Open Access

Correlation of "Google Flu Trends" with Sentinel Surveillance Data for Influenza in 2009 in Japan

Koji Wada*, Hiroshi Ohta and Yoshiharu Aizawa

Department of Public Health, Kitasato University School of Medicine, 1-15-1 Kitasato, Sagamihara, Kanagawa 252-0374, Japan

Abstract: Google Flu Trends (http://www.google.org/flutrends/) (GFT) aggregates Google search data to estimate flu activity in 28 countries. This study explored the correlation of GFT with Japanese national sentinel surveillance data in 2009. We obtained GFT and national sentinel surveillance data for influenza in all 47 Japanese prefectures from 29 June–31 Dec 2009. Pearson correlation coefficients were calculated between GFT and sentinel surveillance data in each prefecture. Multiple regression analysis was also used to determine associations between them. Correlation coefficients were greater than 0.9 between GFT and the sentinel surveillance data in Tokyo, Kanagawa, Aichi, Osaka, and Hyogo, which have relatively large populations. Peaks for GFT were 2 weeks earlier than sentinel surveillance in these prefectures. Multiple regressions showed that only Tokyo ($r = 0.97$; $P < 0.01$) was significantly associated with GFT. People's interest in influenza affects GFT data, which could reflect influenza spread in mega-cities such as Tokyo. Further development of technology to identify prefecture-wide epidemic patterns could offer more accurate information on influenza spread in Japan.

Keywords: Influenza, Google flu trend, sentinel surveillance, Japan.

INTRODUCTION

Internet-based infectious disease surveillance using search engine query data is thought to be a means of estimating disease spread [1]. Google Flu Trends (http://www.google.org/flutrends/) (GFT), developed by Google.org, aggregates Google search data on influenza-like illnesses (ILI) and provides data—free to the public— for 20 countries on a real-time basis, 7–14 days before conventional surveillance systems [2, 3]. This information could be useful in developed countries which have sufficient internet traffic and calibration of search terms. The method is also expected to apply to other infectious diseases [4].

In Japan, the national surveillance system for infectious diseases collects the number of patients who visit sentinel hospitals in all 47 prefectures in a given week, and present with symptoms of designated infectious diseases, such as fever, cough and upper respiratory symptoms [5,6]. Surveillance for influenza is conducted in about 5,000 sentinel hospitals in Japan. It usually takes 2 weeks from initial intake of information for the results to be collated and disseminated. A faster means of identifying infectious disease trends is urgently needed.

Google.org developed GFT based on the assumption that queries about influenza-like illness would correlate with influenza incidence, and in fact, past seasonal influenza queries have correlated well with national influenza surveillance data in the United States [3]. Other studies compared GFT with national surveillance data in New Zealand, Australia, and Europe, showing relatively fair correlation for GFT in these countries for seasonal influenza [7-9]. However, similar studies have not been done for Japan. In 2009, the spread of pandemic 2009 (H1N1) followed a different pattern from seasonal influenzas over the previous decade [5]. This became a matter of wide public attention due to fears of virulence and vaccination priorities; as a result, Google queries on this topic—the foundation of GFT—could be affected. This study explored the correlation of GFT with Japanese national sentinel surveillance data in 2009.

METHODS

We downloaded the available data for Japan in 2009 from the GFT site from 29 June (week 27) to 31 Dec, 2009 (week 53) [2]. We also obtained national sentinel surveillance influenza data for each prefecture [5]. Sentinel medical institutions are required to report weekly on the numbers of ILI outpatients to the Japanese government through local public health centers. Prefectural governments appoint some sentinel medical institutions in the areas as randomly and as representatively as possible. In Japan, there are about 5,000 medical institutions, with about 3,000 for pediatrics and about 2,000 for internal medicine. The sentinel surveillance definition for ILI requires that either patients meet four criteria—acute onset, fever over 38°C, upper respiratory syndrome, and systemic symptoms including general fatigue—or have serologically confirmed detection of pathogens [6]. The sentinel surveillance data are shown as the mean number of patients with the criteria of influenza per hospital in a week.

