Google Trends (GT) related to influenza
Google Trends relacionado à influenza
Google Trends relacionado a la influenza

Viroj Wiwanitkit [1,2,3]

[1] *Surin Rajabhat University, Surin, Thailand.*
[2] *Wiwanitkit House, Bangkok, Thailand.*
[3] *Hainan Medical College, Hainan, China.*

Correspondence
V. Wiwanitkit
Wiwanitkit House, Bangkhae, Bangkok, Thailand.
wviroj@yahoo.com

http://dx.doi.org/10.1590/0102-311XCA020615

Editors,

The recent paper on *Using Google Trends (GT) to Estimate the Incidence of Influenza-Like Illness in Argentina* [1] is very interesting. Orellano et al. studied Google Flu Trends (GFT) and GT with a conclusion regarding "*the utility of GT to complement influenza surveillance*". Indeed, the usefulness of GFT and GT has been mentioned in some earlier reports [2,3]. However, as a computational model, there are several things to be considered in the simulation [4]. Under- or over-estimation can be expected and this is still the present problem in using the Google system for predicting influenza [4]. There is a need for modifications of GT and GFT into a more specific tool that is appropriate for each context. A good example of this is the development of FluBreaks by Pervaiz et al. [5].

1. Orellano PW, Reynoso JI, Antman J, Argibay O. Uso de la herramienta Google Trends para estimar la incidencia de enfermedades tipo influenza en Argentina. Cad Saúde Pública 2015; 31:691-700.
2. Araz OM, Bentley D, Muelleman RL. Using Google Flu Trends data in forecasting influenza-like-illness related ED visits in Omaha, Nebraska. Am J Emerg Med 2014; 32:1016-23.
3. Malik MT, Gumel A, Thompson LH, Strome T, Mahmud SM. "Google flu trends" and emergency department triage data predicted the 2009 pandemic H1N1 waves in Manitoba. Can J Public Health 2011; 102:294-7.
4. Wiwanitkit V. Google Flu for forecasting influenza-like illness. Am J Emerg Med 2014; 32:1417.
5. Pervaiz F, Pervaiz M, Abdur Rehman N, Saif U. FluBreaks: early epidemic detection from Google flu trends. J Med Internet Res 2012; 14:e125.

Submitted on 08/May/2015
Approved on 11/May/2015

The authors reply
Os autores respondem
Los autores responden

Pablo Wenceslao Orellano [1]
Julieta Itatí Reynoso [2]
Julián Antman [3]
Osvaldo Argibay [3]

[1] *Instituto Nacional de Tecnología Industrial, Rosario, Argentina.*
[2] *Hospital Interzonal General de Agudos "San Felipe", Buenos Aires, Argentina.*
[3] *Dirección de Epidemiología, Ministerio de Salud de la Nación, Buenos Aires, Argentina.*

Correspondence
P. W. Orellano
Instituto Nacional de Tecnología Industrial.
Esmeralda y Ocampo, Rosario, Santa Fé 2000, Argentina.
porellano@gmail.com

Expanding the discussion

We appreciate the valuable comments made by Professor Viroj Wiwanitkit about our work and the opportunity to further discuss the Google Flu Trends model (GFT) as well as the methods based on Google Trends (GT) for the estimation of influenza incidence.

The GFT is a model that was developed in 2008 by Google to estimate influenza incidence based on Internet search terms related to the disease. In several countries it has been found to be a good model of performance [1,2,3,4], but problems have arisen when incidence peaks occur [5,6]. In the United States, for example, the original model failed to predict the first peak of the H1N1 pandemic in 2009, and on the other hand overestimated the impact of the 2012/2013 epidemic [5]. Since its original development, there have been numerous changes and updates both in equations and search terms [7]. The reasons for these changes were the differences in timing and intensity between the model results and the actual observations. One possible hypothesis is that these differences may be due to widespread media coverage triggering many flu-related searches by people who were not ill; in any case, there is a clear need to update the current algorithms [8].

To overcome the problems with the incidence estimates, the study by Pervais et al. [9] analyzes the use of different probability distributions such as Poisson or negative binomial, instead of the normal distribution used in the original algorithms. The other approach to the problem is the inclusion of epidemiological surveillance data, developing some form of continuous parameterization of models [10,11].

In countries that do not have GFT estimates, the development of local models based on the GT is possible. In these cases, a model that relates the frequency of Internet search terms and the influenza incidence can be developed locally. After these models are parameterized, estimates can be performed analogously to the GFT. Significant correlations were

observed between the terms of GT and influenza incidence [12,13]. Additionally, the development of local models would enable the inclusion of local variations, and the extension of these methods to other diseases.

Besides models based on Google data alone other methods using big data have been developed. For example, models based on Wikipedia searches have proven to be less sensitive to the increased media reports than models based on web search engines [14]. On the other hand, there are models based on data obtained from Twitter, which have shown good predictive performance [15].

To our knowledge, this is the first published study conducted in Latin America on influenza incidence estimations based on big data analysis models. Both the model based on the GT and the GFT have shown high correlations between the estimated influenza incidence and the epidemiological surveillance data, however regarding the intensity, overestimation was observed with the GFT. In this sense, it should be noted that in Argentina only 6% of private health providers report data to the epidemiological surveillance system [16]. We cannot say for certain whether the GFT is actually overestimating the incidence or seeing cases not detected by the epidemiological surveillance system. What we can say for now is that the dynamics and intensity of the influenza incidence can be adequately modelled using the GT data and the incidence from previous years. The performance of these models is being analyzed for routine use in the epidemiological surveillance system of Argentina as a complement to traditional surveillance activities.

Contributors

All authors contributed equally to the production of the paper.

1. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009; 49:1557-64.
2. 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:pii:19386.
3. Valdivia A, Lopez-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 Surveill 2010; 15:pii:19621.
4. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. PLoS One 2011; 6:e23610.
5. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PLoS Comput Biol 2013; 9:e1003256.
6. Ortiz JR, Zhou H, Shay DK, Neuzil KM, Fowlkes AL, Goss CH. Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends. PLoS One 2011; 6:e18687.
7. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014; 343:1203-5.
8. Butler D. When Google got flu wrong. Nature 2013; 494:155-6.
9. Pervaiz F, Pervaiz M, Abdur Rehman N, Saif U. FluBreaks: early epidemic detection from Google flu trends. J Med Internet Res 2012; 14:e125.
10. Martin LJ, Xu B, Yasui Y. Improving Google Flu Trends estimates for the United States through transformation. PLoS One 2014; 9:e109209.
11. Davidson MW, Haim DA, Radin JM. Using networks to combine "big data" and traditional surveillance to improve influenza predictions. Sci Rep 2015; 5:8154.
12. Cho S, Sohn CH, Jo MW, Shin SY, Lee JH, Ryoo SM, et al. Correlation between national influenza surveillance data and google trends in South Korea. PLoS One 2013; 8:e81422.
13. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013; 8:e55205.
14. McIver DJ, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. PLoS Comput Biol 2014; 10:e1003581.
15. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. PLoS Curr 2014; 6:pii:ecurrents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117.
16. Giovanella L, Feo O, Faria M, Tobar S. Sistemas de Salud en Suramérica: desafíos para la universalidad, la integralidad y la equidad. Rio de Janeiro: Instituto Suramericano de Gobierno en Salud; 2012.

Submitted on 13/May/2015
Approved on 18/May/2015