

Use of Internet Search Data to Monitor Impact of Rotavirus Vaccination in the United States

Rishi Desai,¹ Benjamin A. Lopman,¹ Yair Shimshoni,² John P. Harris,³ Manish M. Patel,¹ and Umesh D. Parashar¹

¹Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia; ²Google Labs, Tel-Aviv, Israel; and ³Health Protection Agency, London, United Kingdom

Google-based Internet query share (IQS) for rotavirus search terms correlated well with US rotavirus laboratory detections from 2004 to 2010 ($r = 0.88$; $P < .001$), capturing the reduction observed during postvaccine years (2008–2010). IQS analysis could become an inexpensive and reliable supplement for monitoring the impact of rotavirus vaccination in the United States.

Over the past 10 years, Internet use has increased dramatically, with the World Bank reporting 231 million Internet users in the United States and 47 million in the United Kingdom as of July 2010 [1]. The majority of users search for information through search engines, with Google attracting 68% of US and 91% of UK search engine market share [2]. Recent studies have shown that Google-based Internet search queries about health correlate strongly with public health surveillance data [3–8]. To our knowledge, Internet search data have not been used to date to monitor the impact of interventions, such as vaccination programs.

The implementation of routine childhood vaccination against rotavirus in 2006 has dramatically reduced severe rotavirus disease in US children [9]. To assess whether search engine data are able to capture the decline in rotavirus disease following vaccine implementation, we compared US Internet query share (IQS) trends for rotavirus-specific terms with national rotavirus laboratory surveillance data from the prevaccine (2004–2006) and postvaccine (2008–2010) years.

As a control for other possible factors that might have resulted in secular trends in IQS, we examined similar data from the United Kingdom, where routine rotavirus vaccination has not been implemented.

METHODS

Rotavirus testing data from sentinel US laboratories ($n = 25$) and UK laboratories ($n = 187$) were examined from 1 January 2004 through 30 December 2010. US laboratories are geographically dispersed (7 in the midwest, 9 in the south, 2 in the northeast, and 7 in the west) and processed an average of 13 208 samples (range, 8979–16,031) annually. The total number of rotavirus tests processed in UK laboratories was unavailable; however, UK laboratories reported an average of 14 291 positive test results (range, 13 020–16,031) annually. The US weekly proportion of positive test results and the UK weekly number of positive test results were reviewed. Google Insights for Search is an Internet application that generates scaled weekly IQS data by dividing all searches performed for selected term(s) by all searches performed in defined regions, and then scaling the weekly IQS data to the highest weekly query share within a defined period [10] (Supplementary Table 1). We hypothesized that caregivers of patients with rotavirus disease would search the Internet for information and that IQS data for the term “rotavirus” and common misspellings identified by IQS (“rota virus,” “roto virus,” “rodo virus,” and “rhoda virus”) would correlate well with US and UK rotavirus laboratory detections.

Weekly IQS trends were aggregated for the 6 rotavirus-related search queries from 1 January 2004 through 30 December 2010 and were compared with national laboratory data in the prevaccine (2004–2006) and postvaccine (2008–2010) years; 2007 was a transitional year with limited vaccine uptake. During the early period after vaccine introduction, the National Immunization Survey does not offer data on uptake of new vaccines recommended for use in infancy. Estimates of rotavirus vaccine coverage are available from other data sources. On the basis of data from 4 US Immunization Information System sentinel sites, after introduction of rotavirus vaccine in February 2006, coverage with ≥ 1 vaccine dose among infants aged 5 months rose quickly to about 50%–60% within the first year (2007) and then steadily thereafter, to 74% by the second quarter of 2009. Pearson correlation coefficients were calculated for IQS data and national laboratory data in the United States and the United

Received 3 November 2011; accepted 17 January 2012; electronically published 14 March 2012.

Correspondence: Rishi Desai, MD, MPH, Centers for Disease Control and Prevention, 1600 Clifton Rd NE, MS-A34, Atlanta, GA 30333 (rdesai1@cdc.gov).

Clinical Infectious Diseases 2012;54(9):e115–8

Published by Oxford University Press on behalf of the Infectious Diseases Society of America 2012.

