The utility of "Google Trends" for epidemiological research: Lyme disease as an example

Ari Seifter[1], Alison Schwarzwalder[1], Kate Geis[1], John Aucott[2]

[1]Lyme Disease Research Foundation of Maryland, 10755 Falls Road, Suite 200, Lutherville, MD 21093, USA; [2]Department of Medicine, Johns Hopkins School of Medicine, 10755 Falls Road, Suite 200, Lutherville, MD 21093, USA

Abstract. Internet search engines have become an increasingly popular resource for accessing health-related information. The key words used as well as the number and geographic location of searches can provide trend data, as have recently been made available by Google Trends. We report briefly on exploring this resource using Lyme disease as an example because it has well-described seasonal and geographic patterns. We found that search traffic for the string "Lyme disease" reflected increased likelihood of exposure during spring and summer months; conversely, the string "cough" had higher relative traffic during winter months. The cities and states with the highest amount of search traffic for "Lyme disease" overlapped considerably with those where Lyme is known to be endemic. Despite limitations to over-interpretation, we found Google Trends to approximate certain trends previously identified in the epidemiology of Lyme disease. The generation of this type of data may have valuable future implications in aiding surveillance of a broad range of diseases.

Keywords: Google Trends, Lyme disease, ticks, epidemiology.

The Internet is an increasingly important source of health information for both patients and physicians (Dickerson et al., 2004). Patients may use search engines to look for key words that reflect their symptoms or to aid initial self-diagnosis, while physicians may use search engines as a portal to available resources on the web. The large number of searches conducted through popular websites creates trend data which can be analyzed over time to identify disease outbreaks and supplement traditional surveillance methods (Eysenbach, 2006; Ginsberg et al., 2009; Wilson and Brownstein, 2009). This area of inquiry falls under the realm of "infodemiology", the study of the distribution and determinants of information on the Internet (Eysenbach, 2009).

Google, a mainstream general search engine which also includes several additional tools and resources, represents an important gateway to online health information for patients and physicians (Wheeler, 2006; Laurent and Vickers, 2009). Trend data generated by the number of Google searches over time in a particular geographic region have recently been made available by Google Trends (http://www.google.com/trends). Google Trends is a feature which allows users to graph the frequency of searches for a term, a string of multiple terms, or a phrase. The data used to generate these graphs are scaled to the average search traffic for the selected term and are also normalized on a relative, rather than absolute, basis (About Google Trends, 2009). These graphs can also be manipulated to restrict results to specific time frames and geographic regions.

In the health-related fields, these types of search trend data have traditionally been examined in the context of disease outbreak surveillance. For example, a Google Trends search for the string "bird flu" reveals a spike in searches from the fall of 2005 to

Corresponding author:
Alison Schwarzwalder
Lyme Disease Research Foundation of Maryland
10755 Falls Road, Suite 200
Lutherville, MD 21093, USA
Tel. +1 410 583 7124; Fax +1 410 583 7128
E-mail: alison.schwarzwalder@gmail.com

the spring of 2006. Likewise, a search for "swine flu" reveals a dramatic spike in the summer of 2009. This increase in search traffic can be useful in identifying both the presence of the disease as well as the confounding effect of media attention on the search habits of healthy individuals (Eysenbach, 2006).

In addition to epidemic surveillance, health searches can also be conducted for endemic health concerns. Further, searching for several health-related terms can allow for interesting comparisons. We chose to explore Google Trends by examining Lyme disease, which is caused by the bite of infected ticks and which remains highly endemic in several areas of the United States: the Northeast and Mid-Atlantic states, the upper Midwest, and the Northwestern states (Bacon et al., 2008). While tick bites can occur throughout the year, there is a strong seasonal likelihood of tick bite exposure during the spring and summer months. The development of symptoms of acute Lyme disease generally occurs within several weeks following a tick bite exposure; therefore, patients are similarly likely to present to clinicians during the spring and summer as well. The histogram in Figure 1 shows Center for Disease Control (CDC) data which illustrate this trend.

We entered the set of terms "Lyme disease, tick bite, cough" at Google Trends' main page (http://www.google.com/trends) on July 16, 2009,
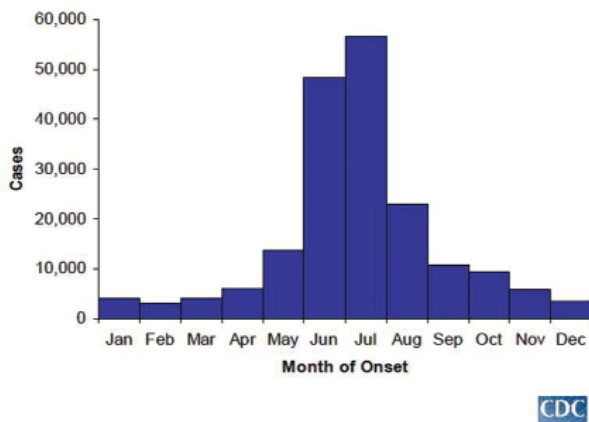

