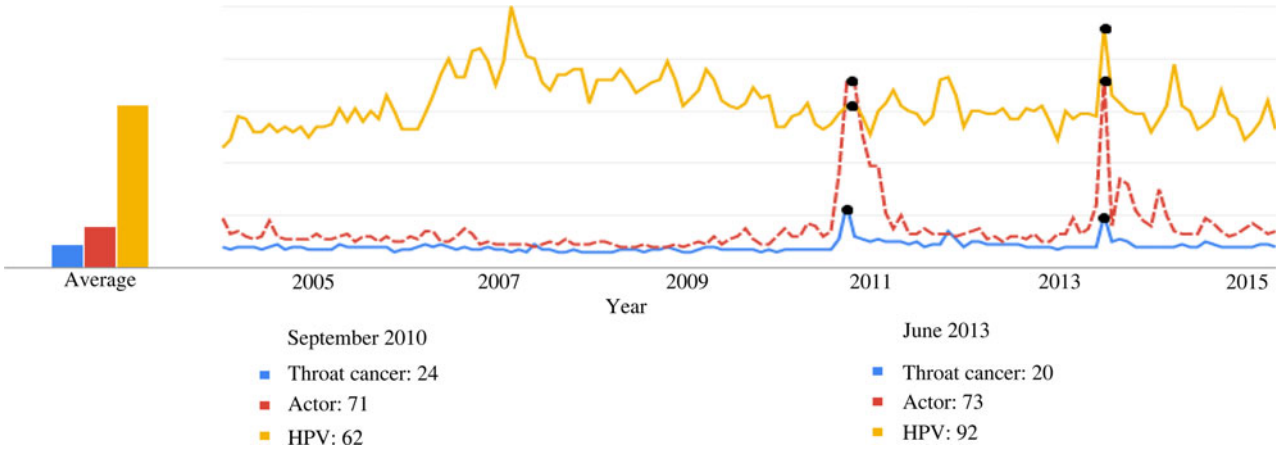

Fig. 2. Graph showing the relationship between Google searches for 'throat cancer' and influencing factors. HPV = human papilloma virus

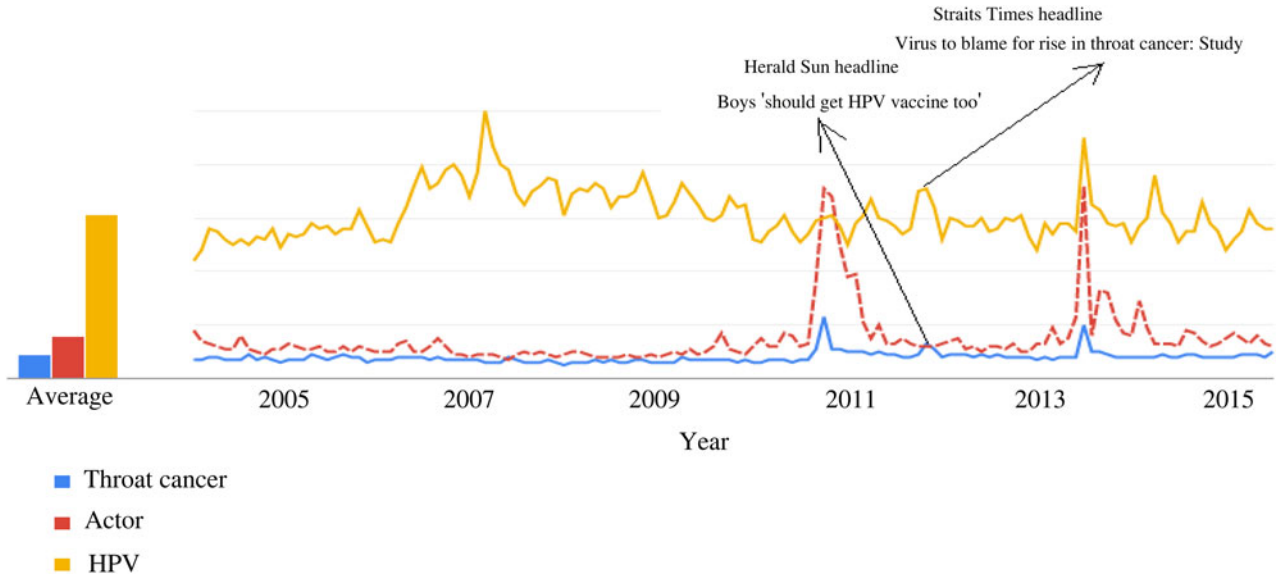

Fig. 3. Graph showing the relationship between Google searches for 'throat cancer' and mass media stories about human papilloma virus (HPV).

