

Prediction of COVID-19 Outbreaks Using Google Trends in India: A Retrospective Analysis

U Venkatesh¹, Periyasamy Aravind Gandhi²

¹Department of Community Medicine, Vardhman Mahavir Medical College (VMMC) and Safdarjung Hospital, New Delhi, India

²Department of Community Medicine and School of Public Health, Postgraduate Institute of Medical Education and Research, Chandigarh, India

Objectives: Considering the rising menace of coronavirus disease 2019 (COVID-19), it is essential to explore the methods and resources that might predict the case numbers expected and identify the locations of outbreaks. Hence, we have done the following study to explore the potential use of Google Trends (GT) in predicting the COVID-19 outbreak in India. **Methods:** The Google search terms used for the analysis were “coronavirus”, “COVID”, “COVID 19”, “corona”, and “virus”. GTs for these terms in Google Web, News, and YouTube, and the data on COVID-19 case numbers were obtained. Spearman correlation and lag correlation were used to determine the correlation between COVID-19 cases and the Google search terms. **Results:** “Coronavirus” and “corona” were the terms most commonly used by Internet surfers in India. Correlation for the GTs of the search terms “coronavirus” and “corona” was high ($r > 0.7$) with the daily cumulative and new COVID-19 cases for a lag period ranging from 9 to 21 days. The maximum lag period for predicting COVID-19 cases was found to be with the News search for the term “coronavirus”, with 21 days, i.e., the search volume for “coronavirus” peaked 21 days before the peak number of cases reported by the disease surveillance system. **Conclusions:** Our study revealed that GTs may predict outbreaks of COVID-19, 2 to 3 weeks earlier than the routine disease surveillance, in India. Google search data may be considered as a supplementary tool in COVID-19 monitoring and planning in India.

Keywords: Public Health Surveillance, Information Technology, COVID-19, Trends, Severe Acute Respiratory Syndrome

Submitted: June 2, 2020

Revised: July 9, 2020

Accepted: July 9, 2020

Corresponding Author

Periyasamy Aravind Gandhi

Department of Community Medicine and School of Public Health, Postgraduate Institute of Medical Education and Research, Chandigarh 160012, India. Tel: +91-9585395395, E-mail: aravindsocialdoc@gmail.com (<https://orcid.org/0000-0003-3898-5450>)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2020 The Korean Society of Medical Informatics

1. Introduction

Coronavirus disease 2019 (COVID-19) is rapidly spreading across the globe and has become a significant public health threat to humankind infecting millions worldwide [1]. India is a low middle-income country in the South-East Asia region with a population of 1.3 billion. India reported its first case of COVID-19 on January 30, 2020 [2]. The case numbers were almost static for over a month and gradually started to increase during early March. As of July 7, 2020, India recorded 719,665 cases and 20,160 people succumbed to COVID-19 [1]. Considering the rising menace of COVID-19, it is essential to explore the methods and resources that might predict the case numbers expected and help in identifying the locations of outbreaks. This will help us understand what to expect and prepare for in terms of caseload

and intensive care requirements.

India has an established disease surveillance system, the Integrated Disease Surveillance Program (IDSP), to identify the signals, suspects, and cases of certain notified diseases [3]. IDSP enables the government to make evidence-based decisions on outbreaks. However, the system captures data only when people access the healthcare service. All around the world, non-conventional, informal data sources such as school absenteeism, over the counter drug disbursement, Internet search engines, and social media are being explored and used to supplement formal disease surveillance systems in predicting outbreaks [4]. Search engines and social media tools include Google Trends (GT; Web search, YouTube search, News search, Image search), Twitter, Wikipedia, Baidu, Weibo, and so forth. GTs have been used over the last decade to provide reliable predictions of outbreaks of influenza and other diseases [4-7].

Internet usage among Indians has been on the rise, reaching about 451 million (36%) active users every month, with two-thirds of them being daily users [8]. Search engines are one of the most commonly used facilities in the Internet for identifying and learning information on a wide range of subjects. Among search engines, Google has a monopoly in India, with 98.8% of the total search engine market share [9]. YouTube is an archive/database of videos uploaded across the world on multiple subjects and topics. India has a major share of people accessing and watching YouTube, with around 265 million active users monthly [10]. Both Google and YouTube are free to use, and the data on search terms and patterns are available in open source.

It has been shown that relative search volumes (RSV) of terms specific to a disease from GTs can predict outbreaks of that particular disease in India [6]. However, it is necessary to determine and confirm the correlation, if any, between GTs and other diseases in the country [6]. COVID-19 is one such disease which has the additional feature of being a novel infection in the current scenario. Hence, we have conducted a study to analyze the potential use of GTs to monitor public concern regarding COVID-19 epidemic infection in India and to evaluate the GTs data in predicting the COVID-19 outbreak in India.

II. Methods

Our study was based on a most common search engine database used in India, Google Trends, using different keywords which the public might have used to access information on COVID-19, from January 30, 2020 to April 15, 2020. All

data used in our study were available in open source, and no explicit permission was required to utilize the data.

