COACTION PUBLISHING

Norovirus disease surveillance using Google search data

Rishi Desai[1]*, Aron Hall[1], Benjamin Lopman[1], Yair Shimshoni[2], Manish Patel[1] and Umesh Parashar[1]

[1]CDC, Atlanta, GA, USA; [2]Google Labs, Tel Aviv, Tel Aviv, Israel

Objective

We compared norovirus syndromic search query trends with existing surveillance data in the United States in order to assess whether GIfS data can be used to monitor norovirus disease.

Introduction

While norovirus is the leading cause of gastroenteritis in the United States, leading to an estimated 21 million illnesses per year (1), timely surveillance data are limited. Google Insights for Search (GIfS) is a new application that allows users to track specific Google search queries during specified time periods and geographic regions (2).

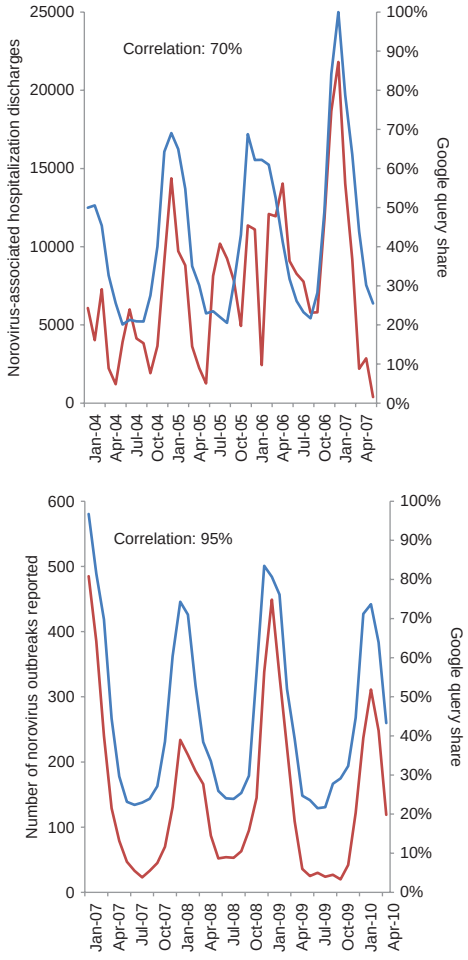

Fig. 1.

Methods

GIfS was used to generate monthly query share totals of norovirus-related searches from January 1, 2004, through April 30, 2010. These totals were compared with estimated U.S. norovirus hospital discharges from January 1, 2004, and June 30, 2007, and with norovirus outbreak surveillance data from January 1, 2007, and April 30, 2010. Peak months were compared for each norovirus season, and trends were compared by means of correlation coefficients.

Results

Between January 1, 2004, and May 1, 2007, correlation between GIfS data and estimated norovirus-associated hospital discharges was moderately strong ($R^2 = 0.70$). Among the 3 full norovirus seasons assessed, GIfS data had the same peak month as hospital discharges in two seasons (2004–2005 and 2006–2007) but differed by 3 months in the 2005–2006 season. Between January 1, 2007, and May 1, 2010, correlation between GIfS data and norovirus outbreak surveillance data was extremely strong ($R^2 = 0.95$). When compared to U.S. norovirus outbreak surveillance data, GIfS data had the same peak month in two seasons (2007–2008 and 2009–2010) and had a peak month 1 month earlier in one season (2008–2009).

Conclusions

GIfS data provide access to real-time information, allowing individual healthcare providers the opportunity to track norovirus trends. Internet-based search data hold promise in settings where traditional surveillance data are limited, absent or simply lack timeliness, but internet access is widespread. As two new surveillance systems for norovirus outbreaks in the United States become better established (1), additional data will become available to help further evaluate GIfS as a complement to traditional surveillance.

Keywords

Norovirus; Google; internet

Acknowledgements

We would like to thank Yossi Matias and NivEfron for their contributions to this work.

References

1. Updated norovirus outbreak management and disease prevention guidelines. MMWR Recomm Rep. 2011;60(RR-3):1–20.
2. Shimshoni Y, Efron N, Matias Y. On the Predictability of Search Trends. Technical report, Google, August 2009. <http://research.google.com/archive/google_trends_predictability.pdf >.

*Rishi Desai
E-mail: butterfinger81@gmail.com

Citation: Emerging Health Threats Journal 2011, **4**: 11012 - DOI: 10.3402/ehtj.v4i0.11012