SCIENTIFIC REP🞿RTS

Early detection of an epidemic erythromelalgia outbreak using Baidu search data

Yuzhou Gu[1], Fengling Chen[2], Tao Liu[1], Xiaojuan Lv[1], Zhaoming Shao[2], Hualiang Lin[1], Chaobin Liang[2], Weilin Zeng[1], Jianpeng Xiao[1], Yonghui Zhang[3], Cunrui Huang[4], Shannon Rutherford[5] & Wenjun Ma[1]

Dozens of epidemic erythromelalgia (EM) outbreaks have been reported in China since the mid-twentieth century, and the most recent happened in Foshan City, Guangdong Province early 2014. This study compared the daily case counts of this recent epidemic EM outbreak from February 11 to March 3 with Baidu search data for the same period. After keyword selection, filtering and composition, the most correlated lag of the EM Search Index was used for comparison and linear regression model development. This study also explored the spatial distribution of epidemic EM in China during this period based on EM Search Index. The EM Search Index at lag 2 was most significantly associated with daily case counts in Foshan ($\rho = 0.863$, $P < 0.001$). It captured an upward trend in the outbreak about one week ahead of official report and the linear regression analysis indicated that every 1.071 increase in the EM Search Index reflected a rise of 1 EM cases 2 days earlier. The spatial analysis found that the number of EM Search Indexes increased in the middle of Guangdong Province and South China during the outbreak period. The EM Search Index may be a good early indicator of an epidemic EM outbreak.

Erythromelalgia (EM) is a clinical syndrome characterized by a triad of erythema, burning pain and increased temperature of feet or hands or both. This syndrome is rare in the western world[1,2], and there are no outbreak reports in developed countries. More than 70 epidemic EM outbreaks and over 80,000 cases have been reported in Chinese literature since the mid-twentieth century. For example, around 10,000 and 19,000 cases were observed in Hunan and Hubei Province respectively during a serious outbreak in 1987, and the number of cases in Fujian Province and Hainan Province were more than 10,000 and 11,000 in 1990 outbreak, respectively[3-9]. Although the onset of epidemic EM is acute, its clinical symptoms are not very serious and usually disappear within a few days[7,8,10]. As China has not developed any traditional disease surveillance to monitor this syndrome, the real situation of epidemic EM outbreaks remains unknown in China.

Most epidemic EM outbreaks in China have been reported between February and March, coinciding with a V-shaped temperature change: namely a sharp temperature decline followed by a rapid temperature rise within a few days. Previous studies have hypothesized that these large temperature fluctuations that occur in South China are associated with epidemic EM outbreaks[6-8,11-14]. Furthermore, Liu et al.[14] recently found that one degree Celsius increment of daily temperature might trigger an average rise of 1.22 EM cases in epidemic EM outbreak. During February 2014, the temperature in Foshan City of Guangdong Province experienced a very large temperature fluctuation, accompanied by an epidemic EM

[1]Guangdong Provincial Institute of Public Health, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, China. [2]Chancheng Prefectural Center for Disease Control and Prevention of Guangdong Province, Foshan, China. [3]Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, China. [4]Public Health School of Sun Yat-sen University, Guangzhou, China. [5]Center for Environment and Population Health, School of Environment, Griffith University, Brisbane, Australia. Correspondence and requests for materials should be addressed to W.M. (email: mawj@gdiph.org.cn)

Date	Case count	Date	Case count	Date	Case count
2014-02-11	1	2014-02-18	17	2014-02-25	65
2014-02-12	0	2014-02-19	9	2014-02-26	71
2014-02-13	1	2014-02-20	26	2014-02-27	65
2014-02-14	5	2014-02-21	21	2014-02-28	37
2014-02-15	4	2014-02-22	14	2014-03-01	4
2014-02-16	4	2014-02-23	50	2014-03-02	3
2014-02-17	15	2014-02-24	80	2014-03-03	2

Table 1. Daily EM case counts during the outbreak period in Foshan City.

outbreak in two high schools. The cases of this outbreak were characterized by burning pain and numbness in toes and feet. As most cases were mild, and epidemic EM is not a notifiable disease in China, it is not clear whether unreported cases occurred elsewhere in Guangdong or China during this period.

The availability and popularity of the Internet has grown greatly in recent years. As at December, 2013, there were 618 million Internet users in China, accounting for about 45.8% of the national population, and the proportion in Guangdong Province was even higher[15]. At the same time, an increasing number of people, including patients and their family members, are inclined to search online for health information before seeking medical service[16,17], making it possible to monitor the health status of the population by tracking changes in frequencies of specific search keywords. Internet search engines are now the most common tool to obtain information for Internet users[18,19], and data from different search engines have been successfully utilized for early detection of diseases such as influenza and dengue[17,18,20–26]. Such studies suggest that Internet search data-base surveillance might be a novel way to monitor epidemic EM outbreaks in near real-time.