1. Data on Google Search Terms

The Google Trends homepage (www.google.com/trends) features clustered topics that Google detects to be related and trending together on either Web search, YouTube, or Google News. Trending keywords are collected based on Google's Knowledge Graph technology, and data is normalized and presented on a scale from 0 to 100, where the highest point, 100, divides each point on the graph [6,11]. On the results page, the user can add topics to compare them simultaneously in the charts by clicking the + Compare button or remove an item by clicking the "x" that appears in its box when the user hovers his or her cursor over it. Using this comparison method, we assessed 15 possible keywords that the Indian population might have used. Among them, the five most commonly used keywords were considered. The Google search terms used for the analysis were "coronavirus", "COVID", "COVID 19", "corona", and "virus".

Web search is a generic search, irrespective of whether the content is images, videos, or text news. News search is specific for articles published in the media. The study period RSVs for each of the search terms were retrieved from the GTs for India [12]. The RSV number represents the proportion of popularity of a term relative to the peak popularity during the reference period for the selected region. Hence, it gives a relative weight in terms of temporal and spatial aspects for search phrases in Google. A value of 100 means the term was at the peak of its popularity, while a value of 25 indicates that the search term was 25% as popular as that of its peak popularity during the specified time in the particular region. The reference period for the RSV data for the search terms was from January 30, 2020 to April 15, 2020. India reported its first case of COVID-19 on January 30, 2020 [2].

2. Data on the Number of COVID-19 Cases

The number of daily new confirmed cases and the cumulative confirmed cases in India were obtained for the period until April 15 from <https://datahub.io/>. The data were sourced from this upstream repository maintained by the team at Johns Hopkins University Center for Systems Science and Engineering (CSSE). The upstream dataset obtains data from the World Health Organization (WHO), for India. A confirmed case is defined as one in which the patient tests positive for COVID-19 in the reverse transcriptase-polymerase chain reaction (RT-PCR) test.

