BMJ Open

Using Non-English Language Google Trends to Assess Epidemic Diseases and Public Opinion through Popular Search Behavior

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-034156
Article Type:	Original research
Date Submitted by the Author:	13-Sep-2019

Complete List of Authors:	Chang, Yu-Wei; Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University; Department of Laboratory, Taitung Hospital, Ministry of Health and Welfare Chiang, Wei-Lun; Pan Media Company; OmnInsight Company Wang, Wen-Hung; School of Medicine, Graduate Institute of Medicine, Sepsis Research Center, Center of Dengue Fever Control and Research, Kaohsiung Medical University; Department of Internal Medicine, Kaohsiung Medical University Hospital Lin, Chun-Yu; Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University; School of Medicine, Graduate Institute of Medicine, Sepsis Research Center, Center of Dengue Fever Control and Research, Kaohsiung Medical University Hung, Ling-Chien; School of Medicine, Graduate Institute of Medicine, Sepsis Research Center, Center of Dengue Fever Control and Research, Kaohsiung Medical University; Department of Internal Medicine, Kaohsiung Medical University Hospital Tsai, Yi-Chang; Department of Laboratory, Chang-Hua Hospital, Ministry of Health and Welfare Chen, Y.-H. ; School of Medicine, Graduate Institute of Medicine, Sepsis Research Center, Center of Dengue Fever Control and Research, Kaohsiung Medical University; Department of Internal Medicine, Kaohsiung Municipal Ta-Tung Hospital

Keywords:	Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, EPIDEMIOLOGY, PUBLIC HEALTH

Title page

Title: Using Non-English Language Google Trends to Assess Epidemic Diseases and Public Opinion through Popular Search Behavior

Authors: Yu-Wei Chang[1,2*]; Wei-Lun Chiang[3,4*]; Wen-Hung Wang[5,6]; Chun-Yu Lin[1,5,6]; Ling-Chien Hung[5,6]; Yi-Chang Tsai[7]; Yen-Hsu Chen[6,8,9], MD, PhD, MBA

Affiliation:

[1]Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

[2]Department of Laboratory, Taitung Hospital, Ministry of Health and Welfare, Taitung, Taiwan

[3]Pan Media Company, Taipei, Taiwan

[4]OmnInsight Company, Taipei, Taiwan

[5]Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan

[6]School of Medicine, Graduate Institute of Medicine, Sepsis Research Center, Center of Dengue Fever Control and Research, Kaohsiung Medical University, Kaohsiung, Taiwan

[7]Department of Laboratory, Chang-Hua Hospital, Ministry of Health and Welfare, Chang Hua, Taiwan

[8]Department of Internal Medicine, Kaohsiung Municipal Ta-Tung Hospital, Kaohsiung, Taiwan

[9]Department of Biological Science and Technology, College of Biological Science and Technology, National Chiao Tung University, HsinChu, Taiwan

*these authors contributed equally

Corresponding author:

Yen-Hsu Chen, MD, PhD, MBA

Department of Internal Medicine

Kaohsiung Municipal Ta-Tung Hospital No.68, Jhonghua 3rd Rd, Cianjin District

Kaohsiung, 80145, Taiwan

Phone: 886 7 291 1101

Fax: 886 7 291 1590

Email: infchen@gmail.com

Abstract

Objective: This study aims to develop a suitable surveillance system for monitoring epidemic outbreak and observing related public opinion in the non-English language countries. We further evaluated whether social media reflected social uneasiness and fear during epidemic outbreaks and natural catastrophes.

Design: Prospective, observational study.

Setting: The freely available epidemic data in Taiwan

Main outcome measure: The weekly epidemic incidence data obtained from Taiwan Center for Disease Control (TCDC) and the web search query data obtained from Google Trends between October 4, 2015, and April 2, 2016. To validate whether the non-English query keywords were the excellent surveillance tools, we estimated the correlation between the web query data and epidemic incidence in Taiwan.

Results: Based on our approach, we found the keywords, "感冒(common cold), 發燒(fever), and 咳嗽(cough)", revealed good to excellent correlation between the Google Trends query data and influenza incidence (r = 0.898, P < .001; r = 0.773, P < .001; r = 0.796, P < .001). Those also displayed a high correlation with the influenza-like illness emergency (r = 0.900, P < .001; r = 0.802, P < .001; r = 0.886, P < .001) and outpatient visits (r = 0.889, P < .001; r = 0.791, P < .001; r = 0.870, P < .001). We further found the query "腸病毒 (enteroviruses)" showed excellent correlation with enterovirus infected patients in the emergency department (r = 0.914, P < .001).

Conclusions: These results suggested that Google Trends can serve as a good surveillance tool for epidemic outbreaks even in non-English language countries. The online search activity indicated people's concerns for epidemic diseases even they do not visit hospitals. It prompted us to develop the effectiveness of epidemic monitoring in web social media, which reflected the infectious trend more timeliness than traditional reporting system.

Keywords: Google Trends; epidemic surveillance tool; non-English language

Strengths and Limitations:

1. This is the potential study to access the association between non-English queries and the incidence of the epidemic outbreaks in the non-English language countries.
2. In public opinion observation, we found that internet search activities of appropriate non-English language (Chinese) keywords not only reflect the epidemic disease surveillance but what people concern in the infectious crisis.

3. The forecasting effects of web queries data in the other seasonal infectious diseases were not clear so far, due to these findings mainly focused on the specific epidemic diseases, such as influenza-like illness and EN71 infection.
4. Apart from Google trends, we need to combine more social media to comprehensively analysis the epidemic information through web-related behaviors.

Introduction

Timeliness and public opinion are critical in acute epidemic disease control.[1 2] Effective disease surveillance systems and public relations crisis management support public healthcare action and disseminate accurate health information messages.[3-6] Thus, to develop an early warning system for epidemics is the real critical work. However, the traditional epidemic surveillance systems in current depended on the information derived from laboratory test results, outpatient reports, and mortality statistics. Those of laboratory results for a real-time response are limited because of a several-week lag in reporting.[1] The previous studies highlighted the prolonged delays in reporting an epidemic situation hinder the prevention of the spread of infectious diseases.[1 7 8] Furthermore, inadequate timeliness induces negative public opinion and causes a public relations crisis for the government.

With the development of the Internet and social media, scientists have used data such as Google Trends, health-related tweets, and self-established cloud platforms to assess the activity of acute epidemic diseases and improve individual health care.[9-13] Estimating the levels of infectious diseases by analyzing Internet activity provides highly sensitive assessments compared with those calculated through hospital reports because online activity indicates people's concerns for epidemic diseases even when they do not visit hospitals.[14-16] Moreover, tracking diseases through Internet activity requires lesser effort than that necessary for evaluating laboratory test results and hospital reports.

Analysis based on the relative intensity of Google keywords research relative intensity provides near real-time data.[17] Internet data analysis has advantages over surveys, providing options to narrow data according to desired countries, times, and languages. However, the studies of establishing a surveillance model to estimate epidemic diseases in non-English language countries are still controversial.[18 19] There are few effective surveillance systems for assessing infectious diseases on the basis of Internet activity, despite the high availability and use of the Internet and social media in Taiwan. Therefore, the objectives of this study are to assess whether non-English language (Chinese) Google Trends can be used as an epidemic surveillance system and further application in monitoring public opinion and managing public relations.

Methods

Setting and study period

National surveillance data of influenza (from October 4, 2015, to April 2, 2016) and enterovirus (from January 1, 2012, to December 29, 2012) were obtained from TCDC, which regularly collects and manages the epidemiological data received from the whole cities and counties in Taiwan.

Data Sources

Epidemiological Surveillance Data

This survey provided by TCDC is responsible for the national emerging disease surveillance and disease prevention.[20] For the influenza analysis, the epidemic data was collected and categorized to the weekly number of positive influenza tests, the ratio of emergency department patients with influenza-like illness (ILI), the ratio of outpatient department patients with ILI, and weekly deaths from pneumonia and ILI. For the enterovirus analysis, the ratio of emergency department patients with enteroviruses infection was obtained. Ethically, the open data obtained from TCDC was anonymous and publicly available.

Query Data from Google Trends

The query data was obtained from Google Trends website, which was provided by Google Inc.[21] As previously described, the query trends obtained from Google Trends indicates the normalized results (0-100), which is compared to the maximal value for the particular queries during the search interval.[3] Based on our approach, 10 non-English influenza and enterovirus related search terms were enrolled in the analysis, which was cataloged in Table 1, including the disease names, symptoms, medical equipment, and drugs. For example, disease names cataloged as query terms such as "common cold", "influenza" and "enterovirus". Table 1 listed the epidemic related categories and query terms; Multimedia Appendix 1 listed these terms represented in Chinese and described in English simultaneously. Other criteria involved in the queries analysis are following by Chinese (language), (search interval), and Taiwan (location). By setting the criteria as followed by location to "Taiwan", the search time interval to "October 2015 to March 2016 for influenza; January 2012 to December 2012 for enterovirus", and the language to "Chinese", we downloaded query information of Google trends via all the search terms in Traditional Chinese.

Patient and Public Involvement

No patient involved in the study. The epidemic information presented in the study was freely available from TCDC, which has been removed any possible identifying features relating to individuals.