We analyzed the Pearson correlation coefficient of GFT in Japan with the data from the national sentinel surveillance in each prefecture. Multiple regression analysis was then

*Address correspondence to this author at the Department of Public Health, Kitasato University School of Medicine, 1-15-1 Kitasato, Sagamihara, Kanagawa 252-0374, Japan; Tel: +81-42-778-9352; Fax: +81-42-778-9257; E-mail: kwada-sgy@umin.ac.jp

Fig. (1). Map of Japan, showing Aichi, Hokkaido, Hyogo, Kanagawa, Okinawa, Osaka, and Tokyo prefectures.

used to explore associations between them. All analyses were performed using the Statistical Package for the Social Sciences (SPSS) for Windows, ver. 15.0 [10].

RESULTS

Prefectures which have Pearson correlation coefficients greater than 0.9 between GFT and the national sentinel surveillance of ILI data based on the number of patients at the designated hospitals in each prefecture per week were Tokyo (r = 0.97), Kanagawa (r = 0.95), Aichi (r = 0.92), Osaka (r = 0.96), and Hyogo (r = 0.93). Fig. (**1**) shows the location of these prefectures on a map of Japan. In the other 42 prefectures, the correlation coefficient range between sentinel surveillance in each prefecture and GFT was 0.31–0.89.

Fig. (**2**) shows the weekly GFT rate for Japan and weekly sentinel surveillance data showing the average number of patients with the defined criteria of influenza per hospital in Tokyo, Kanagawa, Aichi, Osaka, and Hyogo from 29 June–31 Dec in 2009. The initial increase in the weekly GFT rate was very closely tied to the first infection spread, which occurred in Okinawa, Japan in week 33rd (10-16Aug). Peaks of ILI and GFT rates occurred in week 42 (12–18 Oct) and 43 (19–25 Oct). According to sentinel surveillance, in Tokyo, Kanagawa, Osaka, and Hyogo, the peak of ILI among hospital-counted patients occurred at week 44 (26 Oct–1 Nov). In Aichi, ILI patients peaked in week 46 (9 Nov–15 Nov), with the number of hospital-counted patients there nearly as high at week 44. The number of ILI patients in Hokkaido and Okinawa peaked before week 44.

Multiple regression analysis showed that only Tokyo ($\beta = 0.97$; $P < 0.01$) was significantly associated with GFT in Japan. Adjusted r^2 was 0.84. Sentinel surveillance data in other prefectures were not significantly associated with the GFT if the data for Tokyo was included.

DISCUSSION

Google Flu Trends in Japan are congruent with sentinel surveillance data for mega-cities such as Tokyo. Google Flu Trends could signal early warning of infection spread in Japan. However, it did not reflect sentinel surveillance in all of Japan with regards to the trend of influenza spread in 2009. Google Flu Trend's peak in 2009 was two weeks earlier than national sentinel surveillance data in Tokyo, Kanagawa, Aichi, Osaka, and Hyogo.

In week 34 (17-23Aug), influenza patients in Okinawa prefecture peaked; on 15 August the first death due to pandemic (H1N1) 2009 was identified [11].“Google insights for search” (http://www.google.com/insights/search/#) finds search volume patterns across specified terms for each prefecture in Japan [12]. Google queries on influenza-related words peaked in these periods in Okinawa. Media reports all over Japan covered the spread of influenza in Okinawa, because it was the first peak of pandemic (H1N1) 2009 in the nation. This coverage would trigger increased the GFT, even though spread of influenza was not identified in that period in any prefecture except Okinawa.