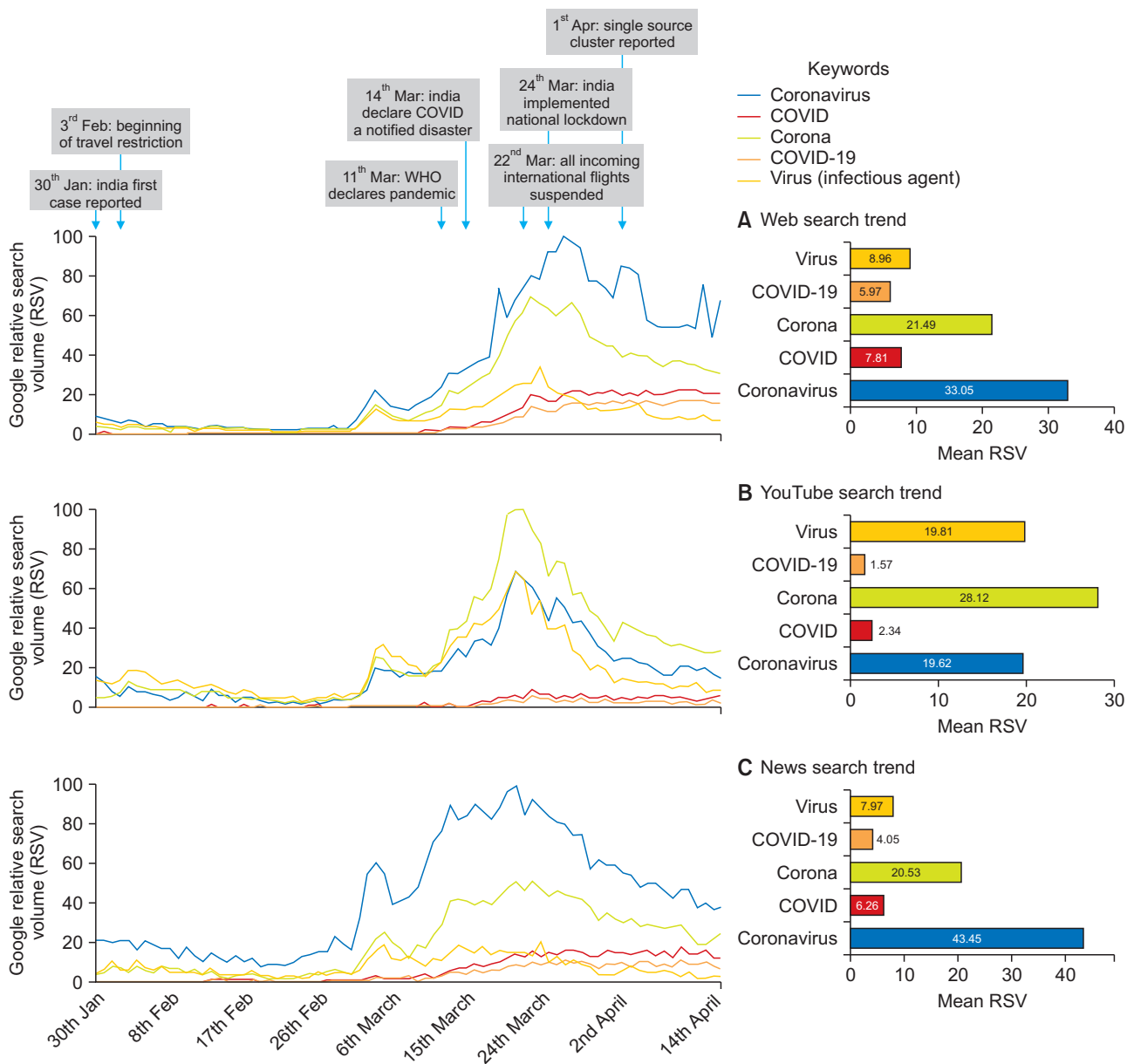


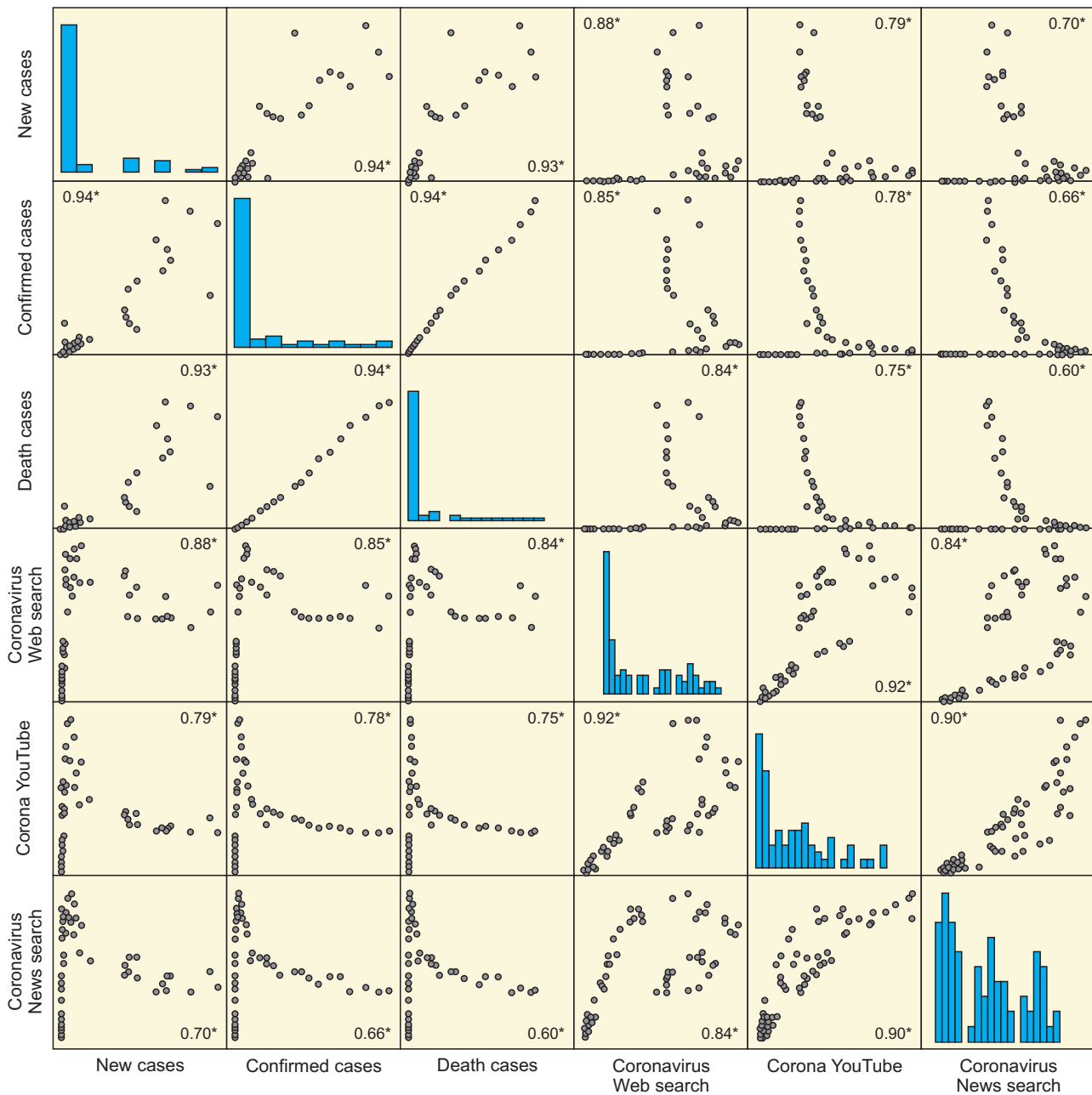
Figure 1. Time series plots of Google Trend relative search volume (RSV) in Web search, YouTube search, and News search.

3. Statistical Analysis

Data were downloaded in Excel format. The analysis was done using SPSS trial version 26.0 (IBM, Armonk, NY, USA). Spearman correlation was used to determine the correlation between the daily new confirmed cases, daily cumulative cases, and the Google search terms. To establish the temporal relationships for up to 30 days, we also did a lag correlation analysis. An r -value of >0.7 is considered as a high correlation, and a p -value of <0.05 is considered as a statistically significant result.

III. Results

Figure 1 shows the overall trends of data from the keyword search for “coronavirus”, “COVID”, “corona”, “COVID 19”, and “virus” (infective agent category) during the selected period and the overall mean RSV of these keywords. It was observed that, among the search terms used, “coronavirus” and “corona” were the terms most commonly used by surfers using Google. Figure 1 also shows that the dynamics of GT data in India were related to public concern at the time of various important announcements and actions taken by the government of India. The spike in search volumes started after the WHO declared COVID-19 as a pandemic



*Correlation is significant with p-value of 0.01 (2-tailed).