Statistical Analysis

Initial analysis was conducted by graphically evaluating the data trends according to time. Furthermore, a cross-correlation analysis using SAS 9.3 (SAS Institute, Cary, NC, USA) was conducted to examine the correlation of Google Trends with influenza-related data. One-week forward forecasting analyses were introduced to assess these relationships temporally (i.e., cross-correlated analysis through Google-based research of relative intensity for N weeks and of influenza-related data for N+1 weeks). Correlation coefficients of > 0.8 were defined as indicating excellent correlation, 0.6–0.8 as good correlation, 0.4–0.6 as moderate correlation, and < 0.4 as poor or no correlation.[22-24]

Results

Influenza-like Illness (ILI) Surveillance Study

TCDC routinely provides open government data for epidemiologic surveillance, which is available for not only infectious disease control but preventive healthcare study.[20] First, we aimed to evaluate the effectiveness in influenza surveillance of anonymous logs derived from Web search engine queries. Based on the proposal, we collected national influenza surveillance data from October 4, 2015 to April 2, 2016. The total of 8 queries related to influenza (Table 1) were chosen to estimate the correlation between query data from Google Trends and the influenza-related data obtained from TCDC, including weekly number of patients with influenza, ratio of emergency department patients with ILI, ratio of outpatient department patients with ILI, and weekly deaths from pneumonia and ILI. As shown in Figure 1, the peak indicating the patients with influenza was observed in February 2016 (weeks 18–21), with a simultaneously elevated level for the other three influenza-related data (Figures 2-4). Of note, Figures 1-4 graphically represent the temporal relationship and illustrate the evident increase in the four influenza-related data with a simultaneous increase in the relative intensities of the Google keywords. After February 2016, the four influenza-related data and all keywords decreased simultaneously from March 2016 (after week 22). Analytically, Table 2 lists the correlation coefficients between Google keywords research relative intensities and influenza-related data for "no forward" and "one week forward". Those indicate the non-English language keywords, such as 感冒 (common cold, $r = 0.898$, $P < .001$), 發燒 (fever, $r = 0.773$, $P < .001$), and 咳嗽 (cough, $r = 0.796$, $P < .001$) had a high correlation with the weekly number of positive influenza tests. When 1-week forward was introduced to the forecasting analysis, the similar results can be seen in these keywords, which presented the good to excellent correlation. Interestingly, these certain keywords research intensities also highly correlated with the ILI related medical requests, including the ratio of emergency (or outpatient) department patients with ILI and weekly deaths due to pneumonia and ILI. For all correlations, the keyword "感冒 (common cold)" showed the highest correlation with all influenza-related data for "on forward" (r = .898, P < 0.001) and "one-week forward" (r = .900, P < 0.001) analysis and indicated excellent correlation. However, the symptom keywords research relative intensities, such as 流鼻水(runny nose, r = 0.076~0.263) and 喉嚨痛 (sore throat, r = 0.639~0.783) showed a weaker correlation with the influenza-related data (Table 2), thus indicating that suitable non-English language (Chinese) keywords reflect the level of influenza. Taken together, these results showed that the suitable non-English language (Chinese, eg, 感冒, 發燒, and 咳嗽) keywords research relative intensities reflected the real-time infectious condition of influenza, including the positive rate and the overall medical requests for the ILI syndrome.

EN71 Infection Surveillance Study

Enterovirus 71 (EN71) was first identified in California, USA, in 1969. From then on, EN71 has been detected worldwide.[25] Especially in Taiwan, EN71 repeatedly caused life-threatening outbreaks of hand-foot-mouth disease with the neurological disorder in childhood.[26 27] To our approach, we then estimated whether the web query data in Google Trends can serve as a surveillance tool of EN71 infection in Taiwan. As can be seen in figure 5 and table 3, the query "腸病毒 (Enterovirus)" revealed excellence correlation (r = 0.91, P < .001) with the ratio of emergency department patients with EN71 infection. However, the following by search terms, such as 水泡(blister) or 發燒(fever), showed poor to moderate correlation. (r= 0.478, P < .001; r = 0.359, P < .001, respectively)

Public Opinion Estimation

Public opinion analysis is critical for acute epidemic disease control. Moreover, public opinion regarding an epidemic disease is influenced by several external factors, which can be classified into various groups such as culture, media, opinion leaders, and major events.[28] The certain keywords include standard names of diseases (influenza), medical equipment (ECMO, Extra-Corporeal Membrane Oxygenation), and drugs (Tamiflu), which were selected for estimating public opinion. As shown in Figure 6, a severe earthquake occurred on the Chinese New

Year holiday (week 18; February 6, 2016). The media and opinion leaders focused on the earthquake, and therefore, these keywords research relative intensities did not increase simultaneously while the peak of epidemic disease was presented. (weeks 18–21; February 2016). After the Chinese New Year holiday, the media and opinion leaders refocused the discussions of influenza, including the issue of the influenza outbreak, the flu-vaccine policy, and the medical resource rearrangement, etc. Those affected public opinion and cause a public relation crisis for the Taiwan government in epidemic policy. Thus, the group C keywords research relative intensities reached the peak from the end of February to mid-March (weeks 20 – 25). There was 4-weeks log between the Internet query data and epidemic advance. Figure 6 also showed the peak of the keyword "流感 (influenza)" was presented in week 23 (early March), with a simultaneously elevated level of two keywords for medical equipment and drugs. Altogether, these indicated that appropriate non-English language (Chinese) keywords reflect the concerns of media and opinion leaders regarding epidemic diseases.

Tables

Table 1. Keywords used in Google Trends for this study[a]

Category	Query terms
Disease terms	common cold; influenza; enterovirus
Symptom terms	fever; cough; runny nose; sore throat; blister
Medical equipment term	ECMO (Extra-Corporeal Membrane Oxygenation)
Drug term	Tamiflu

[a]The query terms are shown in their English equivalents in table 1; see supplementary table for these terms in traditional Chinese.

Table 2. Pearson correlation coefficient values for cross-correlation with the intensity of influenza related query terms in Taiwan.

	Weekly number of positive influenza tests		The ratio of emergency department patients with ILI		The ratio of outpatient department patients with ILI		Weekly deaths from pneumonia and ILI	
Query terms	No forward	One week forward	No forward	One week forward	No forward	One week forward	No forward	One week forward
感冒/common cold	*r* = .898	*r* = .900	*r* = .900	*r* = .899	*r* = .889	*r* = .885	*r* = .936	*r* = .936
	P < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001
	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent
發燒/fever	*r* = .773	*r* = .774	*r* = .802	*r* = .807	*r* = .791	*r* = .798	*r* = .837	*r* = .843
	P < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001
	Good	Good	Excellent	Excellent	Good	Good	Excellent	Excellent
咳嗽/cough	*r* = .796	*r* = .793	*r* = .886	*r* = .883	*r* = .870	*r* = .864	*r* = .913	*r* = .911
	P < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001
	Good	Good	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent
流鼻水/runny nose	*r* = .238	*r* = .212	*r* = .145	*r* = .212	*r* = .119	*r* = .076	*r* = .263	*r* = .230
	P = .24	*P* = .31	*P* = .48	*P* = .61	*P* = .56	*P* = .72	*P* = .19	*P* = .27
	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor
喉嚨痛/sore throat	*r* = .640	*r* = .630	*r* = .766	*r* = .760	*r* = .753	*r* = .744	*r* = .783	*r* = .775
	P < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001	*P* < .001
	Good	Good	Good	Good	Good	Good	Good	Good

Table 3. Pearson correlation coefficient values for cross-correlation with the intensity of enterovirus related query terms in Taiwan.

Query terms	The ratio of emergency department patients with enteroviruses infection
腸病毒/entervirus	*r* = .914
	P < .001
	Excellent
水泡/Blister	*r* = .478
	P < .001
	Moderate
發燒/Fever	*r* = .359
	P < .001
	Poor

Figure legends

Figure 1: Temporal comparison of keywords research relative intensity and weekly number of positive influenza tests. (October 4, 2015, to April 2, 2016)
Figure 2: Temporal comparison of keywords research relative intensity and the ratio of emergency department patients with ILI. (October 4, 2015, to April 2, 2016)
Figure 3: Temporal comparison of keywords research relative intensity and the ratio of outpatient department patients with ILI. (October 4, 2015, to April 2, 2016)
Figure 4: Temporal comparison of keywords research relative intensity and weekly death of pneumonia and ILI patients. (October 4, 2015, to April 2, 2016)
Figure 5: Temporal comparison of keywords research relative intensity and the ratio of emergency department patients with enteroviruses infection in Taiwan. (January 1, 2012, to December 29, 2012)
Figure 6: Temporal comparison of keywords research relative intensity and weekly number of positive influenza tests. (October 4, 2015, to April 2, 2016)

Discussion

Principal findings

This study confirmed that suitable non-English language (Chinese) keywords Google research relative intensities are favorable epidemic disease surveillance tools in non-English language country. Moreover, suitable keywords that were sensitive and specific to the public opinion regarding epidemic disease were identified. Keywords such as standard names of diseases, medical equipment, and drugs [e.g. "流感" (Influenza), "葉克膜" (ECMO), and "克流感" (Tamiflu)] were highly sensitive and effective for estimating public opinion. The web search engine, such as Google trends, have been suggested better suited to be the disease surveillance in developed countries, which have large populations of internet search users.[29] However, there are few forecasting tools for epidemic diseases based on web queries data; despite the Internet usage is really high in Taiwan. Thus, we aimed to assess whether non-English language (Chinese) Google Trends can be used as an epidemic surveillance system and further application in monitoring public opinion and managing public relations. These results highlight the potential use of Internet activity related to epidemic diseases for coordinating the supply of medical resources and managing public opinion during influenza outbreaks. For assessing health-related information on the Internet, Google Trends can combine critical data from a large spectrum of the population with geospatial data to create a selected geographical surveillance system. The previous studies had proved the effectiveness of social media in the infectious disease surveillance in many countries, such as United States, Japan, South Korea, China, Greece, and Italy.[16 19 30-33] These findings revealed suitable queries in different languages were available for the application in the epidemic prediction and clinician study. To our best knowledge, our study is first to estimate the correlation between infectious diseases and Internet search activities in Taiwan. The web user's education level, economic situation, cultural and language backgrounds can influence the local habits of Internet searchers.[29] Comparing to the previous reports,[19 24] we identified certain Chinese query terms with significant correlation with epidemic forecasting system, including "common cold" for influenza (r=.898, P<.001) and "enterovirus" for EN71 infection (r=.914, P<.001). These suggested the web query based surveillance system is available in the local language queries for disease prediction but not for international languages, such as English or Simplified Chinese.

Search engines and social media enable people to share information and their experiences in a crisis, as well as to assess message credibility and receive confirmation.[13 34] Internet data should be incorporated into clinical data for risk and crisis management. Furthermore, Internet activity can provide a quantifiable varying assessment of public opinion during a particular disease outbreak to health authorities, researchers, and the media.

The social amplification of risk explains how public risk perception is formed by psychology, mass communication, and cultural factors that result in enhancing or attenuating the public attention to risk. This study can be extended to quantify social uneasiness and fear during outbreaks and catastrophes and delivered information through social media platforms. Moreover, our approach indicates that a surveillance system based on Internet activity can be an essential tool for assessing epidemic diseases and public opinions during epidemics and catastrophes in non-English language countries.

