

BRIEF COMMUNICATION

Google unveils a glimpse of allergic rhinitis in the real world

M.-G. Kang^{1,2,3}, W.-J. Song^{1,2}, S. Choi^{4,5}, H. Kim⁴, H. Ha⁴, S.-H. Kim^{1,2,3}, S.-H. Cho^{1,2}, K.-U. Min^{1,2}, S. Yoon⁴ & Y.-S. Chang^{1,2,3}

¹Department of Internal Medicine, Seoul National University College of Medicine; ²Institute of Allergy and Clinical Immunology, Seoul National University Medical Research Center, Seoul; ³Division of Allergy and Clinical Immunology, Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Gyeonggi-do; ⁴Department of Electrical and Computer Engineering, Seoul National University; ⁵Department of IT Convergence, Korea University, Seoul, Korea

To cite this article: Kang M-G, Song W-J, Choi S, Kim H, Ha H, Kim S-H, Cho S-H, Min K-U, Yoon S, Chang Y-S. Google unveils a glimpse of allergic rhinitis in the real world. *Allergy* 2015; **70**: 124–128.

Keywords

allergic rhinitis; allergy; epidemiology; Google; pollen count.

Correspondence

Yoon-Seok Chang, MD, PhD, Department of Internal Medicine, Seoul National University Bundang Hospital, 300 Gumi-dong, Bundang-gu, Seongnam, Gyeonggi-do 463-802, Korea.
Tel.: +82-31-787-7023
Fax: +82-31-787-4052
E-mail: addchang@snu.ac.kr

Accepted for publication 28 September 2014

DOI:10.1111/all.12528

Edited by: Marek Sanak

Abstract

Google Trends (GT) is a Web-based surveillance tool used to explore the searching trends of specific queries on Google. Recent studies have suggested the utility of GT in predicting outbreaks of influenza and other diseases. However, this utility has not been thoroughly evaluated for allergic diseases. Therefore, we investigated the utility of GT for predicting the epidemiology of allergic rhinitis. In the USA, GT for allergic rhinitis showed repetitive seasonality that peaked in late April and early May and then rapidly decreased, and a second small peak occurred in September. These trends are highly correlated with the searching trends for other queries such as ‘pollen count’, antihistamines such as loratadine and cetirizine (all $r > 0.88$ and all $P < 0.001$), and even the total pollen count collected from 21 pollen counters across the USA ($r = 0.928$, $P < 0.001$). Google Trends for allergic rhinitis was similar to the monthly changes in rhinitis symptoms according to the US National Health and Nutrition Examination Survey III, sales for Claritin® and all over-the-counter antihistamines, and the number of monthly page views of ‘claritin.com’. In conclusion, GT closely reflects the real-world epidemiology of allergic rhinitis in the USA and could potentially be used as a monitoring tool for allergic rhinitis.

Google Trends (GT) is a Web-based surveillance tool used to visualise the searching trends of specific queries in Google (1). Google Trends has been proved to be effective in predicting outbreaks of influenza (2), later termed Google Flu Trends, and other diseases (3, 4). However, this utility has not been thoroughly evaluated for allergic diseases. Allergic rhinitis (AR) is a common allergic disease characterised by nasal afferent neuronal hypersensitivity to allergen exposure; thus, its seasonal outbreak pattern may be related to pollen exposure (5). Like Google Flu Trends, which are explosively searched during flu outbreaks, the number of Google searches for AR might also fluctuate according to how many allergenic pollens are in the air and how many patients are affected by AR. Therefore, we questioned whether GT has utility in predicting the epidemiology of AR in the real world.

Methods

Google Trends outcomes were extracted for AR search terms in the United States (USA) during the most recent 6 years (2008–2013). First, we examined the seasonal pattern of searching trends for AR on GT in the USA. Next, we compared the outcomes for AR with those of several related terms such as ‘pollen count’, ‘loratadine’ and ‘cetirizine’. Finally, we correlated the GT outcomes for AR with real-world epidemiology data of AR in the USA, such as pollen count data, antihistamine prescription data and the monthly prevalence of rhinitis symptoms from the National Health and Nutrition Examination Survey (NHANES) III database (6). The data for total pollen counts of the 21 pollen stations was generated through the National Allergy Bureau™ of the American Academy of Allergy, Asthma and Immunology.

