



OPEN

Using the Baidu index to understand Chinese interest in thyroid related diseases

Qian Hu, Yuan-lin Mou, Ruo-yun Yin, Lei Tang & Fan Zhang✉

Common thyroid diseases are hyperthyroidism, hypothyroidism, thyroiditis, thyroid tumor and so on. Baidu is currently the most widely used online search tool in China, has developed an internet search trends collection and analysis tool called the Baidu Index. The aim of the present study was to understand the trend and characteristics of public's online attention to thyroid diseases, and to explore the value of Baidu Index in monitoring online retrieval behavior of thyroid-related information. Taking the period from January 1, 2011 to December 31, 2019 as the time range into consideration, we used the big data analysis tool of Baidu Index and took "thyroid nodules", "thyroid cancer", "thyroiditis" "hyperthyroidism" and "hypothyroidism" as the keywords, the data of "search index" and "media index" were recorded on a weekly basis, and all information were aggregated into quarterly and annual to generate the final data which was carried out for secondary analysis. Pearson correlation analysis was used to analyze the correlation between the search index of keywords and the year. One-way Analysis of Variance was used to analyze the differences between search index and media index. Among the five keywords, thyroid nodule search index had the highest growth rate (640%), followed by thyroid cancer (298%). The media's attention to thyroid diseases had been declining year by year. Unlike the public's attention, the media index of hyperthyroidism was significantly higher than other keywords. Over the past nine years, the public's attention to thyroid-related diseases has been increasing gradually. Baidu Index is an effective tool to track the health information query behavior of Chinese internet users, which can provide a cost-effective supplement to traditional monitoring system.

Abbreviations

TDs Thyroid diseases
USI Universal salt iodization

Thyroid diseases (TDs) are a category of non-communicable disease that are easy to neglected, misdiagnosed and poorly managed¹. At both ends of the spectrum, inadequate or excessive iodine intake can lead to thyroid disorders². China was an iodine deficient country with a high prevalence of iodine deficiency disorders³, and in 1996, China implemented Universal Salt Iodization (USI) legislation nationally. During the 20 years of USI enactment, China has experienced excessive iodine intake (defined as median urine iodine concentration (UIC) ≥ 300 $\mu\text{g/L}$) for 5 years (1996–2001), more than adequate iodine intake (defined as median UIC from 200 to 299 $\mu\text{g/L}$) for 10 years (2002–2011), and adequate iodine intake (defined as median UIC from 100 to 199 $\mu\text{g/L}$) for 5 years (2012–2016)⁴. Since TDs are a health care and socio-economic burden in China, there has been increased interest in research on TDs intervention and prevention^{5–7}.

The development of the internet has greatly changed people's lives, especially the expansion of search engines, which has further enhanced the value of the internet as a tool for life, learning and work. According to the 49th Statistical Report on Internet Development in China, there were approximately 1032 million internet users in China by the end of December 2021, and the internet penetration rate reached 73.0%⁸. It was estimated that the utilization rate of search engine among netizens was about 81.3%. 77.3% of users could find the information they need through this service. Baidu search accounted for 90.9% of search engine users, ranking first⁹.

Baidu Index is a big data sharing platform constructed by massive user behavior information, which shows the search trend of the selected keywords, gains insight into the changes in the needs of netizens, monitors the

School of Public Health, Research Center for Medicine and Social Development, Collaborative Innovation Center of Social Risks Governance in Health, Chongqing Medical University, No.61 Daxuecheng Middle Road, Shapingba District, Chongqing 400016, China. ✉email: zhangfan@cqmu.edu.cn

trend of media public opinion, and locates the characteristics of users. The platform can provide data such as search index, demand map, information index, media index and population attributes. Currently, scholars have used Baidu index big data to analyze health data, and the research involves various aspects such as the assessment of online search trends and real demand for Lower urinary tract symptoms¹⁰, evaluation of outbreak monitoring prediction models for COVID-19 epidemics^{11,12}, and prediction of the incidence of HIV/AIDS in China¹³. Through the analysis of these online search trend data, it is possible to reflect the pattern of health information search behavior and interest of internet users on population level. And there are no studies on TDs using the Baidu index yet.

This study used the Baidu Index data platform to obtain data and conduct secondary analysis in order to understand the characteristics of public attention to TDs, information search behavior and the trend of media attention, to explore the value of internet search data in monitoring online information search behavior. It provides a basis for meeting the public's need for understanding TDs, targeting the prevention and treatment of TDs, and complementing the advantages of traditional TDs surveillance systems.