Figure 2. Correlation matrix of “coronavirus” and “corona” keywords used in different sub-searches with cumulative confirmed cases and daily new cases.

on March 11, 2020 and when the Indian government made it a notifiable disease on March 14, 2020. It reached its peak immediately after India instituted a nationwide lockdown on March 24, 2020. Figure 2 presents the correlations matrix between the two most common keywords used in various sub-searches with cumulative confirmed cases, daily new cases, and cumulative deaths. The calculated Spearman correlation coefficient was found to be highly significant with all variables at the *p*-value level of 0.01.

1. Lag Correlation Analysis

Table 1 and Figure 3A–3C show the lag Spearman correlation between the RSV from GTs for various sub-searches (Web search, YouTube search, and News search), and the cumulative laboratory-confirmed COVID-19 cases. Correlation between the News search terms “coronavirus” and “corona” was high ($r > 0.7$) with the daily cumulative case for lag periods of 21 days and 20 days, respectively. The strength of correlation increases as the lag period decreases, reaching the maximum ($r = 0.83$) during lag periods of 11 days

Table 1. Lag correlation coefficients and SE between Google Trends data and cumulative laboratory cases in India, January 30, 2020 to April 15, 2020

Days earlier	Web search				YouTube search				News search			
	Corona-virus	SE	Corona	SE	Corona-virus	SE	Corona	SE	Corona-virus	SE	Corona	SE
30	0.201	0.146	0.171	0.146	0.339	0.146	0.293	0.146	0.454	0.146	0.375	0.146
29	0.232	0.144	0.209	0.144	0.376	0.144	0.332	0.144	0.484	0.144	0.411	0.144
28	0.257	0.143	0.237	0.143	0.402	0.143	0.367	0.143	0.517	0.143	0.448	0.143
27	0.294	0.141	0.268	0.141	0.443	0.141	0.412	0.141	0.552	0.141	0.479	0.141
26	0.328	0.140	0.307	0.140	0.487	0.140	0.46	0.140	0.585	0.140	0.515	0.140
25	0.363	0.139	0.350	0.139	0.535	0.139	0.507	0.139	0.617	0.139	0.551	0.139
24	0.401	0.137	0.394	0.137	0.581	0.137	0.553	0.137	0.644	0.137	0.585	0.137
23	0.439	0.136	0.440	0.136	0.624	0.136	0.595	0.136	0.673	0.136	0.620	0.136
22	0.475	0.135	0.483	0.135	0.662	0.135	0.634	0.135	0.698	0.135	0.651	0.135
21	0.513	0.134	0.524	0.134	0.692	0.134	0.666	0.134	0.722	0.134	0.678	0.134
20	0.551	0.132	0.562	0.132	0.728	0.132	0.699	0.132	0.743	0.132	0.706	0.132
19	0.588	0.131	0.600	0.131	0.761	0.131	0.731	0.131	0.763	0.131	0.731	0.131
18	0.624	0.130	0.638	0.130	0.787	0.130	0.754	0.130	0.779	0.130	0.754	0.130
17	0.658	0.129	0.673	0.129	0.812	0.129	0.777	0.129	0.795	0.129	0.775	0.129
16	0.685	0.128	0.700	0.128	0.830	0.128	0.793	0.128	0.802	0.128	0.792	0.128
15	0.711	0.127	0.725	0.127	0.842	0.127	0.807	0.127	0.810	0.127	0.801	0.127
14	0.735	0.126	0.747	0.126	0.849	0.126	0.816	0.126	0.816	0.126	0.813	0.126
13	0.756	0.125	0.767	0.125	0.852	0.125	0.819	0.125	0.821	0.125	0.822	0.125
12	0.780	0.124	0.784	0.124	0.854	0.124	0.828	0.124	0.824	0.124	0.828	0.124
11	0.804	0.123	0.800	0.123	0.856	0.123	0.835	0.123	0.825	0.123	0.835	0.123
10	0.824	0.122	0.815	0.122	0.854	0.122	0.840	0.122	0.823	0.122	0.838	0.122
9	0.836	0.121	0.828	0.121	0.850	0.121	0.842	0.121	0.820	0.121	0.840	0.121
8	0.845	0.120	0.837	0.120	0.841	0.120	0.842	0.120	0.816	0.120	0.839	0.120
7	0.852	0.120	0.844	0.120	0.829	0.120	0.840	0.120	0.809	0.120	0.837	0.120
6	0.859	0.119	0.852	0.119	0.816	0.119	0.835	0.119	0.800	0.119	0.834	0.119
5	0.864	0.118	0.858	0.118	0.807	0.118	0.829	0.118	0.792	0.118	0.831	0.118
4	0.869	0.117	0.868	0.117	0.788	0.117	0.827	0.117	0.778	0.117	0.830	0.117
3	0.872	0.116	0.872	0.116	0.774	0.116	0.819	0.116	0.763	0.116	0.816	0.116
2	0.880	0.115	0.865	0.115	0.773	0.115	0.801	0.115	0.757	0.115	0.793	0.115
1	0.883	0.115	0.872	0.115	0.758	0.115	0.790	0.115	0.738	0.115	0.782	0.115
0	0.894	0.114	0.891	0.114	0.720	0.114	0.785	0.114	0.702	0.114	0.787	0.114

Value in bold text shows high correlation with $r > 0.7$ at a statistical significance p -value of 0.01.