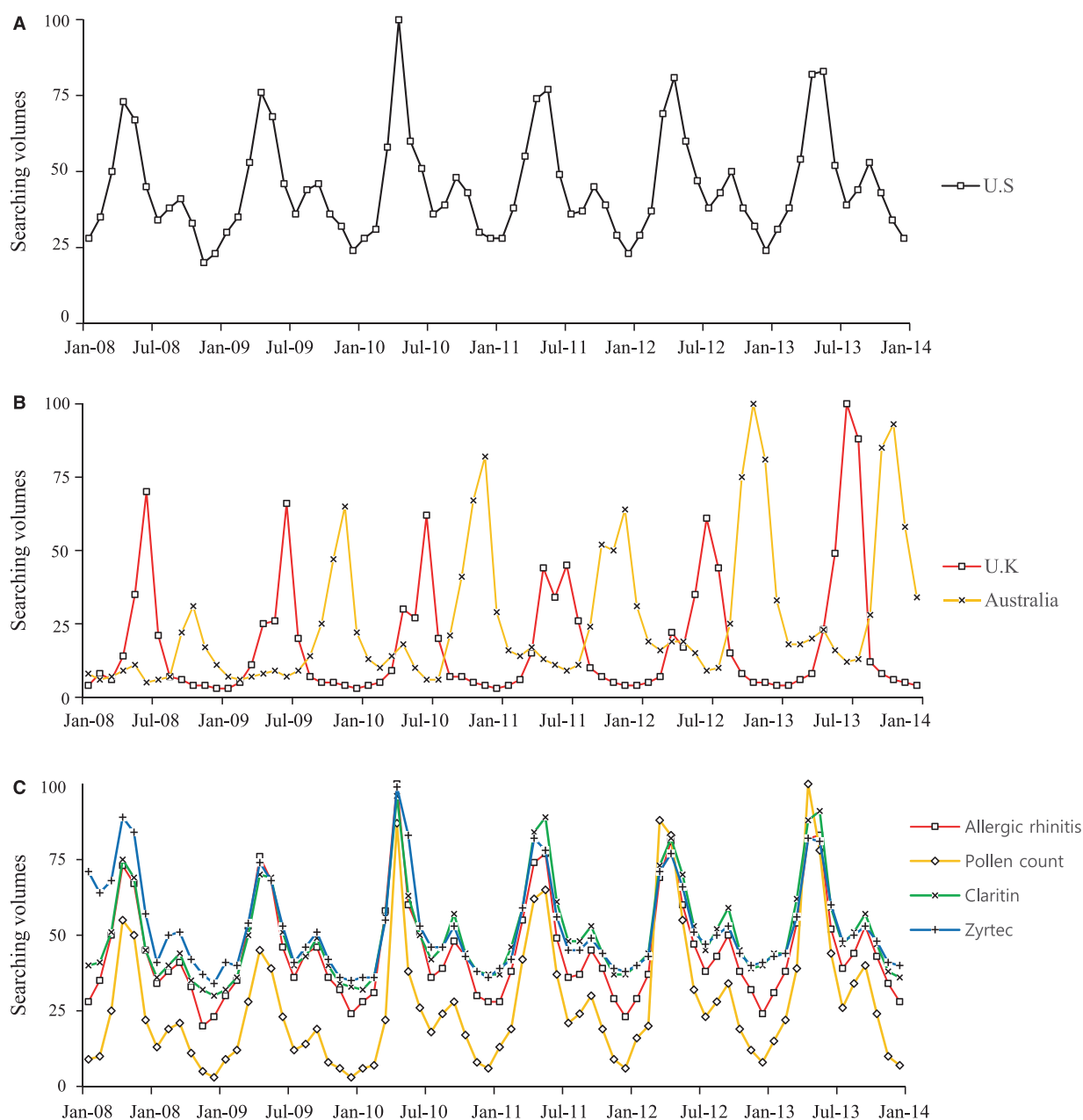


Figure 1 Google Trends for allergic rhinitis. (A) Distinct searching trends for allergic rhinitis in the United States. (B) Searching trends for allergic rhinitis in the United Kingdom and Australia. (C) Similar searching trends for allergic rhinitis and related queries such as pollen count ($r = 0.896$, $P < 0.001$) and antihistamines such as loratadine ($r = 0.914$, $P < 0.001$) and cetirizine ($r = 0.887$, $P < 0.001$) in the United States. From January 2008 to December 2013.

