

Title Page

Type: Original Investigation

Characterization of the Second Wave of the COVID-19 Pandemic in India: A Google Trends Analysis

Aayush Visaria, MD, MPH¹, Pooja Polamarasetti², Shivani Reddy³, Alizah Ali⁴, Fariha R. Hameed⁵, Joel James⁶, Moizz Akhtar⁷, Sumaiya Islam⁶, Priyanka Raju⁸, Rajat Thawani, MD⁹

¹Department of Medicine, Rutgers New Jersey Medical School, Newark, New Jersey, USA

²Wayne State School of Medicine, Detroit, Michigan, USA

³Western Michigan University School of Medicine, Kalamazoo, Michigan, USA

⁴Morehouse School of Medicine, Atlanta, Georgia, USA

⁵University of Texas McGovern Medical School, Houston, Texas, USA

⁶CUNY School of Medicine, New York City, New York, USA

⁷AT Still University School of Osteopathic Medicine, Kirksville, Missouri, USA

⁸Eastern Virginia Medical School, Norfolk, Virginia, USA

⁹Oregon Health & Science University, Portland, Oregon, USA

Emails

Aayush Visaria: aayush.visaria@rutgers.edu

Priyanka Raju: rajup@evms.edu

Sumaiya Islam: sislam011@citymail.cuny.edu

Joel James: jjames001@citymail.cuny.edu

Pooja Polamarasetti: ga0800@wayne.edu

Fariha R. Hameed: fariha.r.hameed@uth.tmc.edu

Moizz Akhtar: sa201424@atsu.edu

Shivani Reddy: shivani.reddy@med.wmich.edu

Alizah Ali: aiali@msm.edu

Rajat Thawani: rajat13@gmail.com

Corresponding author: Aayush Visaria, MD, MPH

Address: 185 South Orange Ave, Newark, NJ, USA 07103

Phone: (609) 216-1277

Email: aayush.visaria@rutgers.edu

Ethics approval and consent to participate: This study was exempt from Rutgers's Institutional Review Board approval.

Competing interests: The authors declare that they have no competing interests.

Funding: No funding provided for this study.

Abstract

Background: The second wave of the COVID-19 pandemic has led to considerable morbidity and mortality in India, in part due to lack of healthcare access, low health literacy, and poor disease surveillance. In this retrospective, descriptive ecological study, we utilized Google Trends (GT) to characterize the second COVID-19 wave and its association with official case counts based on search terms related to symptoms, testing, disease complications, medications, preventive behaviors, and healthcare utilization.

Methods: GT is a publicly available, online tracking system of Google searches. Searches are presented as relative search volumes (RSV) from 0 (least) to 100 (most number of searches). We performed pre-defined Web searches in India from 2/12/2021 to 5/09/2021. We characterized the peak RSV, RSV doubling rates, and Spearman rank correlation of selected search terms with official case counts. We also used date-adjusted linear regression to estimate the association between highly correlated search terms and official case counts. We then qualitatively classified public search queries into thematic groups to better understand public awareness and needs related to COVID-19.

Results: We observed that searches for symptoms (most searched terms in order: fever, cough, headache, fatigue, chest pain), disease states (infection, pneumonia), COVID-19-related medications (remdesivir, ivermectin, azithromycin, Fabiflu, dexamethasone), testing modalities (PCR, CT Scan, D-dimer, C-reactive protein, oxygen saturation), healthcare utilization (oxygen cylinders, hospital, physician), and preventive behaviors (lockdown, mask, pulse oximetry, hand sanitizer, quarantine) all demonstrated increases, in line with increases in official case counts. Symptoms, PCR testing, outpatient medications, and preventive behaviors peaked around April 24th, approximately two weeks prior to the peak RSV in official case counts. Contrarily, healthcare utilization factors, including searches for hospital, physicians, beds, disease states, and inpatient medications did not peak until the first week of May. There were highly significant correlations between ‘Coronavirus Disease 2019’ ($r=0.959$), ‘fever’ ($r=0.935$), ‘pulse oximetry’ ($r=0.952$), ‘oxygen saturation’ ($r=0.944$), ‘C-reactive protein’ ($r=0.955$), ‘D-Dimer’ ($r=0.945$), & ‘Fabiflu’ ($r=0.943$) and official case counts.

Conclusion: GT search terms related to symptoms, testing, and medications are highly correlated with official case counts in India, suggesting need for further studies examining GT’s potential use as a disease surveillance and public informant tool for public health officials.

Introduction

The COVID-19 pandemic has led to considerable morbidity, mortality, and near healthcare system collapse in India. This has been especially evident during the second wave of the COVID-19 pandemic, which has as of May 19th, 2021, resulted in a cumulative 25.5 million cases and over 283,000 reported deaths [1]. A novel B.1.617.2 Indian variant [2], mass gatherings, political rallies, slow governmental response, and considerable healthcare access issues have put India's COVID-19 situation to the forefront of international media [3-7]. In addition, many experts, including the World Health Organization and Lancet Commission [3,4], have scrutinized data reporting in India due to its lack of comparability with crematorium accounts, ambiguity surrounding suspected case handling, and paradoxically low case fatality rates compared to other countries [5-7].