DOI: 10.1093/cid/cis121

Kingdom. We performed an interrupted time series analysis, using STATA software, release 11 (StataCorp; College Station, TX). The analysis assumed a negative binomial distribution of laboratory reports and controlled for background seasonality and secular trends of IQS by including an indicator variable for month and a sequential variable for time. A 2-tailed significance test ($\alpha = .05$) was performed, and association coefficients and 95% confidence intervals (CIs) were estimated for a range of ± 26 -week lead and lag times. We identified the lead or lag time that maximized the association between laboratory detections and IQS data as measured by the incident rate ratio, which indicates the change in laboratory detections for a 1% IQS increase [11]. Ethics approval was not needed for this study.

RESULTS

Rotavirus IQS searches in the United States and United Kingdom correlated strongly with rotavirus laboratory detections ($r = 0.88$ [$P < .001$] and $r = 0.80$ [$P < .001$], respectively). Rotavirus IQS data peaked within 3 weeks of the peak in laboratory detections of rotavirus in 6 of 7 seasons in the United States (except in 2010, when the IQS peak was 11 weeks before the peak in laboratory detections) and in 6 of 7 seasons in the United Kingdom (except in 2008, when the IQS peak was 4 weeks after the peak in laboratory detections).

In the United States, compared with prevaccine years, the regression model estimated a 39% decrease (95% CI, 28%–48%; $P < .001$) in weekly IQS data and 24% decrease (95% CI, 4%–40%; $P = .023$) in weekly laboratory detections in postvaccine years (Figure 1). In postvaccine years in the United States, search volume and laboratory detections had the strongest association (IRR, 1.96 for a 1% increase in IQS; 95% CI, 1.73–2.23; $P < .001$) when a lead of 1 week was introduced into the IQS data (ie, IQS data from week_{x+1} regressed onto rotavirus laboratory data from week_x). In the United Kingdom, compared with prevaccine years, the regression model estimated a 23% increase (95% CI, 1%–50%; $P < .042$) in weekly IQS data and a 38% increase (95% CI, 16%–65%; $P < .001$) in weekly laboratory detections in postvaccine years.

DISCUSSION

US and UK IQS data correlated strongly with rotavirus laboratory surveillance data, demonstrating a similar seasonal magnitude and peak week in both countries. Furthermore, US IQS data and US laboratory data showed a postvaccine decline that compares well with the well-documented decline in rotavirus disease following vaccine implementation [12]. However, rotavirus vaccine has not been adopted in the United Kingdom,

and both UK IQS data and Health Protection Agency data did not show a decreasing trend, further supporting that the decline in rotavirus IQS and laboratory data in the United States after vaccine introduction are attributable to the effects of the vaccination program. Taken together, these findings demonstrate the utility of Internet search data in monitoring rotavirus disease trends and potentially assessing the impact of vaccination.

Rotavirus disease and testing are largely restricted to children younger than 5 years, so parents/guardians are likely performing rotavirus Internet searches. Increasing familiarity with the term “rotavirus” over time, in part due to vaccine introduction, may explain the higher year-round baseline of Internet searches. Optimal correlation between IQS data and laboratory data occurred when a 1-week lead time was included. This may be due to Internet queries being performed after clinical presentation and laboratory confirmation of disease. In the United Kingdom, increases in IQS data in postvaccine years may be attributable to more available information on the Internet about rotavirus subsequent to vaccine introduction. Rotavirus testing practices in the United Kingdom during prevaccine years have been demonstrated to be a consistently reliable indicator of rotavirus disease burden in the community [13]. However, the increase in laboratory detections in postvaccine years in the United Kingdom is difficult to interpret without corresponding information on total laboratory tests performed and may represent changes in testing or reporting practices.

In addition to healthcare-seeking behavior, some search queries may be responses to media or general information seeking. The most notable disconnect between rotavirus search trends and laboratory data occurred in the United States in spring 2010, when the media reported the presence of porcine circovirus genetic material in rotavirus vaccines [14]. In addition, 2010 was the only year in which the rotavirus IQS data did not peak within 3 weeks of the peak in laboratory detections. The discrepancy was a result of the diminished rotavirus season that made identification of a clear seasonal peak unreliable.

