How to Cite:

Dengue fever occurrence in India, Brazil, Paraguay, Philippines and Singapore using Google trends

Manish Kumar Thimmaraju
Balaji Institute of Pharmaceutical Sciences, Narsampet, Warangal, Telangana India
Email: manishcancer@gmail.com

Vijaya Kumar Meher
School of Pharmacy, Centurion University of Technology and Management, Odisha, India

Goje Arjun
Teegala Ram Reddy College of Pharmacy, Hyderabad, Telangana, India

Bhavani Boddeda
Koringa College of Pharmacy, Korangi, East Godavari Andhra Pradesh India

Bodika Thirupathy
Balaji Institute of Pharmaceutical Sciences, Narsampet, Warangal, Telangana India

Anil Kumar Garige
Jayamukhi Institute of Pharmaceutical sciences, Narsampet, Warangal, India

Vijitha Chandupatla
Vagdevi Institute of Pharmaceutical sciences, Bollikunta, Warangal Telangana India

Kotthapally Tejashwini Sharma
Balaji Institute of Pharmaceutical Sciences, Narsampet, Warangal, Telangana India

A. Mohathasim Billah
MRM College of Pharmacy, Hyderabad, Telangana, India
**Abstract**---Background. Dengue is a mosquito-borne disease and its occurrence is a worldwide problem. The health seekers may turn to the Internet based data for possible complications and various therapeutic approaches for the treatment. Therefore, our objective is to identify and characterize the seasonal transition of dengue fever using Google Trends. Methods. This was a cross-sectional study of dengue fever internet searches between 2004 and 2017. The Google was indexed the aggregated web data and ranked based on the search volume with respect to time. The countries selected for the study include India, Brazil, Paraguay, Philippines and Singapore based on the maximum number of cases across world. The selected regions were further classified in to seasonal time points and the relative search volume ranks were categorized to two season variables. Mann–Whitney U test was applied to identify the statistical significance. Results. The summer to monsoon transitions were showed highest number cases beginning from 2004 in the selected regions. Furthermore, dengue searches were peaked in both the seasons of summer and monsoon in selected regions. The analyses of data further revealed that 16.2 % of variation in dengue trends is associated with the time between 2004 and 2017. Conclusions. The popularity of Google search activity is associated with the health seekers looking for symptoms and immediate therapeutic interventions which were correlated with seasons. The seasonal dengue prevalence is helpful for disease tracking, identifying the outbreaks, policy planning and directing the products or services to the end users.

**Keywords**---dengue fever, mosquito-borne disease, dengue prevalence.

**Introduction**

Healthcare analytics is the utilization of health data and further developing insights through application of statistical procedures to facilitate drawing conclusions and decision making for learning, planning, measurement and management in healthcare. IBM Analytic solutions for business (https://www.ibm.com/analytics?lnk=404) declared that healthcare analysis helps organization to enjoy benefits like inventory optimization, streamlined operations, customer insight, and predictive and preventive maintenance. The fourth industrial revolution is changing the paradigm of the healthcare industry by converging the physical, digital, and biological spheres based on the information and communication technology (ICT). For the fourth industrial revolution, data is not only the new oil, it is oxygen for digital economy (http://www.financialexpress.com/industry/data-is-not-only-the-new-oil-it-is-oxygen-for-digital-economy-mukesh-ambani/872897) and playing an important role in various industries beginning from fast moving consumer goods to climate change and playing significant role in health and pharmaceutical sector. Data analysis and interpretation enables the insights of data and helps the analyst in making
decisions. A quick decision after analysis of data is essential for various strategic approaches. Healthcare data can be used to 1) empower individuals to better manage a health condition with which they live, 2) help people to understand and contribute to the care of family and friends, 3) raise standards and positive impact the experience of service-users, 4) Assist professionals in commissioning the most effective interventions to benefit the populations they serve, 5) Allow researchers to develop next generations medicine and treatments and 6) enable innovators to deploy new technologies – including the internet of Things (IoT), machine learning and cognitive services – to transform services for public benefit (https://futurecarecapital.org.uk/wp-content/uploads/2017/07/Full-Report-Unleashing-the-potential-of-health-and-care-data.pdf).

The healthcare industry generates large amounts of data and being collected by scientists in academic organizations, insurance companies, government and industry doubles every year. The global number of internet users in 2005 was 16% of 6.5 billion populations and tripled by the year 2016 and responsible for generation large amounts of data in the form of various searches in Google (https://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users). The historical healthcare data is essential for disease surveillance and facilitates to take necessary preventive actions and population health management. Public health surveillance is the iterative process involves systematic data collection and interpretation of healthcare related data for planning, application and assessment of public health practice and serves as an early warning system for emerging public health risks (http://www.who.int/topics/public_health_surveillance/en/).