Limitations

Some limitations existed in this study. First, our findings mainly focused on ILI and EN71 infection. The forecasting effects of web queries data in the other seasonal infectious diseases were not clear so far. Our future works will persist to develop more prediction models based on Internet big data to improve the effectiveness of epidemics surveillance in Taiwan. Second, although our research had been introduced to the pandemic in the recent 5 years in Taiwan, the evaluated period was too short to well represent the long-term condition. Actually, our approach was to provide the additional application in web search data, such as Google, which is the most-used search engine in Taiwan. After then, it will encourage more searchers to use the "big data" from social media to track and predict the disease. Third, we only enrolled the single queries data for Google Trends, despite it was the mainly tracking resource in Taiwan. In the future, we will evaluate the epidemic predictions in specific region or language approaches, to establish broad benefits for non-English countries [e.g. the search intensity for "erupção" (the Portuguese term for rash) for detecting Zika virus fever in Brazil; see Multimedia Appendix 2], and further utilize other sources of Internet data, including Tweets, Baidu, Yahoo!, or other social media, and to evaluate whether they provide the information in "infodemiology".[35 36]

Conclusions

Our study demonstrated that non-English language (Chinese) keywords Google search relative intensities are highly effective for estimating the level of epidemic diseases on the basis of people's search behavior. These results suggested the medical information derived from the online resource can play a significant role in the current epidemic surveillance system in Taiwan.

Acknowledgements

All authors listed in the manuscript contributed equally. We all participated in study design, data collection, statistics analysis, and manuscript drafting. In addition, this study is based on epidemic data from Taiwan center for disease control, and query data from Google Trends.

Author Contributions

All the listed authors made substantial contributions to the conception and design of the project. Yu-Wei Chang and Wei-Lun Chiang conceived and designed the project; Chun-Yu Lin provided the clinical knowledge; Yu-Wei Chang, Wen-Hung Wang, Ling-Chien Hung, and Yi-Chang Tsai performed the experimental works; Yu-Wei Chang and Wei-Lun Chiang interpreted the analyzed results and drafted the manuscript. Yen-Hsu Chen is the guarantor of integrity of the entire study and responsible to edit and finally review the paper. All authors have read and approved the final vision to be submitted.

Funding

This project received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflicts of Interest

None declared

Abbreviations

ILI: Influenza-like illness
TCDC: Taiwan center for disease control
EN71: enterovirus 71
ECMO: Extra-Corporeal Membrane Oxygenation

Supplementary material

Supplementary table 1: List of the Chinese query terms used in the study and their English equivalents.

	English equivalent	Traditional Chinese query terms
Disease terms		
	common cold	感冒
	influenza	流感
	enterovirus	腸病毒
Symptom terms		
	fever	發燒
	cough	咳嗽
	running nose	流鼻水
	sore throat	喉嚨痛
	blister	水泡
Medical equipment term		
	ECOM	葉克膜
Drug term		
	Tamiflu	克流感

Supplementary figure 1: Temporal comparison of major event of unknown exanthematous illness in Salvador, Brazil and Portuguese keywords research relative intensity (Language: Portuguese; search terms: erupção (rash); artralgia (artralgia); olhos vermelhos (red eyes); location: Brazil; time: January 4, 2015 to September 5, 2015)

References

1. Jajosky RA, Groseclose SL. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. *BMC public health* 2004;4:29. doi: 10.1186/1471-2458-4-29
2. SteelFisher GK, Blendon RJ, Bekheit MM, et al. The public's response to the 2009 H1N1 influenza pandemic. *The New England journal of medicine* 2010;362(22):e65. doi: 10.1056/NEJMp1005102
3. Wang HW, Chen DR, Yu HW, et al. Forecasting the Incidence of Dementia and Dementia-Related Outpatient Visits With Google Trends: Evidence From Taiwan. *Journal of medical Internet research* 2015;17(11):e264. doi: 10.2196/jmir.4516
4. Wang HW, Chen DR. Economic Recession and Obesity-Related Internet Search Behavior in Taiwan: Analysis of Google Trends Data. *JMIR public health and surveillance* 2018;4(2):e37. doi: 10.2196/publichealth.7314
5. Leal Neto OB, Albuquerque CM, Albuquerque JO, et al. The schisto track: a system for gathering and monitoring epidemiological surveys by connecting geographical information systems in real time. *JMIR Mhealth Uhealth* 2014;2(1):e10-e10. doi: 10.2196/mhealth.2859
6. Majumder MS, Santillana M, Mekaru SR, et al. Utilizing Nontraditional Data Sources for Near Real-Time Estimation of Transmission Dynamics During the 2015-2016 Colombian Zika Virus Disease Outbreak. *JMIR public health and surveillance* 2016;2(1):e30. doi: 10.2196/publichealth.5814
7. Bonacic Marinovic A, Swaan C, van Steenbergen J, et al. Quantifying reporting timeliness to improve outbreak control. *Emerging infectious diseases* 2015;21(2):209-16. doi: 10.3201/eid2102.130504
8. Wang MH, Chen HK, Hsu MH, et al. Cloud Computing for Infectious Disease Surveillance and Control: Development and Evaluation of a Hospital Automated Laboratory Reporting System. *Journal of medical Internet research* 2018;20(8):e10886. doi: 10.2196/10886
9. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature* 2009;457(7232):1012-4. doi: 10.1038/nature07634
10. Miller M, Banerjee T, Muppalla R, et al. What Are People Tweeting About Zika? An Exploratory Study Concerning Its Symptoms, Treatment, Transmission, and Prevention. *JMIR public health and surveillance* 2017;3(2):e38. doi: 10.2196/publichealth.7157
11. Du J, Tang L, Xiang Y, et al. Public Perception Analysis of Tweets During the 2015 Measles Outbreak: Comparative Study Using Convolutional Neural Network Models. *Journal of medical Internet research* 2018;20(7):e236. doi: 10.2196/jmir.9413
12. Schootman M, Toor A, Cavazos-Rehg P, et al. The utility of Google Trends data to examine interest in cancer screening. *BMJ Open* 2015;5(6):e006678. doi: 10.1136/bmjopen-2014-006678

13. Asch JM, Asch DA, Klinger EV, et al. Google search histories of patients presenting to an emergency department: an observational study. *BMJ Open* 2019;9(2):e024791. doi: 10.1136/bmjopen-2018-024791
14. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? *American journal of infection control* 2015;43(6):563-71. doi: 10.1016/j.ajic.2015.02.023
15. Fung IC, Fu KW, Ying Y, et al. Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. *Infectious diseases of poverty* 2013;2(1):31. doi: 10.1186/2049-9957-2-31
16. Woo H, Cho Y, Shim E, et al. Estimating Influenza Outbreaks Using Both Search Engine Query Data and Social Media Data in South Korea. *Journal of medical Internet research* 2016;18(7):e177. doi: 10.2196/jmir.4955
17. Klembczyk JJ, Jalalpour M, Levin S, et al. Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits. *Journal of medical Internet research* 2016;18(6):e175. doi: 10.2196/jmir.5585
18. Cho S, Sohn CH, Jo MW, et al. Correlation between national influenza surveillance data and google trends in South Korea. *PloS one* 2013;8(12):e81422. doi: 10.1371/journal.pone.0081422
19. Kang M, Zhong H, He J, et al. Using Google Trends for Influenza Surveillance in South China. *PloS one* 2013;8(1):e55205. doi: 10.1371/journal.pone.0055205
20. Yang J-R, Teng H-J, Liu M-T, et al. Taiwan's Public Health National Laboratory System: Success in Influenza Diagnosis and Surveillance. *Health Secur* 2017;15(2):154-64. doi: 10.1089/hs.2016.0104
21. Google Trends. URL: https://www.google.com/trends/ [Accessed: 2018-12-04] (Archived by WebCite® at http://www.webcitation.org/74PPqrP4j)
22. Dugas AF, Hsieh YH, Levin SR, et al. Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2012;54(4):463-9. doi: 10.1093/cid/cir883
23. Kadry B, Chu LF, Kadry B, et al. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. *Journal of medical Internet research* 2011;13(4):e95. doi: 10.2196/jmir.1960
24. Ortiz JR, Zhou H, Shay DK, et al. Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends. *PloS one* 2011;6(4):e18687. doi: 10.1371/journal.pone.0018687
25. Schmidt NJ, Lennette EH, Ho HH. An apparently new enterovirus isolated from patients with disease of the central nervous system. *The Journal of infectious diseases* 1974;129(3):304-9. doi: 10.1093/infdis/129.3.304
26. Chia MY, Chiang PS, Chung WY, et al. Epidemiology of enterovirus 71 infections in Taiwan. *Pediatrics and neonatology* 2014;55(4):243-9. doi: 10.1016/j.pedneo.2013.07.007
27. Li K, Li X, Si W, et al. Identifying risk factors for neurological complications and monitoring long-term neurological sequelae: protocol for the Guangzhou prospective cohort study on hand-foot-and-mouth disease. *BMJ Open* 2019;9(2):e027224. doi: 10.1136/bmjopen-2018-027224

28. Young SD, Konda K, Caceres C, et al. Effect of a community popular opinion leader HIV/STI intervention on stigma in urban, coastal Peru. *AIDS and behavior* 2011;15(5):930-7. doi: 10.1007/s10461-010-9826-5
29. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2009;49(10):1557-64. doi: 10.1086/630200
30. Santillana M, Nguyen AT, Dredze M, et al. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. *PLoS computational biology* 2015;11(10):e1004513. doi: 10.1371/journal.pcbi.1004513 [published Online First: 2015/10/30]
31. Sugawara T, Ohkusa Y, Ibuka Y, et al. Real-time prescription surveillance and its application to monitoring seasonal influenza activity in Japan. *Journal of medical Internet research* 2012;14(1):e14. doi: 10.2196/jmir.1881
32. Samaras L, Garcia-Barriocanal E, Sicilia MA. Syndromic Surveillance Models Using Web Data: The Case of Influenza in Greece and Italy Using Google Trends. *JMIR public health and surveillance* 2017;3(4):e90. doi: 10.2196/publichealth.8015
33. Hagihara A, Onozuka D, Miyazaki S, et al. Influenza newspaper reports and the influenza epidemic: an observational study in Fukuoka City, Japan. *BMJ Open* 2015;5(12):e009900. doi: 10.1136/bmjopen-2015-009900
34. Kim Y, Huang J, Emery S. Garbage in, Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection. *Journal of medical Internet research* 2016;18(2):e41. doi: 10.2196/jmir.4738
35. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *Journal of medical Internet research* 2009;11(1):e11. doi: 10.2196/jmir.1157
36. Petersen J, Simons H, Patel D, et al. Early detection of perceived risk among users of a UK travel health website compared with internet search activity and media coverage during the 2015–2016 Zika virus outbreak: an observational study. *BMJ Open* 2017;7(8):e015831. doi: 10.1136/bmjopen-2017-015831