The 21 pollen stations used in this study were described in the supplementary materials. Antihistamine prescription data were retrieved from market sales data for Claritin® (Merck & Co., Inc., Whitehouse Station, NJ, USA) and total over-the-counter (OTC) antihistamine sales obtained from Information Resources, Inc., and the number of monthly page views of the website 'claritin.com' were retrieved from the Compete Site Analytics (Compete, Inc., Boston, MA, USA).

Detailed methods are described in the Supporting Information (Appendix S1).

Results

The search trends for AR on GT showed a distinct seasonal pattern during the most recent 6 years, showing the first peak in late April to early May and the second peak

in September (Fig. 1A). The seasonal outbreak pattern showed geographical differences compared with the United Kingdom or Australia (Fig. 1B). Further analysis of the seasonality of GT for AR in the USA is described in Fig. S2. Next, the seasonal pattern of AR on GT in the USA showed high correlations with several related terms, such as 'pollen count' ($r = 0.896$, $P < 0.001$), 'loratadine' ($r = 0.914$, $P < 0.001$) and 'cetirizine' ($r = 0.887$, $P < 0.001$) (Fig. 1C). Finally, GT outcomes for AR were tested for correlations with real-world epidemiological data. From January 2013 to December 2013, the mean serial total pollen count of the 21 pollen counters across the USA showed significant correlations with GT for AR ($r = 0.928$, $P < 0.001$) (Fig. 2A). The monthly prevalence of nasal symptoms in patients with AR as reported by Singh et al. (6) in the US NHANES III ($r = 0.958$, $P = 0.001$) (Fig. 2B) showed seasonal features similar to those of GT for AR in the USA (Fig. 1A). The monthly sales data of Claritin® or total OTC antihistamines (US data, 2013) also showed significant correlations with GT outcomes for AR (all $r = 0.968$, $P < 0.001$) (Fig. 2C). In addition, GT for AR outcomes showed a significant correlation with the number of

monthly page views on 'claritin.com' during the same period ($r = 0.769$, $P = 0.003$) (Fig. 2D).

Discussion

The key finding of this study is that the Web-searching behaviours for AR collected by GT highly reflect the real-world epidemiology of AR, suggesting the potential utility of GT for predicting the outbreak of AR. The outcomes for AR on GT showed repetitively distinct seasonal patterns in the USA. The term 'allergic rhinitis' was highly correlated with several related terms such as 'pollen count' and various antihistamines on Google search outcomes, suggesting the representative nature of the search term 'allergic rhinitis'. Finally, the practical utility of GT analyses was examined in relation to the real-world epidemiology data of AR.

Google Flu Trends has been proved to be useful in both the USA and the United Kingdom (7), Spain (8) and even China (9). The effectiveness of Google Flu Trends in predicting outbreaks of influenza might be related to the fact that on the Web, people are seeking information about influenza in proportion to the occurrence of influenza epidemics.

