

Can Internet Searches Provide Useful Data For Public Health Surveillance?

Farzad Mostashari, MD, MSPH

New York City (NYC) Department of Health and Mental Hygiene (DOHMH), New York, NY

Objective: To evaluate whether trends in internet searches might provide useful data for public health surveillance.

Background: Internet-based technologies have been used to assist in disease surveillance and reporting. The Public Health Agency of Canada operates the Global Public Health Information Network, credited with early notification of many outbreaks (including SARS) through automated multilingual analysis of internet media sources such as news wires and web pages (www.phac-aspc.gc.ca/media/nr-rp/2004/2004_gphin-rmispbk_e.html). An innovative web-based forum (www.RUsick2.msu.edu) collects foodborne illness reports from visitors to a web site and has been used to identify foodborne outbreaks in Michigan (1). Health-related topics are among the most popular Internet searches. Many individuals experiencing symptoms of illness conduct Internet searches prior to seeking medical attention. An early site-based study found limited utility to monitoring of Internet queries (2), but recent developments merit re-examination of the potential of internet searches for public health surveillance purposes.

Methods: The recently released “Google Trends” application (www.google.com/trends) uses a sample of Google web searches to show the trends in the volume of searches for specific terms over time, relative to all internet searches. It permits Boolean searches and negation, and also uses IP addresses from server logs to estimate the cities and regions which had the highest rate of searches for the specified term(s), relative to all searches. The available time period is from January 2004 to the present. Several searches described below illustrate the potential utility of this application for health surveillance.

Results: As an example of the system’s ability to detect enduring differences in disease rates in different communities, a search for the term “leptospirosis” in the United States finds dramatically higher rates of searches from Honolulu, Hawaii, consistent with the epidemiology of the illness in the United States (more than half of all national cases are reported from Hawaii). Atlanta, Georgia had the second highest rate of searches for this term in the United States.

www.google.com/trends?q=leptospirosis&ctab=0&geo=US&date=all

Increases in web searches following publicized disease outbreaks are easily demonstrated. For

example, following a series of nursing home norovirus outbreaks in the Pacific Northwest in April, 2006, a spike in searches for “norovirus” from this region is seen:

www.google.com/trends?q=norovirus&ctab=0&geo=US&date=2006-4

Global surveillance might offer early warning of some health concerns prior to their widespread recognition in the United States. For example, internet searches for “contact lens” increased in Singapore in February 2006, prior to the notification of CDC of the first US cases of contact lens--associated Fusarium keratitis in March, and widespread coverage in April, 2006.

www.google.com/trends?q=contact+lens&ctab=0&geo=SG&date=2006
www.google.com/trends?q=contact+lens&ctab=0&geo=US&date=2006

Even prior to widespread news coverage,

In some instances, it appears that concern regarding a possible bioterrorist event may have resulted in a detectable event. Following large anti-war protests on the Mall in Washington DC in late September, 2005, multiple environmental sensors detected the presence of *Francisella tularensis*. Interestingly, the volume of queries may have increased prior to notification of public health officials and the public on September 30th.

www.google.com/trends?q=tularemia&ctab=0&geo=US&date=2005

Seasonal events can also be observed through this lens. Searches for allergies and ragweed increase in the spring, and allergy and pollen searches increase prominently in the fall. It would also appear that Texas and Oklahoma are leading locales for ragweed:

www.google.com/trends?q=ragweed%2C+allergy+%2C+pollen&ctab=0&geo=US&date=all

Use of this tool requires some care, however, while uncommon words like “croup” readily reveal the expected seasonal pattern of searches, more common words like “cough” or “throat” require modifiers to rule out more common search phrases:

www.google.com/trends?q=cough+-syrup%2C+throat+-deep&ctab=0&date=all&geo=US

Conclusions: Internet searches are a promising tool for public health surveillance. However, prospective spatial cluster analysis for syndromes of interest would require access to data that is not publicly available at the current time.

(1) Wethington H, Bartlett P. Usage and data collection patterns for a novel web-based foodborne-disease surveillance system. *J Environ Health*. 2006 Mar;68(7):25-9

(2) Li CS, Aggarwal C, Campbell M, et al. Site-Based Biosurveillance. *MMWR* September 24, 2004 / 53(Suppl):249