

# Analysis of Correlation between Google Search Trends and Dengue Outbreaks from India

NASIR SALAM<sup>1</sup>, FARAH DEEBA<sup>2</sup>, FAIZAN QADIR<sup>3</sup>, FRAS AL-HIJLI<sup>4</sup>, YASER NAIF AL-OTAIBI<sup>5</sup>

## ABSTRACT

**Introduction:** Dengue has become an endemic problem in India with frequent outbreaks reported from several parts of the country every year. A passive surveillance system is unable to deal with the mounting number of cases every year. In the last couple of years the internet has become a valuable tool to access healthcare related information.

**Aim:** To analyse the correlation between dengue cases reported every year and annual Internet search data for the term “dengue” obtained via Google trends.

**Materials and Methods:** Dengue incidence data was collected from the National Vector Borne Disease Control Programme (NVBDPC) database alongside relative search volumes from Google trends for the term “dengue” and correlation was estimated by calculating Pearson’s correlation coefficient.

**Results:** The data was analysed from the year 2004 up to year 2017. In this period, a total of 693,318 cases of dengue were reported from India. Relative search volume for dengue searches on Google trends was found to be highly correlated with dengue incidence data. Google trends indicate a seasonal pattern showing maximum search volume in the monsoon months, which also coincides with most dengue cases. Internet based searches and Google trends can be used in addition to traditional surveillance methods for predicting disease outbreaks.

**Conclusion:** The analysis shows an overall strong positive correlation between incidence of dengue cases and Google trends indicating the usefulness of internet searches in gathering healthcare related information during the time of outbreaks. Internet based searches could be an additional tool along with classical surveillance methods for accurately predicting disease outbreaks in resource poor settings.

**Keywords:** Disease surveillance, Epidemiology, Internet trends

## INTRODUCTION

Dengue is an arboviral disease that continues to expand its geographical realm. Its global expansion has reached nearly 30 folds in the last 50 years [1]. The disease is primarily spread by *Aedes aegypti* mosquitoes, which transmit any of the four serotypes of dengue virus [2]. There are 390 million cases reported every year with 3.9 billion people living in risk zones in 128 countries [3]. The disease often results in flu-like illness with some cases progressing to Dengue Haemorrhagic Fever (DHF) and Dengue Shock Syndrome (DSS), which can result in death. As of 2019, United States Food and Drug Administration have approved Dengvaxia, a live recombinant tetravalent vaccine produced by Sanofi-Pasteur and licensed in 20 countries but yet to be introduced in India [4]. Earlier Clinicians relied primarily on supportive therapy for treatment of dengue [5]. The spread of dengue is influenced by several human factors like large-scale migration of populations from rural to urban areas that might result in overcrowding in small spaces without development of proper civic infrastructure [6]. The disease shows a clear seasonal pattern with most cases being reported around the monsoon season, which is conducive for breeding of mosquitoes [7]. India is one of the most endemic countries in the world that witnesses dengue outbreaks almost every monsoon season. Nearly, 190,000 cases were reported in 2017 alone [8]. These cases represent the tip of the iceberg, as nearly 99% of the cases are not recorded by the NVBDPC, the official government body for collection of disease surveillance data [9]. There are several lacunae in correctly diagnosing and reporting cases of dengue infection. Passive surveillance system as established in India is unable to detect asymptomatic cases and those with undifferentiated febrile illnesses [10]. In several cases with mild symptoms, the infection is resolved within a week ruling out the visit to a health care provider [11].

India has seen a rise in socio-economic status of its population indicated by better living standards and access to modern technology like internet usage and smartphones. This has changed the way people handle day-to-day life, including access to health related information. Nearly, 30% of India’s population is using the internet with maximum usage in the cities of Mumbai and Delhi [12]. Web queries and internet based searches might be able to give an idea about the extent of disease and provide supportive information in addition to lab based diagnostic methods [13]. Analysing localised search patterns for disease-associated keywords could be a faster approach as compared to traditional surveillance method for that particular region. Several reports have been published that employs Google trends for such analysis giving an indication about their spread [14-17].

The aim of this study was to analyse the correlation between annual Google trends for dengue in India and disease incidence for the time period starting from the year 2004 to 2017.

## MATERIALS AND METHODS

The dengue incidence data was collected for the time period of 1<sup>st</sup> January 2004 to 31<sup>st</sup> December 2017. The data for dengue incidence for India was obtained in 2018 from NVBDPC (<https://www.nvbdpc.gov.in>) which is the repository of all epidemiological data pertaining to vector borne diseases from India and is updated on a monthly basis. Retrospective yearly data for the past 14 years from 2004 to 2017 was collected. Yearly Google trends data associated with the selected keyword “dengue” in English was obtained by setting the location parameter to India and time period from 2004 to 2017. The output was provided as Relative Search Volume (RSV) for each month for every year in Comma-Separated Values (CSV) file. The monthly RSV was added for each year to provide total RSV. The data was finally compiled in Excel sheets.