Baidu is the most popular search engine in China, with 86.7% of Internet users preferring it[19]. This wide use makes it the most representative for analyzing Chinese online behavior[26]. Further, the search volume of Baidu users are released daily on Baidu's Index website (http://index.baidu.com), which allows for timely capture in the changes in search keywords. Although more and more studies are investigating the relationship between search data and some infectious diseases, no study has yet focused on epidemic EM. Due to the lack of a traditional surveillance system, the Internet surveillance approach for early detection of epidemic EM outbreaks is a promising one.

The present study compared Baidu search data with case counts reported in the Foshan outbreak during the same period, in order to identify whether there was an association between epidemic EM and Internet search behavior, and develop an Internet search data-based surveillance method which would be useful to detect an outbreak of epidemic EM in the future.

Materials and Methods
Data sources.
Outbreak data. This study used daily case counts over the entire 21 day outbreak period from February 11 to March 3, 2014 in Foshan City of Guangdong Province, China. The definition of a case is that a student in an outbreak high school reported an onset of pain, redness or numbness in toes or feet with no obvious cause after February 10, 2014. We eliminated those cases induced by injury. The first case was reported on February 27 and 494 cases were retrospectively confirmed by epidemiologists and clinical experts after a systematic field investigation. Daily case counts are shown in Table 1.

Baidu search data. The Baidu index website (http://index.baidu.com) contains search volumes for numerous keywords keyed in by Baidu users from June 2006. Data are available on a daily basis, at a city, province and national level. Considering the time lags between symptoms onset and online searching, we collected the data for 24 days from February 11 to March 6, 2014. The search volume for the same period in 2013 was collected for comparison.

Meteorological data. Due to the hypothesis of an association between epidemic EM and large temperature fluctuation we collected the daily maximum temperature in Foshan City from February 6 to March 3, 2014 (a total of 26 days), which contains the entire period of a large temperature change. Meteorological data were obtained from a free weather query website in Chinese (http://www.tianqihoubao.com).

No ethics committee approval or written consent from patients were required to obtain since only daily count data was obtained, and no information about the identity of any case was revealed.

Keyword selection and filtering.
Keyword selection is the critical issue in Internet search data-based surveillance, as it directly affects the ability and detective accuracy of the surveillance method. Different people may type in entirely different words when searching the same information, especially when searching in Chinese language, where one meaning can be expressed in several ways. Consequently,

diverse results can be obtained by selecting different keywords. Despite the significance of this, there are no principles or standards for guidance[18,26,27]. Previous studies generally chose the names or clinical symptoms of target diseases as their core keywords[22,23,25,26]. As EM is a little-known disease within the lay Chinese community, insufficient search volume of this word leads to Baidu's failure in calculating its search information. Therefore, we chose primary keywords which represent the major clinical characteristics of cases in 2014 and previous outbreaks (see Supplementary Table S1). A Chinese website (http://tool.chinaz.com/baidu/words.aspx) was used for further obtaining related keywords. Related key-word recommendations in the website not only include suggestions from Baidu, but also mining from portal websites, blogs, and online reports using semantic correlation analysis[26]. Upon typing in the 19 primary terms respectively, we obtained 62 related keywords (see Supplementary Table S1).

However, more keywords do not necessarily lead to a better result[17,28] since some recommended keywords are not closely related to EM, which could reduce the detective ability of the surveillance system. Hence, we collected the search data in Foshan City from Baidu and filtered keywords following two steps:

1. We eliminated the words irrelevant to EM and those with a search volume of zero during the outbreak period, and 32 keywords remained (see Supplementary Table S1).
2. Spearman's rank correlation coefficients (ρ) were then calculated between daily case counts and daily search volumes for each keyword using different time lags. We deleted the words with maximum correlation coefficients less than 0.4 in each time lag and those correlations that were statistically insignificant. Taking into account the remaining number, as well as strength of the correlation of keywords that met the criteria above, we considered time lags of 0 to 3 days. The remaining keywords for each of the four time lags were 14, 15, 17 and 17, respectively (Table 2).

EM Search Index composition.   Following selection and filtering, the remaining keywords were used for composition of an EM Search Index for each time lag. Weights of keywords were defined by the strength of the correlation coefficient[26,27]. The weights calculation and EM Search Index composition formulae are as follows:

$$weight_{li} = \frac{\rho_{li}}{\sum_{i=1}^{n}\rho_{li}} \tag{1}$$

$$EM\ Search\ Index_l = \sum_{i=1}^{n} weight_{li}\ keyword_{li} \tag{2}$$

In the above formulae, l denotes time lag of the search data, n is the number of keywords at each time lag, $keyword_{li}$ and $weight_{li}$ represent the ith keyword and the weight of it.