Methods

Data from Baidu index. The data from Baidu Index (<http://index.baidu.com/Helper/?tpl=help&word=#pdesc>) was used. Baidu Index is a big data sharing platform constructed by massive user behavior information, which shows the search trend of the selected keywords, gains insight into the changes in the needs of netizens, monitors the trend of media public opinion, and locates the characteristics of users. The platform can provide data such as search index, demand map, information index, media index and population attributes.

The data used in this study included: (1) search index: the data based on the search volume of netizens in Baidu, with keywords as statistical objects, scientifically analyze and calculate the weighted search frequency of each keyword in Baidu web search. (2) media index: the number of news reported by major internet media related to keywords and included by Baidu News Channel. (3) annual netizen search rate: search index/annual number of netizens (the annual number of netizens comes from the Statistical Report on Internet Development in that year).

The keyword “thyroid” was searched through the demand map of Baidu Index platform, and the weekly keyword demand map was collected in December 2019. The keywords related to TDs with the highest demand were selected: “thyroid nodule”, “thyroid cancer”, “thyroiditis”, “hyperthyroidism” and “hypothyroidism”. The two nouns of non-thyroid-related diseases: “what are the symptoms of thyroid” and “thyroid function” were excluded. The search index and media index for each keyword from January 1, 2011 to December 31, 2019 were obtained, a total of nine complete years. At the same time, due to the limitations of Baidu Index tools and the needs of research and analysis, this study recorded the data of search index and media index with weekly as the smallest unit, and summarized them to the quarter and year as the basis for subsequent data analysis.

Statistical methods. In order to understand the characteristics and trends of public and media attention to TDs, we conducted an analysis by the following statistical methods. We added up the five keyword search indexes of each year to get the annual search index; the differences of annual search index, quarterly search index and annual media index of each keyword were analyzed by one-way ANOVA; the correlation between search index and year was analyzed by Pearson correlation analysis. After drawing the scatter plot and the regression line of the netizens' search rate in each year, the covariance analysis was conducted to test the statistical difference of the slope of the regression line among each group. $P < 0.05$ (two-tailed) was considered statistically significant. Microsoft Office Excel 365 (Microsoft, Redmond, WA, USA) and SPSS version 20.0 (SPSS, Inc., Chicago, IL, USA) were used to draw figures, and all statistics analyses were performed with SPSS.

Results

Changes in search index. Over the past nine years, the sum of the annual search index of each keyword showed an upward trend and was positively correlated with the year (Pearson's correlation = 0.983, $P < 0.001$). The Fig. 1 showed the changing trend of the annual search index. Each keyword was also positively correlated with the year (thyroid nodule: Pearson's correlation = 0.981, $P < 0.001$, thyroid cancer: Pearson's correlation = 0.956, $P < 0.001$, thyroiditis: Pearson's correlation = 0.934, $P < 0.001$, hyperthyroidism: Pearson's correlation = 0.784, $P = 0.012$; hypothyroidism: Pearson's correlation = 0.954, $P < 0.001$). In terms of search index growth, the absolute increase of thyroid nodule search index was the highest (4,236,537), followed by hyperthyroidism (1,845,562). The growth rate of thyroid nodule was the highest (640%), followed by thyroid cancer (298%). The changes of each search index over the nine years and their correlation with years were represented in Table 1.

Using the least-significant difference method, we found that there was a statistical difference between the search index of thyroid nodule and thyroid cancer, thyroiditis and hypothyroidism ($P < 0.001$), and between hyperthyroidism and thyroid cancer, thyroiditis and hypothyroidism ($P < 0.001$). However, there was no statistical difference between thyroid nodule and hyperthyroidism ($P = 0.838$). The search index of thyroid nodule surpassed that of hyperthyroidism for the first time in April 2015 and was higher than that of hyperthyroidism for four consecutive years; the search index of thyroid nodule and hyperthyroidism was always higher than that of the other three keywords in nine years. The specific results were shown in Table 2.

As shown in Fig. 2, in the past nine years, the annual search rate of netizens showed an upward trend, and the regression linear slope of the five keywords was all greater than 0. The results of the covariance analysis showed that there was a statistical difference in the linear regression slope between different groups ($F = 16.876$, $P < 0.001$).

