International Journal of Community Medicine and Public Health
Thamaraikannan R et al. Int J Community Med Public Health. 2022 Oct;9(10):3917-3923
http://www.ijcmph.com
pISSN 2394-6032 | eISSN 2394-6040

DOI: https://dx.doi.org/10.18203/2394-6040.ijcmph20222587

Review Article

Using "Google trends" for dengue surveillance and epidemiological research

Ravichandran Thamaraikannan[1], Vijayakumar Munisamy[2]*

[1]Medical Officer, Government Erode Medical College, Perundurai, Erode, Tamil Nadu, India
[2]Institute of Community Medicine, Madras Medical College, Chennai, Tamil Nadu, India

Received: 26 July 2022
Accepted: 02 September 2022

***Correspondence:**
Dr. Vijayakumar Munisamy,
E-mail: drvijaym12@gmail.com

ABSTRACT

Millions of people worldwide search online for health-related information and search engines have become an increasingly popular resource for accessing health-related information and provides valuable source. Key words used as well as the number and geographic location of searches can provide trend data, available by Google trends. In this study exploring this resource using dengue disease as an example. Objectives were to use Google trends data for comparison across different locations in India for the past 5 years, and to assess the specific search terms used in Google trends data and to correlate the real time dengue outbreak of Tamil Nadu with Google trend search. It was a cross sectional study. Data collection was done via Google search queries and record was included. Weekly trends were accessed from Google Trends. Data is a randomly collected sample of real time and non-real time Google search queries. Search traffic for the string "dengue fever" reflected increased likelihood of exposure and the string "dengue symptoms and treatment" had higher relative traffic during rainy season. Cities and states with the highest amount of search traffic for "dengue disease" overlapped where dengue is endemic. Found that search trend data produced by Google to approximate the seasonality, spikes at September to November every year and geographic distribution also identified in dengue disease. Web search query data were found to be capable of tracking dengue activity and predict periods of large incidence of dengue with high accuracy and may prove useful.

Keywords: Google trends, Dengue, Public health surveillance, Epidemiology, Information technology

INTRODUCTION

Dengue is a mosquito-borne viral infection.[1] The virus responsible for causing dengue, is called dengue virus (DENV). There are four DENV serotypes, meaning that it is possible to be infected four times. Severe dengue is a leading cause of serious illness and death in some states in India and many other countries.[2] Dengue virus is transmitted by female mosquitoes mainly of the species *Aedes aegypti* and, to a lesser extent, *Aedes albopictus*.[2] Dengue has distinct epidemiological patterns, associated with the four serotypes of the virus. The global incidence of dengue has grown dramatically in recent decades. About half of the world's population is now at risk. There are an estimated 100-400 million infections each year.[2] In India, dengue is of special concern because the area affected by dengue has expanded and the incidence has steadily increased in recent years. According to the report from national vector borne disease control programme (NVBDCP), Ministry of Health and Family Welfare, Government of India, the dengue incidence was increased during the past decade.[3,4] In particular, India experienced an unprecedented outbreak of dengue in 2017, and the number of cases reached the highest level over the past 25 years.

There was a major outbreak in the state of Tamil Nadu in 2017.[3,4] Dengue is the major public health problems in

Tamil Nadu.[5,6] Aedes species are endophagic and a daytime biter; its biology, geographical range and transmissibility are dependent upon climatic variables, temperature, humidity and wind velocity.[7] Studies from India have correlated increased dengue incidence with the monsoon and post-monsoon season.[8]

Developing countries are facing surveillance system problems like delay and data loss. Access and the availability of health-related information on the internet have changed what people seek on the web.[9] Accurate real-time tracking of dengue outbreaks helps public health officials make timely and meaningful decisions that could save lives. In 2004 Google developed google dengue trends (GDT) based on the number of search terms related with the disease in a determined time and place.[9-11] Recent studies have exploited internet search queries to monitor in a timely fashion infectious diseases that have substantial seasonal and geographic variation. Google trends can be used as tool for prediction of disease outbreaks and to estimate the patterns of disease and population behavior.[10,11]

However, it still remains unclear what is the degree of correlation between frequencies of searches for key queries and dengue fever dynamics in India.[13] This study aimed to analyze the correlation between the real time dengue outbreaks of Tamil Nadu with Google trend search. Then, use Google trends data for comparison of dengue related search terms across different locations in India for the past 5 years and also to assess the specific search terms used in Google trends data in Tamil Nadu.