Real-world, data-driven approaches to understand public response and conduct disease surveillance are of paramount importance. One such tool, Google Trends™ (GT), has been successfully used to describe disease epidemics including Ebola, Influenza, and early COVID-19 responses in India [8]. In addition to nearly universal access to Google and real-time data collection, many individuals search Google before accessing healthcare [9], suggesting that we may be able to better capture disease burden, disease processes, and societal response via the GT tool, especially in countries with inadequate public health infrastructure.

The purpose of our retrospective, descriptive, ecological study of India's second COVID-19 wave is to utilize GT to describe disease burden, symptomatology, and complications, and their associations with public interest in preventive measures, COVID-19 testing, and vaccination. We also determined the correlation of disease surveillance via GT with government-reported case and death rates. Lastly, we qualitatively describe themes of search queries to help clinicians and public officials better inform and address the general public's questions.

Methods

Data source

We utilized GT, a publicly available, online tracking system of Google searches by search terms, topics, geographic region, and date. Rather than presenting absolute numbers of searches, GT displays the relative search volume (RSV) for pre-specified geographic, time, and search terms. The RSV for a time interval is determined as the search volume in that time interval divided by the largest search volume within the pre-specified criteria. It is then indexed from 0 to 100, where 100 is the reference, largest RSV and 0 is used when there is insufficient data. Further information about GT can be found at <https://trends.google.com>.

Search Strategy

As our primary interest was to describe the second wave of India's COVID-19 pandemic, we conducted Google web searches in India, from February 12, 2021 to May 9th, 2021. The analysis was completed from May 11th to May 18th, 2021.

We initially compiled a list of 141 search terms within the following categories: General COVID-19 terms, Symptoms, Disease Process/Complications, Testing, Healthcare Utilization, Medication Use, Festivals/Holidays, Preventive Behaviors, and Control Searches (see Table 1).

Search terms were compiled with the help of physicians, researchers, and patients from India. Based on comparative searches within the GT platform, we narrowed down search terms to the top three to five searches within each category. We also utilized generalized ‘topic’ forms of each search term rather than the exact search term in order to capture variations in related searches such as non-English searches, abbreviations, and cultural and geographic colloquialisms. The final terms used are listed in Table 2.

Data Management

Once individual searches were completed, the corresponding CSV files were downloaded and combined in order to allow for further data analysis and visualization. This study was exempt from Rutgers IRB approval as all data was publicly available, only at the regional level, and de-identified. In order to compare results with official COVID-19 case and fatality data, we utilized raw data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE), available publicly at <https://github.com/CSSEGISandData/COVID-19/>. Our specific dataset is available upon reasonable request.

Statistical Analysis

This retrospective, ecological study was primarily meant to be a descriptive analysis of GT data. We determined relative proportional increases in RSV throughout the 90-day time period for all search terms using 2/12/21 as the reference, determined peaks in searches, and described common COVID-19 related questions related to the searches. We also determined initial and second doubling dates: initial or first doubling date was defined as the date during which the three-day average RSV doubled from the average RSV from 2/12/21 to 2/14/21. The second doubling date was the date where the three-day average RSV doubled from the initial doubling date.

We then determined the Spearman rank correlation between official COVID-19 case data and search terms in order to determine whether searches were comparable to official data. Where they were comparable, we subsequently estimated beta coefficients using time-adjusted linear regression to determine significant search covariates. We also performed multivariate linear regression to estimate future caseload; however, due to significant multicollinearity among search terms, this was not statistically sound and thus results are not presented. All analyses were done using SAS 9.4 with a significance level of 0.05.

Results

Clinically Relevant Search Terms

Out of 141 search queries, the top topics searched are presented in Table 2. In order from most to least common, searches for symptoms included fever, cough, headache, chest pain, and fatigue (Figure 1a). Whereas the peak RSVs for presenting symptoms (e.g. fever, cough) and associated symptoms (e.g. headache, fatigue) were generally in the last week of April, ‘chest pain’ peaked in the last week of the analysis (Table 3; first week of May). This was in line with official COVID-19 case counts, which also peaked from May 5th - May 7th within the time period of our analysis. Searches for ‘fever’ doubled by March 29th and doubled again 15 days later. This was nearly two weeks after official case counts doubled, whereas the peak RSV for ‘fever’ was two weeks earlier than official counts. The date by which ‘cough’ searches doubled

was delayed by approximately 2 weeks compared to ‘fever’ but also demonstrated a second doubling time of 11 days, identical to official case count doubling times (Table 3). Both ‘headache’ and ‘fatigue’ searches did not have a second doubling time due to a high baseline RSV.

Among disease complications, most individuals searched for pneumonia, infection, myocardial infarction, and stroke, in that order (Figure 1b). Searches for ‘pneumonia’ peaked on May 1st and had a first doubling date of April 1st and a second doubling date of April 18th. This is approximately 4 weeks after official case doubling times. However, the peak RSV for ‘pneumonia’ was five days prior to the peak in official cases (Table 3).