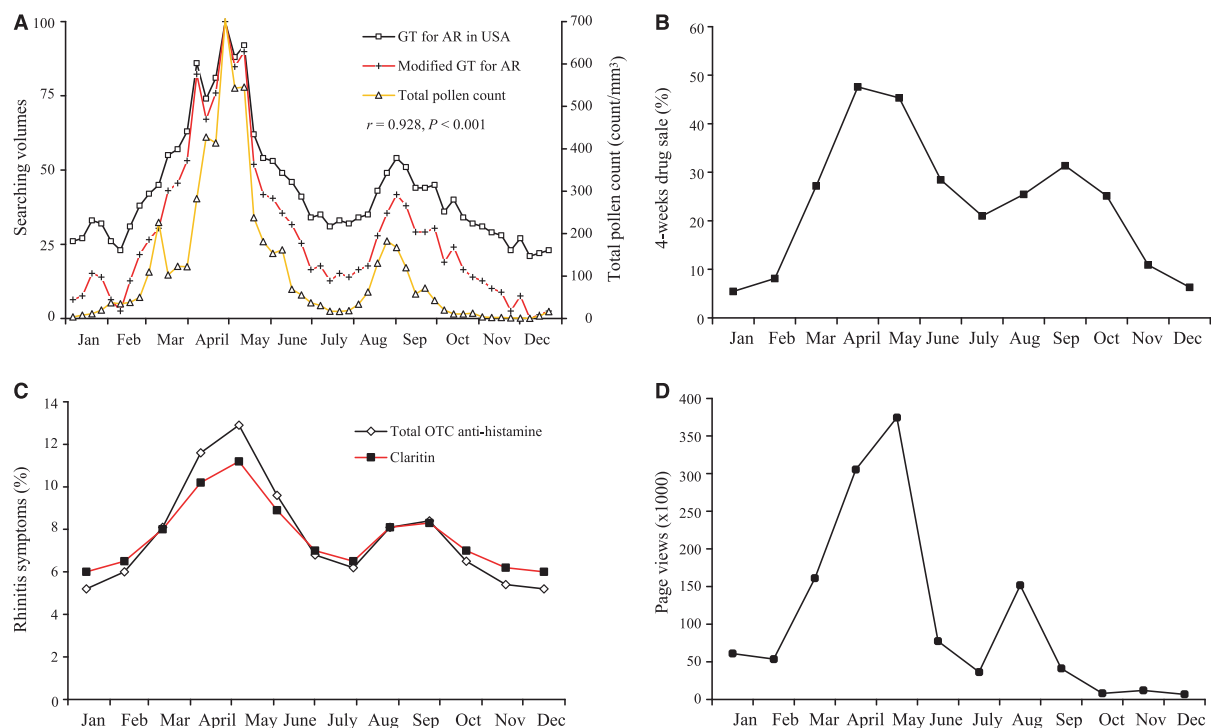


Figure 2 Time series similarity between Google Trends and real-world epidemiologic data for allergic rhinitis in the United States in 2013. (A) Correlations between Google Trends for allergic rhinitis and the total pollen count from 21 centres across the United States. Modified Google Trends for allergic rhinitis was rescaled from 0 to 100, in which minimal searching volume for allergic rhinitis was 21 at 2nd week in December 2013 and was changed to 0; (B) Monthly percentage of patients who experienced symptoms of allergic rhinitis in 2013

according to the third National Health and Nutritional Examination Survey in the United States. [Source: Reproduced from [6]. Reprinted from *Journal of Allergy and Clinical Immunology*, Vol. 126, Singh et al, 'The epidemiology of ocular and nasal allergy in the United States, 1988–1994', Pages 778–783, Copyright 2010, with permission from Elsevier]. (C) Sales of total over-the-counter antihistamines and Claritin® during 2013 in the United States. (D) Monthly number of page views on 'claritin.com' during 2013 in the United States.

In addition, there have been continuing reports that GT is also useful in diseases such as dengue fever (3) and psychological phenomena including depression (10) and suicide (11).

The searching trends for AR reflect the epidemiology of AR in the corresponding countries. Singh et al. (6) reported monthly changes in AR symptoms using data from the US NHANES III. Similar to the situation with GR for AR in the USA, the highest number of patients complaining of AR symptoms occurred between April and May; this number subsequently decreased and was followed by a smaller peak in September (Fig. 2B). OTC medication sales are useful in predicting the epidemiology or outbreaks of specific diseases in a community (12). The epidemiology of AR cannot be accurately estimated by hospital admission or clinical visits alone (13). Because many antihistamine drugs are regarded as OTC drugs, more patients with AR are purchasing OTC antihistamines without prescriptions. In this regard, the patterns of market sales of OTC antihistamines could serve to predict the seasonal trends of AR (14). The monthly market sales data for Claritin® and total OTC antihistamines in the USA showed features similar to those of GT for AR in the USA as well as monthly changes in the nasal symptoms according to the NHANES III.

These close relationships between the GT outcomes and the real-world epidemiology of AR may be due to the health-seeking behaviours of patients. The Web is now a highly accessible tool with which to seek health information, and it was estimated that about 100 million people searched 'allergy' on Google in 2013 (15). Like Google Flu Trends, it may be expected that GT could be further developed as a predictive tool for outbreaks of AR.

Conclusion

To our knowledge, this is the first study to evaluate the potential utility of GT for predicting AR. By analysing large databases such as Google, we can monitor outbreaks of AR in a real-time manner and predict health-seeking behaviours of patients with AR.

References

- Google Trends Lab (31 May 2014). About Google Trends (cited 1 October 2009). Available at: <http://support.google.com/trends>.
- Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009;**457**:1012–1014.
- Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl Trop Dis* 2011;**5**:e1206.
- Rossignol L, Pelat C, Lambert B, Flahault A, Chartier-Kastler E, Hanslik T. A method to assess seasonality of urinary tract infections based on medication sales and Google trends. *PLoS ONE* 2013;**8**:e76020.
- D'Amato G, Cecchi L. Effects of climate change on environmental factors in respiratory allergic diseases. *Clin Exp Allergy* 2008;**38**:1264–1274.
- Singh K, Axelrod S, Bielory L. The epidemiology of ocular and nasal allergy in the United States, 1988–1994. *J Allergy Clin Immunol* 2010;**126**:778–783.
- 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**:2–7.
- Valdivia A, Monge-Corella S. Diseases tracked by using Google trends, Spain. *Emerg Infect Dis* 2010;**16**:168.
- Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. *PLoS ONE* 2013;**8**:e55205.
- Yang AC, Huang NE, Peng CK, Tsai SJ. Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect. *PLoS ONE* 2010;**5**:e13728.