Epidemic EM outbreak detection.   In order to compare the epidemic situation with temperature change, we first graphed a line figure to depict the relationship between daily maximum temperature and daily case counts.

Spearman's rank correlation coefficients were then calculated between outbreak data and the EM Search Index for each time lag of 0 to 3 days. The time lag that has the largest coefficient was selected for further analysis. Based on this, we further developed a linear regression model as follows:

$$Daily\ case\ counts = \beta_0 + \beta_1 EM\ Search\ Index_l + \varepsilon \tag{3}$$

$EM\ Search\ Index_l$ denotes the lag EM Search Index with the largest correlation, β_1 as the regression coefficient. The model estimates the case count l days before, based on the Baidu search data for the current day.

Though the temperature fluctuation between February and March 2014 was widespread in South China, there were no other reports of epidemic EM outbreaks from other cities or provinces. Therefore, it was not clear whether EM cases occurred elsewhere during this period. We calculated the EM Search Index from February 11 to March 3, 2014 for Guangdong Province and 33 other provinces/municipalities of China. By plotting these data on maps, we aimed to roughly explore whether similar outbreaks of epidemic EM occurred in other parts of China during this period. In order to understand the influence of regional difference, Internet search data from February 11 to March 3 2013 were collected for comparison.

All analyses were performed using SPSS 19.0, and the maps were plotted with ArcGIS 9.3 (ESRI).

Results
A large fluctuation in temperature in Foshan City between February and March in 2014 was observed. Daily maximum temperature suddenly dropped about 10 °C on February 8 and continuously declined to the lowest (6 °C) on February 13, then slowly returned to a relatively high level afterwards (Fig. 1). When temperature began to increase, EM cases occurred (Fig. 1). Spearman's rank correlation coefficient

Keywords of lag 0 (n=14)	Keywords of lag 1 (n=15)	Keywords of lag 2 (n=17)	Keywords of lag 3 (n=17)
脚痛怎么办 (what to do with foot pain)	脚痛怎么办 (what to do with foot pain)	脚痛 (foot pain)	脚痛 (foot pain)
脚痛的原因 (cause of foot pain)	脚痛是怎么回事 (foot pain, what happened)	脚痛怎么办 (what to do with foot pain)	脚痛怎么办 (what to do with foot pain)
脚掌疼是怎么回事 (sole pain, what happened)	脚痛的原因 (cause of foot pain)	脚痛是怎么回事 (foot pain, what happened)	脚痛是怎么回事 (foot pain, what happened)
脚底痛 (thenar pain)	脚掌疼是怎么回事 (sole pain, what happened)	脚掌疼是怎么回事 (sole pain, what happened)	脚掌疼是怎么回事 (sole pain, what happened)
脚底痛是怎么回事 (thenar pain, what happened)	脚底痛 (thenar pain)	脚底痛 (thenar pain)	脚底痛 (thenar pain)
脚底痛是什么原因 (what causes the thenar pain)	脚底痛是怎么回事 (thenar pain, what happened)	脚底痛是怎么回事 (thenar pain, what happened)	脚底痛是怎么回事 (thenar pain, what happened)
脚底板疼是怎么回事 (pain in sole of foot, what happened)	脚底痛是什么原因 (what causes the thenar pain)	脚底痛是什么原因 (what causes the thenar pain)	脚底痛是什么原因 (what causes the the pain)
脚跟痛 (heel pain)	脚底板疼是怎么回事 (pain in sole of foot, what happened)	脚底板疼是怎么回事 (pain in sole of foot, what happened)	脚底板疼是怎么回事 (pain in sole of foot, what happened)
脚后跟疼 (ache in heel of foot)	脚跟痛 (heel pain)	脚跟痛 (heel pain)	脚后跟疼是怎么回事 (ache in heel of foot, what happened)
脚后跟疼是怎么回事 (ache in heel of foot, what happened)	脚跟疼 (heel ache)	脚跟疼 (heel ache)	脚后跟疼是什么原因 (what causes the ache in heel of foot)
脚后跟疼是什么原因 (what causes the ache in heel of foot)	脚后跟痛 (pain in heel of foot)	脚后跟痛是什么原因 (what causes the pain in heel of foot)	脚趾痛 (toe pain)
脚趾痛 (toe pain)	脚后跟疼是什么原因 (what causes the ache in heel of foot)	脚后跟疼是怎么回事 (ache in heel of foot, what happened)	脚麻 (foot numbness)
脚麻 (foot numbness)	脚趾痛 (toe pain)	脚后跟疼是什么原因 (what causes the ache in heel of foot)	脚麻是怎么回事 (foot numbness, what happened)
脚麻是什么原因 (what causes the foot numbness)	脚麻 (foot numbness)	脚趾痛 (toe pain)	脚麻是什么原因 (what causes the foot numbness)
	脚麻是什么原因 (what causes the foot numbness)	脚麻是怎么回事 (foot numbness, what happened)	脚麻怎么办 (what to do with foot numbness)
		脚麻是什么原因 (what causes the foot numbness)	脚发麻 (numb foot)
		脚发麻 (numb foot)	脚趾发麻 (toe numbness)

Table 2.   Keywords under time lags of 0 to 3 days after second step filtering. Baidu users search information in Chinese and the corresponding translation of each Chinese keywords are listed.

analysis showed that the daily case counts were positively associated with daily maximum temperature during the temperature increase ($\rho = 0.650, P = 0.001$).

