

ABSTRACT

Comparison of the US outpatient influenza-like illness surveillance network and the Google Flu Trends from 2008 to 2010

A Fowlkes¹, M Mohebbi², S Cook², C Conrad², L Brammer¹, and L Finelli¹

¹Influenza Division, Centers for Disease Control and Prevention, Atlanta, GA, USA; and ²Google.org, New York City, NY, USA
 E-mail: ahl4@cdc.gov

Objective

This paper compares estimates of influenza-like illness (ILI) activity with estimates from the Centers for Disease Control’s ILINet from October 2008 through March 2010.

Introduction

http://Google.org developed a regression model that used the volume of influenza-related search queries best correlated with the proportion of outpatient visits related to ILI model to estimate the level of ILI activity.¹ For calibration, the model used ILINet data from October 2003 to 2009, which report weekly ILI activity as the percentage of patient visits to health care providers for ILI from the total number

patient visits for the week. Estimates of ILI in 121 cities were added in January 2010.

Methods

For the 2008–2009 and 2009–2010 influenza seasons, we used ILINet data to compare the weekly percentage of outpatient ILI visits with the web query-based model estimates of ILI from Google Flu Trends. Data from Google Flu Trends are publicly available through the website. The time period for evaluation was parsed according to the emergence of pandemic influenza A (H1N1) virus (pH1N1): ‘pre-pandemic’ was defined as October 2008 through March 2009, ‘wave one’ as April through July 2009, and ‘wave two’

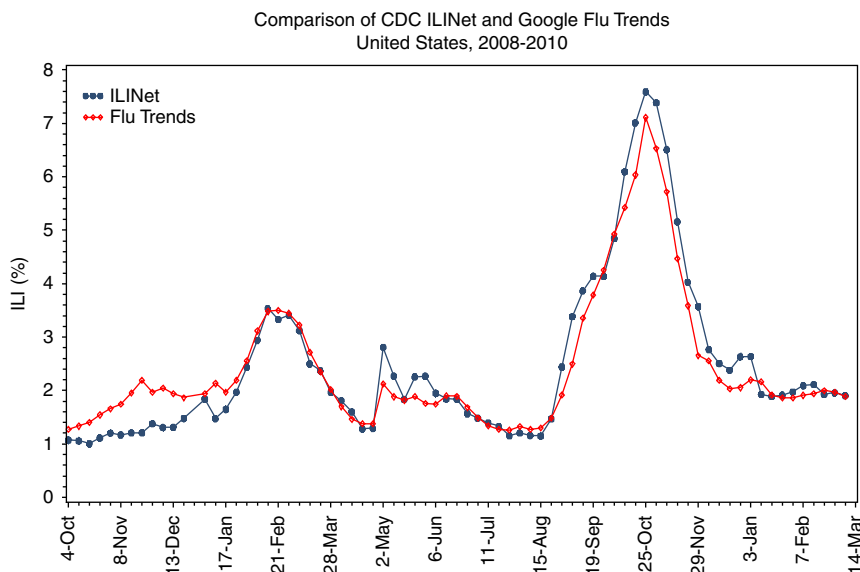


Figure 1 Comparison of Google Flu Trends estimates of influenza-like illness (red) with ILINet estimates from October 2008 to March 2010.

as August 2009 through April 2010. A correlation coefficient was calculated for the nation, US Department of Health and Human Resources-defined regions, and states, and for 47 major metropolitan areas. Correlations below zero were displayed as zero.

Results

From October 2008 through April 2010, all national and regional ILI percentage estimates reported by ILINet and Google Flu Trends correlated highly (Figure 1). During each of the study time periods, the median correlation was 0.93 for national and regional comparisons. In the state-specific comparisons, the median correlation during the pre-pandemic season was 0.82, including two states with zero correlation (range 0–0.94). During wave one the median state correlation was 0.65, including four states with zero correlation (range 0–0.99). During wave two the median state correlation was 0.94 (range 0.68–0.99). In the major metropolitan areas, the median city correlation was 0.73 (range 0–0.95) during the pre-pandemic period, 0.48 (range 0–0.93) during wave one, and 0.89 (range 0.18–0.99) during wave two.

Conclusions

During periods of widespread influenza circulation, the percentage of ILI-related outpatient visits determined by ILINet and estimated by Google Flu Trends correlated well for national, regional, state and many major metropolitan-area levels. After the introduction of pH1N1, but before widespread circulation, state and metropolitan area estimates of ILI activity were less correlated. Google's experimental estimates of ILI activity in major metropolitan areas correlated well with provider-based surveillance data, but are less well correlated than the more robust national and regional data.

Acknowledgements

This paper was presented as a poster at the 2010 International Society for Disease Surveillance Conference, held in Park City, UT, USA, on 1–2 December 2010.

Reference

- 1 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.