Tokyo, Kanagawa, Aichi, Osaka, and Hyogo, which have a Pearson correlation coefficient greater than 0.9 between GFT in Japan and prefecture sentinel surveillance data, are the most populous prefectures in Japan; Tokyo (12.9 million), Kanagawa (9.0 million), Osaka (8.8 million), Aichi (7.4 million), and Hyogo (5.6 million) in 2009 [13]. Since GFT counts queries on influenza, the data could depend on population size, or size of population with Internet access. Only Okinawa, the southern-most prefecture, and Hokkaido, the northern-most prefecture, had peaks of ILI patients before week 44 (26 Oct - 1 Nov). Thirty-nine out of the 47 prefectures had peaks of ILI patients after week 45 (2 Nov- 8 Nov) week. Infectious diseases often starts to spread in high population density areas, and the 2009 influenza infection

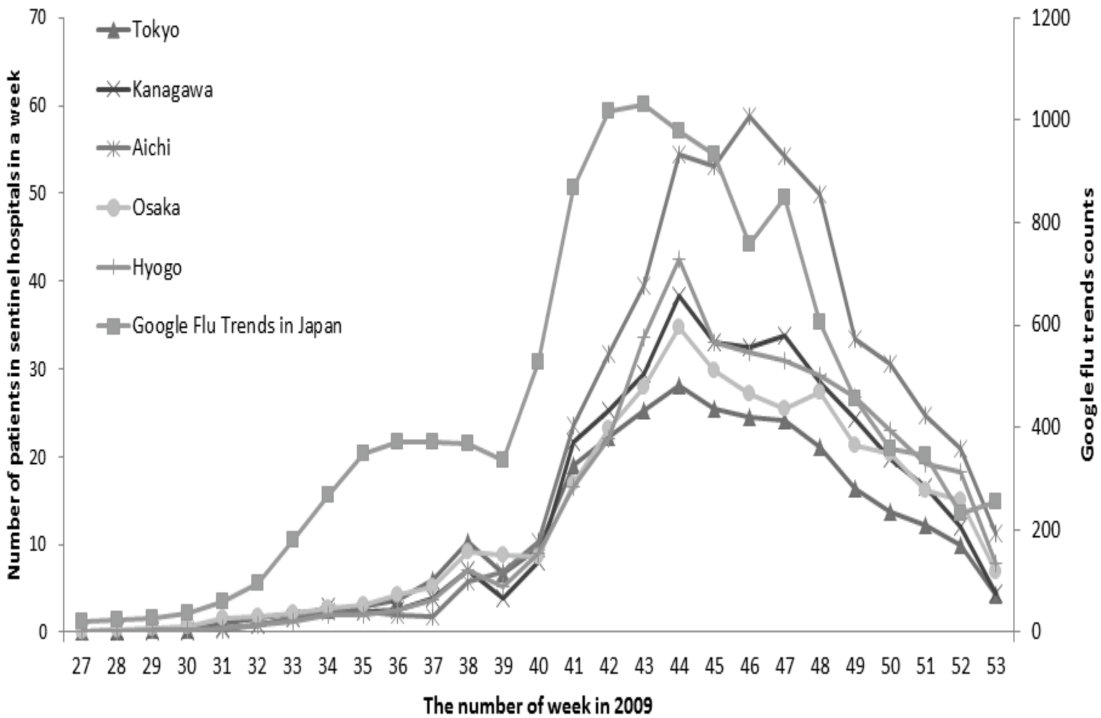

Fig. (2). Comparison of Google Flu Trends and sentinel surveillance data in Tokyo, Kanagawa, Aichi, Osaka, and Hyogo from the week 27 to week 53 in 2009.

pattern started from these areas. In addition, Tokyo and Osaka are the two biggest cities in Japan, while Kanagawa, Aichi, and Hyogo are suburb prefectures. Many people commute between these areas via airplane and Shinkansen bullet train, possibly contributing to ILI patient peaks in the same periods.

Google Flu Trends showed peaks in weeks 42 and 43 (12–25 Oct) in 2009, even though the national sentinel surveillance data in Tokyo, Kanagawa, Osaka, and Hyogo showed the peak of pandemic (H1N1) 2009 occurring in week 44 (26 Oct–1 Nov). The time lag of the peak could be explained by peoples' interests and behaviors [1]; mass media reports began widespread coverage of influenza in these major prefectures, after which people checked for influenza information on web sites. In New Zealand, GFT inquiries peaked a week before and a week after case peaks identified by two independent national surveillance systems for ILI patients [7]. Another possible reason of the time lag of the peak is that "words" selected on influenza for GFT. The words on influenza for GFT were selected in the study for ensuring validation with the national sentinel surveillance data in the United States [3]. We contacted the Google.org to inquire as to the Japanese words selected for GFT in Japan; however, we were unable to obtain the answer. Follow-up research will be needed to validate GFT's ability to identify the same peak pattern for ILI patients.

