Interpreting "Google Flu Trends" data for pandemic H1N1 influenza: The New Zealand experience

N Wilson (nick.wilson@otago.ac.nz)[1], K Mason[2], M Tobias[2], M Peacey[3], Q S Huang[3], M Baker[1]
1. Department of Public Health, University of Otago, Wellington, New Zealand
2. New Zealand Ministry of Health, Wellington, New Zealand
3. WHO National Influenza Centre, Institute of Environmental Science and Research Limited (ESR), Wellington, New Zealand

This article was published on 5 November 2009.
Citation style for this article: 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. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19386

For the period of the spread of pandemic H1N1 influenza in New Zealand during 2009, we compared results from *Google Flu Trends* with data from existing surveillance systems. The patterns from *Google Flu Trends* were closely aligned with (peaking a week before and a week after) two independent national surveillance systems for influenza-like illness (ILI) cases. It was much less congruent with (delayed by three weeks) data from ILI-related calls to a national free-phone Healthline and with media coverage of pandemic influenza. Some patterns were unique to *Google Flu Trends* and may not have reflected the actual ILI burden in the community. Overall, *Google Flu Trends* appears to provide a useful free surveillance system but it should probably be seen as supplementary rather than as an alternative.

The website *Google Flu Trends*, developed by Google.org, uses aggregated Google search data on influenza-like illness (ILI) symptoms to estimate influenza activity "up to two weeks faster than traditional systems" [1]. As of mid-October 2009, the site graphically presents data for Australia, New Zealand, Mexico (selected regions only), the United States (US) and 14 European countries [2]. An analysis of this surveillance system for seasonal influenza data in the US indicated that it was able to "accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day" [3]. For the Australian state of Victoria, the data from *Google Flu Trends* showed a "remarkable correlation" with ILI surveillance data from sentinel practices and the Melbourne Medical Deputising Service [4]. This was for data from May and June 2009 – the time of the spread of new pandemic H1N1 influenza in that state. In fact, the Google data showed an increase in ILI activity five to six weeks prior to the actual increase in reported ILI cases.

As New Zealand has a number of different influenza surveillance systems in operation [5-7], we aimed to further explore the possible utility of *Google Flu Trends* in the setting of an influenza pandemic.

Methods

We downloaded the freely available data for New Zealand in 2009 from the *Google Flu Trends* website [1] from the week beginning 29 March (week 14) to the week beginning 4 October 2009. Data were for the 'Google search ratio', a metric developed by Google and based on Google searches for ILI symptoms that were calibrated against past seasonal influenza data reported through the specific surveillance system(s) in a given country. These data were then compared graphically with ILI data from a national network of sentinel general practices (Sentinel GP system) and another much larger national network of computerised general practices (HealthStat). A comparison was also made with ILI data from a national free-phone Healthline. These systems have all previously been described in *Eurosurveillance* [5]. Of note is that in the graphs the 'weeks' are shifted by one day against those used for *Google Flu Trends:* the reporting week in *Google Flu Trends* starts on Sunday, while the HealthStat week starts on the day before (Saturday) and the reporting weeks in the Sentinel GP system and Healthline start on the day after (Monday).

In addition we obtained a weekly tally of media reports relating to the H1N1 influenza pandemic in New Zealand in 2009 by searching the news archive of 'Google news (New Zealand)' [8]. The search used all the following terms together: 'swine' AND 'flu' AND 'Zealand' AND (the phrase) 'Ministry of Health'. Less specific search strategies (e.g. without the phrase 'Ministry of Health') did not return results that were sufficiently specific for local news media reports from New Zealand because there was extensive international media reporting of some early events relating to New Zealand, such as the arrival of a group of symptomatic students in Auckland on a flight from Mexico in late April 2009.

Results

The initial increase in the weekly rate of ILI cases reported from the Sentinel GP system and the increase in the Google search ratio (representing internet searches for ILI symptoms) were very similar and were noted between week 19 (starting 3 May) and week 24, 2009 (Figure 1). However, the Google search ratio peaked a week earlier, in week 28 (starting 5 July) versus week 29.

The comparison with computerised general practice (HealthStat) ILI data gave some indication that the Google search ratio increased initially before the increase in the ILI data (Figure 2). After that, it seemed to lag behind and peaked a week later, in week 28 versus week 27 for HealthStat data.

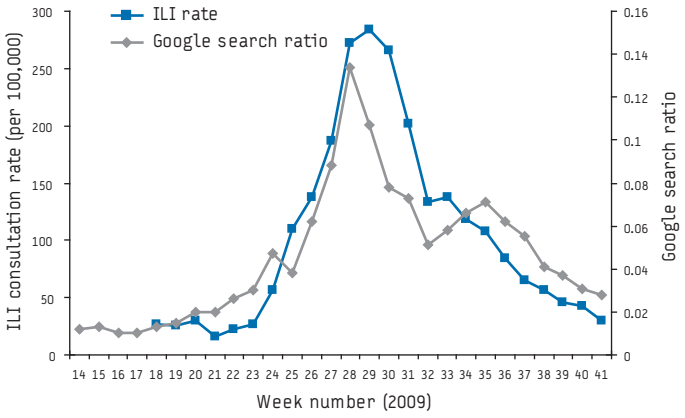
When compared to the ILI calls to the Healthline, there was a similar pattern initially and then a growing gap with the Google search ratio following behind (Figure 3). Indeed, the latter peaked 3 weeks after the peak in ILI Healthline calls (which peaked in week 25 [starting 14 June]).

