

Wikipedia: a tool to monitor seasonal diseases trends?

Pascal Vilain*¹, Sophie Larrieu¹, Sébastien Cossin³, Céline Caserio-Schönemann² and Laurent Filleul¹

¹French national public health agency, Regional unit (Cire) Océan Indien, Saint-Denis, Réunion; ²French national public health agency, Paris, France; ³French National Institute of Health and Medical Research, Unit 1219, Bordeaux, France

Objective

To explore the interest of Wikipedia as a data source to monitor seasonal diseases trends in metropolitan France.

Introduction

Today, Internet, especially Wikipedia, is an important part of everyday life. People can notably use this popular free online encyclopedia to search health-related information. Recent studies showed that Wikipedia data can be used to monitor and to forecast influenza-like illnesses in near real time in the United States [1,2]. We carried out a study to explore whether French Wikipedia data allow to monitor the trends of five seasonal diseases in metropolitan France: influenza-like illness, gastroenteritis, bronchiolitis, chickenpox and asthma.

Methods

To collect Wikipedia data, we used two free web applications (<https://stats.grok.se> and <https://tools.wmflabs.org/pageviews>), which aggregate daily views for each French entry of the encyclopedia.

As some articles have several entries (redirects), we collected view statistics for all the article entries and added them to make time series from January 1st, 2009 to June 30, 2016 (Figure 1). Then, we compared these data to those of OSCOUR[®] network, which is a robust national surveillance system based on the emergency departments. For each disease, we modeled daily variations in Wikipedia views according to daily visits in ED using Poisson regression models allowing for overdispersion. The following adjustment variables were included in the model: long-term trend, seasonality, day of the week. We tested several lags (day-7 to day+7) in order to explore whether one of the two indicators (Wikipedia view or ED visits) varied earlier than the other.

Results

The mean number of daily views was 764 [16-8271] for influenza-like illness, 202 [6-1660] for bronchiolitis, 1228 [59-10030] for gastroenteritis, 475 [21-2729] for asthma and 879 [25-4081] for chickenpox. Times series analyses showed a positive association between page views and ED visits for each seasonal disease (Figure 2). For each increase in 100 Wikipedia views, the number of ED visits the same day increased by 2.9% (95% CI=[2.5-3.3]) for influenza, 1.8 (95% CI=[1.4-2.2]) for bronchiolitis, 2.4% (95% CI=[2.2-2.7]) for gastroenteritis, 1.4% (95% CI=[1.0-1.7]) for asthma and 2.9% (95% CI=[1.7-4.1]) for chickenpox. Globally, the highest relative risks were observed for lag-1 (day-1) to lag0.

Conclusions

This study allowed to show that French Wikipedia data can be useful to monitor the trends of seasonal diseases. Indeed, they were significantly associated with data from a robust surveillance system, with a maximum lag of one day. Wikipedia can therefore be considered as an interesting complementary data source, notably when traditional surveillance systems are not available in real time.

Further works will be necessary to elaborate forecasting models for these seasonal diseases.

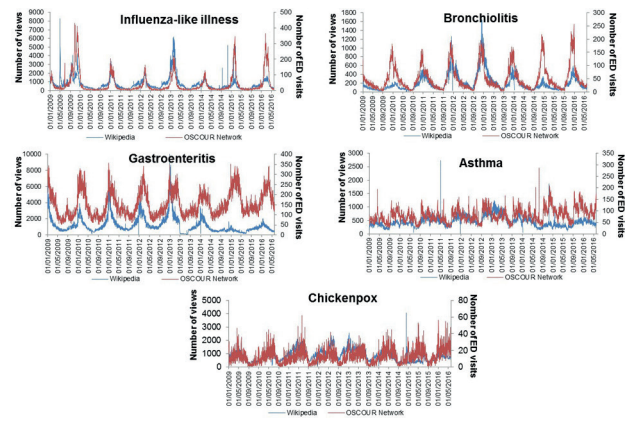


Figure1. Daily number of page views and ED visits for seasonal diseases, January 1st, 2009 to June 30, 2016

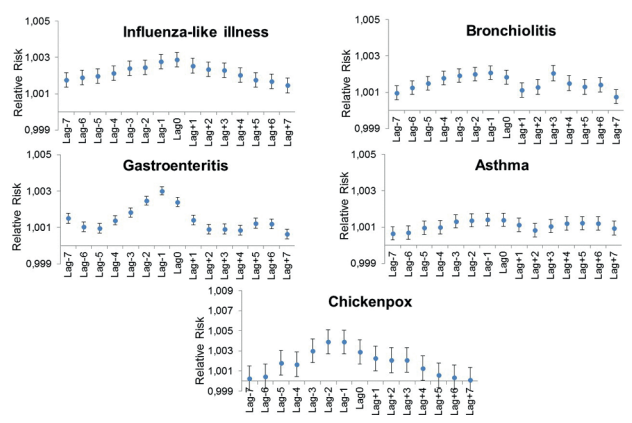


Figure2. Relative risk between Wikipedia page views and ED visits for seasonal diseases by several lags

Keywords

wikipedia; surveillance; seasonal diseases

References

- [1] 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(4):e1003581.
- [2] Hickmann KS, Fairchild G, Priedhorsky R, Generous N, Hyman JM, Deshpande A, Del Valle SY. Forecasting the 2013-2014 influenza season using Wikipedia. *PLoS Comput Biol.* 2015;11(5):e1004239.

*Pascal Vilain
E-mail: pascal.vilain@ars.sante.fr