EM Search Indexes for time lags of 0 to 3 days were composed of 14, 15, 17 and 17 keywords respectively (Table 2), and the correlation coefficients between EM Search Indexes and outbreak data are listed in Table 3. We found the correlation getting closer with the increase of lag days before reaching a peak at lag 2 ($\rho = 0.863, P < 0.001$). Therefore, EM Search Index at lag 2 was chosen for further analysis.

We then graphed the curves of daily case counts and EM Search Index at lag 2 over the outbreak period (Fig. 2). Obviously, the search data accurately captured the change in daily case counts. Particularly, we found an apparent increase in EM cases after February 20, followed by a similar uptrend of search volume after February 21. However, the first case hadn't been reported until February 27 by the local Center for Disease Control and Prevention, which suggested that, although there was a lag between the EM outbreak and the EM Search Index, the EM Search Index still had the ability to detect the epidemic about 1 week before the outbreak was reported.
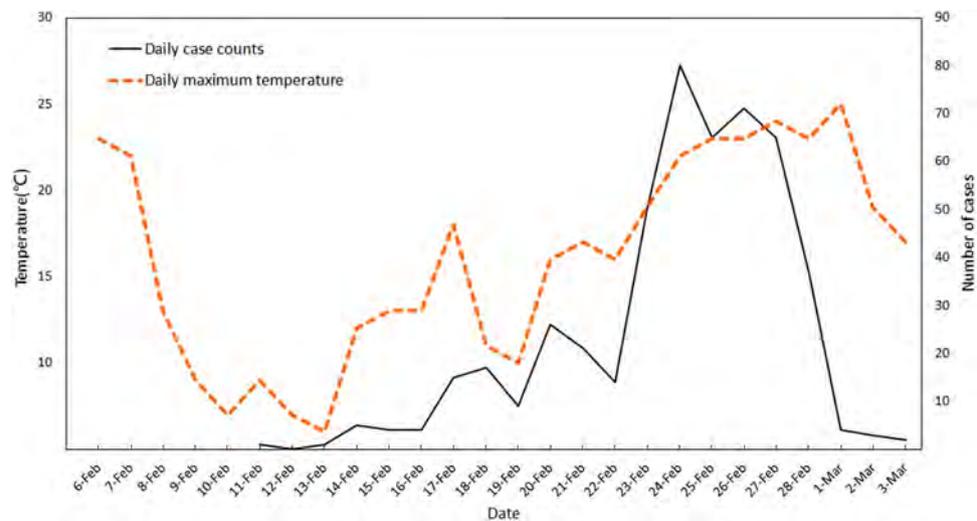
Figure 1. Daily case counts and daily maximum temperature. This figure displays the pattern of temperature change in Foshan City between February and March 2014 and provides the trend in daily EM case counts within this period.

	Lag 0	Lag 1	Lag 2	Lag 3
Spearman's rank correlation coefficient	0.747*	0.846*	0.863*	0.843*

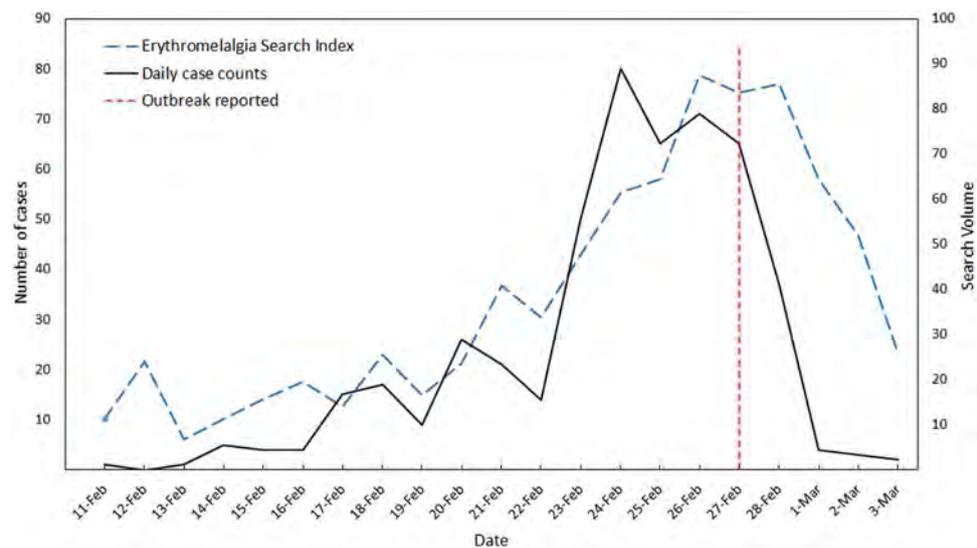**Table 3. Correlation between outbreak data and EM Search Index (lags of 0 to 3 days).** *P < 0.0001.