Infodemiology is known as the science of sharing of contents in the Internet, or in a population, with the purpose to advise public policy and public health. The real time need of information surge (web activity, tweets and news info, etc) at globally or regionally is essential to the public health professionals to address ‘epidemics of fear’. The data is produced from the search and click (web navigation) is the key infodemiology indicators which correlates the behavior of the people. An escalated search queries are deemed to be the primary forecaster of any disease outbreak. The primary forecaster can be observed using Google Trends (https://trends.google.com/trends/). The web users looking for health information frequently initiate using Google search, which account for 81% of market share search by the year 2017 (https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&pcustomid=0). The input queries in Google search are being logically analyzed by Google Trends which is an online tool shows how frequently an exact search-term is entered relative to the gross search-volume across diverse regions of the world, and in different languages in real time. Traditional methods of reporting to physicians, healthcare centers and public dengue surveillance has several drawbacks as these methods take many weeks to months to gather, sort and publish the surveillance data from various hierarchy levels of any country. Recently, we have observed a rise in the use of analysis of web activity to determine
the users interest on wide variety of medical condition. 9-14 In this context, the cognizance of the quantity and details of dengue fever searches observed on Google Trends from various regions globally which may contribute to the identification of societal health needs in order to direct health authorities in the advancement of further action plans, delivering critical healthcare information to affected area. This could be helpful in preliminary epidemiological investigation of disease. It would be very important to track data on real time in regard to dengue fever since the diagnosis depends on the subject who is experiencing fever, headache joint pains, skin rashes, worsening abdominal pain, vomiting, enlarged of the liver, mucosal bleeding, high hematocrit with low platelet count, lethargy, restlessness and serosal effusion. 15

The aim of the proposed study was to characterize the dengue fever interests of Google users from the Paraguay, Brazil, Singapore, Philippines and various states in India with respect to change of seasons. We hypothesize that the surveillance of web user behaviors during change of seasons would provide corresponding data about the impact of dengue in various population groups in different weather conditions at different demographics and assisting in the reduction of damage caused by the self-management of that disease condition. 7

**Materials and methods Countries selection**

The regions selected for the analysis of dengue fever was based on the search interest in Google Trends, great internet diffusion and more than 3 million internet users. The following countries such as India, Brazil, Paraguay, Philippines and Singapore were selected, which were able to meet the requirements for statistical analyses.

**Study design**

The seasonal variation of occurrence of dengue fever in various countries were observed in Google Trends and performed cross-sectional study after querying with specific words like dengue and dengue fever. The relative search volume (RSV) and the Interest over time of dengue fever data was freely obtained from Google Trends, in two separate seasons which includes seasonal transition or between two seasons. The RSV consisting of numbers which represents Google search interest relative to the highest value on the selected region and time. A value of 100 indicates peak popularity and a value of 50 is half of the popular search interest. Furthermore, a value 0 represents there exist insufficient data for the selected term. The RSV is the ratio of fraction of search interest term divided by total searches at unique place and time. 16

**Season selection**

The selection of seasons or seasonal transition is having greater importance in analysis an interpretation of data. The reason behind the seasonal transition or change of season was due to a rapid rise of RSV during the period between 2004 and 2017. The common rises in each year between 2004 and 2017 were identified and further correlated to seasons in various parts of world as mentioned earlier.
The seasonal impact was very high in those selected regions and occurring often in Google Trends. The selected terms for search include dengue and dengue fever. In India, summer is between April and June, monsoon is between July and September and autumn is from October to November (https://en.wikipedia.org/wiki/Climate_of_India). The two data points selected for India were April to June and September to November. The latter period in Google Trends was showed higher RSV and assumed as seasonal transition. Furthermore, states in India were considered for the study instead India alone due to the fact that Delhi (National Capital Region) alone has recorded first worst outbreak over 15000 between 2006 and 2015, which has driven us to study various states in India than India alone (http://www.who.int/mediacentre/factsheets/fs117/en/). Although, in Brazil and Paraguay, the summer is between December and March, the high temperatures are generally observed between January and March (https://en.wikipedia.org/wiki/Climate_of_Brazil). and Monsoon is between April and October (https://en.wikipedia.org/wiki/Monsoon). The two data points selected for Brazil and Paraguay were January to March and April to October. In Philippines, the rainy season is between June and October and summer is between March and May. Furthermore, in Singapore, Northeast Monsoon Season is between December and early March, Southwest Monsoon Season is between June and September (http://www.weather.gov.sg/climate-climate-of-singapore/).