## STATISTICAL ANALYSIS

Pearson correlation coefficient and coefficient of determination was calculated using StatPlus software to identify correlation between annual dengue cases and annual dengue trends on Google.

## RESULTS

Retrospective data obtained from year 2004-2017 indicated a total of 693,318 cases with 2272 deaths over a period of 14 years. Maximum number of cases (188,401) and maximum number of deaths (325) were reported from the year 2017 [Table/Fig-1].

Year	Dengue cases	Dengue deaths	Google trends data (RSV)
2004	4153	45	48
2005	11985	157	73
2006	12317	184	164
2007	5534	69	58
2008	12561	80	62
2009	15535	96	72
2010	28292	110	147
2011	18860	169	56
2012	50222	242	138
2013	75808	193	131
2014	40571	137	95
2015	99913	220	198
2016	129166	245	181
2017	188401	325	213
Total	693318	2272	1636

**[Table/Fig-1]:** Annual number (N) of dengue cases and dengue deaths as reported by NVBDCP from India and RSV as reported by Google trends.

Data from internet searches using Google trends data indicated an overlapping pattern with the number of cases reported [Table/Fig-2]. In any given year, the searches peaked in the months of September and October, which is also the time when most cases of dengue are reported. Statistical analysis between Google trends and annually reported cases of dengue indicated a Pearson correlation coefficient of 0.81 with  $p < 0.05$ , indicating a strong overall positive correlation between the two data sets.

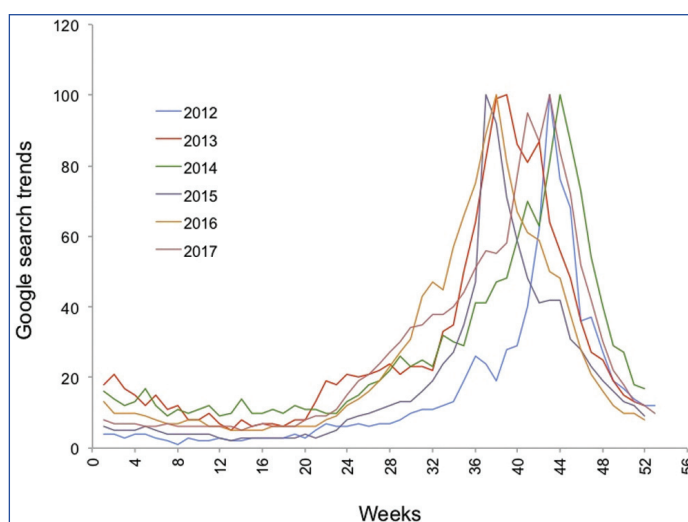
	Dengue cases vs. google search data	Dengue deaths vs. Google search data
<i>R</i> (Pearson correlation coefficient)	0.81	0.82
<i>R-squared</i> (Coefficient of determination)	0.76	0.94
<i>Adjusted R-squared</i>	0.76	0.94
<i>N</i>	14	14
<i>p-value</i>	0.0004	0.0003

**[Table/Fig-2]:** Statistical analysis between dengue cases/deaths and Google search trends.

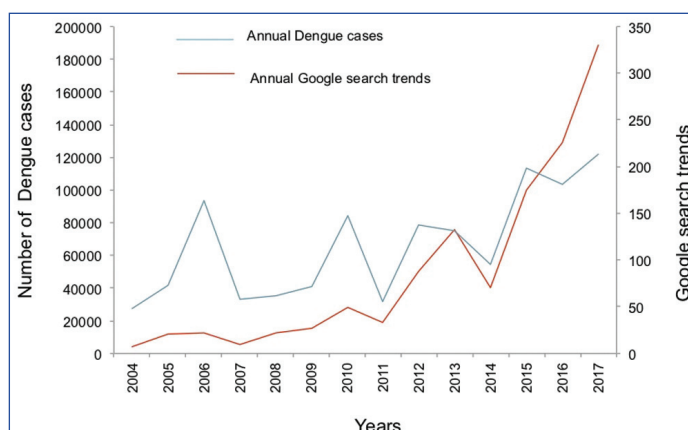
The overall coefficient of determination for this calculation was 0.76. Analysis of Pearson correlation coefficient and coefficient of determination between deaths due to dengue and Google trends was also done. Here also, a strong overall positive correlation between internet searches and dengue deaths was found with Pearson correlation coefficient as 0.82 with  $p < 0.05$ . The coefficient of determination for this analysis was 0.94. [Table/Fig-3] describes weekly google search patterns for the last six years. [Table/Fig-4,5] shows patterns of dengue incidence and deaths.