SE: standard error.

and 9 days for “coronavirus” and “corona”, respectively. The correlation fluctuates and falls, thereafter. GTs for the search terms “coronavirus” and “corona” in Web searches were found to be highly correlated ($r > 0.7$) with the daily cumulative cases, for a lag period of 15 days from the peak of the cumulative case numbers. Similar to the News search, the

strength of correlation increases as the lag period decreases, and it reaches the maximum ($r = 0.89$), on the zero-day i.e., the day on which the cases peak. With regards to YouTube search, a high correlation exists between the terms “coronavirus”, “corona”, and cumulative cases with lag periods of 20 days and 19 days, respectively. The strength of correlation

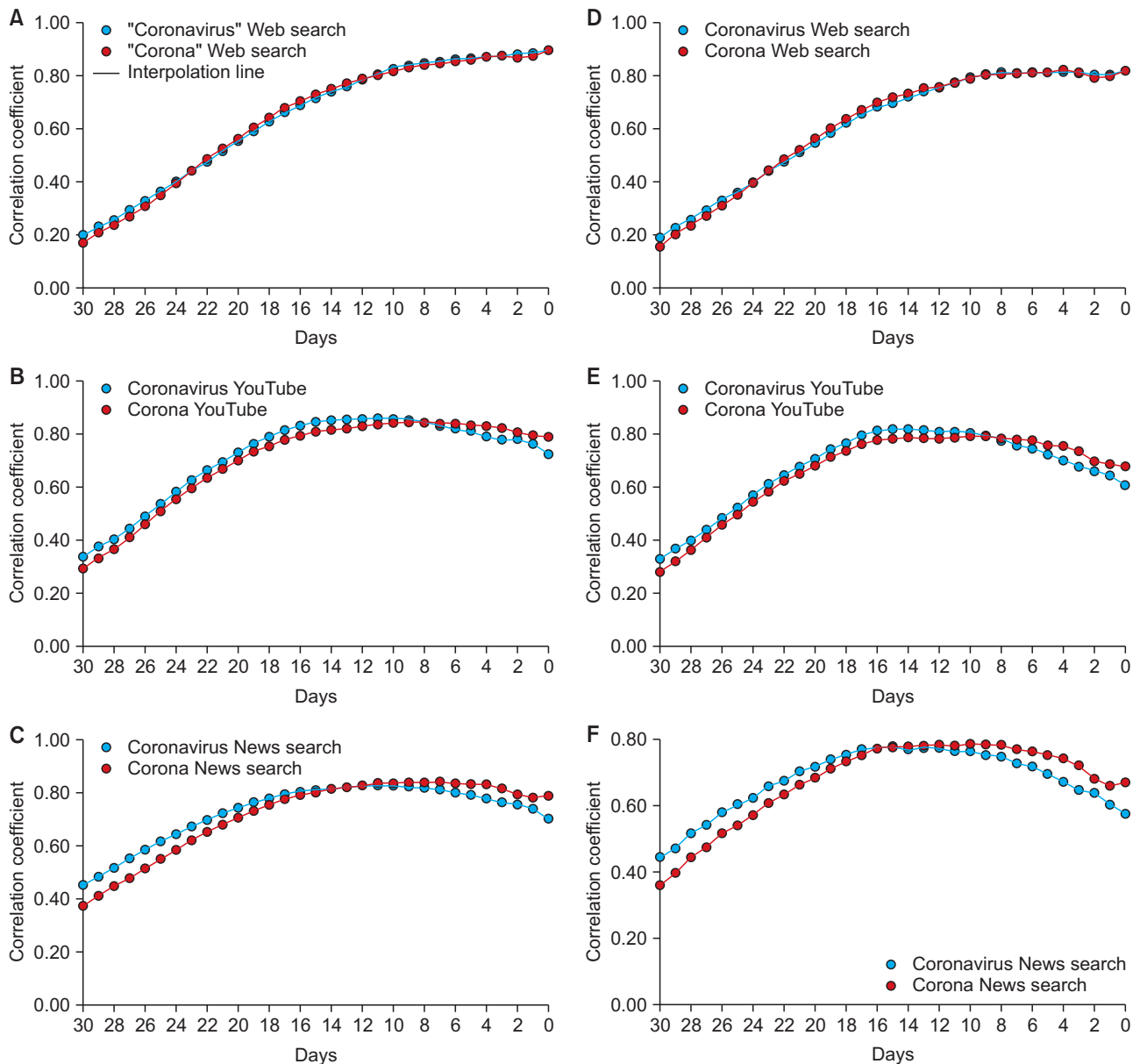


Figure 3. Lag correlation of cumulative confirmed cases (A, B, C) and daily new cases (D, E, F) with data from Google Trends.

reaches the maximum for the terms “coronavirus” ($r = 0.86$) and “corona” ($r = 0.84$), 11 days and 9 days, respectively, before the day the cases peak.