METHODS

A cross-sectional study was done using Google trends data for dengue fever and related search queries from 2014–2019 for India and Tamil Nadu. Tamil Nadu is the one of the southern states in India. Dengue has been rampant in parts of Tamil Nadu in the past two decades. The prevalence of dengue vector and silent circulation of dengue viruses have been detected in rural and urban Tamil Nadu, which is ever increasing.

We used official dengue reports for real time outbreaks, weekly data from the IDSP website and NVBDCP website, under the Ministry of Health and Family welfare.[3,4] Over the study period included the 29 states of India for the time trend plots and access the Google search terms. Time trend plots to assess specific search terms and compare across different locations in India for the past 5 years. Weekly trends can be accessed from Google trends, a special open-access domain of Google. Data were downloaded in comma separated values (CSV) file from Google trends' website (https://trends.google.com/trends/) and are available on a weekly basis. Data were obtained using variable search terms related to disease definition, symptom, treatment, and vector of disease. Search terms were collected from Google trends (search terms listing the most frequently used) and Google correlate (search terms which have a similar pattern with the search term).

In Google trends each data point is divided by the total searches in a specific geographical area over a period of time to compare relative popularity and scaled on a range of 0 to 100 based on a topic's proportion. Pearson correlation and scatter plots used to correlate Google trends and official data for dengue fever in Tamil Nadu. Pearson correlation was performed for search terms with the highest pattern of similarity with official dengue reports. The correlation strength was defined as a correlation coefficient r value of 0.7 ($p \leq 0.05$). Statistical analysis was done using statistical package for the social sciences (SPSS) version 16.

RESULTS

Google trends search volume data of keywords were extracted from the Google trends website. We first determined primary search terms by choosing names, symptoms, vectors of dengue and treatment by highest number of searches and topics proportion.

Results of data analysis in Figure 1 show the time series of Google trends index for dengue cases in India from 2015 until 2019. There were five peaks of dengue cases with the highest peak in 2015 and 2017. Figure 1 shows the Google search index for dengue outbreaks per period in India which tended to increase between August to November.

From the Figure 2, it shows the states with the highest amount of search traffic of Google trends index for the search term "dengue fever" during the period of 2015-2019. Nagaland and Uttarakhand were the top two states which showed the highest search trafficking. Tamil Nadu were in the seventh place.

Figure 3 shows that the various dengue related search terms in Google trends which trends in India during the period 2015-2019. From the graph it clearly showed that the related search terms were overlapping each other. Most commonly search term 'dengue mosquito', 'dengue treatment' followed by 'dengue fever symptoms'. In August 2019, the search term of 'dengue vaccine' showed the spike and trends in the Google, which may due the arrival of dengue vaccine in some places in world.

Line diagram from the Figure 4 depicted the Google trends index of 'dengue fever' in Tamil Nadu (2015-2019). There were the four spikes during the time period of 2015-2019. There was a peak and highest spike during the August 2017 to November 2017 in Tamil Nadu.

From the Figure 5, showed the districts of Tamil Nadu with the highest amount of search traffic of Google trends index for the search term "dengue fever" during the period of 2015-2019. Salem and Vellore were the top two districts which showed the highest search trafficking.

From the Figure 6, the various dengue related search terms in Google trends which trends in Tamil Nadu during the period 2015-2019. From the graph it clearly showed that the related search terms were overlapping each other. Most commonly search term 'dengue symptoms', 'Nilavembu' followed by 'prevention of dengue'. There was high spike during the August 2017 – November 2017.

In Tamil Nadu, the search term 'prevention of dengue' was constantly used over the intervals of time periods.