Rotavirus vaccine is known to be most effective at preventing severe disease [15–17], which may result in a trend toward fewer laboratory diagnostic evaluations and more outpatient management. If this shift is occurring, then laboratory surveillance data, which are largely based on testing on inpatients with rotavirus, will increasingly become an unreliable indicator of community rotavirus disease burden. Identifying the determinants of laboratory testing and the motivations for individuals to perform specific searches will be useful in understanding how IQS can serve as an adjunct to traditional rotavirus surveillance.

In conclusion, we show for the first time to our knowledge that Internet search data correlate well with trends in laboratory detections of rotavirus and could be used to assess declines in

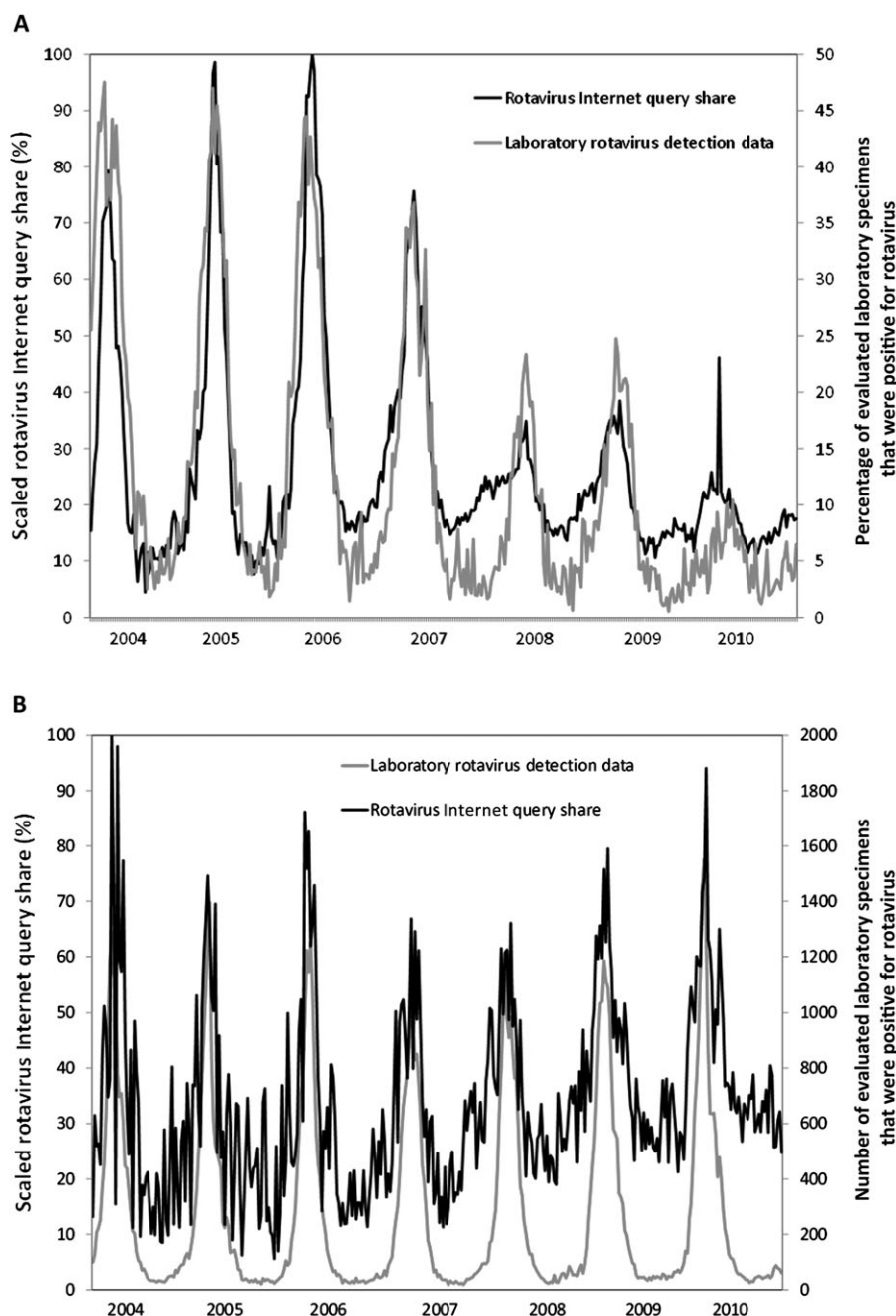


Figure 1. A, Weekly US laboratory data on the percentage of evaluated specimens in which rotavirus was detected (black line) and weekly Internet query share (IQS) rotavirus data for the United States (gray line). B, Weekly UK laboratory data on the number of evaluated specimens in which rotavirus was detected (black line) and weekly IQS rotavirus data for the United Kingdom (gray line). All IQS data are scaled such that the highest point equals 100%, and the remaining data are a proportion of that value.