Table 2 and Figure 3D–3F show the lag Spearman correlation between the RSV from GTs for various sub-searches and the daily new laboratory-confirmed COVID-19 cases. Correlation between the Web search terms “coronavirus” and “corona” was high ($r > 0.7$) with the daily new cases, from lag periods of 14 days and 15 days, respectively. The strength of correlation increases as the lag period decreases, and it reaches the maximum during lag periods of 4 days for “corona” ($r = 0.82$), 4 days and zero-days for “coronavirus” ($r = 0.81$). News search GTs for the terms “coronavirus”

and “corona” were found to be highly correlated ($r > 0.7$) with the daily new cases 21 days and 19 days before the cases peak, respectively. The strength of correlation reaches the maximum during lag periods of 13 days and 9 days for “coronavirus” ($r = 0.77$) and “corona” ($r = 0.78$), respectively. In YouTube search, a high correlation exists between the terms “coronavirus”, “corona”, and new case numbers with lag periods of 20 days and 19 days, respectively. The strength of correlation reaches the maximum 15 days for the term “coronavirus” ($r = 0.82$) and 10 days for “corona” ($r = 0.79$) before the day the cases peak.

Table 2. Lag correlation coefficients and SE between Google Trends data and daily new cases in India, January 30, 2020 to April 15, 2020

Days earlier	Web search				YouTube search				News search			
	Corona-virus	SE	Corona	SE	Corona-virus	SE	Corona	SE	Corona-virus	SE	Corona	SE
30	0.188	0.146	0.155	0.146	0.328	0.146	0.28	0.146	0.443	0.146	0.360	0.146
29	0.225	0.144	0.202	0.144	0.367	0.144	0.322	0.144	0.470	0.144	0.396	0.144
28	0.256	0.143	0.235	0.143	0.397	0.143	0.363	0.143	0.516	0.143	0.444	0.143
27	0.293	0.141	0.271	0.141	0.438	0.141	0.409	0.141	0.540	0.141	0.473	0.141
26	0.327	0.140	0.309	0.140	0.481	0.140	0.457	0.140	0.578	0.140	0.516	0.140
25	0.355	0.139	0.341	0.139	0.520	0.139	0.495	0.139	0.602	0.139	0.537	0.139
24	0.391	0.137	0.386	0.137	0.566	0.137	0.543	0.137	0.621	0.137	0.570	0.137
23	0.431	0.136	0.434	0.136	0.609	0.136	0.583	0.136	0.655	0.136	0.605	0.136
22	0.466	0.135	0.478	0.135	0.645	0.135	0.622	0.135	0.672	0.135	0.631	0.135
21	0.503	0.134	0.514	0.134	0.674	0.134	0.649	0.134	0.700	0.134	0.661	0.134
20	0.537	0.132	0.552	0.132	0.703	0.132	0.68	0.132	0.715	0.132	0.681	0.132
19	0.575	0.131	0.589	0.131	0.740	0.131	0.713	0.131	0.736	0.131	0.708	0.131
18	0.612	0.130	0.626	0.130	0.762	0.130	0.734	0.130	0.748	0.130	0.731	0.130
17	0.644	0.129	0.661	0.129	0.791	0.129	0.76	0.129	0.766	0.129	0.750	0.129
16	0.672	0.128	0.689	0.128	0.811	0.128	0.775	0.128	0.769	0.128	0.767	0.128
15	0.686	0.127	0.707	0.127	0.816	0.127	0.782	0.127	0.776	0.127	0.771	0.127
14	0.712	0.126	0.727	0.126	0.815	0.126	0.785	0.126	0.767	0.126	0.775	0.126
13	0.731	0.125	0.742	0.125	0.812	0.125	0.782	0.125	0.772	0.125	0.779	0.125
12	0.747	0.124	0.75	0.124	0.805	0.124	0.781	0.124	0.770	0.124	0.779	0.124
11	0.768	0.123	0.765	0.123	0.805	0.123	0.784	0.123	0.760	0.123	0.777	0.123
10	0.787	0.122	0.781	0.122	0.800	0.122	0.789	0.122	0.761	0.122	0.783	0.122
9	0.796	0.121	0.793	0.121	0.792	0.121	0.788	0.121	0.750	0.121	0.781	0.121
8	0.804	0.120	0.798	0.120	0.772	0.120	0.782	0.120	0.745	0.120	0.779	0.120
7	0.803	0.120	0.800	0.120	0.754	0.120	0.775	0.120	0.725	0.120	0.767	0.120
6	0.808	0.119	0.801	0.119	0.743	0.119	0.773	0.119	0.715	0.119	0.761	0.119
5	0.804	0.118	0.803	0.118	0.722	0.118	0.756	0.118	0.693	0.118	0.749	0.118
4	0.808	0.117	0.816	0.117	0.698	0.117	0.752	0.117	0.669	0.117	0.740	0.117
3	0.802	0.116	0.807	0.116	0.676	0.116	0.733	0.116	0.646	0.116	0.718	0.116
2	0.798	0.115	0.783	0.115	0.659	0.115	0.695	0.115	0.635	0.115	0.677	0.115
1	0.798	0.115	0.789	0.115	0.640	0.115	0.684	0.115	0.599	0.115	0.657	0.115
0	0.809	0.114	0.812	0.114	0.606	0.114	0.676	0.114	0.574	0.114	0.667	0.114

Value in bold text shows high correlation with $r > 0.7$ at a statistical significance p -value of 0.01.