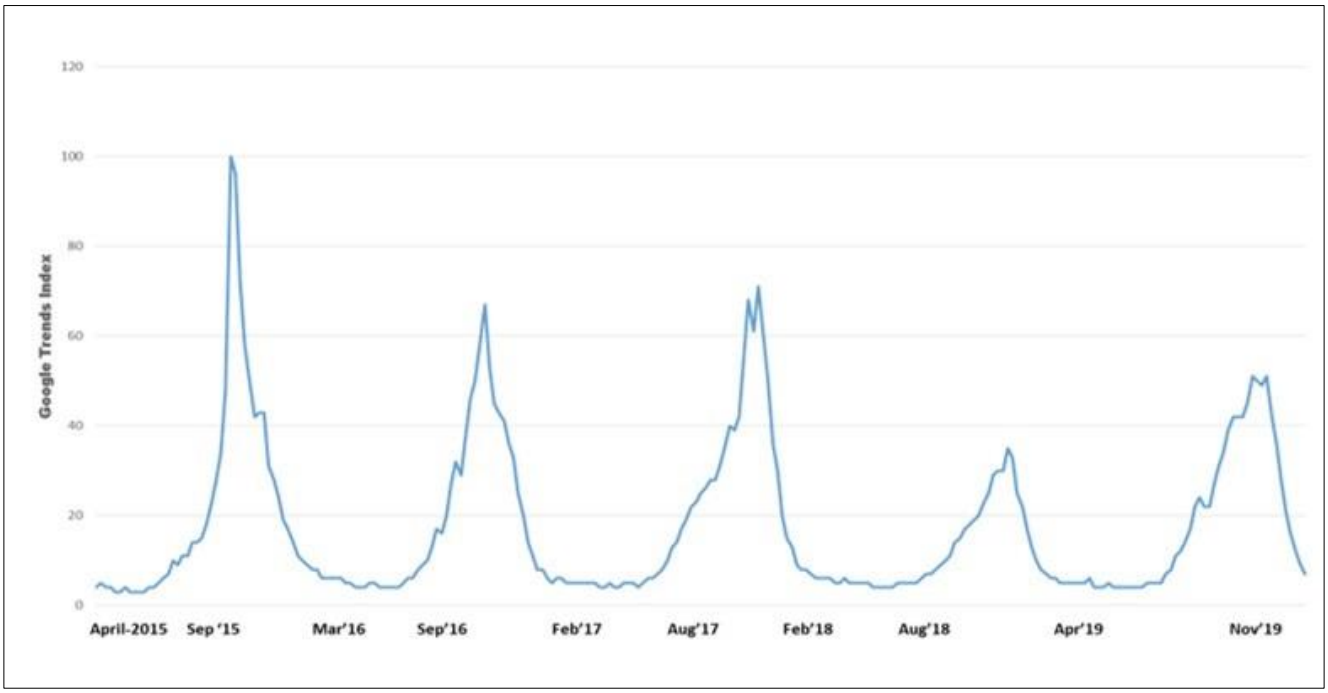


Figure 1: Line diagram depicting Google trends index of dengue cases in India (2015-2019).