Figure 2. EM Search Index and daily case counts. This figure describes the changes in daily EM case counts and the EM Search Index at lag 2 during the outbreak period (February 11–March 3) for Foshan City. The report date of outbreak is clearly indicated.

The coefficient (β_1) for the linear regression model between outbreak data and the EM Search Index was 0.934 (*P < 0.001*), indicating that during the outbreak period, every 1.071 increase in EM Search Index reflected a rise of 1 case 2 days before. The R^2 was 0.83, suggesting that the Search Index could explain 83% of the variation in daily case counts.

The EM Search Index from February 11 to March 3, 2014 for each city in Guangdong Province and 34 provinces/municipalities in China were plotted on maps, in contrast with the same period of 2013 (Figs 3 and 4). As demonstrated in Fig. 3, most cities of Guangdong Province showed low search frequencies
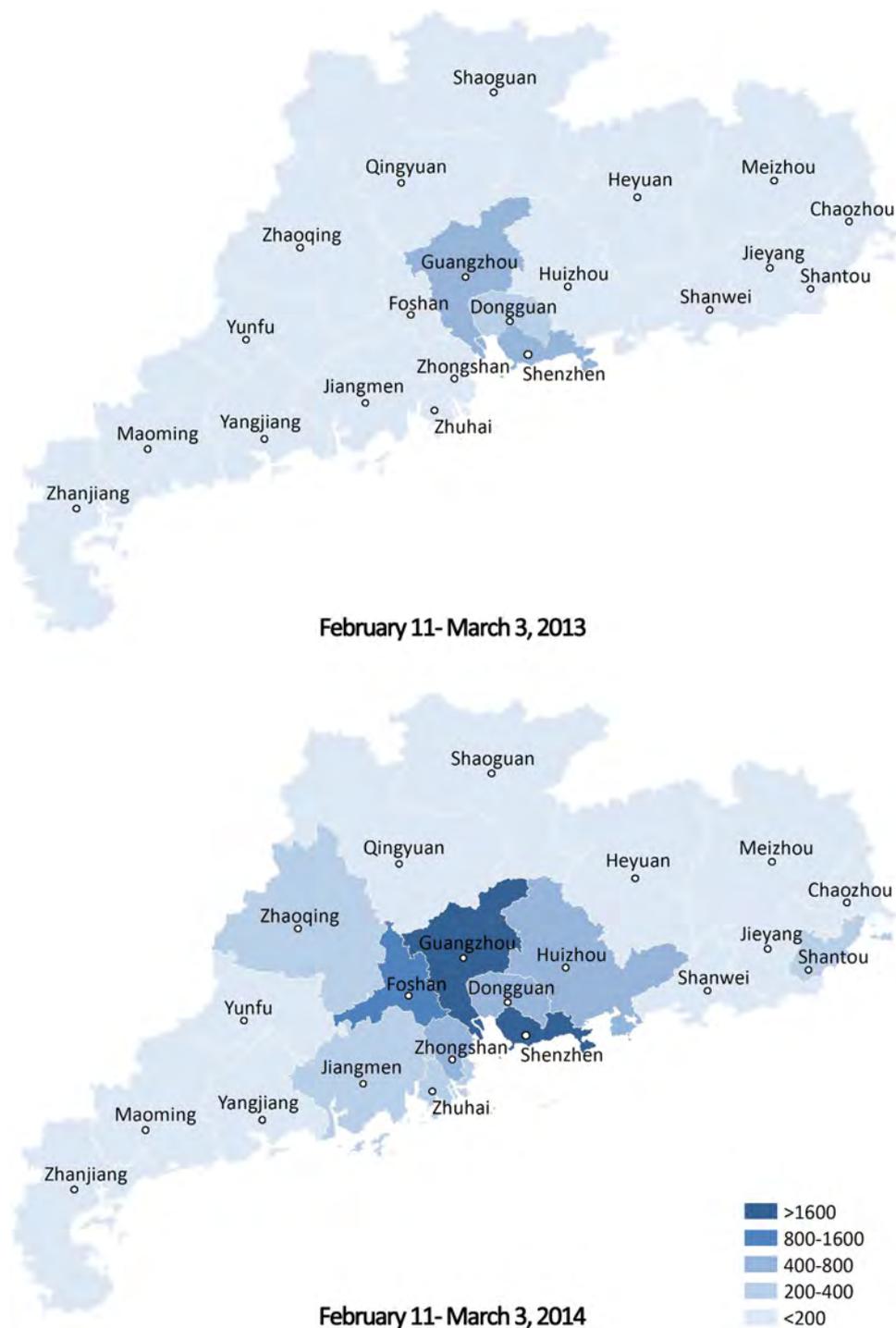
Figure 3. The spatial distribution of EM Search Index in Guangdong Province, China. This figure depicts the spatial distribution of EM Search Index counts across Guangdong Province during the outbreak period in Foshan City in 2014 by filling different colour depth for the cities through ArcGIS 9.3 (ESRI). Distribution of the same period in 2013 was plotted for comparison.