The summary of data points for study was showed in Table 1.

<table>
<thead>
<tr>
<th>Country</th>
<th>Data point-I</th>
<th>Season term</th>
<th>Data point-II</th>
<th>Season term</th>
<th>Selected data point-II</th>
<th>Seasonal transition in the data points</th>
</tr>
</thead>
<tbody>
<tr>
<td>India*</td>
<td>April–June</td>
<td>Summer</td>
<td>July–September</td>
<td>Monsoon</td>
<td>September–November</td>
<td>Yes</td>
</tr>
<tr>
<td>Brazil</td>
<td>January–March</td>
<td>Summer</td>
<td>April–October</td>
<td>Monsoon</td>
<td>April–October</td>
<td>No</td>
</tr>
<tr>
<td>Paraguay</td>
<td>January–March</td>
<td>Summer</td>
<td>April–October</td>
<td>Monsoon</td>
<td>April–October</td>
<td>No</td>
</tr>
<tr>
<td>Philippines</td>
<td>March–May</td>
<td>Summer</td>
<td>June–October</td>
<td>Monsoon</td>
<td>June–October</td>
<td>No</td>
</tr>
<tr>
<td>Singapore</td>
<td>December–March</td>
<td>N/E Monsoon</td>
<td>June–September</td>
<td>S/W Monsoon</td>
<td>June–September</td>
<td>No</td>
</tr>
</tbody>
</table>

(Data point-1 and Selected data point-II were selected for the basis of statistical analyses)

Various states in India were analyzed separately

**Statistical Analysis**

The selected two data point variables based on season change were ranked by Google Trends. These two variables are independent, ranked based search volume (i.e., ordinal responses) and assuming both population distributions are equal. These assumptions were lead us to proceed for determining the statistical difference using Mann–Whitney U (MWU) test. 17 SigmaStat 4.0 (Ref: SC/SSI-IND/GST/2017- 18/325, Starcom Information Technology Limited, Bangalore)
software was used to determine the significance of the data. The data was retrieved from Google Trends for various countries such as India, Brazil, Paraguay, Philippines and Singapore. For India, prospective data from Google Trends were downloaded for various states. The ranked data obtained for various states were not available for each state uniformly between 2004 and 2017. Furthermore, the obtained data was fit into the analyses were inclusive and we did not exclude any zero value data or missing data for each time point. The months of seasons for each country were different in both hemispheres and we assumed that there was a chance of search outbreak which may be an early predictor of epidemic.

Results

Every state in India was selected and ranked data was downloaded. Only seven states out of thirty six states and union territories have showed a significant difference in the occurrence of dengue with respect to season transition. Though, it was possible to select the period between July and September for the data analyses but the unusual search outbreak was observed during the preliminary data forecast between September and November and which has given an indication to select these periods. In the Indian states, the highest median summer value for the data observed for Kerala was 52 and the lowest median summer value for the data observed for Jammu & Kashmir was 8. In addition to the above, the highest median for the selected data point-II was Haryana having 41 and the lowest was Kerala having 12.5. Brazil has the highest median summer value and Paraguay has the lowest median while, Philippines showed highest median dengue ranks during monsoon and the lowest median monsoon ranks were showed by Paraguay. The seasonal increase in percent was calculated and the positive value (seasonal increase,%) is an indication of dengue search outbreaks which are highest during monsoon and negative indicates the outbreaks, which are highest during summer. In Indian states, except Kerala, the numbers of dengue search outbreaks were mostly happened during seasonal transition. The highest dengue fever search outbreak was found in West Bengal and the least was in Maharashtra. The results for Indian states indicate that all the regions were dengue outbreaks have raised from 75% to 238%. In South America, Paraguay has highest summer search outbreak than Brazil by seasonal increase 10% in summer. Philippines and Singapore was showed similar trends like Brazil and Paraguay where summer has more outbreaks. The results of analyses between 2004 and 2017 were showed in the table 2 and figure 1.