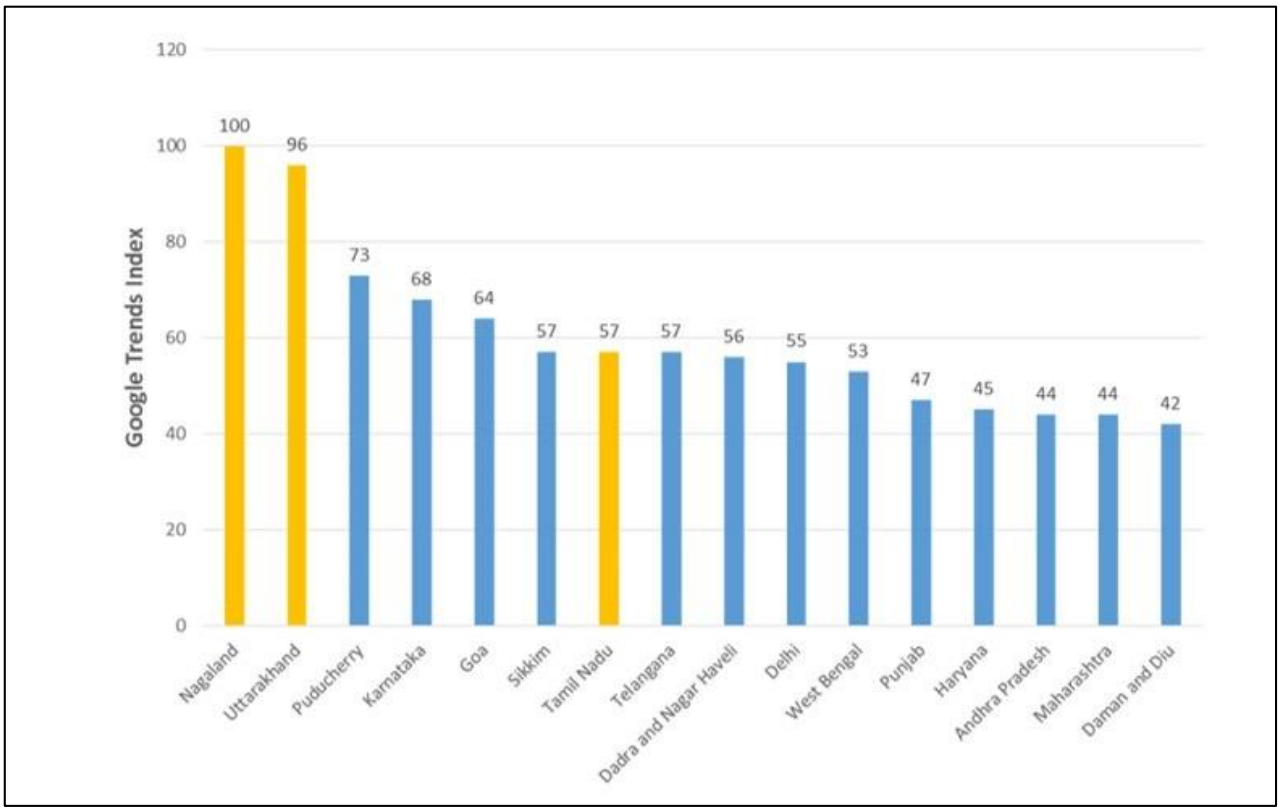


Figure 2: States with the highest amount of search traffic for "dengue fever" (2015-2019).