Acknowledgment

The authors thank Tae-Hoon Kim and Ki-Young Park in MSD Korea for providing the IRI (Information Resources, Inc.) sales data for Claritin® and total OTC antihistamine in 2013 in USA. For providing the pollen count of US, the authors thank Jerome Schultz, the board of the National Allergy Bureau of the AAAAI and the head of the pollen counting stations of Madison (WI), Melrose Park (IL), Onalaska (WI), Pleasanton (CA), Rochester (NY), Seattle (WA), Tulsa (OK), Twin Falls (ID), Washington (DC), Waterbury (CT) and York (PA).

Author contributions

Yoon-Seok Chang, Sungroh Youn and Min-Gyu Kang involved in the conception and design of this study. Min-Gyu Kang, Woo-Jung Song and Yoon-Seok Chang involved in the drafting of the project outline. Min-Gyu Kang, Sungwoon Choi, Hanjoo Kim, Heonseok Ha and Woo-Jung Song involved in the analysis and interpretation of data. Woo-Jung Song, Sae-Hoon Kim, Sang-Heon Cho and Kyung-Up Min involved in the critical revision for important intellectual content of the script.

Funding

This work was supported by grant from the Seoul National University Bundang Hospital Research Fund (12-2013-009).

Conflicts of interest

The authors declare that they have no relevant conflicts of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Supplementary methods.

Figure S1. Categorical searches for allergic rhinitis in Google.

Figure S2. Time-series searching trends for allergic rhinitis.

11. Ayers JW, Althouse BM, Allem JP, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. *Am J Prev Med* 2013;**44**:520–525.
12. Andersson T, Bjelkmar P, Hulth A, Lindh J, Stenmark S, Widerstrom M. Syndromic surveillance for local outbreak detection and awareness: evaluating outbreak signals of acute gastroenteritis in telephone triage, web-based queries and over-the-counter pharmacy sales. *Epidemiol Infect* 2014;**142**:303–313.
13. Das D, Metzger K, Heffernan R, Balter S, Weiss D, Mostashari F. Monitoring over-the-counter medication sales for early detection of disease outbreaks—New York City. *MMWR Morb Mortal Wkly Rep* 2005;**54**(Suppl):41–46.
14. Sheffield PE, Weinberger KR, Ito K, Matte TD, Mathes RW, Robinson GS et al. The association of tree pollen concentration peaks and allergy medication sales in New York City: 2003–2008. *ISRN Allergy* 2011;**2011**: 537194.
15. Ryan O. Searching for Allergy Relief: How Marketers Can Help (updated 31 May 2014; cited October 2013). Available at: <http://www.thinkwithgoogle.com/articles/allergy-relief-how-marketers-can-help.html>.

Supplementary methods

1) Google Trends

Google Trends (GT) can visualize the changes of searching volumes of an interested query on a numerical scale (on the Y-axis) for a specific region and over a defined periods (on the X-axis)(1). The numbers on the graph for GT reflect how many searches have been submitted each week for specific query compared to the total searches for all relevant queries during that time. In a given period, the searching trends are represented from 0 to 100, which the searching volumes each week are normalized as a proportion of the highest searching volume which were regarded as 100. GT can localize the searching trends for specific query. By using of internet protocol, it is possible to track the general physical locations where the searches have been done. It is possible to compare the searching trends in different countries and in a city-level if there are sufficient number of searches. Google offers data for searching trends weekly from January 2004, which can be downloaded as a csv files for a given search term, over time period and by location.

2) Categorical search

Google Flu Trends predicts the outbreak of influenza by monitoring the searching volumes of top 45 queries related to influenza. Similarly, Google provides the trend for categorical searches for specific query (2). For example, the categorical search for 'Tokyo' will not only show the result for the keyword 'Tokyo' but will also include the '東京', 'Токио' as well as 'capital of Japan'. Likewise, allergic rhinitis is more commonly called as 'hay fever' in U.K(3), 'heuschnupfen' in German, and '花粉症' in Chinese (4). Therefore, The result of categorical searching for 'allergic rhinitis' not only includes query for 'allergic rhinitis', but also that in other languages, and other descriptions of similar meaning (supplementary figure 1). In this study, we used category search results for specific queries if categorical searches were available.