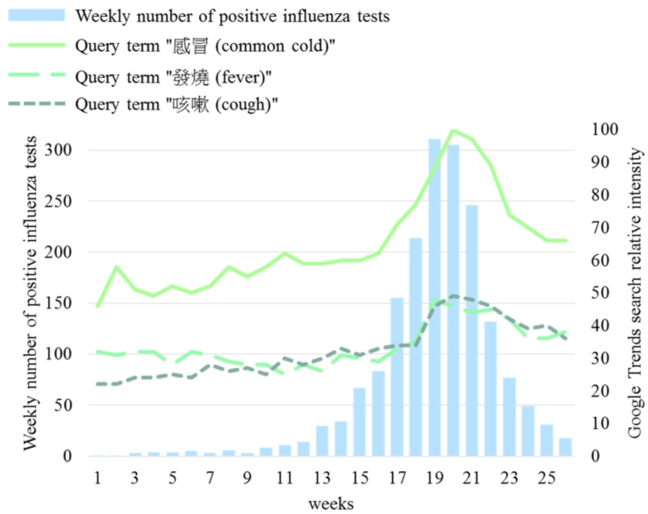


Figure 1. Temporal comparison of keywords research relative intensity and weekly number of positive influenza tests. (October 4, 2015, to April 2, 2016)

338x190mm (300 x 300 DPI)

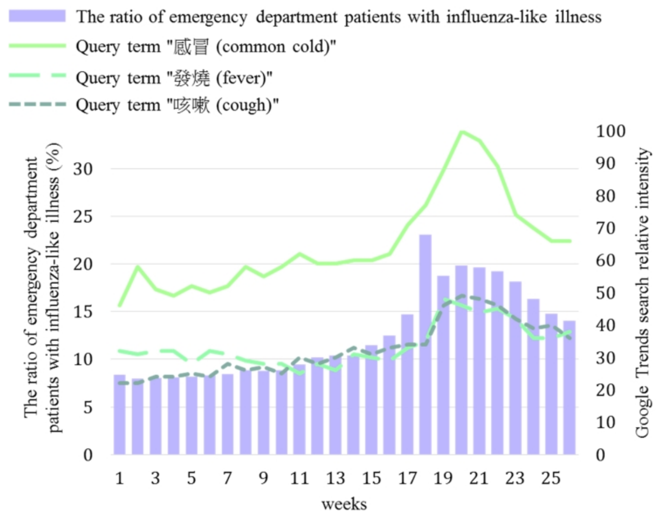


Figure 2. Temporal comparison of keywords research relative intensity and the ratio of emergency department patients with ILI. (October 4, 2015, to April 2, 2016)

338x190mm (300 x 300 DPI)

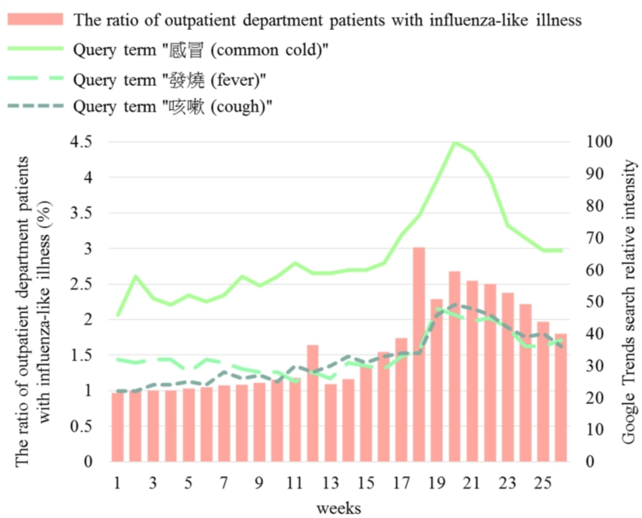


Figure 3. Temporal comparison of keywords research relative intensity and the ratio of outpatient department patients with ILI. (October 4, 2015, to April 2, 2016)

338x190mm (300 x 300 DPI)

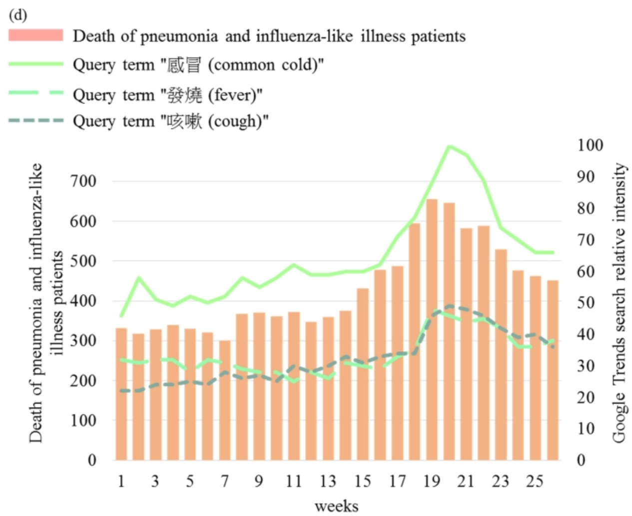


Figure 4. Temporal comparison of keywords research relative intensity and weekly death of pneumonia and ILI patients. (October 4, 2015, to April 2, 2016)

338x190mm (300 x 300 DPI)

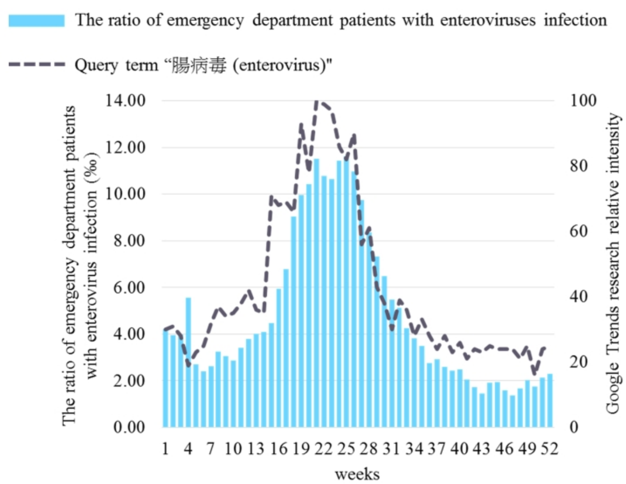


Figure 5. Temporal comparison of keywords research relative intensity and the ratio of emergency department patients with enteroviruses infection in Taiwan. (January 1, 2012, to December 29, 2012)

338x190mm (300 x 300 DPI)

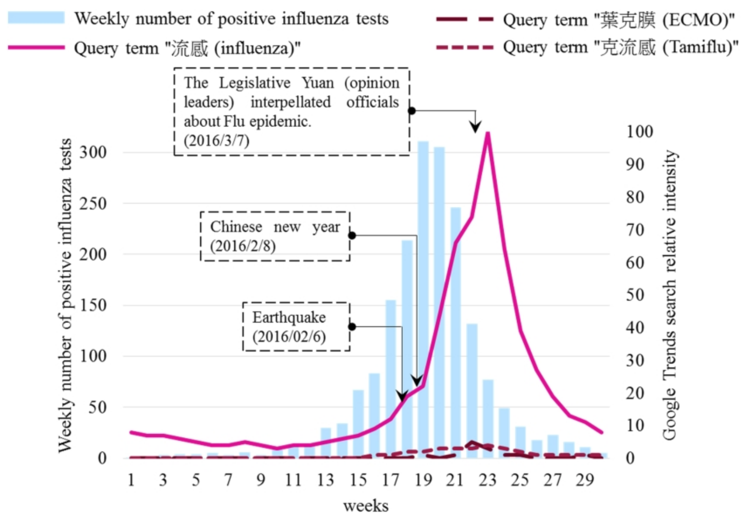
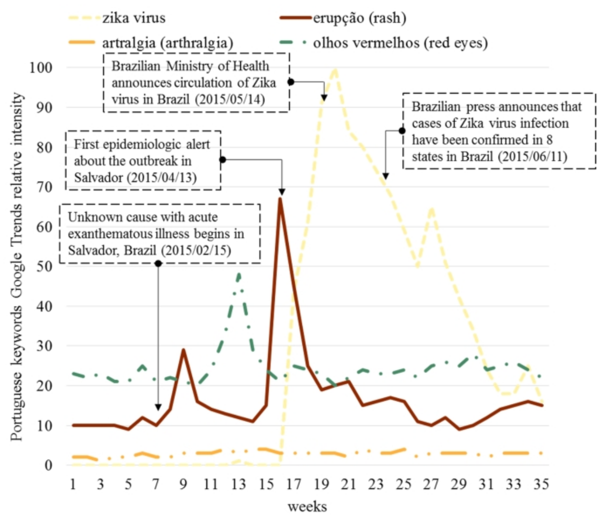


Figure 6. Temporal comparison of keywords research relative intensity and weekly number of positive influenza tests. (October 4, 2015, to April 2, 2016)

338x190mm (300 x 300 DPI)



338x190mm (300 x 300 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	4
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*4*
Bias	9	Describe any efforts to address potential sources of bias	NA
Study size	10	Explain how the study size was arrived at	4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5
		(*b*) Describe any methods used to examine subgroups and interactions	5
		(*c*) Explain how missing data were addressed	NA
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	5
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5-7
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	5-7
		(b) Indicate number of participants with missing data for each variable of interest	NA
Outcome data	15*	Report numbers of outcome events or summary measures	5-7
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	5-7