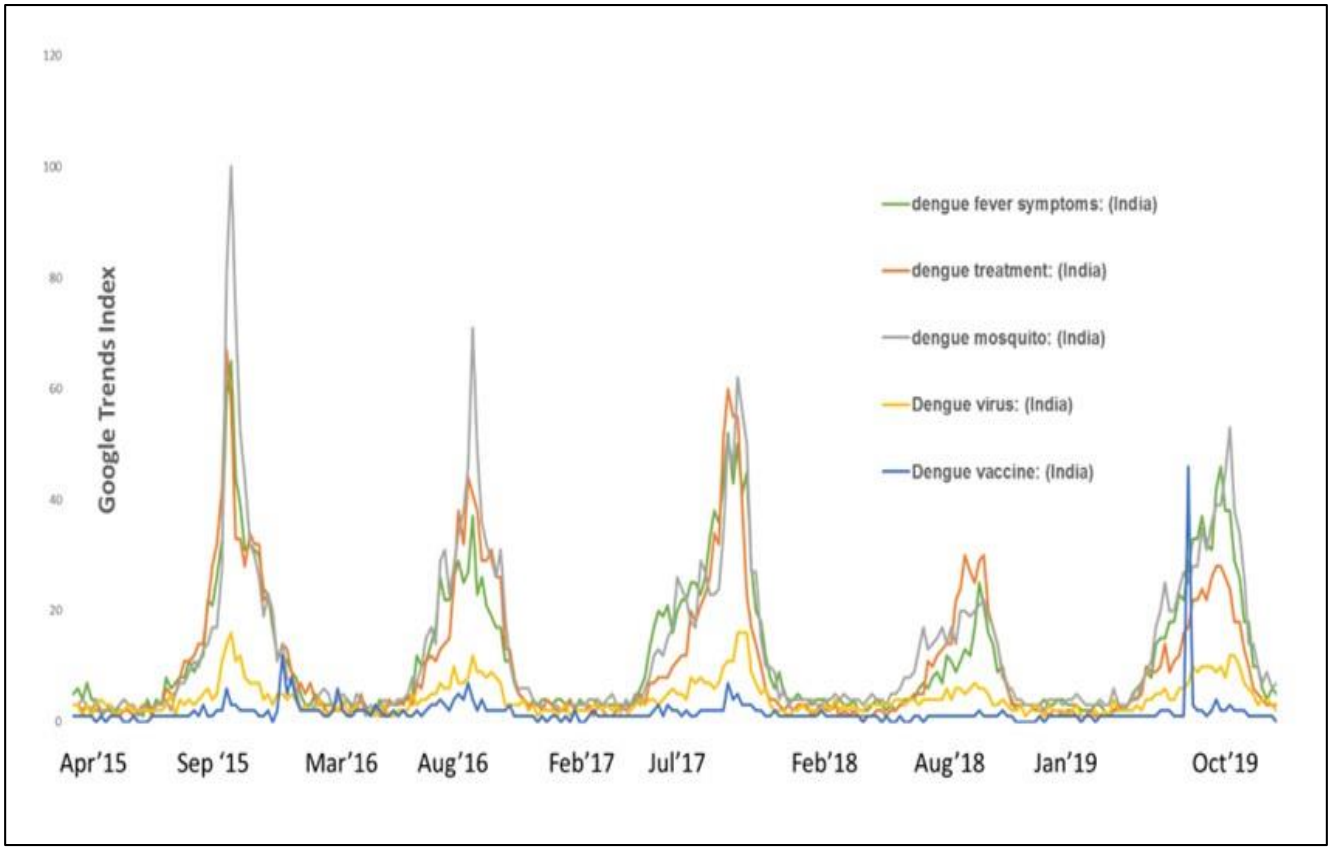


Figure 3: Google trends graph depicting tendency over time to related search terms India (2015-2019).

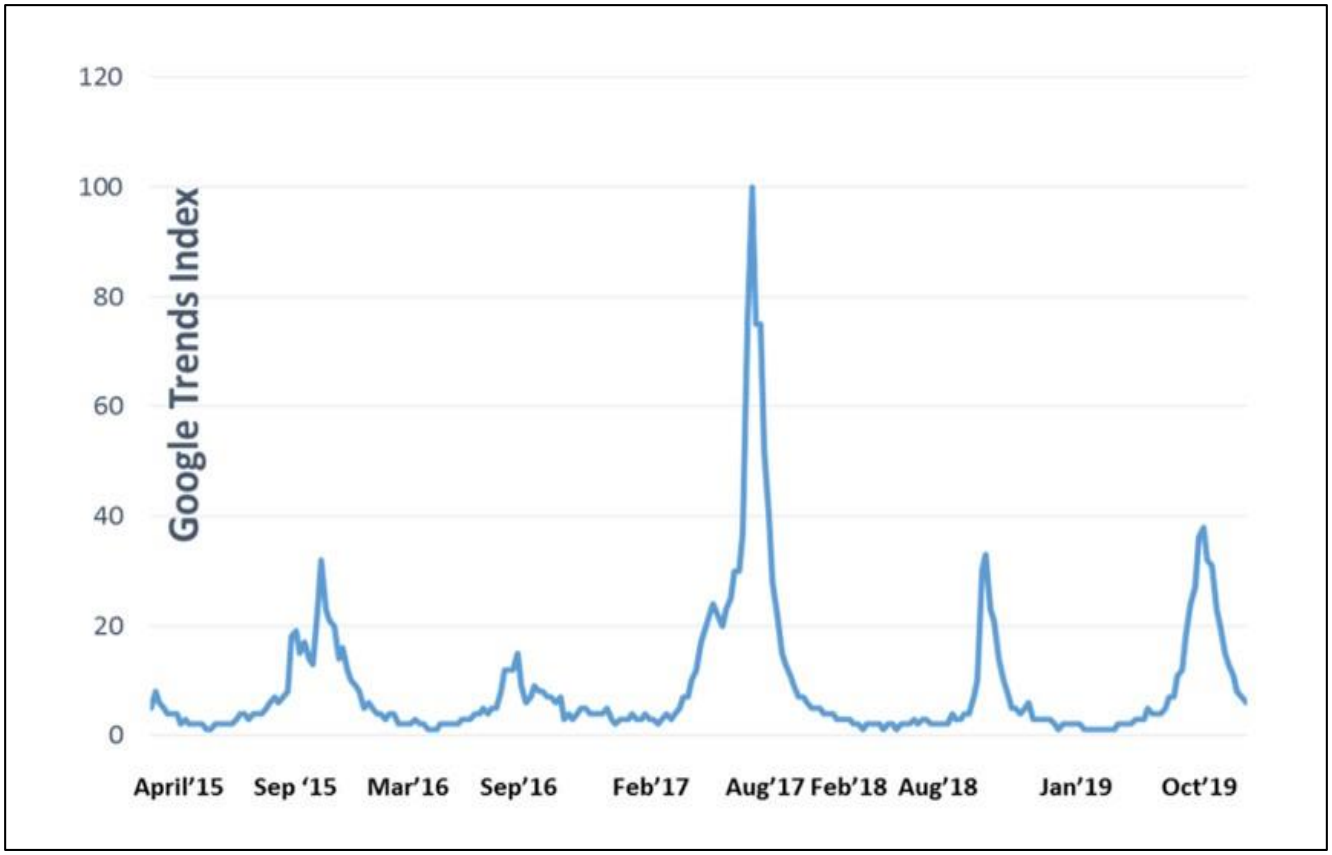


Figure 4: Line diagram depicting Google trends index of dengue fever in Tamil Nadu (2015-2019).