Changes in media index. Unlike the keywords search index, the media index showed a downward trend in nine years (Pearson's correlation = -0.835 , $P = 0.005$). The Table 3 showed the changes in the media index for each keyword over the nine years. Among them, the media index of hyperthyroidism was statistically different

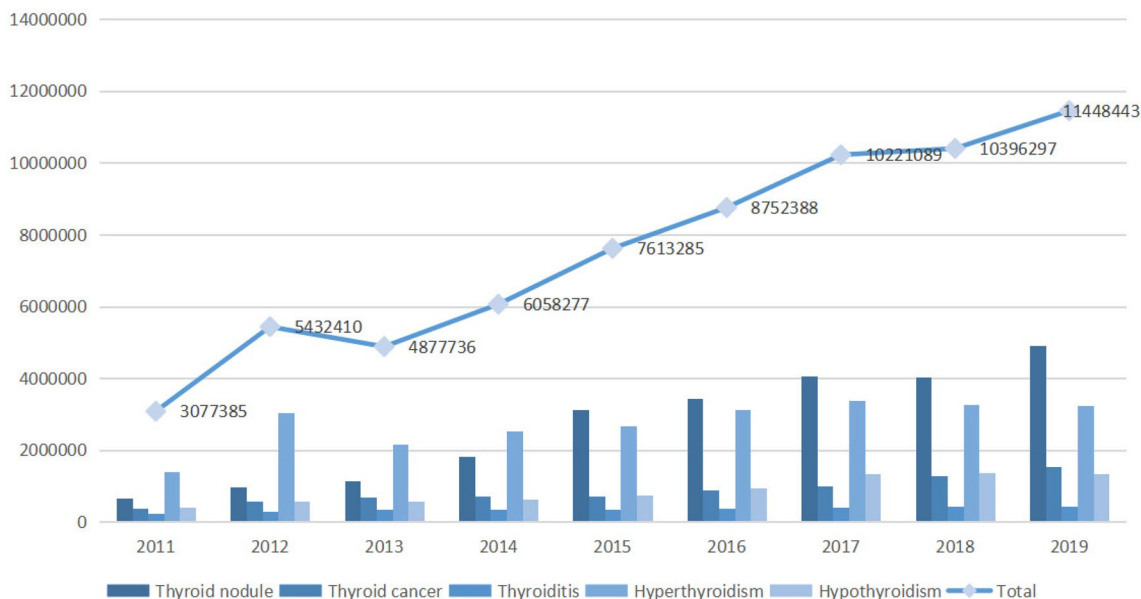


Figure 1. Changes in the annual search index.

Keywords	2011	2012	2013	2014	2015	2016	2017	2018	2019	Increment	Increment rate (%)	Correlation coefficient	P
Thyroid nodule	662,208	964,825	1,140,236	1,808,814	3,122,765	3,425,706	4,071,499	4,034,230	4,898,745	4,236,537	640	0.981	<0.001
Thyroid cancer	388,935	561,461	677,555	712,290	723,164	877,078	1,007,796	1,297,074	1,549,871	1,160,936	298	0.956	<0.001
Thyroiditis	221,298	293,053	344,820	357,781	348,763	378,601	409,109	429,130	421,535	200,237	90	0.934	<0.001
Hyperthyroidism	1,402,019	3,027,818	2,150,914	2,536,677	2,670,551	3,136,113	3,393,895	3,258,544	3,247,581	1,845,562	132	0.784	0.012
Hypothyroidism	402,925	585,253	564,211	642,715	748,042	934,890	1,338,790	1,377,319	1,330,711	927,786	230	0.954	<0.001
Annual search index	3,077,385	5,432,410	4,877,736	6,058,277	7,613,285	8,752,388	10,221,089	10,396,297	11,448,443	8,371,058	272	0.983	<0.001

Table 1. Basic situation and correlation analysis of search index from 2011 to 2019.

Keywords		P	95% CI	
Thyroid nodule	Thyroid cancer	<0.001	1,058,300.69	2,571,433.53
	Thyroiditis	<0.001	1,568,426.69	3,081,559.53
	Hyperthyroidism	0.838	-833,797.98	679,334.87
	Hypothyroidism	<0.001	1,043,897.13	2,557,029.98
Thyroid cancer	Thyroiditis	0.181	-246,440.42	1,266,692.42
	Hyperthyroidism	<0.001	-2,648,665.09	-1,135,532.24
	Hypothyroidism	0.969	-770,969.98	742,162.87
Thyroiditis	Hyperthyroidism	<0.001	-3,158,791.09	-1,645,658.24
	Hypothyroidism	0.169	-1,281,095.98	232,036.87
Hyperthyroidism	Hypothyroidism	<0.001	1,121,128.69	2,634,261.53

Table 2. Multiple comparisons between keywords (search index).

from that of thyroid nodule ($P=0.039$), thyroiditis ($P<0.001$), and hypothyroidism ($P=0.010$). The relationship between other keywords were shown in Table 4.