		(*b*) Report category boundaries when continuous variables were categorized	NA
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	5-7
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10-11
Generalisability	21	Discuss the generalisability (external validity) of the study results	10-11
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	NA

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Google Trends-based non-English-language query data and epidemic diseases: a cross-sectional study of the popular search behavior in Taiwan

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-034156.R1
Article Type:	Original research
Date Submitted by the Author:	20-Feb-2020

Complete List of Authors:	Chang, Yu-Wei; Kaohsiung Medical University; Taitung Hospital Chiang, Wei-Lun; Pan Media Company; OmnInsight Company Wang, Wen-Hung; Kaohsiung Medical University; Kaohsiung Medical University Chung Ho Memorial Hospital Lin, Chun-Yu; Kaohsiung Medical University; Kaohsiung Medical University Chung Ho Memorial Hospital Hung, Ling-Chien; Kaohsiung Medical University; Kaohsiung Medical University Chung Ho Memorial Hospital Tsai, Yi-Chang; Chang-Hua Hospital Ministry of Health and Welfare Suen, Jau-Ling; Kaohsiung Medical University Hospital, Department of Medical Research Chen, Y.-H. ; Kaohsiung Medical University; Kaohsiung Medical University Chung Ho Memorial Hospital

Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology, Health informatics, Infectious diseases, Public health
Keywords:	Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, EPIDEMIOLOGY, PUBLIC HEALTH

SCHOLARONE™
Manuscripts

TITLE PAGE

Title: Google Trends-based non-English-language query data and epidemic diseases: a cross-sectional study of the popular search behavior in Taiwan

Authors: Yu-Wei Chang[1,2*]; Wei-Lun Chiang[3,4*]; Wen-Hung Wang[5,6]; Chun-Yu Lin[1,5,6]; Ling-Chien Hung[5,6]; Yi-Chang Tsai[7]; Jau-Ling Suen[1,8,9]; Yen-Hsu Chen[5,6,10,11], MD, PhD, MBA

Affiliation:

[1]Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

[2]Department of Laboratory, Taitung Hospital, Ministry of Health and Welfare, Taitung, Taiwan

[3]Pan Media Company, Taipei, Taiwan

[4]OmnInsight Company, Taipei, Taiwan

[5]Center for Tropical Medicine and Infectious Disease Research, Kaohsiung Medical University, Kaohsiung, Taiwan

[6]Division of Infectious Disease, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan

[7]Department of Laboratory, Chang-Hua Hospital, Ministry of Health and Welfare, Chang Hua, Taiwan

[8]Research Center of Environmental Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

[9]Department of Medical Research, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan

[10]Department of Internal Medicine, Kaohsiung Municipal Ta-Tung Hospital, Kaohsiung, Taiwan

[11]Department of Biological Science and Technology, College of Biological Science and Technology, National Chiao Tung University, HsinChu, Taiwan

*These authors contributed equally

Correspondence to Yen-Hsu Chen MD, PhD, MBA; infchen@gmail.com

ABSTRACT

Objective: This study developed a surveillance system suitable for monitoring epidemic outbreaks and assessing public opinion in non-English-speaking countries. We evaluated whether social media reflects social uneasiness and fear during epidemic outbreaks and natural catastrophes.

Design: Cross-sectional study.

Setting: Freely available epidemic data in Taiwan.

Main Outcome Measure: We used weekly epidemic incidence data obtained from the Taiwan Centers for Disease Control and online search query data obtained from Google Trends between October 4, 2015 and April 2, 2016. To validate whether non-English query keywords were useful surveillance tools, we estimated the correlation between online query data and epidemic incidence in Taiwan.

Results: With our approach, we noted that keywords 感冒 ("common cold"), 發燒 ("fever"), and 咳嗽 ("cough") exhibited good-to-excellent correlation between Google Trends query data and influenza incidence ($r = 0.898$, $P < 0.001$; $r = 0.773$, $P < 0.001$; $r = 0.796$, $P < 0.001$, respectively). They also displayed high correlation with influenza-like illness emergencies ($r = 0.900$, $P < 0.001$; $r = 0.802$, $P < 0.001$; $r = 0.886$, $P < 0.001$, respectively) and outpatient visits ($r = 0.889$, $P < 0.001$; $r = 0.791$, $P < 0.001$; $r = 0.870$, $P < 0.001$, respectively). We noted that the query 腸病毒 ("enterovirus") exhibited excellent correlation with the number of enterovirus-infected patients in emergency departments ($r = 0.914$, $P < 0.001$).

Conclusions: These results suggested that Google Trends can be a good surveillance tool for epidemic outbreaks, even in Taiwan, the non-English-speaking country. Online search activity indicates that people are concerned about epidemic diseases, even if they do not visit hospitals. This prompted us to develop useful tools to monitor social media during an epidemic because such media usage reflects infectious disease trends more quickly than does traditional reporting.

Keywords: Google Trends; epidemic surveillance tool; non-English language

Strengths and Limitations of this study:

1. This study analyzed the association between non-English-language queries and epidemic outbreak incidence in a non-English-speaking country.
2. Public opinion during infectious outbreaks was assessed in the study.
3. This study mainly focused on influenza and enterovirus infections, and other seasonal infectious diseases were not evaluated.

4. Confounders such as educational level, age, and economic conditions should be considered in the future.
5. More big data are required to comprehensively study "infodemiology."

INTRODUCTION

Timeliness of response and public opinion are critical in acute epidemic disease control.[1 2] Effective disease surveillance systems and crisis management public relations support public health care efforts and the dissemination of accurate health information.[3-6] Thus, developing an early warning system for epidemics is critical. However, current epidemic surveillance systems depend on information from laboratory test results, outpatient reports, and mortality statistics. Using laboratory results to develop real-time responses involves several weeks of lag before the results are reported.[1] Studies have reported that prolonged delays in reporting during epidemic situations hinder efforts to prevent the spread of infectious diseases.[1 7 8] Furthermore, inadequate timeliness induces negative public opinion and may cause public relations crises for governments.

With the development of the Internet and social media, scientists have used data from Google Trends, health-related tweets, and self-established cloud platforms to assess the spread of acute epidemic disease activity and improve individual health care.[9-13] Estimating infectious disease levels by analyzing Internet activity enables more sensitive assessments than doing so by studying hospital reports because online activity indicates how concerned people are about epidemic diseases, even when they do not visit hospitals.[14-16] Moreover, tracking diseases through Internet activity requires less effort than is necessary to evaluate laboratory test results and hospital reports.

Human infection diseases may be characterized by a ubiquitous feature of the seasonal cyclicity, indicating each acute infection has the specific seasonal window of occurrence. However, the seasonality of infectious diseases may vary among geographic locations and differ from other diseases within the same location.[17] For instance, influenza is the major seasonal disease and remains a serious public health threat in Taiwan. It has been well defined that the influenza season in Taiwan usually starts from December, and peaks in January to February of the following year.[18] In addition, enterovirus infection, a significant cause of neurological disorder and death in children, generally has caused outbreaks during the summer months in Taiwan, and epidemics recur with a seasonal pattern.[19]

Analysis based on the relative intensity of Google keyword searches can provide near-real-time data to be particularly useful in epidemic surveillance and control.[20 21] Internet data analysis has certain advantages over surveys and provides options for narrowing data by country, time period, and language. However, studies on the value of establishing a surveillance model to estimate epidemic diseases in non-English-speaking countries have not reached a consensus.[22-24] Few effective surveillance systems for assessing infectious diseases based on Internet activity have been developed, despite the ready availability and use of the Internet and social media in Taiwan. Therefore, the objectives of this study were to assess whether Google Trends for non-English words, specifically Chinese words, can be used for an epidemic surveillance system and for monitoring public opinion and managing public relations.

METHODS

Setting and Study Period

National surveillance data on influenza (October 4, 2015 to April 2, 2016) and enterovirus (January 1, 2012 to December 29, 2012) were obtained from the Taiwan Centers for Disease Control (TCDC), which regularly collects and manages epidemiological data from all cities and counties in Taiwan.

Data Sources

Epidemiological Surveillance Data

A survey by the TCDC is employed for national emerging disease surveillance and disease prevention.[25] For the influenza survey, epidemic data were collected and categorized by the weekly number of positive influenza tests, the ratio of emergency department patients with influenza-like illness (ILI), the ratio of outpatient department patients with ILI, and weekly deaths from pneumonia and ILI. For the enterovirus survey, the ratio of emergency department patients with enterovirus infections was obtained. With respect to ethical considerations, the open data obtained from the TCDC was anonymous and publicly available.

Query Data from Google Trends

Query data were obtained from the Google Trends website provided by Google Inc.[26] Query trends from Google Trends indicate normalized results (0–100), which are compared with the maximum value for particular queries during search intervals.[3] Based on our approach, 10 non-English influenza- and enterovirus-related search terms were enrolled in the analysis (Table 1); these were related to names and symptoms of diseases, medical equipment, and drugs. For example, disease names categorized as query terms included “common cold,” “influenza,” and “enterovirus.” Table 1 lists epidemic-related categories and query terms, and Supplementary Table 1 simultaneously lists these terms in Chinese with English descriptions. Other criteria for analysis of queries were the following: Chinese (language), search interval described as above, and Taiwan (location). We set the location to "Taiwan," search time interval to "October 2015 to March 2016" for influenza, "January 2012 to December 2012" for enterovirus, and the language to "Chinese" and downloaded query information from Google Trends with all search terms in traditional Chinese.