3) Epidemiologic data for AR

In this study, several epidemiologic data were introduced to prove the effectiveness of GT for predicting the dynamic nature of allergic rhinitis. At first, we used the IRI (Information Resources,

Inc) sales data of Claritin® (MSD Consumers) and total over-the-counter antihistamines in 2013 in U.S. This data were freely provided from MSD (Merck Sharp &Dohme Corp) Korea for the research purpose. Second, questionnaire data for nasal symptom by month (“During which months of the year dose pollen make your symptoms worse?”) were used, which were previously reported from Singh et al. by analyzing the U.S. National Health And Nutrition Examination Survey (NHANES) III(5). Third, we looked the monthly number of page-view for ‘claritin.com’ in 2013, which is the official website of claritin® which were managed by MSD for advertising claritin®.

4) Total pollen count in U.S

We were officially obtained these data from the National Allergy Bureau of the American Academy of Allergy Asthma and Immunology (6). The weekly sum of the total pollen counts of the 21 stations across the U.S were compared with the time-series searching trends for ‘allergic rhinitis in 2013. The 21 pollen stations analysed in figure 2 (a) are following: Baltimore (MD), Bellevue (NE), Berkeley (MO), Colorado Springs (CO), Dayton (OH), Erie (PA), Eugene (OR), Kansas City (MO), Knoxville (TN), Louisville (KY), Madison (WI), Melrose Park (IL), Onalaska (WI), Pleasanton (CA), Rochester (NY), Seattle (WA), Twin Falls (ID), Washington (DC), Waterbury (CT) and York (PA).

4) Statistical analysis

The figures in this article and correlations between queries were drawn with Excel (MicrosoftInc., Bellevue, WA, USA). Spearman correlation coefficients were used to evaluate the potential between GT and total pollen counts in US, the epidemiologic data described above. The seasonal trends of each queries were analyzed by using statistical functions such as autocorrelation, partial autocorrelation and periodogram which were provided from the season packages of the R software (version 3.0.3) (7).