in 2013, but a much higher EM Search Index was observed in Guangzhou, Foshan and Shenzhen in 2014. South China and East China showed relatively high EM Search Index in 2014, with the highest in Guangdong. In contrast, no region showed a high EM Search Index during the same period in 2013 (Fig. 4).

Discussion

Since Eysenbach et al.[20] set the important precedent for disease surveillance using Internet search data, there have been more and more studies on this topic. Most existing studies have focused on infectious

Figure 4. **The spatial distribution of EM Search Index in China.** This figure depicts the spatial distribution of EM Search Index counts across China during the outbreak period in Foshan City in 2014 by filling different colour depth for the provinces/municipalities through ArcGIS 9.3 (ESRI). Distribution of the same period in 2013 was plotted for comparison.

diseases such as influenza and dengue fever[17,21–26]. This study is the first that has investigated the application of Internet search data in the early detection of outbreaks of epidemic EM.

In this study, we compared Baidu search index counts and daily case counts of a recent epidemic EM outbreak in Foshan City, Guangdong Province, China, and found that the EM Search Index at 2 lag days was significantly associated with an EM outbreak, and every 1.071 increase in EM Search Index might reflect a rise of 1 case 2 days before. These findings indicate that the onset of EM symptoms were associated with an increase in Internet search behavior for keywords relating to the illness after 2 days. Even though a 2 days lag was identified, the EM Search Index captured the sharp uptrend of daily case counts about a week ahead of the official report because of the delayed reports from the local Center for Disease Control and Prevention, This suggests that EM Search Index may be a good predictor for early detection of epidemic EM outbreaks. Due to little attention to EM by the public, no mass media reported EM during this outbreak, which adds weight to the utility of Internet search data and how it reflects individual's health concerns and issues[29]. However, epidemic EM have mainly occurred in students living in schools, which makes our results useful for extrapolating to a similar population rather than general population[14].

EM is little known within the Chinese ordinary people and insufficient knowledge might result in more pain and more panic during an epidemic EM outbreak. An early detection system could help to facilitate the timely treatment of cases and ease public concerns about the health symptoms. Previous studies reported the phenomenon of epidemic EM outbreaks accompanied by a large temperature fluctuation[6–8,11–14], and this is confirmed by the outbreak that happened in Foshan in February and March, 2014. Certainly, conducting EM surveillance in schools, communities or hospitals during large temperature fluctuations is a direct way to detect EM outbreaks. However, it may be more cost-effective to monitor the changes of temperature and EM Search Index simultaneously. For example, coinciding with the sharp drop in ambient temperature followed by a temperature rise within a short period, we observed an obviously increasing trend of EM Search Index, which could be a strong signal for the occurrence of an EM outbreak. Therefore, Internet search data provides an opportunity for government or the public to early detect epidemic EM outbreaks and consequently take measures in time.

According to the Chinese literature, most epidemic EM outbreaks have coincided with large temperature changes in many provinces of South China[3–6,8,9,30–35]. For example, the outbreak in 1987, which affected Hubei, Henan, Hunan, Jiangxi and Zhejiang Province of South China[4–7,31,32], and up to six provinces including Fujian, Anhui, Guangdong, Guangxi, Guizhou and Hainan of South China were involved in the outbreak[3,4,8,9,30,33–36]. Thus it is possible that the EM outbreak might not only limited in Foshan City in 2014 because ambient temperature fluctuation could be observed in many parts of South China between February and March. Therefore, we tried to retrospectively explore spatial distribution of epidemic EM using EM Search Index. Our results showed that cities in the middle of Guangdong Province and some provinces in South China had relative high search frequencies on symptoms of EM during the outbreak period. Some of the cities or provinces with high EM Search Index have ever occurred one or even more than one epidemic EM outbreaks in previous studies, such as Guangzhou[13,36], Shenzhen[37], Zhongshan[38] of Guangdong Province and Zhejiang[31], Jiangsu[39], Henan[4], Fujian[3,9], Hebei[40], Hubei[4,6] and Hunan[4,5,12]. On the other hand, the whole country showed a low search frequency during the same period in 2013, when there was no large temperature fluctuation, suggesting that there was a real epidemic during this time. From our findings, we speculate that these cities or provinces with greater EM Search Index counts might have experienced epidemic EM outbreaks during this period in 2014, when the temperature experienced a large fluctuation.