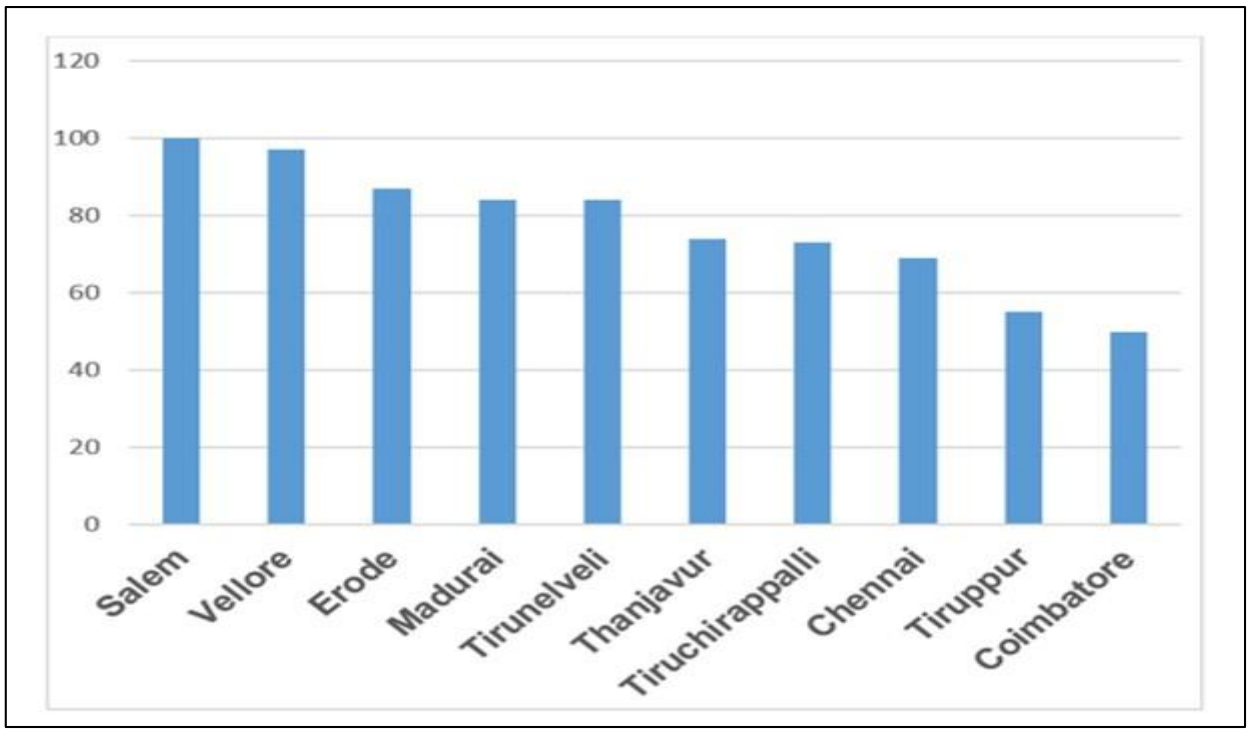


Figure 5: Districts with the highest amount of search traffic for "dengue fever" in Tamil Nadu (2015-2019).