Patient and Public Involvement

No patients or members of the public were involved in the design and conduct of this study

Statistical Analysis

Initial analysis was conducted by graphically evaluating data trends according to time. Furthermore, a Pearson correlation analysis was conducted using SAS 9.3 (SAS Institute, Cary, NC, USA) to examine the correlation of Google Trends with influenza-related data. 1-week lag forecasting analyses were used to assess these relationships temporally (i.e., correlation analysis through Google-based search of relative intensity for N weeks and of influenza-related data for $N + 1$ weeks). Correlation coefficients >0.8 indicated excellent correlation, 0.6–0.8 indicated good correlation, 0.4–0.6 indicated moderate correlation, and <0.4 indicated poor or no correlation.[27-29]

RESULTS

Influenza-Like Illness Surveillance Study

The TCDC routinely provides open government data for epidemiologic surveillance that is available for infectious disease control and preventive health care studies.[25] First, we evaluated the benefits of influenza surveillance from anonymous logs from online search engine queries. Through this, we collected national influenza surveillance data from October 4, 2015 to April 2, 2016. During this interval, an influenza outbreak, a natural disaster, and an earthquake, were occurred in Taiwan. This period was suitable to be the research target, which we evaluated whether the Google Trends was suitable to be the surveillance tool of the epidemic disease and the public opinion. A total of 10 queries related to influenza (Table 1) were chosen for use in estimating the correlation between query data from Google Trends and influenza-related data obtained from the TCDC concerning the weekly number of patients with influenza, ratio of emergency department patients with ILI, ratio of outpatient department patients with ILI, and weekly deaths from pneumonia and ILI. As shown in Figure 1, a peak indicating the number of patients with influenza was observed in February 2016 (Weeks 6–9); simultaneously elevated levels were evident for three other influenza-related data categories (Figures 2–4). Figures 1–4 graphically represent the temporal relationship between and illustrate evident increases in four influenza-related data categories with a simultaneous increase in the relative intensity of Google keywords. After February 2016, the four influenza-related data categories and all keywords decreased simultaneously from March 2016 (after Week 10). Table 2 lists the correlation coefficients between Google keyword search intensity and influenza-related data for "no lag" and "1-week lag." These indicate that non-English keywords, such as 感冒 ("common cold," $r = 0.898$, $P < 0.001$), 發燒 ("fever," $r = 0.773$, $P < 0.001$), and 咳嗽 ("cough," $r = 0.796$, $P < 0.001$), had a high correlation with the weekly number of positive influenza test results. When 1-week lag was introduced to the forecasting analysis, similar results were observed for these keywords and presented good-to-excellent correlation. Keyword search intensity was also highly correlated with ILI-related medical requests, including the ratio of emergency (or outpatient) department patients with ILI and weekly deaths from pneumonia and ILI. For all correlations, the keyword 感冒 ("common cold") exhibited the highest correlation of all influenza-related data for "no lag" ($r = 0.898$, $P < 0.001$) and "1-week lag" ($r = 0.900$, $P < 0.001$) analysis and indicated excellent correlation. However, the search intensity of symptom keywords, such as 流鼻水 ("runny nose," $r = 0.076–0.263$) and 喉嚨痛 ("sore throat," $r = 0.639–0.783$) exhibited weaker correlation with influenza-related data (Table 2), which indicated that appropriate non-English (Chinese) keywords reflect influenza levels. Altogether, these results revealed that appropriate non-English (Chinese, such as 感冒, 發燒, and 咳嗽) keyword search intensity can reflect the real-time infectious condition of influenza, including positive rates and overall medical requests for ILI.

EN71 Infection Surveillance Study

Enterovirus 71 (EN71) was first identified in California in the United States in 1969. Since then, EN71 has been detected worldwide.[30] For Taiwan in particular, EN71 repeatedly causes life-threatening outbreaks of hand, foot, and mouth disease and neurological disorders in children.[31 32] Using our approach, we estimated whether query data from Google Trends can serve as a surveillance tool for EN71 infections in Taiwan. Figure 5 and Table 3 indicate that the query 腸病毒 ("Enterovirus") exhibited an excellent correlation ($r = 0.914$, $P < 0.001$) with the ratio of emergency department patients with EN71 infection. However, using search terms such as 水泡 ("blister") or 發燒 ("fever") exhibited poor-to-moderate correlation ($r = 0.478$, $P < 0.001$; $r = 0.359$, $P < 0.001$, respectively).

Public Opinion Estimation

Public opinion analysis is critical for acute epidemic disease control. Moreover, public opinion regarding epidemic diseases is influenced by several external factors that can be classified into the categories of culture, media, opinion leaders, and major events.[33] Keywords were standard disease names (influenza), medical equipment (extracorporeal membrane oxygenation; ECMO), and drugs (Tamiflu), which were selected to estimate public opinion. Figure 6 illustrates a severe earthquake that occurred during the Chinese New Year holiday (Week 6; February 6, 2016). Media and opinion leaders focused on the earthquake; therefore, the search intensity of these keywords did not increase with the peak in epidemic diseases (Weeks 6–9; February 2016). After the Chinese New Year holiday, media and opinion leaders refocused on influenza, discussing influenza outbreaks, influenza vaccine policies, and medical resource logistics, among other topics. These discussions by thought leaders and media affected public opinion and caused a public relations crisis for the Taiwanese government regarding epidemic policy. Thus, keyword search intensity peaked from the end of February to mid-March (Weeks 9–12). A 4-week lag appeared between Internet query data and epidemic advancement. Figure 6 shows the peak of the keyword 流感 (“influenza”) in Week 11 (early March 2016), with simultaneously elevated levels of two keywords for medical equipment and drugs. Altogether, these indicated that appropriate non-English (Chinese) keywords reflect the concerns of media and opinion leaders regarding epidemic diseases.

Tables

Table 1. Google Trends keywords in this study.[a]

Category	Query terms
Disease terms	common cold, influenza, enterovirus
Symptom terms	fever, cough, runny nose, sore throat, blister
Medical equipment term	ECMO (extra-corporeal membrane oxygenation)
Drug term	Tamiflu

[a]English equivalents of query terms. See Supplementary Table 1 for the traditional Chinese terms.

Table 2. Pearson correlation coefficient values for the intensity of influenza-related query terms in Taiwan.

	Weekly number of positive influenza tests		The ratio of emergency department patients with ILI		The ratio of outpatient department patients with ILI		Weekly deaths from pneumonia and ILI	
Query terms	No lag	1-week lag	No lag	1-week lag	No lag	1-week lag	No lag	1-week lag
感冒/common cold	$r = 0.898$	$r = 0.900$	$r = 0.900$	$r = 0.899$	$r = 0.889$	$r = 0.885$	$r = 0.936$	$r = 0.936$
	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$
	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent
發燒/fever	$r = 0.773$	$r = 0.774$	$r = 0.802$	$r = 0.807$	$r = 0.791$	$r = 0.798$	$r = 0.837$	$r = 0.843$
	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$
	Good	Good	Excellent	Excellent	Good	Good	Excellent	Excellent
咳嗽/cough	$r = 0.796$	$r = 0.793$	$r = 0.886$	$r = 0.883$	$r = 0.870$	$r = 0.864$	$r = 0.913$	$r = 0.911$
	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$
	Good	Good	Excellent	Excellent	Excellent	Excellent	Excellent	Excellent
流鼻水/runny nose	$r = 0.238$	$r = 0.212$	$r = 0.145$	$r = 0.212$	$r = 0.119$	$r = 0.076$	$r = 0.263$	$r = 0.230$
	$P = 0.24$	$P = 0.31$	$P = 0.48$	$P = 0.61$	$P = 0.56$	$P = 0.72$	$P = 0.19$	$P = 0.27$
	Poor	Poor	Poor	Poor	Poor	Poor	Poor	Poor
喉嚨痛/sore throat	$r = 0.640$	$r = 0.630$	$r = 0.766$	$r = 0.760$	$r = 0.753$	$r = 0.744$	$r = 0.783$	$r = 0.775$
	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$	$P < 0.001$
	Good	Good	Good	Good	Good	Good	Good	Good

Table 3. Pearson correlation coefficient values for the intensity of enterovirus-related query terms in Taiwan.

Query terms	The ratio of emergency department patients with enterovirus infection
腸病毒/enterovirus	*r* = 0.914
	P < 0.001
	Excellent
水泡/Blister	*r* = 0.478
	P < 0.001
	Moderate
發燒/Fever	*r* = 0.359
	P < 0.001
	Poor

Figure Legends

Figure 1: Temporal comparison of Google Trends search relative intensity and weekly number of positive influenza tests. (October 4, 2015 to April 2, 2016).
Figure 2: Temporal comparison of Google Trends search relative intensity and the ratio of emergency department patients with ILI (October 4, 2015 to April 2, 2016).
Figure 3: Temporal comparison of Google Trends search relative intensity and the ratio of outpatient department patients with ILI (October 4, 2015 to April 2, 2016).
Figure 4: Temporal comparison of Google Trends search relative intensity and weekly deaths of pneumonia and ILI patients. (October 4, 2015 to April 2, 2016).
Figure 5: Temporal comparison of Google Trends search relative intensity and the ratio of emergency department patients with enterovirus infection in Taiwan (January 1, 2012 to December 29, 2012).
Figure 6: Temporal comparison of Google Trends search relative intensity and weekly number of positive influenza tests. (October 4, 2015 to April 2, 2016).

DISCUSSION

Principal Findings

This study confirmed that the Google search intensity of appropriate non-English (Chinese) keywords is a favorable epidemic disease surveillance tool in non-English-speaking countries. Moreover, suitable keywords related to public opinion regarding epidemic diseases, such as disease names (流感, "influenza"), medical equipment (葉克膜, "ECMO"), and drugs (克流感, "Tamiflu"), were useful for estimating public opinion. Online search engine data, such as those of Google Trends, are well-suited for disease surveillance in developed countries, which have large populations of Internet search users.[34] However, few forecasting tools for epidemic diseases are based on online query data, despite the high Internet usage in Taiwan. Thus, we assessed whether non-English (Chinese) keywords that appear in Google Trends can be used for an epidemic surveillance system and to monitor public opinion and manage public relations.

These results highlighted the potential use of Internet activity related to epidemic diseases to coordinate supplies of medical resources and manage public opinion during influenza outbreaks. To assess online health-related information, Google Trends can combine critical data from a large spectrum of the population with geospatial data to create a surveillance system for a selected geographical area.

Studies had demonstrated the benefits of using social media in infectious disease surveillance in many countries, such as the United States, Japan, South Korea, China, Greece, and Italy.[16 23 35-38] These findings have revealed suitable queries in various languages for epidemic prediction and clinical studies. Based on our evidence, certain queries showed a higher correlation with epidemic data (e.g., common cold, fever, and cough in ILI), which may reflect what people concerned about and their web search behaviors in the epidemic outbreak. To the best of our knowledge, our study is the first to estimate the correlation between infectious diseases and Internet searches in Taiwan. However, some possible intrinsic limitations regarding the use of big data on epidemic disease surveillance should be concerned in the study. Algorithms and computational techniques, which are built and rely on the analysis, still need to be carefully refined, tuned, and calibrated to avoid the overfitting risk in Big data inference.[24] For instance, web users' educational level, economic situation, and cultural and language backgrounds can influence users' habits.[34] Comparing to the previous reports,[23] we identified Traditional Chinese query terms that were significantly correlated with epidemic forecasting, including "common cold" for influenza (r = 0.898, P < 0.001) and "enterovirus" for EN71 infection (r = 0.914, P < 0.001). These findings suggested an online query-based surveillance system can be available in Taiwan local language queries for disease prediction but not in Simplified Chinese.