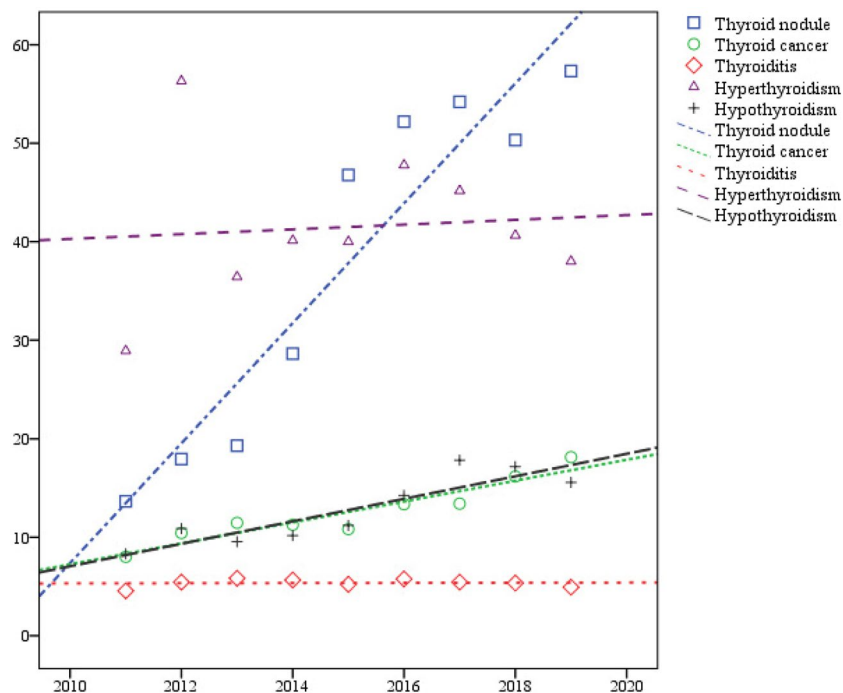


Figure 2. Scatter plot of annual netizen search rate.

Keywords	2011	2012	2013	2014	2015	2016	2017	2018	2019
Thyroid nodule	327	662	322	1336	638	201	170	134	250
Thyroid cancer	624	1382	371	481	651	445	204	199	218
Thyroiditis	173	282	129	280	13	0	0	0	0
Hyperthyroidism	1818	1237	1442	1960	672	380	126	120	190
Hypothyroidism	739	523	232	784	385	118	67	67	76
Media index	3681	4086	2496	4841	2359	1144	567	520	734

Table 3. Basic situation of media index from 2011 to 2019.

Keywords		P	95% CI	
Thyroid nodule	Thyroid cancer	0.772	-471.04	352.15
	Thyroiditis	0.092	-60.15	763.04
	Hyperthyroidism	0.039	-845.49	-22.29
	Hypothyroidism	0.570	-295.04	528.15
Thyroid cancer	Thyroiditis	0.050	-0.71	822.49
	Hyperthyroidism	0.073	-786.04	37.15
	Hypothyroidism	0.393	-235.60	587.60
Thyroiditis	Hyperthyroidism	<0.001	-1196.93	-373.73
	Hypothyroidism	0.256	-646.49	176.71
Hyperthyroidism	Hypothyroidism	0.010	138.85	962.04

Table 4. Multiple comparisons between keywords (media index).

Discussion

The results of this study showed that public attention to TDs had increased in the past nine years, but there were differences in different diseases. The attention of thyroid nodule and hyperthyroidism was significantly higher than that of hypothyroidism, thyroid cancer and thyroiditis, and the growth rate of thyroid nodule search index was more than twice that of the second place. Although all keywords showed an upward trend, the rising trend

of thyroid nodule was more obvious than the other four keywords. This might be related to the increase in the prevalence of thyroid nodule in recent years¹⁴. The incidence of thyroid nodules was insidious, and most patients were asymptomatic in the early stage, and patients were more likely to inquire relevant information on their own after detecting discomfort¹⁵. As the largest search tool in China, Baidu's search results can reflect people's needs well. The disease prediction product jointly developed by Baidu and the Chinese Center for Disease Control and Prevention can provide real-time data on infectious diseases¹⁶. At the same time, it can also be used to predict the epidemic trend of diseases, as a powerful complement to the traditional detection system¹⁷. Baidu Index has not been used for TDs related research in China. Our study is the first attempt to explore the behavior and interest of Chinese netizens in TDs, confirming the potential of using online search trend data to represent the real situation of TDs patients in China.