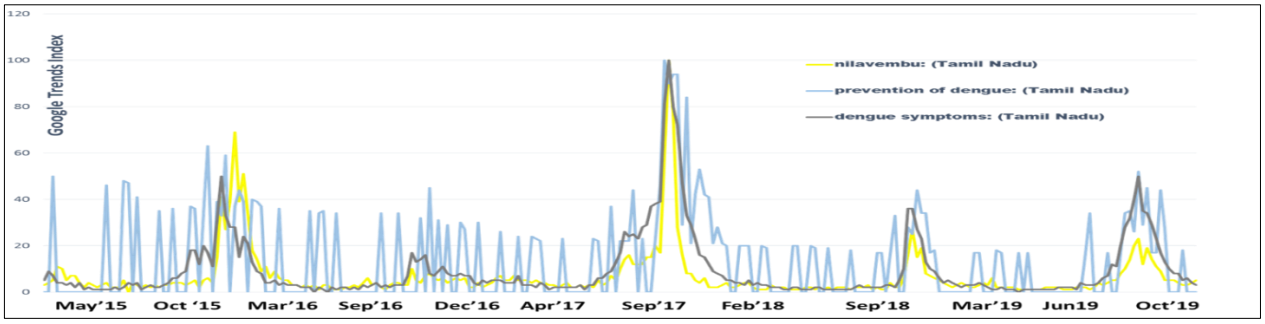
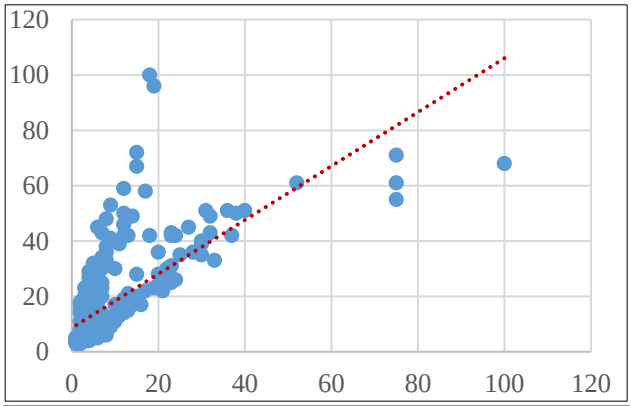


Figure 6: Google Trends graph depicting tendency over time to related search terms (Tamil Nadu 2015-2019).



Correlation	P value	Correlation coefficient
Pearson correlation	0.03	0.7

Figure 7: Scatter plot between Google trends and official report for dengue fever in Tamil Nadu (2015-2019).

Results of Pearson correlation in Figure 7, show high correlation (r value ≥0.7 and p≤0.05) between official dengue reports and the Google trends data. During the time period, R-value seems to be increased in every epidemic period and search terms 'dengue fever' and 'dengue symptoms' seem to have stable r value.

DISCUSSION

Google, a search engine which also includes several featured tools and resources, represents an important way to seek online information about health care. Google trend based predicting surveillance made a robust use in seasonal flu. Accurate estimation of influenza epidemic may predict by using google flu trends (GFT).[14,15] Since the launch of GFT in 2009, much attention and respect has been devoted to the new evolving branch of 'digital epidemiology'.[16] Application and investigation of internet-based surveillance is widely recognized. The present study demonstrates that an Internet-search based surveillance

system has the potential to effectively contribute to the control of dengue. However, correlations alone should not be viewed as definitive evidence of epidemics or impending outbreaks as the analyses. Interpretation of this study results with caution keeping in mind the biological plausibility and concerned with the natural history of the disease.

Furthermore, Google trends use as the epidemiological approach to find the outbreak of dengue and surveillance. From the Figure 1-3, the diagram showed that the Google trends search terms about dengue and related searches in India. Figure 2 depicted that the highest states with search traffic for 'dengue fever' in India 2015-2019. When compare these figures with official data from government, there were an epidemic outbreak during that spike period.[3,4] The Google trends searches indirectly predicted that the ongoing break.

From the Figures 4-6, the diagram showed that the Google trends search terms about dengue and related searches in Tamil Nadu. Figure 5 showed the districts with highest search term for 'dengue fever' during the period of 2015-2019. By comparing this data with the official report of Tamil Nadu government data on dengue fever from IDSP website.[3,4] The months were spike showed in Google trends which correlate with the official report as epidemics during that period.

From the Figure 7, scatter plot between Google trends and official report for dengue fever in Tamil Nadu (2015-2019). Pearson correlation indicated high correlation for defined search terms. The present study which was similar to the study conducted by Verma in India and Husnayain in Indonesia.[13,17] During the outbreak season in 2017, we found an 99% increase in search volume data across the entire keyword set we analyzed (Figures 1, 3, 4, and 6). We found that the constructed dengue search index on Google showed an obviously similar changing trend with dengue dynamics, suggesting that Google trend is a good indicator for estimating intensity and peak incidence of dengue fever in India.

Google dengue trends is a valuable indicator and timely available tool that can be used by public health authorities to monitor the behavior of the disease, and potentially guide the application of control measures against dengue in epidemic countries like India.[18] Google trends has very modest reliability for delineating the true population epidemiology. Unique and innovative technology takes us one step closer to true real-time outbreak surveillance. Search trend data produced by Google approximates the seasonality and geographic distribution. Useful in identifying more subtle geographic spread in incidence of emerging diseases and for its control. Google trends does not provide data to identify the exact place of an outbreak at intra-district level or block level. This tool could be useful in other countries to track the behavior of other infectious diseases.