Search engines and social media enable people to share information and their experiences during crises, assess message credibility, and receive confirmation of information.[13 39] Internet data should be incorporated into clinical data for risk and crisis management. Furthermore, Internet activity can provide quantifiable assessments of public opinion during disease outbreaks for health authorities, researchers, and the media.

The social amplification of risk explains how public risk perception is formed by psychology, mass communication, and cultural factors that enhance or attenuate public attention to risk. This study can be extended to quantify social uneasiness and fear during outbreaks and catastrophes and the delivery of information through social media platforms. Moreover, our approach indicates that a surveillance system based on Internet activity can be an essential tool for assessing epidemic diseases and public opinion during epidemics and catastrophes in non-English-speaking countries.

Limitations

This study contains some limitations. First, our findings mainly focused on ILI and EN71 infections. The forecasting effects of online query data for the other seasonal infectious diseases remain unclear. In the future, we will develop more prediction models from Internet-derived big data to optimize the predictive accuracy of epidemic surveillance. Second, although our research related to pandemics occurring in the past 5 years in Taiwan, the evaluated period was too short to represent long-term conditions well. Our approach provided additional applications for online search data collected by companies such as Google, the most-used search engine in Taiwan. Our study may encourage researchers to use "big data" from social media to track and predict diseases. Third, we only enrolled single-query data for Google Trends, despite this being the main tracking resource in Taiwan. In the future, we will evaluate epidemic predictions in specific regions or language approaches to establish broad benefits for other non-English-speaking countries and use multiple big data sources including other social media (Facebook, Twitter, Baidu, or Yahoo!), local meteorology, and resident consumption behavior to evaluate whether they provide information for "infodemiology."[40 41]

CONCLUSIONS

Our study demonstrated that non-English (Chinese) keyword Google search intensity is related to epidemic disease levels as evident in people's search behavior. These results suggested that medical information derived from online resources could be a crucial for addition to the current epidemic surveillance system in Taiwan.

ACKNOWLEDGEMENTS

This study is supported partially by Kaohsiung Medical University Research Center Grant (KMU-TC108B03), Ministry of Science and Technology, Taiwan (MOST 106-2314-B-037 -087 and MOST 107-2314-B-037 -079 to Y-H Chen) and Ministry of Health and Welfare, Taiwan (Project No. 10965 to Y-W Chang).

CONTRIBUTORS

YW-C and WL-C conceived and designed the project; CY-L provided the clinical knowledge; YW-C, WH-W, LC-H, and YC-T performed the experimental works; YW-C and WL-C interpreted the analyzed results and drafted the manuscript. YH-C is the guarantor of integrity of the entire study and responsible to edit and finally review the paper. All authors have read and approved the final vision to be submitted

FUNDING

This project received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

CONFLICTS OF INTEREST

None declared

DATA SHARING STATEMENT

No additional data are available.

ABBREVIATIONS

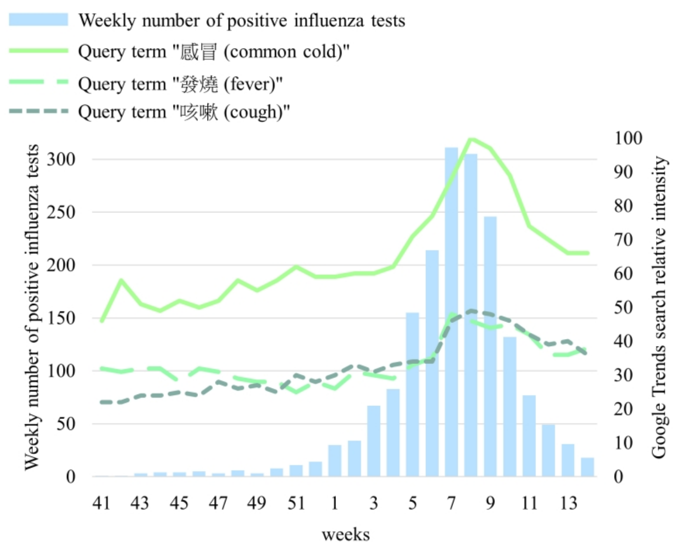
ILI: Influenza-like illness
TCDC: Taiwan center for disease control
EN71: enterovirus 71
ECMO: Extra-Corporeal Membrane Oxygenation

REFERENCE

1. Jajosky RA, Groseclose SL. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. *BMC Public Health* 2004;4:29-29.
2. SteelFisher GK, Blendon RJ, Bekheit MM, et al. The Public's Response to the 2009 H1N1 Influenza Pandemic. *New England Journal of Medicine* 2010;362(22):e65.
3. Wang H-W, Chen D-R, Yu H-W, et al. Forecasting the Incidence of Dementia and Dementia-Related Outpatient Visits With Google Trends: Evidence From Taiwan. *Journal of medical Internet research* 2015;17(11):e264.
4. Wang H-W, Chen D-R. Economic Recession and Obesity-Related Internet Search Behavior in Taiwan: Analysis of Google Trends Data. *JMIR Public Health Surveill* 2018;4(2):e37.
5. Leal Neto OB, Albuquerque CM, Albuquerque JO, et al. The schisto track: a system for gathering and monitoring epidemiological surveys by connecting geographical information systems in real time. *JMIR Mhealth Uhealth* 2014;2(1):e10-e10.
6. Majumder MS, Santillana M, Mekaru SR, et al. Utilizing Nontraditional Data Sources for Near Real-Time Estimation of Transmission Dynamics During the 2015-2016 Colombian Zika Virus Disease Outbreak. *JMIR Public Health Surveill* 2016;2(1):e30.
7. A M, C S, J vS, et al. Quantifying Reporting Timeliness to Improve Outbreak Control. *Emerg Infect Dis* 2015;21(2):209-16.
8. Wang MH, Chen HK, Hsu MH, et al. Cloud Computing for Infectious Disease Surveillance and Control: Development and Evaluation of a Hospital Automated Laboratory Reporting System. *Journal of medical Internet research* 2018;20(8):e10886.
9. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature* 2008;457:1012.
10. Miller M, Banerjee T, Muppalla R, et al. What Are People Tweeting About Zika? An Exploratory Study Concerning Its Symptoms, Treatment, Transmission, and Prevention. *JMIR Public Health Surveill* 2017;3(2):e38.

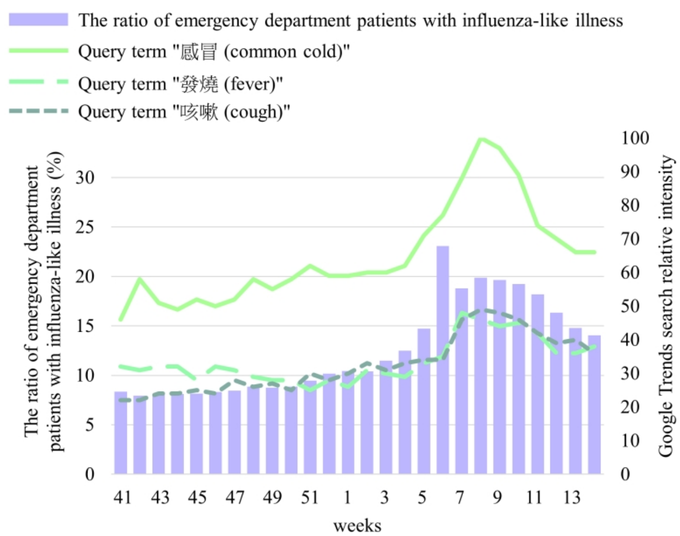
11. Du J, Tang L, Xiang Y, et al. Public Perception Analysis of Tweets During the 2015 Measles Outbreak: Comparative Study Using Convolutional Neural Network Models. *Journal of medical Internet research* 2018;20(7):e236.
12. Schootman M, Toor A, Cavazos-Rehg P, et al. The utility of Google Trends data to examine interest in cancer screening. *BMJ open* 2015;5(6):e006678.
13. Asch JM, Asch DA, Klinger EV, et al. Google search histories of patients presenting to an emergency department: an observational study. *BMJ open* 2019;9(2):e024791.
14. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? *American journal of infection control* 2015;43(6):563-71.
15. Fung IC, Fu KW, Ying Y, et al. Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. *Infectious diseases of poverty* 2013;2(1):31.
16. Woo H, Cho Y, Shim E, et al. Estimating Influenza Outbreaks Using Both Search Engine Query Data and Social Media Data in South Korea. *Journal of medical Internet research* 2016;18(7):e177.
17. Martinez ME. The calendar of epidemics: Seasonal cycles of infectious diseases. *PLoS Pathog* 2018;14(11):e1007327-e27.
18. Su C-p, Tsou T-P, Chen C-H, et al. Seasonal influenza prevention and control in Taiwan—Strategies revisited. *Journal of the Formosan Medical Association* 2019;118(3):657-63.
19. Chang H-L, Chio C-P, Su H-J, et al. The Association between Enterovirus 71 Infections and Meteorological Parameters in Taiwan. *PLOS ONE* 2012;7(10):e46845.
20. Klembczyk JJ, Jalalpour M, Levin S, et al. Google Flu Trends Spatial Variability Validated Against Emergency Department Influenza-Related Visits. *Journal of medical Internet research* 2016;18(6):e175-e75.
21. Gianfredi V, Bragazzi NL, Mahamid M, et al. Monitoring public interest toward pertussis outbreaks: an extensive Google Trends-based analysis. *Public health* 2018;165:9-15.
22. Cho S, Sohn CH, Jo MW, et al. Correlation between national influenza surveillance data and google trends in South Korea. *PLoS One* 2013;8(12):e81422.
23. Kang M, Zhong H, He J, et al. Using Google Trends for influenza surveillance in South China. *PloS one* 2013;8(1):e55205-e05.
24. Gianfredi V, Bragazzi NL, Nucci D, et al. Harnessing Big Data for Communicable Tropical and Sub-Tropical Disorders: Implications From a Systematic Review of the Literature. *Front Public Health* 2018;6:90-90.
25. Yang J-R, Teng H-J, Liu M-T, et al. Taiwan's Public Health National Laboratory System: Success in Influenza Diagnosis and Surveillance. *Health Secur* 2017;15(2):154-64.
26. Google Trends.
27. Dugas AF, Hsieh YH, Levin SR, et al. Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2012;54(4):463-9.
28. Kadry B, Chu LF, Kadry B, et al. Analysis of 4999 online physician ratings indicates that most patients give physicians a favorable rating. *Journal of medical Internet research* 2011;13(4):e95-e95.