The comparison with news item media coverage is shown in Figure 4. There appears to be little congruence, especially around the massive peak in media coverage associated with week 18 (starting 26 April) when a group of symptomatic school students returned to New Zealand on a plane from Mexico, the first confirmed cases in New Zealand. There was some similarity in the pattern of increase in week 24 when official reports were of cases first exceeding a total of 1000. But there was no similarity after that point except where both levels declined from week 29 onward.

While a second, smaller peak appears in the Google search ratio in week 35 (starting 23 August), no such peak was seen in the Sentinel GP and HealthStat systems, in the Healthline calls data, or in media items (Figures 1–4).

Figure 1

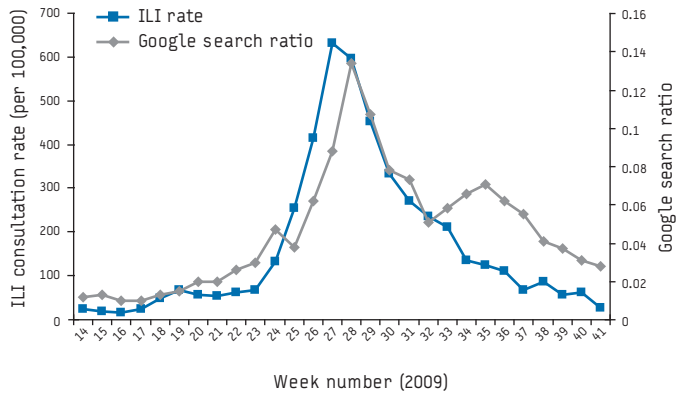
Weekly rate of ILI per 100,000 registered population from the national Sentinel General Practice Surveillance System, compared to the Google search ratio, New Zealand, 29 March – 4 October 2009



ILI: influenza-like illness.

Figure 2

Weekly rate of ILI per 100,000 registered population from the national computerised general practice (HealthStat) surveillance system, compared to the Google search ratio, New Zealand, 29 March – 4 October 2009



ILI: influenza-like illness.

Discussion

Key findings and interpretation

These results suggest that the patterns from the G*oogle Flu Trends* system are fairly congruent with actual surveillance systems for ILI cases in New Zealand. For 2009, these ILI cases were representative of mainly pandemic H1N1 influenza activity, albeit with some minor contribution of seasonal influenza [5]. Furthermore, the week in which the Google search ratio peaked (week 28, starting 5 July) was also the peak week for hospitalisations and admissions for pandemic H1N1 influenza to intensive care units in New Zealand (as detailed elsewhere [5]). Nevertheless, *Google Flu Trends* would not have provided any advance warning of ILI cases compared to the weekly reporting of HealthStat data (neither of the major increase nor the timing of the peak).

The overall similar results with primary care data on ILI are not surprising in that *Google Flu Trends* for New Zealand was initially calibrated on the Sentinel GP surveillance data for seasonal influenza in previous years. But of course the congruence of the two systems with regards to pandemic influenza, has never before been examined for New Zealand.

The fact that *Google Flu Trends* data lagged behind the increase in Healthline ILI-related call levels may reflect the design of the former, being originally calibrated on Sentinel GP surveillance. Another contributing factor could be that symptomatic people used the Healthline before thinking of performing Google searches. This could reflect Ministry of Health promotion (e.g. in media statements) of this national free service as an alternative to people consulting their general practitioner. It might also reflect social patterning of disease spread: If lower-income New Zealanders were at increased risk of influenza early in the pandemic (e.g. household crowding and family size are influenced by socio-economic status), then this group may prefer using Healthline as they have better telephone access than internet access. Healthline callers may also represent individuals who were influenced more by media coverage, but in fact, the major increase in Healthline calls occurred several weeks before the week when the first death attributed to pandemic H1N1 influenza in New Zealand was officially announced (in week 27, starting 28 June) [9]. In the same week, the regular (at least daily) Ministry of Health media release first referred to hospitalised cases of pandemic H1N1 influenza.

Google Flu Trends data might also produce spurious minor patterns that are not mirrored by other systems e.g. the second peak identified in week 35, starting 23 August. This second peak was probably not due to the return to school, as this appears to have occurred earlier during the holiday period and was identified through increased HealthStat consultation rates for school age groups (5–14 years) in weeks 30–32 (the weeks starting 19 June to 2 August) [5].