For the media index part, the results showed that the media's attention to the hyperthyroidism was higher than other keywords. This suggested that the media had pushed and reported more information about hyperthyroidism to the public in the past nine years. Overall, the media attention of TDs was on the decline. The reason might be related to the rapid development of the internet, the scattered news points, the shortage of media practitioners and the declined in the number of media concerned about TDs.

Despite the huge medical expenditure imposed on China by TDs¹⁸, due to China's vast territory and large population, it is difficult to evaluate the true prevalence rate of TDs and to understand the characteristics and needs of TDs patients. With the wide application of the internet and the increasing reliance of the public on search engines as the main way to query health information, some online digital diseases surveillance tools has been explored in recent years^{19–21}. As a query tool, search engine can provide sensitive information on the disease before the diagnosis of the disease is reported, thus improving disease control. Internet big data has a broad application prospect in the medical field, which may be a supplement and an expansion of the current clinical and epidemiological data. Today, with the rapid development of the internet services and search engines, combined with network data analysis can be regarded as an auxiliary means of traditional disease monitoring.

Limitations. This study also has several limitations. First, we only focused on the attention of Baidu search engine users to TDs, without considering the public attention on other search engines or social media, which can only reflect part of the public's attention to TDs. Second, there might be sampling biases in Baidu Index. Although the internet penetration rate in China had been greatly improved, the characteristics of internet users were obviously skewed to those with higher socioeconomic level and better educated segments. Meanwhile, although the target population of this study was Chinese, it was unavoidable that a small number of foreigners were included in the data. Third, although the data from the Baidu index were processed by a weighted filtering algorithm, the specific algorithm of Baidu Index has not been made public, so its validity and reliability cannot be assessed yet. Therefore, future research should consider including multiple search engines or social media for analysis to ensure the richness of the data. At the same time, the data should be mined in depth to control the influence of confounding factors on the study results and make the results more objective.

Conclusion

Between 2011 and 2019, the online search rate of TDs maintained a sustained growth while the media index showed a downward trend. The Baidu Index can be used to track Chinese netizens' online behavior and interest in TDs. This may help to improve our understanding of the incidence of disease, patient education and the use of online resources. Internet search trend data is a valuable source for monitoring the search behavior of TDs-related information. It can be used as an exploratory tool to better understand the characteristics and preferences of patients and provide a scientific evidence for the control and prevention of TDs in China.

Data availability

The data analyzed in this study are available in Baidu Index Data Platform (<http://index.baidu.com/Helper?tpl=helpandword=#pdsc>).

Received: 21 December 2021; Accepted: 27 September 2022

Published online: 13 October 2022

References

- Fualal, J. & Ehrenkranz, J. Access, availability, and infrastructure deficiency: The current management of thyroid disease in the developing world. *Rev Endocr Metab Dis.* **17**, 583–589. <https://doi.org/10.1007/s11154-016-9376-x> (2016).
- Zimmermann, M. B. & Boelaert, K. Iodine deficiency and thyroid disorders. *Lancet. Diabetes Endocrinol.* **3**, 286–295. [https://doi.org/10.1016/S2213-8587\(14\)70225-6](https://doi.org/10.1016/S2213-8587(14)70225-6) (2015).
- Ma, T., Guo, J. & Wang, F. The epidemiology of iodine-deficiency diseases in China. *Am J Clinical Nutrition.* **57**(2), 264S–S266. <https://doi.org/10.1093/ajcn/57.2.264S> (1993).
- Li, Y. *et al.* Efficacy and safety of long-term universal salt iodization on thyroid disorders: Epidemiological evidence from 31 provinces of mainland China. *Thyroid.* **30**, 568–79. <https://doi.org/10.1089/thy.2019.0067> (2020).
- Shan, Z. *et al.* The iodine status and prevalence of thyroid disorders after introduction of mandatory universal salt iodization for 16 years in China: A cross-sectional study in 10 cities. *Thyroid.* **26**, 1125–1130. <https://doi.org/10.1089/thy.2015.0613> (2016).
- Liang, Z., Xu, C. & Luo, Y. Association of iodized salt with goiter prevalence in Chinese populations: A continuity analysis over time. *Mil Med Res.* **4**, 8. <https://doi.org/10.1186/s40779-017-0118-5> (2017).
- Gu, F. *et al.* Incidence of thyroid diseases in Zhejiang Province, China, after 15 years of salt iodization. *J Trace Elem Med Bio.* **36**, 57–64. <https://doi.org/10.1016/j.jtemb.2016.04.003> (2016).
- China Internet Network Information Center (CNNIC). The 49th China Statistical Report on Internet Development. http://www.cnnic.net.cn/hlwfyj/hlwzbg/hlwjtbg/202202/t20220225_71727.htm#, Accessed 19 July 2022.
- China Internet Network Information Center (CNNIC). China statistical report on search engine, 2019. (2019). http://www.cnnic.cn/hlwfyj/hlwzbg/ssbg/201910/t20191025_70843.htm. Accessed 19 July 2022.