rotavirus disease after vaccine implementation. Optimal association between IQS and laboratory data occurred with a 1-week IQS lead time, suggesting that Internet searches may occur soon after laboratory confirmation of disease. IQS may complement traditional surveillance systems by offering information that tracks rotavirus disease trends. However, because the availability, robustness, and quality of IQS data

may differ by setting, additional validation of our findings in other countries using rotavirus vaccine is desirable.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online (http://www.oxfordjournals.org/our_journals/cid/). Supplementary materials consist of data provided by the author that are published to benefit the reader.

The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. We thank Jeffrey D. Mutuc, MPH, from the Centers for Disease Control and Prevention (CDC; Atlanta, GA), for his critical review of the manuscript; and Niv Efron and Yossi Matias, from Google Labs (Tel-Aviv, Israel), for their contributions to the design of the Google Insights for Search database. No compensation was given to any of the aforementioned individuals or Google Incorporated. No funding body or sponsor had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. R. D. had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Disclaimer. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the CDC.

Potential conflicts of interest. Y. S. works at Google Incorporated. All other authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. World Bank. World indicators data. Available at: http://data.worldbank.org/data-catalog/world-development-indicators?cid=GPD_WDI. Accessed 26 July 2010.
2. Search market share. Available at: http://thekeyword.co.uk/info/search_engine_market.html. Accessed 15 December 2011.
3. Dukic VM, David MZ, Lauderdale DS. Internet queries and methicillin-resistant *Staphylococcus aureus* surveillance. *Emerg Infect Dis* **2011**; 17: 1068–70.
4. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis* **2009**; 49: 1557–64.
5. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. *CMAJ* **2009**; 180:829–31.
6. 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.
7. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the Web for public health surveillance. *N Engl J Med* **2009**; 360:2153–5, 2157.
8. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. *Clin Infect Dis* **2008**; 47:1443–8.
9. Cortes JC, Tate J, Cortese M, Patel M, Zhou F, Parashar U. Rotavirus vaccine and health care utilization for diarrhea in U.S. Children. *N Engl J Med* **2011**; 365:1108–17.
10. Shimshoni Y, Efron N, Matias Y. On the predictability of search trends. Technical report, Google. **2009** http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en/us/archive/google_trends_predictability.pdf. Accessed 9 February 2012.
11. Armstrong B. Models for the relationship between ambient temperature and daily mortality. *Epidemiology* **2006**; 17:624–31.
12. Centers for Disease Control and Prevention. Reduction in rotavirus after vaccine introduction—United States, 2000–2009. *MMWR Morb Mortal Wkly Rep* **2009**; 58:1146–9.
13. Atchison CJ, Lopman BA, Harris CJ, Tam CC, Iturriza Gómara M, Gray JJ. Clinical laboratory practices for the detection of rotavirus in England and Wales: can surveillance based on routine laboratory testing data be used to evaluate the impact of vaccination? *Euro Surveill* **2009**; 14:19217.
14. Victoria JG, Wang C, Jones MS, et al. Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. *J Virol* **2010**; 84:6033–40.
15. Staat MA, Payne DC, Donauer S, et al. Effectiveness of pentavalent rotavirus vaccine against severe disease. *Pediatrics* **2011**; 128: e267–75.
16. Boom JA, Tate JE, Sahni LC, et al. Effectiveness of pentavalent rotavirus vaccine in a large urban population in the United States. *Pediatrics* **2010**; 125:e199–207.
17. Cortese MM, Leblanc J, White KE, et al. Leveraging state immunization information systems to measure the effectiveness of rotavirus vaccine. *Pediatrics* **2011**; 128:e1474–81.