Using the same search terms, the countries with the highest volume of searches were the UK, the US, Ireland, Canada, Australia, Kenya and Ghana (Figure 4).

Worldwide, the top five cities with the highest volume of searches for 'throat cancer' were London (normalised data = 100), Manchester (normalised data = 97), Birmingham (normalised data = 92), Atlanta (normalised data = 91) and Chicago (normalised data = 90). In the UK, England had the most regional interest (normalised data = 100) followed by Wales (normalised data = 90), Scotland (normalised data = 87) and Northern Ireland (normalised data = 85).

When comparing 'throat cancer' to 'laryngeal cancer' and 'head and neck cancer', it was found that there was a significant correlation between the search terms 'throat cancer' and

https://doi.org/10.1017/S0022215119001348 Published online by Cambridge University Press

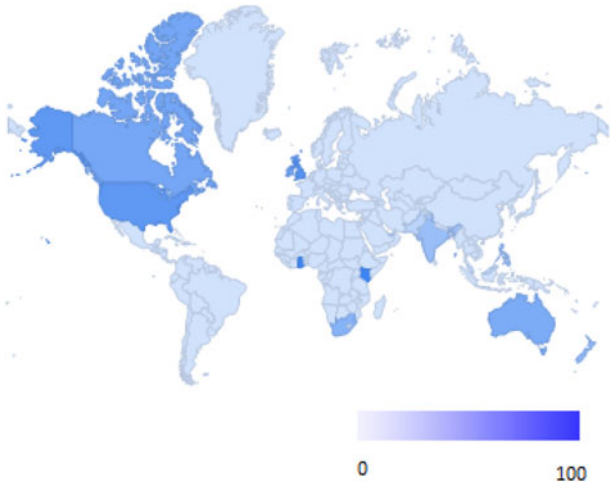

Fig. 4. Map showing countries with the highest volume of Google searches for 'throat cancer' and human papilloma virus.

'laryngeal cancer' in terms of peaks and timeline changes. On the contrary, the search term 'head and neck cancer' showed no correlation (Figure 5).

Comparing the search term 'throat cancer' and 'laryngeal cancer' geographically during the same time period, the countries that showed the highest search volume using the term 'throat cancer' were the UK, the US, Ireland, Canada, Australia, Kenya and Ghana as mentioned above. The term 'laryngeal cancer' was used in entirely different countries, which were Costa Rica, Uruguay, Japan, Vietnam, China, Turkey and Germany (Figure 6).

Discussion

A rising number of internet users visit websites every day. Each question, request or search can be seen as a personal vote. By using search engines, information about our interests, codified as search terms, is left behind, meaning search engines can gather our interests on the smallest possible scale – the scale of individual requests.[13] On longer time scales, our interests form trends, and aggregated search volume data can be used to show the trends that affect our life on larger scales.

The US National Cancer Institute stated that behavioural measures are needed in the healthcare environment and in public health planning, where national indices of progress on behaviour measures could guide policy and conveyance planning.[8] Using internet big data analysis in healthcare research holds promise, and it may complement and extend the current data foundations available for healthcare research according to the National Academy of Medicine.[12,14]

To date, Google Trends has proven to be a valuable and accessible tool. It has mostly been used for monitoring and surveillance of communicable diseases and epidemics, and it has also been used to study consumer behaviour interest changes. Using Google Trends to analyse healthcare-related topics is still largely unexplored, and it is believed that huge amounts of data can be extracted. We conducted this study to show a potential research use for Google Trends by exploring the information-seeking behaviour changes relating to throat cancer and its relation to other factors including mass media. Mass media campaigns can directly and indirectly produce positive changes or prevent negative changes in health-related behaviours across large populations.[15]