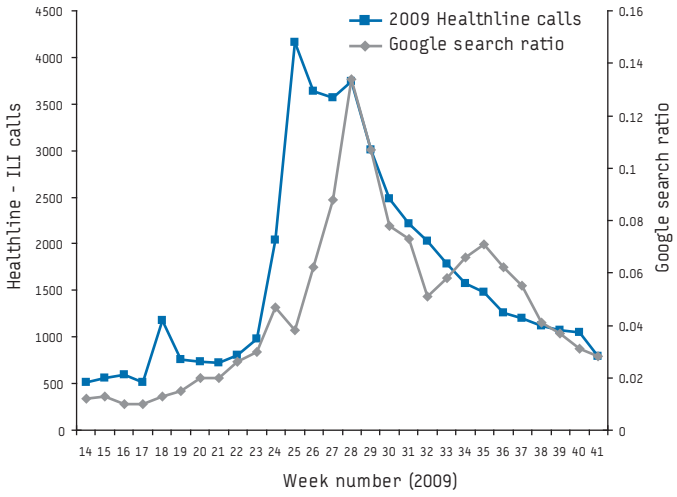
Implications for surveillance and research

A major benefit of *Google Flu Trends* is that it is free and that it is likely to provide some indication of when the incidence of ILI has started to increase in the community and is likely to have peaked. This system also provides daily graphical data and weekly total data that are immediately available to download at the end of each reporting week. This contrasts with an average delay of four days for the GP Sentinel system and four days for HealthStat data (the time for national health authorities to report these data to the rest of the health sector at the end of the data collection week).

Google Flu Trends could be particularly useful for countries where other influenza surveillance systems are poorly developed, though it would probably be less reliable if it had not been calibrated with a robust existing surveillance system for the country in question. Countries with well-established surveillance systems can also potentially profit from *Google Flu Trends* as a supplementary and partial backup surveillance system. In particular, it could assume an important role if the normal systems were disrupted (e.g. in a particularly severe pandemic where health systems are overburdened), or when people with mild illness are discouraged from visiting doctors. *Google Flu Trends* should therefore continue to be closely studied. One question to be addressed is, for example: Does the area under the *Google Flu Trends* epidemic curve reflect the total disease burden in the community (as validated by sero-surveys) better than other surveillance systems?

FIGURE 3

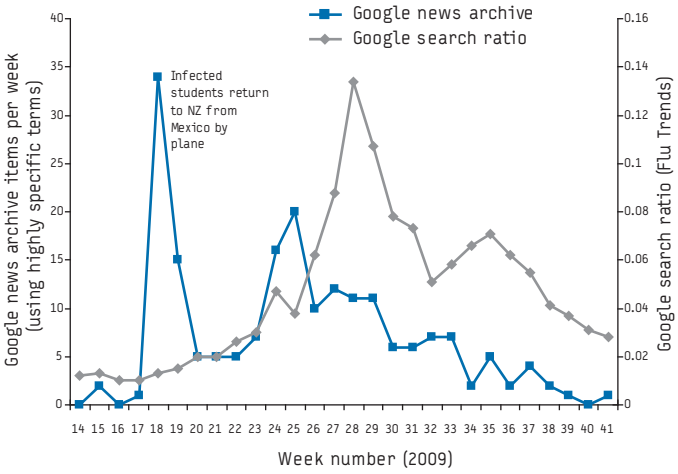
Weekly ILI-related calls to the national Healthline, compared to the Google search ratio, New Zealand, 29 March – 4 October 2009



ILI: influenza-like illness.

FIGURE 4

Weekly news items from the Google news archive related to pandemic H1N1 influenza in New Zealand*, compared to the Google search ratio, New Zealand, 29 March – 4 October 2009



* Retrieved in a search for 'swine AND flu AND Zealand AND "Ministry of Health" '.

Acknowledgements

We thank the many health workers in New Zealand who collect ILI data via the different surveillance systems referred to in this study.

References

1. Google flu trends. Homepage on the internet. Google.org; 2009. Available from: http://www.google.org/flutrends/.
2. Eurosurveillance editorial team. Google flu trends includes 14 European countries. Euro Surveill. 2009;14(40):pii=19352. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19352
3. 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.
4. 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. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19288
5. Baker MG, Wilson N, Huang QS, Paine S, Lopez L, Bandaranayake D, et al. Pandemic influenza A(H1N1)v in New Zealand: the experience from April to August 2009. Euro Surveill. 2009;14(34):pii=19319. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19319
6. Huang QS, Bandaranayake D, Lopez L, Pirie R, Peacey M, Hall R, et al. Surveillance for the 2009 pandemic influenza A (H1N1) virus and seasonal influenza viruses - New Zealand, 2009. MMWR Morb Mortal Wkly Rep. 2009;58(33):918-21.
7. Jackson G, Thornley S. Burden of novel influenza A virus (H1N1) in Auckland and Counties Manukau DHBs (July 2009): a capture-recapture analysis. N Z Med J. 2009;122(1301):66-9.
8. Google news (New Zealand). Advanced news archive search. Homepage on the internet. Available from: http://news.google.com/archivesearch/advanced_search?ned=nz&hl=en
9. Ministry of Health: Influenza A (H1N1) Swine Flu: Media updates. Wellington: Ministry of Health; 2009. Available from: http://www.moh.govt.nz/moh.nsf/indexmh/influenza-a-h1n1-news-media.