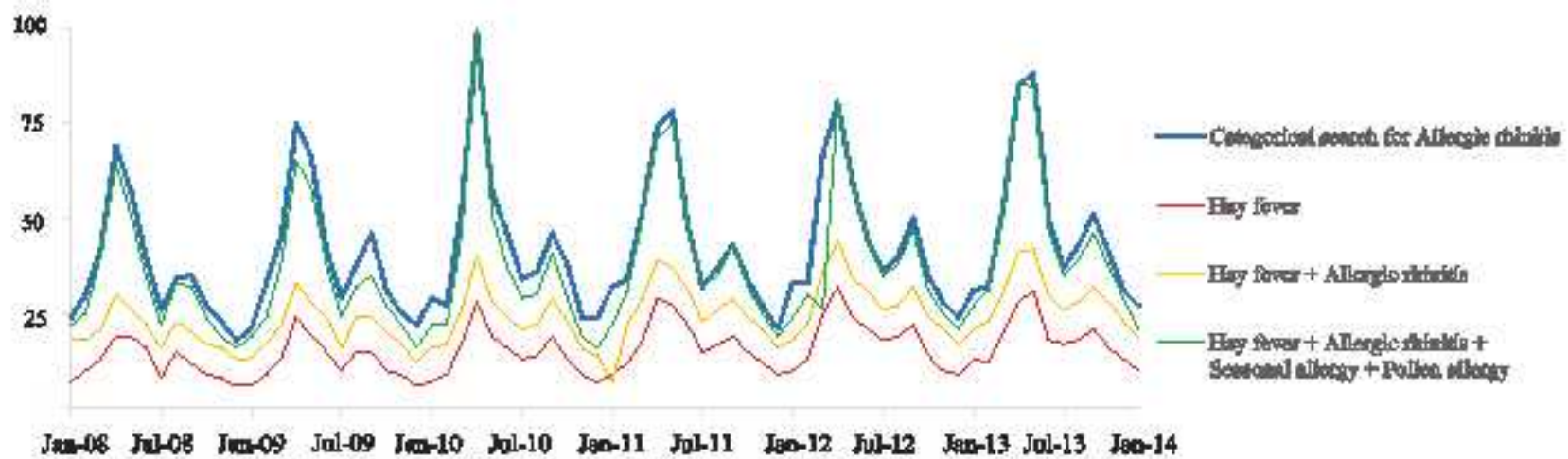
Reference

1. Google Trends Lab (31 May 2014). About Google Trends [cited Oct 1 2009]. Available from: <http://support.google.com/trends>.
2. 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.
3. American Academy of Allergy AaI. Allergic rhinitis (hay fever) [cited 2014 March 15]. Available from: <http://www.aaaai.org/allergist/allergies/types/rhinitis/Pages/default.aspx>.
4. Hay fever translation [cited 2014 March 15]. Available from: [http://translation.babylon.com/english/to-chinese%20\(s\)/Babylon-English-Chinese-\(S\)-Dictionary/](http://translation.babylon.com/english/to-chinese%20(s)/Babylon-English-Chinese-(S)-Dictionary/).
5. Singh K, Axelrod S, Bielory L. The epidemiology of ocular and nasal allergy in the United States, 1988-1994. *The Journal of allergy and clinical immunology*. 2010;126(4):778-83.e6.
6. American Academy of Allergy AaI. National Allergy Bureau [updated 31 May 2014]. Available from: <http://www.aaaai.org/global/nab-pollen-counts.aspx>.
7. Barnett A, Baker P, Barnett MA. Package 'season'. 2014.

Supplementary Figure1.Categorical searches for allergic rhinitis in Google. The result of each categorical search for allergic rhinitis not only includes searches for allergic rhinitis but also for synonyms such as hay fever, seasonal allergy, and pollen allergy. Searching volumes for (a) 'allergic rhinitis' as a categorical query, (b) 'hay fever', (c) the sum of 'hay fever' and 'allergic rhinitis' and (d) the sum of 'hay fever', 'allergic rhinitis', 'seasonal allergy' and 'pollen allergy'. Data Source: [Google Trends \(www.google.com/trends\)](http://www.google.com/trends).

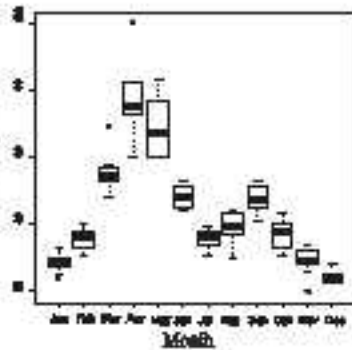
Supplementary Figure2.Time-series searching trends for allergic rhinitis. (a) Box-and-whisker plot. (b) Autocorrelation. (c) Partial autocorrelation. (d) Periodogram from January 2004 to December 2013. Autocorrelation is a useful statistical tool with which to show repetitive seasonal patterns. In these graphs, the lag k in the X axis indicates the association between observations K distance apart. The lag 0 indicates the correlation between the same points, which is always perfectly correlated ($r_0=1$). In (b) and (c), the largest peak is present at lag 12, indicating that the searching volumes for each month are well correlated with the same months in the next year. The periodogram is another tool used to uncover important frequencies in time-series data. In (d), there are only three peaks at a frequency of 0.0833, 0.167, and 0.25, in order, giving dominant cycles of 12 ($1/0.0833$), 6 ($1/0.167$), and 4 ($1/0.25$) months, respectively. This means that the strongest cycle of searching trends for allergic rhinitis is 12 months and that the searching trends thus exhibit annual repetition. The next dominant cycle of 6 months means that there are biannual valleys of searching volumes in January and July in (a). The third but smallest cycle of 4 months indicates that there are two peaks of searching volumes at intervals of 4 months.

Searching volumes



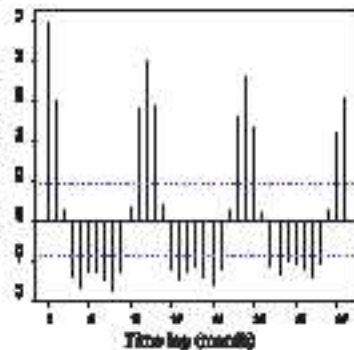
(a)

Sampling values



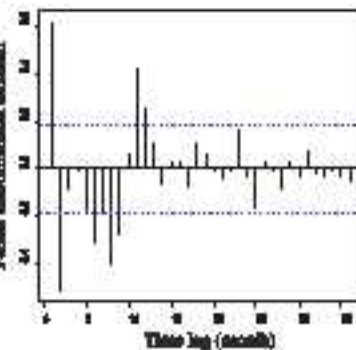
(b)