Our data analysis suggests that any news stories, new interventions or aetiology related to throat cancer can manifest as an increase in information-seeking behaviour for 'throat cancer' on Google. Additionally, information-seeking behaviour regarding throat cancer varies geographically, and the data suggest that big cities were the most influenced by mass media or 'big titles'. It was evident that developed countries and areas where the internet is more commonly provided showed bigger spikes than other countries. This strengthens the idea that information-seeking behaviour is influenced by the level of awareness exposure; however, we were not able to explain why Kenya and Ghana showed high search volumes. This will need further research and could be related to Google Trends' limitations. It was interesting to find a significant correlation between search peaks for 'throat cancer' and 'laryngeal cancer'. Both were linked to the same influencing factors and happened during the same period of time. However, people used different terms in different countries: the countries that used 'throat cancer' as a search term did not use 'laryngeal cancer' and vice versa. Google Trends is therefore a powerful tool to understand the terminology used by patients in different parts of the world.

The use of Google Trends is limited by some factors. Google Trends may have sampling biases, as it deletes repeated searches from the same user over a short period of time to reduce counts of continued searching.[16] Google frequently optimises their search algorithms, which may lead to changes in search results despite the use of the same key terms for a search. This will lead to challenging reproducibility as this type of research depends on stable results. It is unclear if using Google for health information leads to an increase in public knowledge and awareness of health issues. This is because Google Trends gives the number of times certain information has been searched but does not assess the quality of accessed information, which would allow evaluation of potential knowledge gained.

- Internet search analysis can provide an insight into the information-seeking behaviour of the public
- Understanding this behaviour can help inform healthcare delivery
- This study shows the relationship between peaks in searches for 'throat cancer' and media coverage of related stories
- Possessing tools to investigate and understand information-seeking behaviour may help to deliver better healthcare and respond to health concerns and information needs of the general population
- The information provided by Google Trends can help public health organisations to improve the accuracy and accessibility of their online health information sources
- More research is needed on the possible uses of Google Trends in the healthcare system

Despite these limitations, the use of the internet for healthcare information is still increasing. Understanding healthcare information seeking behaviour is essential in order to control and plan the quality of knowledge provided by health organisations, advocacy groups and health

https://doi.org/10.1017/S0022215119001348 Published online by Cambridge University Press

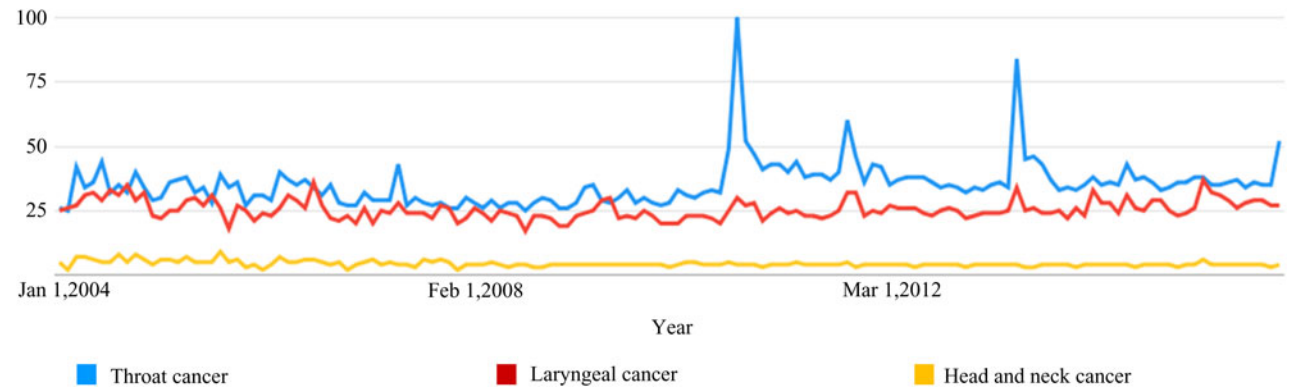

Fig. 5. Graph showing the relationship between Google searches for 'throat cancer', 'laryngeal cancer' and 'head and neck cancer'.