SE: standard error.

IV. Discussion

Search queries have been widely used to predict disease outbreaks all over the world [13,14]. The fundamental principle behind this theory is that symptomatic and soon to be symptomatic people, among others, will search for details

about the disease on the internet before reaching a health facility or accessing healthcare [15]. This will cause a spike in search queries for the particular disease before the patients are captured by the routine disease surveillance system of the health authorities. Our analysis revealed that the terms “coronavirus” and “corona” were the most popular terms

used for Google search in India. Li et al. [14] in their study from China included the term “pneumonia” as well because, during the early stages of the pandemic, COVID-19 was identified as “pneumonia of unknown etiology”. However, by the time the first case emerged in India, it was established to be caused by a coronavirus [16].

We found that the GTs from the Google Web, Google News, and YouTube strongly correlate with the cumulative and new COVID-19 case numbers. The maximum lag period for predicting COVID-19 cases was found to be 21 days with the News search for the term “coronavirus”, that is, the search volume for “coronavirus” peaked 21 days before the peak number of cases. Li et al. [14] reported that search engines were able to predict the COVID-19 outbreak 1 to 2 weeks earlier than that of India.

The greater lag time for India may be attributed to the fact that Indians were sensitized to the corona disease by news from China and other countries, which could have influenced their search behavior. The Internet search pattern and behavior of the population depend on the influence of various factors, such as peer groups, mass media bulletins, government actions, social media interactions, and so forth. They are among the determinants of health-seeking behavior [17]. The series of disease control measures by India, such as suspending international travel and countrywide lockdown to establish physical distancing, may also have played a role in a gradual increase rather than rapid spiking of the COVID-19 case number [18]. However, Li et al. [14] compared the search terms with the new suspected and new confirmed cases, whereas we considered cumulative confirmed and new confirmed cases. The maximum strength of correlation for new confirmed case numbers was found with the term “coronavirus” in Google Web search ($r = 0.82$) and YouTube search ($r = 0.82$), while the strength of correlation was higher ($r = 0.96$) in China [14].

In recent years, GTs have been widely explored as an option to predict various diseases. Shin et al. [5] found in their study in Korea that GTs were useful in predicting Middle East respiratory syndrome coronavirus (MERS-CoV) outbreaks 4 days in advance of the routine disease surveillance system, which is a shorter lag period than our findings. The greater lag period in our study could have been due to the curiosity associated with the novel infection, COVID-19. Santangelo et al. [19] reported that GTs could predict a measles outbreak as early as 4 weeks before the conventional surveillance data in Italy.

In contrast, Provenzano et al. [20] reported no advance prediction capability for Wikipedia trends with maximum

correlation happening on day zero. Carneiro and Mylonakis [15] reported the ability of GTs to predict influenza outbreaks 7 to 10 days earlier than conventional systems. Wilson et al. [21] concluded that GTs could only be explored as supplementary to conventional systems because the Google Flu Trends system did not offer any early prediction and its predictions were in line with the formal surveillance systems for influenza-like illness (ILI) cases in New Zealand. GTs are recommended for countries that do not have well-established and robust disease surveillance systems. Not only the prediction of cases but also the effectiveness of disease control measures have been assessed using GTs. Google searches of COVID-19 control-related terms like “handwashing” have been found to be negatively correlated with the increase in the number of COVID-19 cases, thus acting as an indicator of the effectiveness of COVID-19 prevention strategies [22].

However, our study based on GTs should be cautiously interpreted because it had the following limitations. We included only search terms used in the English language. India is a multi-linguistic country, but the search terms in the other major Indian languages were not accounted for in our study. The fundamental measure of association studied here is correlation, and even a strong correlation per se cannot be used as sufficient evidence for making GTs a primary tool of surveillance [23].

The details of the algorithm of the methodology by which this search data is generated by Google is also unclear. GTs require a large proportion of regular internet users in the country for it to be an effective predictor [15]. However, the exact quantification of this proportion is not available from the literature. Hence, the data obtained by GTs is from one segment of the population only. GTs are more influenced by the media popularity of a particular disease [24], as people will be inclined to look into a disease or condition that is actively displayed and discussed in the popular media.

This phenomenon might have occurred in our study, as we saw a spike in searches using keywords related to COVID-19 whenever a landmark decision was taken by the WHO or the Indian government, which might have had greater media dissemination. It might have caused a disproportionate swing among the public in their internet searching patterns, and may have led to overestimation of the ground reality of the disease. On the other hand, if the general public has poor knowledge about a disease, then the epidemiological burden of that particular disease tends to be underestimated by GTs [24]. Ours was a retrospective study. Real-time prediction of lag time of a disease and outbreaks requires mathematical modelling in addition to internet search data such as the

RSVs from GTs, which is used to correlate the search terms with the disease burden, are calculated based on retrospective data. Hence, future research should focus on strategies to improve the reliability of GTs in disease prediction by formulating mathematical models incorporating internet search data. In the meantime, GTs should not be used as a replacement for robust disease surveillance; rather, it should be explored only to supplement it [21].