Table 2. Summary of seasonal variation of dengue between summer and monsoon

<table>
<thead>
<tr>
<th>Country</th>
<th>Data Point 1</th>
<th>Selected data point-II</th>
<th>Seasonal increase, %</th>
<th>P value MWU Test*</th>
</tr>
</thead>
<tbody>
<tr>
<td>India</td>
<td>Median†</td>
<td>25% 25% 75%</td>
<td>Median† 25% 75%</td>
<td></td>
</tr>
<tr>
<td>Delhi</td>
<td>15.5</td>
<td>9.25 25.25</td>
<td>35.5 17.75 51.25</td>
<td>129</td>
</tr>
<tr>
<td>Haryana</td>
<td>16</td>
<td>10 27</td>
<td>41 28 75</td>
<td>156</td>
</tr>
<tr>
<td>J &amp; K‡</td>
<td>8</td>
<td>5 13</td>
<td>27 17.5 32</td>
<td>238</td>
</tr>
</tbody>
</table>
The difference in the median values between the two groups is greater than would be expected by chance; there is a statistically significant difference (P<0.05)

The negative values indicate Data point-I (summer) has highest RSV occurrence than Selected data point-II (monsoon).

**Figure 1.** The overall time series of dengue fever searches in Google between 2004 and 2017 in India, Brazil, Paraguay, Philippines, Singapore and Worldwide. The vertical axis represents RSV, which is the relative search volume ranked between 0 and 100 by Google internal logic. The short dashed lines indicate confidence interval, the line between dashed lines is regression line and long dashed lines are prediction interval lines. Data plots were generated using SigmaStat 4.0. The regression data was showed in the table 2.
Table 3. Summary of linear equation analyses of dengue fever RSV by country between 2004 and 2017

<table>
<thead>
<tr>
<th>S. No</th>
<th>Country</th>
<th>r²</th>
<th>Linear equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>India</td>
<td>0.081</td>
<td>$y = 0.0028x - 101.48$</td>
</tr>
<tr>
<td>2</td>
<td>Brazil</td>
<td>0.0029</td>
<td>$y = 0.0004x - 6.5078$</td>
</tr>
<tr>
<td>3</td>
<td>Paraguay</td>
<td>0.00002</td>
<td>$y = 4E-05x + 6.7483$</td>
</tr>
<tr>
<td>4</td>
<td>Philippines</td>
<td>0.053</td>
<td>$y = -0.0023x + 117.31$</td>
</tr>
<tr>
<td>5</td>
<td>Singapore</td>
<td>0.0007</td>
<td>$y = -0.0002x + 22.864$</td>
</tr>
<tr>
<td>6</td>
<td>Worldwide</td>
<td>0.1629</td>
<td>$y = 0.0041x - 138.17$</td>
</tr>
</tbody>
</table>

$y = $RSV and $x = $year

**Discussion**

The linear regression was applied for the selected countries to determine the association of RSV with respect to time. The $r^2$ value of each country was determined and the values were 0.081, 0.0029, 0.00002, 0.053, 0.0007, and 0.162 for India, Brazil, Paraguay, Philippines, Singapore, and Worldwide respectively. The values suggest that 8.1%, 0.29%, 0.0015%, 5.3%, 0.07%, and 16.2% of the variation in dengue trends is associated with the time between 2004 and 2017 in India, Brazil, Paraguay, Philippines, Singapore, and Worldwide. The positive and negative slope indicates the up and down trends for the each year regression line data. India, Brazil, Paraguay were showed positive slope and the RSV will increase with each year, whereas Philippines and Singapore has negative slope, the proposed data showed that the RSV trends will decrease for the future years. Worldwide data with positive slope and high $r^2$ value from table 3 suggest that the RSV outbreaks will continue in each year. Confidence interval was mentioned in the figure 1 indicates that the value of a parameter exists within the limit. The prediction limits in the figure 1 was showed to indicate future observations may fall within a certain probability.

The time series Google Trends for India between 2004 and 2017 was correlated with the actual outbreaks recorded during the year 2005-2006 with genotype III of DENV-3, has showed similar patterns. Furthermore, 2007-2009 had predominant genotype DENV-1 by replacing DENV-2 and DENV-3. The strong evidence from Google Trends data in the table 2 which was correlated with actual data for Maharashtra was surprising: Pune is the second largest city in the Indian state of Maharashtra after Mumbai and the seventh most populous city in India, in which dengue is endemic. 18 The seasonality of dengue for Pune, a city of Maharashtra was almost matching with literature data. 19 Recently published report revealed that dengue is a climate sensitive infectious disease, which is very common in Delhi. 20 Similar to the mentioned above, Haryana has also high incidence of RSV outbreaks which are comparable to seasonality. 21 The occurrence of dengue was very common between July and December during the period 2011-2015 in Jammu, a province of J&K. 22 Over 18,700 cases of dengue were reported in 2017, of which Kerala with 9104, accounting 48.68% and followed by Tamilnadu with 4174 by the end of July (http://www.deepikaglobal.com/Kerala.aspx?Dengue-cases-in-India:-Kerala-
India was registered 75808 cases and 193 deaths in 2013, which was the highest number of dengue cases and deaths between 2009 and 2014. The total number of dengue cases and deaths registered in India between 2009 and 2014 were 229288 and 947 respectively. Furthermore, Punjab alone was accounted approximately 6% of dengue cases and 9.6% of deaths during this period. The arrival of rainy season in the month of July at West Bengal is the major causes of dengue outbreaks in this region.