Regression analysis showed that surveillance data in Tokyo was significantly associated with GFT in Japan, possibly because of the limited amount of sample data. However, in Tokyo, there are many people who can search for influenza on web sites, possibly causing GFT to be affected by the number of persons who can access to the Internet. Further technological development to pinpoint areas accessed could improve GFT's accuracy in each area of Japan. In the United States, GFT shows trends for each state. It therefore seems technologically possible for GFT to identify trends for each prefecture in Japan.

There are a few limitations of this study. We assessed the data only for 2009. The pandemic 2009 (H1N1) showed an unusual pattern of infection spread in Japan. For at least the prior 10 years before 2009, seasonal influenza cases peaked in January or February. Moreover, in 2009, people were more interested in influenza because of uncertainty of the virulence of pandemic 2009 (H1N1). Further follow-up is needed to determine the correlation of GFT. The correlation of the national sentinel surveillance for infectious diseases in Japan should also be evaluated.

In conclusion, GFT could reflect influenza spread in mega-cities such as Tokyo. Public interest in influenza could affect GFT data. Further technological development of means to identify prefecture-wide epidemic patterns could offer more accurate information on how influenza spreads in Japan.

ACKNOWLEDGEMENT

None Declared.

CONFLICT OF INTEREST

None Declared.

REFERENCES

[1] Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009; 457(7232): 1012-4.

[2] Google.org flu trends. Available from: http://www.google.org/flutrends/intl/en_gb/jp/data.txt [Accessed 10 May 2010].

[3] Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009; 49: 1557-64.

[4] Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron A. More diseases tracked by using Google Trends. Emerg Infect Dis 2009; 15: 1327-8.

[5] Infectious Disease Surveillance Center, Japan. Available from: http://idsc.nih.go.jp/iasr/index.html [Accessed 10 May 2010].

[6] Yamashita K. Influenza surveillance in Japan. Nippon Rinsho 2006; 64: 1789-94.

[7] Wilson N, Mason K, Tobias M, Peacey M, Huang QS, Baker M. Interpreting Google flu trends data for pandemic H1N1 influenza: the New Zealand experience. Euro Surveill 2009; 14(44) pii: 19386.

[8] Kelly H, Grant K. Interim analysis of pandemic influenza (H1N1) 2009 in Australia: surveillance trends, age of infection and effectiveness of seasonal vaccination. Euro Surveill 2009; 14(31) pii: 19288.

[9] Valdivia A, Lpez-Alcalde J, Vicente M, Pichiule M, Ruiz M, Ordobas M. Monitoring influenza activity in Europe with Google Flu Trends: comparison with the findings of sentinel physician networks - results for 2009-10. Euro Surveillance, 2010;15(29): pii=19621. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19621

[10] SPSS: SPSS for Windows version 15.0. Chicago: SPSS Inc; 2007.

[11] Shimada T, Toyokawa K, Sunagawa T, Taniguchi K, Okabe N. Epidemiological survey of pandemic (H1N1) 2009 in Okinawa-prefecture (in Japanese). Available from: http://idsc.nih.go.jp/disease/swine_influenza/pdf09/Okinawa_091409.pdf [Accessed 10 May 2010].

[12] Google Insights for Search. Available from: http://www.google.com/insights/search. [Accessed 10 May 2010].

[13] Statistical Bureau, Ministry of Internal Affairs and Communications. Population and households. Available from: http://www.stat.go.jp/english/data/nenkan/1431-02.htm [Accessed 10 May 2010].

Received: July 6, 2011 Revised: August 17, 2011 Accepted: August 17, 2011