In conclusion, our study revealed that Google Web, YouTube, and News might be useful to predict outbreaks of COVID-19 2 to 3 weeks earlier than the routine disease surveillance or reporting system in India. This can be further explored and tested for each state in India, using the search terms in the state specific languages. However, Google search data may be considered only as a supplementary tool in COVID-19 monitoring and planning in India until more evidence is generated on its reliability and real-time prediction efficacy. Further, positive search terms, such as “handwashing” and “masks”, which are related to public awareness, can be explored for their usefulness in assessing the effectiveness of COVID-19 transmission prevention measures at large.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

ORCID

U Venkatesh (<https://orcid.org/0000-0001-8059-9364>)

Periyasamy Aravind Gandhi (<https://orcid.org/0000-0003-3898-5450>)

References

1. World Health Organization. Coronavirus disease (COVID-19) situation report 169 [Internet]. Geneva, Switzerland: World Health Organization; 2020 [cited at 2020 Jul 17]. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200707-covid-19-sitrep-169.pdf?sfvrsn=c6c69c88_2.
2. Gandhi P A, Kathirvel S. Epidemiological studies on COVID-19 pandemic in India: too little and too late? *Med J Armed Forces India*. 2020 May 12 [Epub]. <http://doi.org/10.1016/j.mjafi.2020.05.003>.
3. Government of India, Ministry of Health & Family Welfare. Integrated Disease Surveillance Programme (IDSP) [Internet]. Delhi, India: Ministry of Health & Family

Welfare; c2020 [cited at 2020 Jul 8]. Available from: <https://idsp.nic.in/index1.php?lang=1&level=1&sublinkid=5772&lid=3698>.

4. Seo DW, Jo MW, Sohn CH, Shin SY, Lee J, Yu M, et al. Cumulative query method for influenza surveillance using search engine data. *J Med Internet Res* 2014;16(12):e289.
5. 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.
6. Verma M, Kishore K, Kumar M, Sondh AR, Aggarwal G, Kathirvel S. Google Search Trends predicting disease outbreaks: an analysis from India. *Healthc Inform Res* 2018;24(4):300-8.
7. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in china with search query from baidu. *PLoS One* 2013;8(5):e64323.
8. Mandavia M. India has second highest number of Internet users after China: Report [Internet]. Mumbai, India: The Economic Times; 2019 [cited at 2020 Jul 17]. Available from: <https://economictimes.indiatimes.com/tech/internet/india-has-second-highest-number-of-internet-users-after-china-report/articleshow/71311705.cms?from=mdr>.
9. StatCounter. Search engine market share India (June 2019-June 2020) [Internet]. Dublin, Ireland: StatCounter; c2020 [cited at 2020 Jul 17]. Available from: <https://gs.statcounter.com/search-engine-market-share/all/india>.
10. Laghate G. YouTube in India has over 265 mn monthly active users 1200+ channels with 1 mn+ subs [Internet]. Mumbai, India: The Economic Times; 2019 [cited at 2020 Jul 17]. Available from: <https://economictimes.indiatimes.com/industry/media/entertainment/youtube-in-india-has-over-265-mn-monthly-active-users-1200-channels-with-1-mn-subs/articleshow/72456212.cms?from=mdr>.
11. Seo DW, Shin SY. Methods using social media and search queries to predict infectious disease outbreaks. *Healthc Inform Res* 2017;23(4):343-8.
12. Google Trends India [Internet]. Menlo Park (CA): Google Trends; c2020 [cited at 2020 Jul 17]. Available from: <https://trends.google.com/trends/?geo=IN>.
13. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009;457(7232):1012-4.

14. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. *Euro Surveill* 2020;25(10):2000199.
15. Carneiro HA, Mylonakis E. Google Trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis* 2009;49(10):1557-64.
16. World Health Organization. WHO Timeline: COVID-19 [Internet]. Geneva, Switzerland: World Health Organization; c2020 [cited at 2020 Jul 17]. Available from: <https://www.who.int/news-room/detail/08-04-2020-who-timeline---covid-19>.
17. Woo H, Cho Y, Shim E, Lee JK, Lee CG, Kim SH. Estimating influenza outbreaks using both search engine query data and social media data in South Korea. *J Med Internet Res* 2016;18(7):e177.
18. Chatterjee K, Chatterjee K, Kumar A, Shankar S. Healthcare impact of COVID-19 epidemic in India: a stochastic mathematical model. Version 2. *Med J Armed Forces India* 2020;76(2):147-55.
19. Santangelo OE, Provenzano S, Piazza D, Giordano D, Calamusa G, Firenze A. Digital epidemiology: assessment of measles infection through Google Trends mechanism in Italy. *Ann Ig* 2019;31(4):385-91.
20. Provenzano S, Santangelo OE, Giordano D, Alagna E, Piazza D, Genovese D, et al. Predicting disease outbreaks: evaluating measles infection with Wikipedia Trends. *Recenti Prog Med* 2019;110:292-6.
21. Wilson N, Mason K, Tobias M, Peacey M, Huang QS, Baker M. Interpreting Google flu trends data for pandemic H1N1 influenza: the New Zealand experience. *Euro Surveill* 2009;14(44):19386.
22. Lin YH, Liu CH, Chiu YC. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. *Brain Behav Immun* 2020;87:30-2.
23. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of Google Trends in health care research: a systematic review. *PLoS One* 2014;9(10):e109583.
24. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J Epidemiol Glob Health* 2017;7(3):185-9.