Fig. 1. Number of reported Lyme disease cases, by month of illness onset, United States, 1992-2006 (from Bacon et al., 2008).

to assess search trends in these three key words. Cough was included for comparison as a non-Lyme disease symptom to capture respiratory viral infections more likely to manifest during the winter months. Figure 2 shows the Google Trends graph depicting the frequency of searches for these three terms over a 5.5 year period. "Cough" had 2.5 times more relative traffic than "Lyme disease" while "Lyme disease" had 10 times more relative traffic than "tick bite". A minority of the patients with Lyme disease recall a tick bite, so it might be expected that searches for "Lyme disease" would outnumber searches for "tick bite". The well-known clustering of Lyme cases in May through August depicted in Figure 1 is also captured in the trend data in Figure 2. Conversely, the trend for "cough" searches peaks in the winter months of each year followed by valleys in the summer months. In addition, the small peaks for "tick bite" searches appear to occur around May of each year, at the start of a several month increase in search traffic for "Lyme disease". Google Trends also lists the cities and the states with the highest amount of search traffic for a given term. For "Lyme disease", six of the 10 states listed were among the 10 reference states with the highest number of cases reported to the CDC and where Lyme disease is considered endemic (Bacon et al., 2008). Among cities, seven of the 10 with the highest amount of search traffic are located in these same 10 reference states.

As explained on its website, there remain several limitations to over-interpretation of the graphs and data provided by Google Trends. Available data is based on a sample of Google web searches, with the potential for non-representative sampling bias. Furthermore, the calculation of the search value index presented in the results is dependent on several mathematical assumptions and approximations which may obscure true trends in search traffic (About Google Trends, 2009). Given these limitations, we found the search trend data produced by Google to approximate the seasonality and geographic distribution previously identified in Lyme disease. Further refinement of technologies to cap-
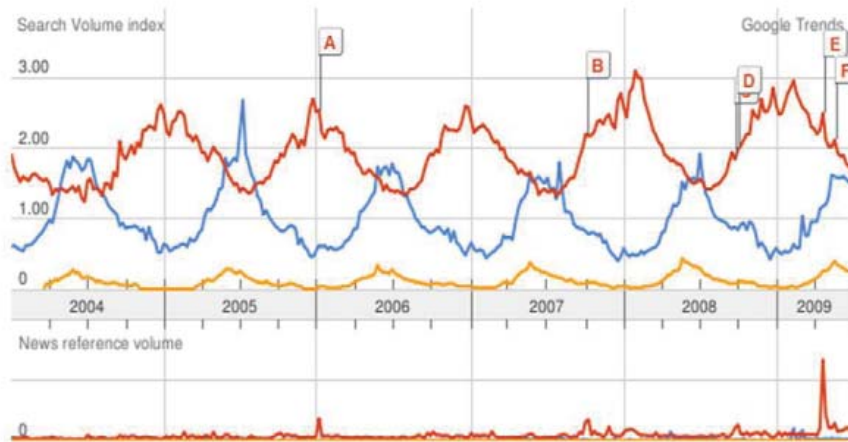
Fig. 2. Google Trends graph depicting tendency over time to search for "Lyme disease", "tick bite", and "cough" (http://www.google.com/trends)[a,b,c]

[a]Lyme disease (blue), tick bite (orange) and cough (red); [b]Search volume index refers to the number of searches for these terms, relative to the total number of searches done on Google over time. News reference volume refers to the number of times these terms appeared in a Google News story (About Google Trends, 2009); [c]A-F represent links to selected Google News stories (information not shown).

ture and utilize search trend data may not only aid outbreak surveillance but also identify more subtle geographic spread in incidence of endemic diseases such as Lyme, particularly within the context of shifting tick habitats as a result of global warming. In Lyme disease, this type of data could inform other research avenues, such as the seasonal variability of tick bite exposure and the relationship between the timing of events such as tick bites and symptom onset. The popularity of search engines such as Google and the increasing availability of health-related information on the Internet have the potential to create valuable resources for further monitoring a broad range of public health concerns.

References

About Google Trends, 2009. http://www.google.com/intl/en/trends/about.html#7 (accessed on 16 July 2009).

Bacon RM, Kugeler KJ, Mead PS; Centers for Disease Control and Prevention (CDC), 2008. Surveillance for Lyme disease - United State, 1992-2006. MMWR Surveill Summ 57, 1-9.

Dickerson S, Reinhart AM, Feeley TH, Bidani R, Rich E, Garg VK, Hershey CO, 2004. Patient Internet use for health information at three urban primary care clinics. J Am Med Inform Assoc 11, 499-504.

Eysenbach G, 2006. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc, 244-248.

Eysenbach G, 2009. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 11, e11.

Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L, 2009. Detecting influenza epidemics using search engine query data. Nature 457, 1012-1014.

Laurent MR, Vickers TJ, 2009. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 16, 471-479.

Wheeler D, 2006. Google as a pathology portal. Adv Anat Pathol 13, 275-276.

Wilson K, Brownstein JS, 2009. Early detection of disease outbreaks using the Internet. CMAJ 180, 829-831.