10. Wei, S. *et al.* Using search trends to analyze web-based interest in lower urinary tract symptoms-related inquiries, diagnoses, and treatments in Mainland China: Infodemiology study of Baidu index data. *J. Med. Internet Res.* **23**(7), e27029. <https://doi.org/10.2196/27029> [publishedOnlineFirst:2021/07/14] (2021).
11. Fang, J. *et al.* Baidu index and COVID-19 epidemic forecast: Evidence from China. *Front. Public Health* **9**, 685141. <https://doi.org/10.3389/fpubh.2021.685141> (2021).
12. Tu, B. *et al.* Using Baidu search values to monitor and predict the confirmed cases of COVID-19 in China: Evidence from Baidu index. *BMC Infect. Dis.* **21**(1), 98. <https://doi.org/10.1186/s12879-020-05740-x> (2021).
13. He, G. *et al.* Using the Baidu search index to predict the incidence of HIV/AIDS in China. *Sci. Rep.* **8**(1), 9038. <https://doi.org/10.1038/s41598-018-27413-1> (2018).
14. Yu, C. C. & Wang, Q. An initial analysis on thyroid nodules prevalence and influencing factors of Chinese healthy adults in. *J. Environ. Health.* **33**, 440–43 (2016).
15. Zou, B., Wang, X., Sun, L., Zhou, Y. & Chen, Z. T. Prevalence of thyroid nodules in health checkups and its relationship with metabolic diseases. *Chin. Gen. Pract.* **23**, 2423–28 (2020).
16. Li, K. *et al.* Using Baidu search engine to monitor AIDS epidemics inform for targeted intervention of HIV/AIDS in China. *Sci Rep.* **9**, 320. <https://doi.org/10.1038/s41598-018-35685-w> (2019).
17. Chen, S. Y. Big data applications and practices of Baidu. *Big Data Res.* **1**, 104–114 (2015).
18. Xie, D. Y., Wang, Q., Li, C., Wang, Z. & Song, C. Y. Analysis of the influencing factors of hospitalization expense of patients with thyroid carcinoma in a hospital. *China Health Insur.* **11**(1), 46–50 (2018).
19. Tu, B. Z., Wei, L. F., Jia, Y. Y. & Qian, J. Using Baidu search values to monitor and predict the confirmed cases of COVID-19 in China: Evidence from Baidu index. *BMC Infect. Dis.* **21**, 98. <https://doi.org/10.1186/s12879-020-05740-x> (2021).
20. Nuti, S. V. *et al.* The use of google trends in health care research: A systematic review. *PLoS ONE* **9**, e109583. <https://doi.org/10.1371/journal.pone.0109583> (2014).
21. Gluskin, R. T., Johansson, M. A., Santillana, M. & Brownstein, J. S. Evaluation of internet-based dengue query data: Google Dengue trends. *PloS Negl. Trop. Dis.* **8**, e2713. <https://doi.org/10.1371/journal.pntd.000271> (2014).

Author contributions

All authors contributed to the study conception and design. The manuscript writing and data analysis were completed by Q.H., Y.-L.M. gave the paper revision and format adjustment work. The rest of the authors gave the paper revision and grammar editing work. R.-Y.Y. and L.T. gave the entire process technical and paper writing guidance support. The corresponding author F.Z. reviewed the article and was responsible for communication with the editors. All authors read and approved the final manuscript.

Funding

This study was funded by the 14th Five-year Plan of Chongqing Education Science in 2021, China (No. 2021-GX-312).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022