There are some limitations of this current study. First of all, Baidu doesn't release the search data of keywords without sufficient search volume, which might result in an underestimation of correlation. Additionally, although the selected keywords captured the trend of outbreak data very well, there still may be some omission due to the diversity of online search habits, and we haven't got other data for model validation. Thirdly, a number of factors affect the individual search behavior thereby influencing the sustainability of our detection model[20,23,25]. Also, Internet access is uneven throughout China, with the lowest provincial Internet penetration of 32.6% in Jiangxi Province and the highest (75.2%) in Beijing (Internet penetration of each province/municipality of mainland China in 2013 see Supplementary Table S2), and the population sizes of different regions are also different. Thus, the accuracy of comparison of actual search index counts between cities or provinces should be considered with caution.

In conclusion, the EM Search Index using Baidu search term methodology may be a good indicator for early detection of an epidemic EM outbreak, especially when combined with temperature change monitoring.

References

1. Kvernebo, K. Erythromelalgia: A condition caused by microvascular arteriovenous shunting. *Vasa-j Vascular Dis.* **51,** 1–39 (1998).
2. Reed, K. B. & Davis, M. Incidence of erythromelalgia: a population-based study in Olmsted County, Minnesota. *J Eur Acad Dermatol.* **23,** 13–15 (2009).
3. Dai, Y. S., Xu, M. H., Zheng, J. Q., Chi, Y. D. & Chen, K. C. Survey of epidemic erythromelalgia in Sanming City. *Chinese Journal of School Doctor.* **5,** 69–71 (1991). (in Chinese).
4. Li, Q. G. & Gao, Q. Epidemic erythromelalgia in middle school students. *Chinese Journal of Rural Medicine.* **6,** 12–14 (1992). (in Chinese).

5. Luo, Z. B. *et al.* Analysis on risk factors of epidemic erythromelalgia. *Chinese Journal of School Health*. **11,** 42–43 (1990). (in Chinese).

6. Mo, Y. M. *et al.* Unique effect between epidemic of erythromelalgia and EL Nino. *Central China Medical Journal*. **12,** 116–117 (1988). (in Chinese).

7. Wang, X. J. & Dong, C. R. Primary discussion of the relationship between epidemic erythromelalgia and El Nino. *Medical Journal of Wuhan University*. **8,** 208–213 (1987). (in Chinese).

8. Wang, C. Z. *et al.* Survey of epidemic situation about erythromelalgia in Hainan Province. *Hainan Medical Journal*. **3,** 40–44 (1992). (in Chinese).

9. Yan, Z. P. Epidemiological survey of erythromelalgia in the west of Fujian. *Chinese Public Health Acta*. **9,** 260 (1990). (in Chinese).

10. Long, T., Yang, Z. L., Wang, H., Jin, D. M. & Qian, M. Investigation on epidemic erythromelalgia. *Occuption and Health*. **21,** 713–714 (2005). (in Chinese).

11. Wang, Q. X. & Li, Y. Relationship between meteorological condition and epidemic erythromelalgia. *Journal of Shandong Meteorology*. **8,** 47–49 (1988). (in Chinese).

12. Xie, F. G., Ning, H. W., Lei, Y. M. & Lei, S. B. Epidemiological survey on an outbreak of epidemic erythromelalgia. *Practical Preventive Medicine*. **17,** 1921–1922 (2010). (in Chinese).

13. Zhu, S. H., Huang, Z. Z., Liu, Z. L. & Li, Z. W. Clinical analysis of 433 cases with erythromelalgia in Guangzhou City. *National Medical Journal of China*. **46,** 365–367 (1960). (in Chinese).

14. Liu, T. *et al.* A large temperature fluctuation may trigger an epidemic erythromelalgia outbreak in China. *Sci Rep*. **5,** doi: 10.1038/srep09525 (2015).

15. China Internet Network Information Center. *The 33rd statistical report on internet development in China*. (2014) (Date of access: 05/03/2014) Available at: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201403/P020140305346585959798.pdf. (in Chinese).

16. Diaz, J. A. *et al.* Patients' use of the Internet for medical information. *J Gen Intern Med*. **17,** 180–185 (2002).

17. Ginsberg, J. *et al.* Detecting influenza epidemics using search engine query data. *Nature*. **457,** 1012–1014 (2009).

18. Bao, J.-x., Lv, B.-f., Peng, G. & Li, N. Gonorrhea incidence forecasting research based on Baidu search data. in *2013 International Conference on Management Science and Engineering*, Harbin, China. Piscataway, NJ, USA: IEEE, 2013 Published.