The limitation was the available data was based on a sample of Google web searches, with the potential for non-representative sampling bias. To investigate the epidemiological trends of some specific diseases, reliability of this approach remains largely speculative. This was a retrospective analysis of correlation, and the prospective performance of internet-based surveillance systems for dengue needs a further evaluation. Moreover, although this work showed a strong correlation between the Google trends and the official report of dengue case counts, future studies should analyze the difference between the two distributions of the notification date of dengue cases and that of people seeking information in order to develop an early warning system.

CONCLUSION

Our study represents the first attempt demonstrating a strong correlation between internet search trends and dengue epidemics in India. These findings will be useful for the government in identifying initiatives needed to strengthen the capacity of traditional surveillance systems for dengue.

Recommendations

Recommend that internet-based surveillance system, cannot be used as an alternative but can be used as a supplement to the existing traditional surveillance systems.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Park K. Textbook of Preventive and Social Medicine. 25th Edition. Jabalpur: Bhanot; 2019;269.
2. Dengue and severe dengue. 2020 [cited 4 November 2020]. Available from: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue
3. Ministry of Health & Family Welfare, Government of India. Integrated Disease Surveillance Programme [Internet].New Delhi: Ministry of Health & Family Welfare; [cited at 2020 Jul 4]. Available from: http://www.idsp.nic.in/.
4. Ministry of health & family welfare. National vector borne disease control programme. Available at: https://nvbdcp.gov.in/index4.php?lang=1&level=0&linkid=431&lid=3715. Accessed on 8 July 2020.
5. Gunasekaran P, Kaveri K, Mohana S, et al. Dengue disease status in Chennai (2006-2008): a retrospective analysis. Indian J Med Res. 2011;133(3):322-5.
6. Tewari SC, Thenmozhil V, Katholi CR, Manavalan1 R, Munirathinam A, Gajanana A. Dengue vector prevalence and virus infection in a rural area in south India. Trop Med Int Health. 2004;9:499–507.
7. Lee HS, Nguyen-Viet H, Nam VS, Lee M, Won S, Duc PP, Grace D. Seasonal patterns of dengue fever

and associated climate factors in 4 provinces in Vietnam from 1994 to 2013. BMC Infect Dis. 2017;17(1):218.
8. Chandy S, Ramanathan K, Manoharan A, Mathai D, Baruah K. Assessing effect of climate on the incidence of dengue in Tamil Nadu. Indian J Med Microbiol 2013;31:283-6
9. Gluskin R, Johansson M, Santillana M, Brownstein J. Evaluation of Internet-Based Dengue Query Data: Google Dengue Trends. PLoS Neglected Tropical Diseases. 2014;8(2):e2713.
10. Chan EH, Sahai V, Conrad C, Brownstein JS (2011) Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLoS Negl Trop Dis 5: e1206.
11. Althouse BM, Ng YY, Cummings DA (2011) Prediction of dengue incidence using search query surveillance. PLoS Negl Trop Dis 5: e1258.
12. Chan EH, Sahai V, Conrad C, et al. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLoS Negl Trop Dis. 2011;5.
13. Verma M, Kishore K, Kumar M, Sondh A, Aggarwal G, Kathirvel S. Google Search Trends Predicting Disease Outbreaks: An Analysis from India. Healthcare Informatics Research. 2018;24(4):300.
14. Yang S, Santillana M, Kou S. Accurate estimation of influenza epidemics using Google search data via ARGO. Proceedings of the National Academy of Sciences. 2015;112(47):14473-8.
15. Polgreen PM, Chen Y, Pennock DM, Nelson FD, Weinstein RA Using Internet searches for influenza surveillance. Clin Infect Dis 2008; 47(11):1443–8.
16. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. Journal of Epidemiology and Global Health. 2017;7(3):185.
17. Husnayain A, Fuad A, Lazuardi L. Correlation between Google Trends on dengue fever and national surveillance report in Indonesia. Global Health Action. 2019;12(1):1552652.

Cite this article as: Thamaraikannan R, Munisamy V. Using "Google trends" for dengue surveillance and epidemiological research. Int J Community Med Public Health 2022;9:3917-23.