## DISCUSSION

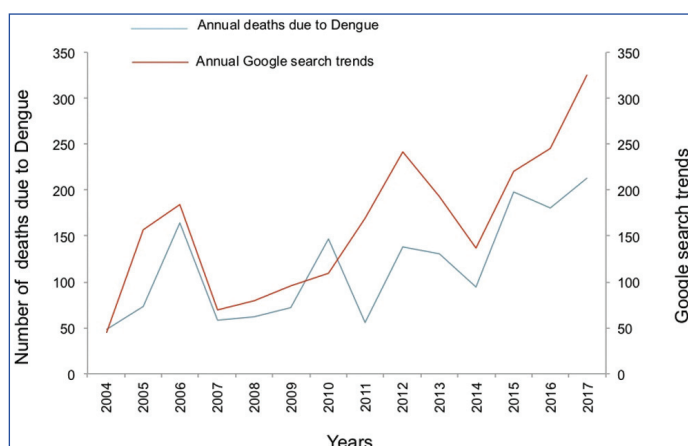
Dengue incidence has increased manifold during the last several years with India contributing significantly to global disease burden. In the absence of any therapeutic intervention or vaccines, disease management relies on supportive therapy and control of vector population [11]. For this to become effective, a robust and active



**[Table/Fig-3]:** Weekly Google search patterns for the last six years (2012-2017).



**[Table/Fig-4]:** Overlapping patterns of dengue incidence and Google search trends.



**[Table/Fig-5]:** Overlapping patterns of deaths due to dengue and Google search trends.

surveillance system is required that targets symptomatic and asymptomatic febrile illnesses during monsoon season. Recent reports suggest that nearly 99% cases of dengue in India go unreported and the true scale of economic and social cost of disease could be much higher [9,10]. Disease surveillance is still dependent on traditional laboratory based methods that relies on the sera of symptomatic individuals showing the presence of IgM antibodies. Asymptomatic individuals may not visit health centres thus remaining unnoticed by the surveillance methods currently employed. Also, many Indians rely on traditional medicine like Ayurveda for several ailments and might remain undetected by a passive surveillance system.

India is gradually expanding its internet coverage with nearly 51% penetration in urban areas and 16% penetration in rural areas. Sixty-five percent of Indians below the age of 35 years are increasingly relying on the internet for travel, trade and availing basic amenities

of life [18]. India is also one of the largest users of smartphones. Health information seeking behaviour has also changed with many using internet for health related queries before visiting a doctor. Several studies have looked at the correlation of disease outbreaks and internet searches analysing and comparing spread of disease with health information seeking behaviour [19,20]. Such studies have shown a strong correlation between the spread of disease and internet traffic about the keywords associated with that disease. The current study also indicates a strong positive correlation between dengue Google trends and annual dengue cases reported from India. Also, the Google searches peaked at the time of maximum disease spread i.e., from the months of August to October, possibly due to people searching for these items and dengue related stories that are covered extensively by the media. Despite an overall 30% user base for the internet, in a country of 1.3 billion people, Google based searches quite impressively reflected the dengue incidence. Keeping an eye on the internet data traffic for monitoring and surveillance of disease outbreak can help prepare health agencies in advance when lab based diagnostic tests might take time.

## LIMITATION

The present study has some limitation as Google trends only captures search patterns of a segment of the population that is using the internet and know how to communicate in English. Though Google is one of the most widely used search engines, the present study did not look at other search engines. Only annual cases of dengue are available publicly from NVBDCP. A weekly break-up of reported cases would have been accurate in predicting disease outbreak at least a few days in advance.

## CONCLUSION

Our study indicates a strong correlation between internet searches and number of dengue cases being reported, highlighting the role of internet based searches as an effective tool in disease surveillance and predicting disease outbreaks in addition to conventionally available methods.