19. China Internet Network Information Center. *2013 Chinese search engine market research report*. (2014) (Date of access: 01/01/2014) Available at: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/ssbg/201401/P020140127366465515288.pdf. (in Chinese).

20. Eysenbach, G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. in *AMIA Annual Symposium Proceedings*, Washington, DC, USA. Washington, DC, USA: AMIA, 2006 Published.

21. Polgreen, P. M., Chen, Y., Pennock, D. M., Nelson, F. D. & Weinstein, R. A. Using internet searches for influenza surveillance. *Clin Infect Dis*. **47,** 1443–1448 (2008).

22. Luo, Y. *et al.* Using multi-source web data for epidemic surveillance: A case study of the 2009 Influenza A (H1N1) pandemic in Beijing. in *2010 IEEE International Conference on Service Operations and Logistics and Informatics*, Qingdao, China. Piscataway, NJ, USA: IEEE, 2010 Published.

23. Zhou, X.-c. & Shen, H.-b. Notifiable infectious disease surveillance with data collected by search engine. *Journal of Zhejiang University-SCIENCE C*. **11,** 241–248 (2010).

24. Althouse, B. M., Ng, Y. Y. & Cummings, D. A. Prediction of dengue incidence using search query surveillance. *PLos Negl Trop Dis*. **5,** e1258 (2011).

25. Kang, M., Zhong, H., He, J., Rutherford, S. & Yang, F. Using google trends for influenza surveillance in South China. *PLoS ONE*. **8,** e55205 (2013).

26. Yuan, Q. *et al.* Monitoring influenza epidemics in china with search query from baidu. *PLoS ONE*. **8,** e64323 (2013).

27. Liu, Y., Lv, B., Peng, G. & Yuan, Q. A preprocessing method of internet search data for prediction improvement: application to Chinese stock market. in *Proceedings of the Data Mining and Intelligent Knowledge Management Workshop*, Beijing, China. NewYork, NY, USA: ACM, 2012 Published.

28. Hulth, A., Rydevik, G. & Linde, A. Web queries as a source for syndromic surveillance. *PLoS ONE*. **4,** e4378 (2009).

29. Cook, A. R., Chen, M. I. C. & Lin, R. T. P. Internet search limitations and pandemic influenza, Singapore. *Emerg Infect Dis*. **16,** 1647–1649 (2010).

30. Chen, K. D. *et al.* Survey of erythromelalgia outbreak in students in Dangshan Country high school *Chinese Journal of School Health*. **15,** 291–292 (1994). (in Chinese).

31. Ke, Z. G., Zhang, W. Z. & Li, H. J. Outbreak of erythromelalgia in Lanju middle school. *Zhejiang Journal of Preventive Medicine*. **2,** 18–19 (1990). (in Chinese).

32. Luo, G. S. *et al.* Epidemiological survey of erythromelalgia outbreak with 296 cases. *Jiangxi Medical Journal*. **22,** 329–330 (1987). (in Chinese).

33. Wang, R. H., Li, G. C., Han, K. Q. & Wang, X. H. First epidemic of erythromelalgia in Zhenning minzu middle school. *Guizhou Medical Journal*. **16,** 111–112 (1992). (in Chinese).

34. Xie, Z. C. *et al.* Epidemiological survey of erythromelalgia. *South China Journal of Preventive Medicine*. **17,** 19–22 (1991). (in Chinese).

35. Yang, D. H. Epidemiological survey of erythromelalgia in Qinzhou City. *Journal of Preventive Medicine Information*. **6,** 193–195 (1990). (in Chinese).

36. Feng, Y. Q., Yuan, A. Q., Zeng, J. Q., Feng, Y. C. & Deng, T. H. Epidemiological survey of erythromelalgia in Panyu health school. *South China Journal of Preventive Medicine*. **4,** 24–25 (1990). (in Chinese).

37. Huang, H. M. Clinical and epidemiological survey of 34 cases with erythromelalgia. *Academic Journal of Guangdong College of Pharmacy*. **11,** 197 (1995). (in Chinese).

38. Chen, X. J., You, H. Q., Yang, Y. M. & Zhou, S. W. Clinical analysis of 102 cases with erythromelalgia. *Academic Journal of Guangdong College of Pharmacy*. **19,** 91–93 (2003). (in Chinese).

39. Chen, X. S. & Lu, L. G. Erythromelalgia and clinical analysis of 35 cases. *Chinese Journal of Psychiatry*. 117–125 (1956). (in Chinese).

40. Li, S. M., Wang, H. Y., Cao, G. J. & Zhao, Z. Z. Survey on an epidemic of erythromelalgia. *Chinese Journal of School Health*. **27,** 634 (2006). (in Chinese).

Author Contributions

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Gu, Y. *et al.* Early detection of an epidemic erythromelalgia outbreak using Baidu search data. *Sci. Rep.* **5**, 12649; doi: 10.1038/srep12649 (2015).