Autocorrelation function



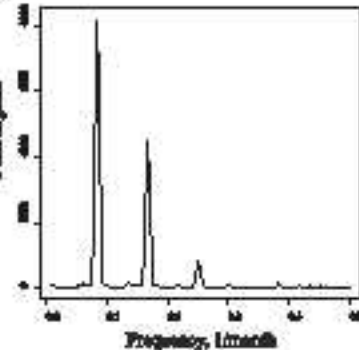
(c)

Partial autocorrelation function

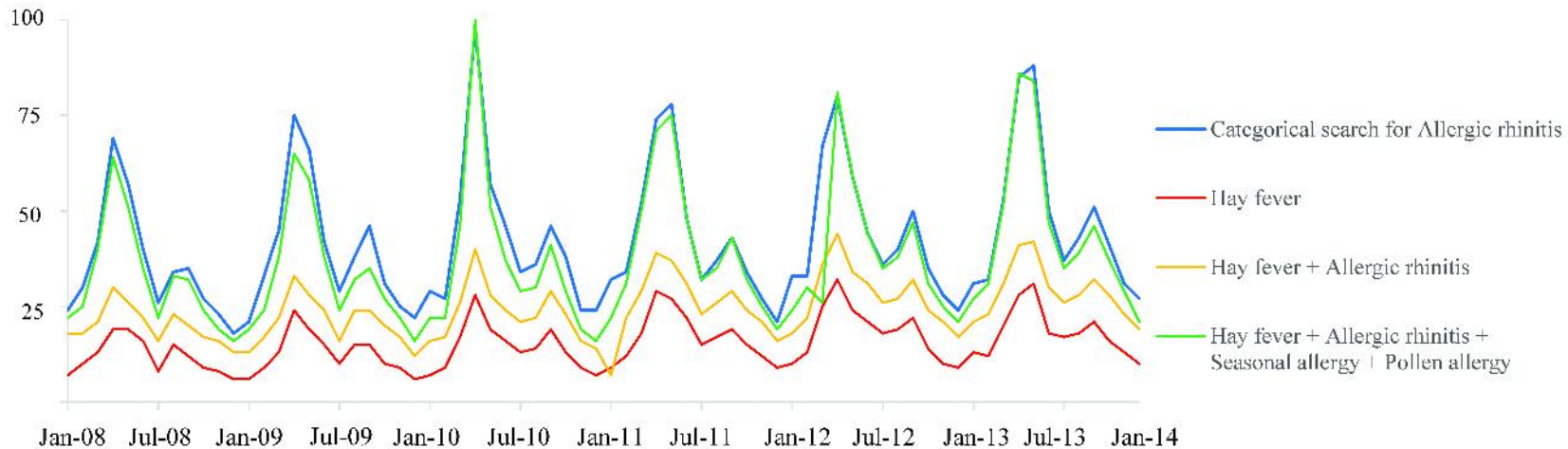


(d)

Periodogram

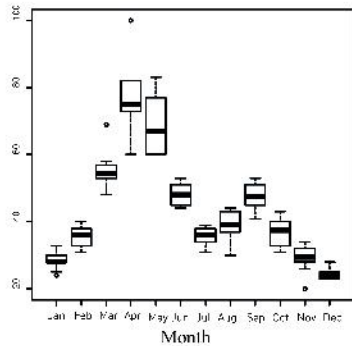


Searching volumes



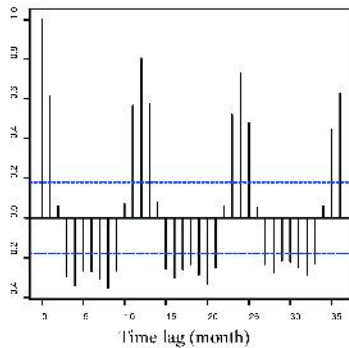
(a)

Searching volume



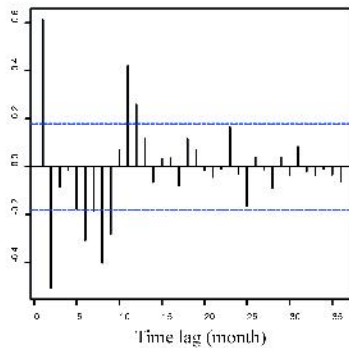
(b)

Autocorrelation function



(c)

Partial autocorrelation function



(d)

Periodogram