## REFERENCES

- [1] Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. 2013;496(7446):504-07.
- [2] Kilpatrick AM, Randolph SE. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet*. 2012;380(9857):1946-55.
- [3] Dengue and severe dengue. World Health Organization, Geneva, Switzerland. 15 April 2019. Available at: <http://www.who.int/mediacentre/factsheets/fs117/en/>; Accessed on 3<sup>rd</sup> November 2018.
- [4] Halstead SB, Dans LF. Dengue infection and advances in dengue vaccines for children. *Lancet Child Adolesc Health*. 2019;3(10):734-41.
- [5] Kularatne SA. Dengue fever. *Br Med J*. 2015;351:h4661.
- [6] Alirol E, Getaz L, Stoll B, Chappuis F, Loutan L. Urbanisation and infectious diseases in a globalised world. *Lancet Infect Dis*. 2011;11(2):131-41.
- [7] Colon-Gonzalez FJ, Fezzi C, Lake IR, Hunter PR. The effects of weather and climate change on dengue. *PLoS Negl Trop Dis*. 2013;7(11):e2503.
- [8] <https://www.nvbdcp.gov.in/index4.php?lang=1&level=0&linkid=431&lid=3715> accessed on 30<sup>th</sup> September, 2019.
- [9] Shepard DS, Halasa YA, Tyagi BK, Adhish SV, Nandan D, Karthiga KS, et al. Economic and disease burden of dengue illness in India. *Am J Trop Med Hyg*. 2014;91(6):1235-42.
- [10] Bagchi S. Dengue surveillance poor in India. *Lancet*. 2015;386(10000):1228.
- [11] Teixeira MG, Barreto ML. Diagnosis and management of dengue. *Br Med J*. 2009;339:b4338.
- [12] Statistics. International Telecommunications Union (ITU). Available at: <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>; Accessed on 3<sup>rd</sup> November, 2018.
- [13] Lotto M, Ayala Aguirre PE, Rios D, Andrade Moreira Machado MA, Pereira Cruvinel AF, Cruvinel T. Analysis of the interests of Google users on toothache information. *PLoS ONE*. 2017;12(10): e0186059.
- [14] Yang S, Kou SC, Lu F, Brownstein JS, Brooke N, Santillana M. Advances in using Internet searches to track dengue. *PLoS Comput Biol*. 2017;13(7):e1005607.
- [15] Strauss RA, Castro JS, Reintjes R, Torres JR. Google dengue trends: An indicator of epidemic behavior. The Venezuelan Case. *Int J Med Inform*. 2017;104:26-30.
- [16] Shin SY, Seo DW, An J, Kwak H, Kim SH, Gwack J, et al. High correlation of Middle East Respiratory Syndrome spread with Google search and Twitter trends in Korea. *Sci Rep*. 2016;6:32920.
- [17] Gluskin RT, Johansson MA, Santillana M, Brownstein JS. Evaluation of internet-based dengue query data: Google dengue trends. *PLoS Negl Trop Dis*. 2014;8(2):e2713.
- [18] Agarwal S. Internet users to touch 420 million by June 2017: IAMAI report. *Economic Times* 2017; 2nd May available at: <https://economictimes.indiatimes.com/tech/internet/420-million-to-access-internet-on-mobile-in-india-by-june-iamai/articleshow/58475622.cms> Accessed on 30<sup>th</sup> September, 2019.
- [19] Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EH, Olsen JM, et al. Using social media for actionable disease surveillance and outbreak management: A systematic literature review. *PLoS ONE*. 2015;10(10):e0139701.
- [20] Milinovich GJ, Williams GM, Clements AC, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. *Lancet Infect Dis*. 2014;14(2):160-68.

### PARTICULARS OF CONTRIBUTORS:

1. Assistant Professor, College of Medicine, Imam University, Riyadh, Saudi Arabia.
2. Research Associate, Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, Delhi, India.
3. Researcher, Epsilon Learning Solution, Delhi, India.
4. Student, College of Medicine, Imam University, Riyadh, Saudi Arabia.
5. Student, College of Medicine, Imam University, Riyadh, Saudi Arabia.

### NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:

Dr. Nasir Salam,  
College of Medicine, Al Imam Muhammad Ibn Saud Islamic University Riyadh,  
Kingdom of Saudi Arabia, Riyadh, Saudi Arabia.  
E-mail: [salamnasir@gmail.com](mailto:salamnasir@gmail.com)

### AUTHOR DECLARATION:

- Financial or Other Competing Interests: No
- Was Ethics Committee Approval obtained for this study? NA
- Was informed consent obtained from the subjects involved in the study? NA
- For any images presented appropriate consent has been obtained from the subjects. NA

### PLAGIARISM CHECKING METHODS: [Jain H et al.]

- Plagiarism X-checker: Aug 10, 2019
- Manual Googling: Oct 09, 2019
- iThenticate Software: Oct 28, 2019 (12%)

### ETYMOLOGY: Author Origin

Date of Submission: **Aug 10, 2019**  
Date of Peer Review: **Aug 12, 2019**  
Date of Acceptance: **Oct 16, 2019**  
Date of Publishing: **Nov 01, 2019**