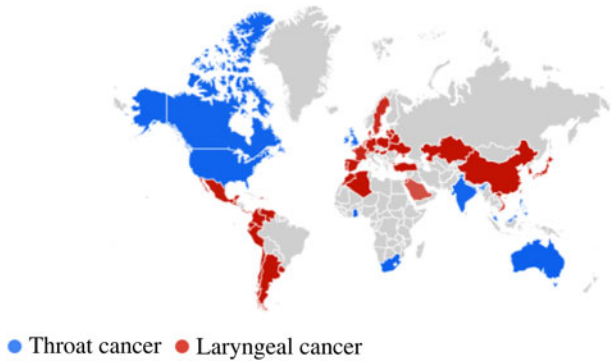

Fig. 6. Map comparing the Google search term 'throat cancer' and 'laryngeal cancer' geographically within the same time period.

professionals regarding head and neck cancers including throat cancer. Google Trends could also be used for research into other health-care related topics. Further research is needed to understand how healthcare providers can utilise Google Trends to understand health information seeking behaviour and its effect on public knowledge, awareness, disease-related anxiety and the interaction between patients and healthcare information providers.

Conclusion

Internet search analysis can provide an insight into the information-seeking behaviour of the public. Using Google Trends to analyse the search term 'throat cancer' during our study period showed the peaks of interest related to a story of the disease in a celebrity and other influencing factors (such as HPV infection). The story of human experience is a powerful driver of interest. Mass media can hugely affect information-seeking behaviour. Using tools such as Google Trends to investigate and understand information-seeking behaviour could help to deliver healthcare and respond to the health concerns and information needs of the general population. The information provided by Google Trends can also lead public health organisations to improve the accuracy and accessibility of their online health information sources.

Competing interests. None declared

References

1 Hilger C, Otto I, Hill C, Huber T, Kendel F. "Dr. Google"-information-seeking behavior and disease-specific anxiety among men with localized prostate cancer [in German]. *Urologe A* 2018;**1**:1–6
2 Willard SD, Nguyen M. Internet search trends analysis tools can provide real time data on kidney stone disease in the United States. *Urology* 2013;**81**:37–42
3 The social life of health information. In: http://www.pewinternet.org [11 June 2009]
4 Search engine market share worldwide. In: http://gs.statcounter.com/search-engine-market-share [25 July 2018]
5 Alexa top sites by category – Search Engine Ranking. In: https://www.alexa.com/topsites/category/Regional [May 16, 2013]
6 Google search statistics. Internet live stats. In: www.internetlivestats.com [30 May 2018]
7 Edosomwan J, Edosomwan TO. Comparative analysis of some search engines. *S Afr J Sci* 2010;**106**:11–12
8 Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behaviour on the internet. *J Med Internet Res* 2009;**11**:e11
9 Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009;**457**:1012–14
10 Valdivia A, Lopez-Alcalde J, Vicente M, Pichiule M, Ruiz M, Ordobas M. Monitoring influenza activity in Europe with Google Flu Trends: comparison with the findings of sentinel physician networks - results for 2009–10. *Euro Surveill* 2010;**15**:2–7
11 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;**9**:e109583
12 How to use Google Trends to gain a competitive edge. In: https://www.hallaminternet.com/google-trends-introduction-business [6 December 2016]
13 Mondria J, Wu T. Imperfect financial integration and asymmetric information: competing explanations of the home bias puzzle? *Canadian Journal of Economics* 2013;**46**:310–37
14 Institute of Medicine (US) committee on a national surveillance system for cardiovascular and select chronic diseases. *A Nationwide Framework for Surveillance of Cardiovascular and Chronic Lung Diseases*. Washington, DC: National Academies Press, 2011
15 Wakefield M.A, Loken B, Hornik R.C. Use of mass media campaigns to change health behaviour. *Lancet* 2010;**376**:1261–71
16 Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J *et al.* The utility of google trends data to examine interest in cancer screening. *BMJ Open* 2015;**5**:e006678

https://doi.org/10.1017/S0022215119001348 Published online by Cambridge University Press