The recent epidemiological review was indicated that Brazil has the highest incidence rates of dengue cases among the Latin American regions. In addition to the above, 60% of world dengue cases occur in Latin Americas and frequent outbreaks are very common from 1980 with numerous hospitalizations and deaths. The brief period was selected to demonstrate the impact of seasonality was described: during the timeline of January and August of 2014 and 2015, 0.59 and 1.4 million cases and 1100 deaths. The Google Trends data for the Brazil was showed an exact concordance of RSV outbreaks for the term Jan-Mar with the recent literature review indicating highest outbreaks between Jan-Aug in which seasonality has higher impact on surge in the disease searches.

Paraguay, a region adjacent to Brazil declared a national epidemiologic alert following the confirmation of 390 cases of dengue in the month of January 2007. By March 2007, 19000 dengue cases and 10 deaths were reported and the similar pattern was observed using Google Trends for Paraguay as showed in figure 1. The number of dengue cases reported during 2004-2017 in Philippines was revealed that, a spike in the number of search queries in Google were dramatically increased till 2010 (actual dengue cases reported were 171138, the highest till 2010) and further decrease and increase was observed for the next three consecutive years. An outbreak of disease was further observed in Google Trends during the year 2015 which is concordance with the reported number of cases. Our data is in agreement with the actual data and Google Trends data between 2004 and 2017. Singapore had the worst health crisis due to dengue in the year 2005 and registered 14209 cases and 25 deaths (http://www.todayonline.com/singapore/dengue-cases-down-4th-week-row). The next outbreak was recorded by claiming 8826 cases with 24 deaths were reported during the year 2007. 32 Google Trends data for Singapore in the figure 1 reveals that the RSV is about 100 in 2005 and approximately 50 by the year 2007. This is a clear indication of surge in the searches in 2005 and become half by 2007 when correlated to the number of cases, which turned slightly above the half from during the year 2007. The subsequent third outbreak in Singapore was occurred in 2013 with 22318 cases and 8 deaths (https://en.wikipedia.org/wiki/2013_dengue_outbreak_in_Singapore). The number of cases reported in 2013 was about two and half times the cases of 2005. As per WHO, dengue has prevalence of 390 million infections per year and of which 96 million are capable of manifesting the symptoms of dengue. Dengue is endemic in more than 100 countries of Africa, the Eastern Mediterranean, South-East Asia, the Americas and the Western Pacific.

The Google Trends analyses of dengue searches reveal the information about the occurrence of infections as per seasonal change between 2004 and 2017. We
hypothesize that the data insights for dengue helps the healthcare providers, policy makers, healthcare industries for proactive identification and control of disease. The pharmaceutical industries with the analytics approach give a way for planning the clinical research, identifying the potential subjects for clinical trials, conducting the effective clinical research for the welfare of humanity, targeting the products to specific locations for marketing purposes and promoting awareness campaigns for prevention of dengue diseases. Furthermore, we presume that various strategies can be implemented with the analytics approaches such as identifying the dengue affected areas and motivating the public by 1) Improving the awareness of dengue to the people by dissemination of information about its symptoms such as fever, vomiting, headache, muscle and joint pains, with skin rashes 2) the cause and consequences of disease such as hemorrhagic fever, low platelets count 3) bring the awareness of Nonsteroidal anti-inflammatory drugs (NSAIDs) like aspirin or Platelet inhibitors 4) Maintenance of best sanitation practices and dispersing the stagnant water to prevent mosquito proliferation 5) Recommending use of mosquito nets and repellents and finally 6) Planning for emergency preparedness and anticipated response from governing and non-governing bodies.

Conclusions

The seasonal variation on dengue occurrence was determined using Google Trends, which is an efficient tool for disease surveillance and also facilitates various public and private organizations to implement strategic approaches to counter the disease outbreaks and helpful for the effective management of public health

References