29. Ortiz JR, Zhou H, Shay DK, et al. Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends. *PLoS One* 2011;6(4):e18687.
30. Schmidt NJ, Lennette EH, Ho HH. An apparently new enterovirus isolated from patients with disease of the central nervous system. *The Journal of infectious diseases* 1974;129(3):304-9.
31. Chia MY, Chiang PS, Chung WY, et al. Epidemiology of enterovirus 71 infections in Taiwan. *Pediatrics and neonatology* 2014;55(4):243-9.
32. Li K, Li X, Si W, et al. Identifying risk factors for neurological complications and monitoring long-term neurological sequelae: protocol for the Guangzhou prospective cohort study on hand-foot-and-mouth disease. *BMJ open* 2019;9(2):e027224.
33. Young SD, Konda K, Caceres C, et al. Effect of a community popular opinion leader HIV/STI intervention on stigma in urban, coastal Peru. *AIDS and behavior* 2011;15(5):930-7.
34. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2009;49(10):1557-64.
35. Santillana M, Nguyen AT, Dredze M, et al. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. *PLOS Computational Biology* 2015;11(10):e1004513.
36. Sugawara T, Ohkusa Y, Ibuka Y, et al. Real-time Prescription Surveillance and its Application to Monitoring Seasonal Influenza Activity in Japan. *Journal of medical Internet research* 2012;14(1):e14.
37. Samaras L, García-Barriocanal E, Sicilia M-A. Syndromic Surveillance Models Using Web Data: The Case of Influenza in Greece and Italy Using Google Trends. *JMIR Public Health Surveill* 2017;3(4):e90.
38. Hagihara A, Onozuka D, Miyazaki S, et al. Influenza newspaper reports and the influenza epidemic: an observational study in Fukuoka City, Japan. *BMJ open* 2015;5(12):e009900.
39. Kim Y, Huang J, Emery S. Garbage in, Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection. *Journal of medical Internet research* 2016;18(2):e41.
40. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *Journal of medical Internet research* 2009;11(1):e11.
41. Petersen J, Simons H, Patel D, et al. Early detection of perceived risk among users of a UK travel health website compared with internet search activity and media coverage during the 2015-2016 Zika virus outbreak: an observational study. *BMJ open* 2017;7(8):e015831.



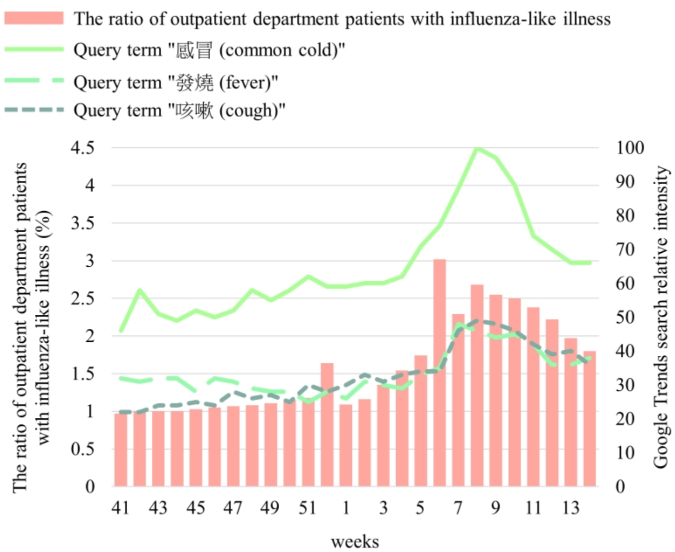
Temporal comparison of Google Trends search relative intensity and weekly number of positive influenza tests. (October 4, 2015 to April 2, 2016).

254x190mm (300 x 300 DPI)



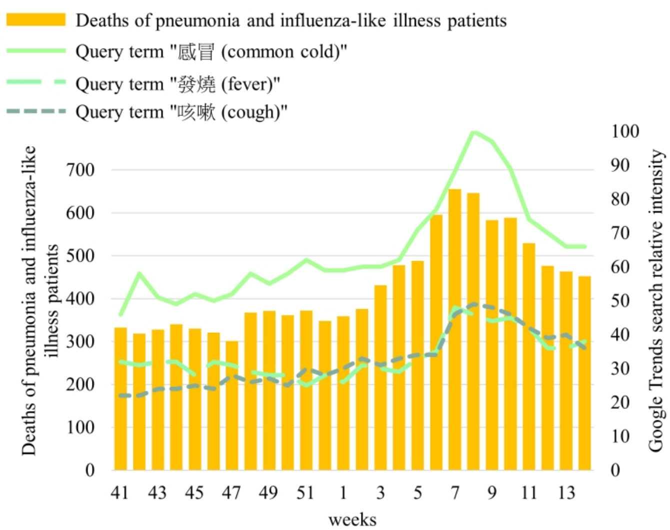
Temporal comparison of Google Trends search relative intensity and the ratio of emergency department patients with ILI (October 4, 2015 to April 2, 2016).

254x190mm (300 x 300 DPI)



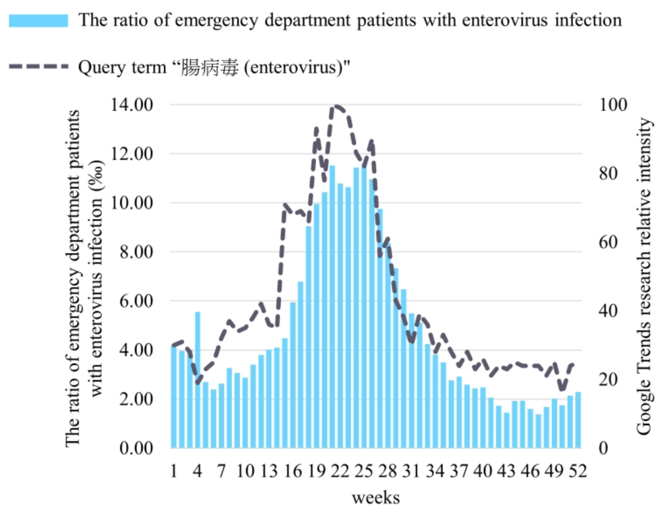
Temporal comparison of Google Trends search relative intensity and the ratio of outpatient department patients with ILI (October 4, 2015 to April 2, 2016).

254x190mm (300 x 300 DPI)



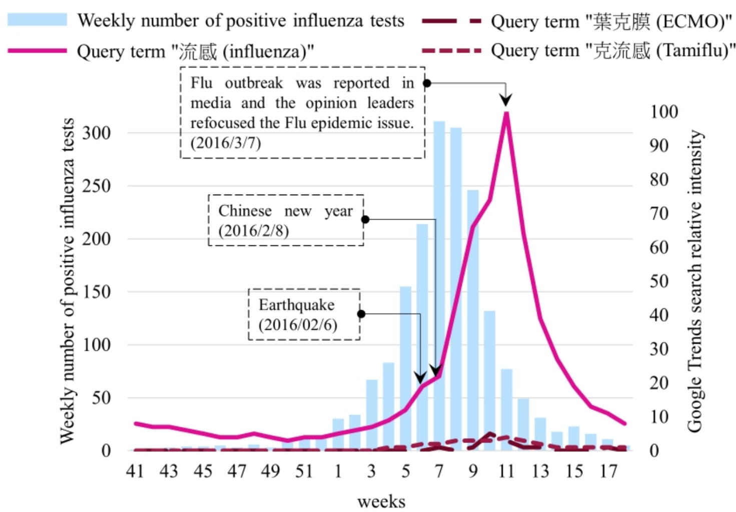
Temporal comparison of Google Trends search relative intensity and weekly deaths of pneumonia and ILI patients. (October 4, 2015 to April 2, 2016).

254x190mm (300 x 300 DPI)



Temporal comparison of Google Trends search relative intensity and the ratio of emergency department patients with enterovirus infection in Taiwan (January 1, 2012 to December 29, 2012).

254x190mm (300 x 300 DPI)



Temporal comparison of Google Trends search relative intensity and weekly number of positive influenza tests. (October 4, 2015 to April 2, 2016).

254x190mm (300 x 300 DPI)

SUPPLEMENTARY MATERIALS

Supplementary Table 1: List of the Chinese query terms used in the study and their English equivalents.

	English equivalent	Traditional Chinese query terms
Disease terms		
	common cold	感冒
	influenza	流感
	enterovirus	腸病毒
Symptom terms		
	fever	發燒
	cough	咳嗽
	running nose	流鼻水
	sore throat	喉嚨痛
	blister	水泡
Medical equipment term		
	ECMO	葉克膜
Drug term		
	Tamiflu	克流感

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1-2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3-4
Objectives	3	State specific objectives, including any prespecified hypotheses	3-4
Methods			
Study design	4	Present key elements of study design early in the paper	4-5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4-5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	4-5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4-5
Bias	9	Describe any efforts to address potential sources of bias	NA
Study size	10	Explain how the study size was arrived at	4-5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4-5
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	5
		(*b*) Describe any methods used to examine subgroups and interactions	5
		(*c*) Explain how missing data were addressed	NA
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	NA
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	NA
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	NA
		(b) Indicate number of participants with missing data for each variable of interest	NA
Outcome data	15*	Report numbers of outcome events or summary measures	5-7
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	5-7

		(*b*) Report category boundaries when continuous variables were categorized	NA
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	NA
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	11-12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	NA

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.