Among testing modalities, searches for ‘polymerase chain reaction’ (PCR) were of the highest RSV, followed by ‘computed tomography’ (CT Scan), ‘C-reactive Protein’, ‘D-dimer’, and ‘Oxygen Saturation’ (Figure 1c). The peak RSV was two weeks earlier for PCR searches than other testing modalities. Compared to official case counts, PCR searches peaked two weeks earlier and also had a first doubling date two weeks earlier (Table 3).

Among medication searches, the majority of searches were for ‘remdesivir’, followed by ‘ivermectin’, ‘azithromycin’, ‘fabiflu’, and ‘dexamethasone’ (Figure 1d). Ivermectin and Fabiflu, both medications used in the outpatient setting, peaked around April 23-25th, with an initial doubling date from March 15-20th. Largely inpatient medications, such as ‘remdesivir’ and ‘dexamethasone’ were more variable in their peak RSV and doubling times, with ‘remdesivir’ parameters two weeks earlier than ‘dexamethasone’ parameters (Table 3).

Socially-relevant Search Terms

General, healthcare utilization, preventive behavior, and festival/holiday searches also demonstrated similar peak RSVs and doubling times as clinically-relevant search terms (Table 3). General searches for the ‘Coronavirus Disease 2019’ topic peaked on May 1st, had a first doubling date of February 28th, approximately two weeks prior to official case count doubling, and a second doubling date of April 2nd, approximately two weeks after official case count doubling.

Among healthcare utilization searches, searches related to ‘oxygen’ were most prevalent, peaking on April 24th and having a second doubling time within one week of first doubling date. Searches for ‘physician’, ‘hospital’, and ‘bed’ peaked in the first week of May and did not show any doubling (Figure 2b, Table 3).

Among preventive behavior searches, we observed a general increase in searches for ‘lockdown’, as well as ‘mask’ and ‘pulse oximetry’ (Figure 1c). Nearly all preventive behavior searches peaked from April 18th to April 28th but had variable doubling dates (Table 3).

Among searches for festivals/mass gatherings, we see a tremendous increase in searches for Holi on March 29th and smaller, but significant, increases in other gatherings (Figure 2d). Peak searches of Eid-al-Fitr, Holi, Easter and Indian New Year occurred in the days leading up to the holiday. No searches under these general categories included queries regarding COVID-19. However, when searching for queries associated with Kumbh Mela, a Hindu festival that was held in Haridwar, Uttarakhand during April 2021 known to have been a “super-spreader” event, had increasing queries regarding the festival and “covid” or “covid cases” during this time frame.

When stratifying Indian New Year by regional festivities such as Ugadi, Gudi Padwa, Vaisakhi, and Vishu, there were no searches associated with “covid”.

Correlation Between Searches and Official Case Counts

We determined the Spearman rank correlation between search terms and official case counts (Table 4). The searches yielding the highest correlation included ‘Coronavirus Disease 2019’ ($r=0.959$), ‘fever’ ($r=0.935$), ‘pulse oximetry’ ($r=0.952$), ‘oxygen saturation’ ($r=0.944$), ‘C-reactive protein’ ($r=0.955$), ‘D-Dimer’ ($r=0.945$), and ‘fabiflu’ ($r=0.943$). This can also be seen visually in Figure 3. Using date-adjusted linear regression to estimate the association between search terms and official case counts, we found that every 1 unit increase in the RSV for ‘fever’ resulted in a mean increase of 4,854 cases (95% CI: 4,538 - 5,170; $p<0.0001$). Every 1 unit increase in ‘pulse oximetry’ yielded a mean increase of 3,141 (95% CI: 2,863 - 3,420; $p<0.0001$) cases. Every 1 unit increase in ‘fabiflu’ resulted in a mean increase of 2,318 (95% CI: 1,821 - 2,815; $p<0.0001$) cases. When looking at the correlation between official case counts and control search terms, we found that ‘film’ and ‘election’ had no correlation, but other control terms including ‘cricket’, ‘bank’, and ‘animal’ were significantly negatively correlated (Table 4). When we determined the correlation of search terms with official case counts stratified by time period (first 21 days, middle period, and last 21 days), we observed that correlations were of lesser significance in the first and last 21 days (data not shown).

We also determined the correlation between search terms and official death counts, which yielded similar but lower magnitude correlations (Table 5).

Qualitative Theme Analysis

In order to understand the underlying questions for specific search topics, we grouped related search queries into observed themes, namely administrative (testing, registration, location, costs), symptom expectations (disease progression, symptom management), and health literacy (meaning of terms, etiology, normal levels). Many individuals seemingly are unsure of testing locations, vaccine registration protocol, location of available hospitals, and costs of various tests, medications and healthcare utilization factors. Individuals are also unsure of whether presenting symptoms are related to COVID-19 and have questions about expected length of fever, definition of fever, normal values for common laboratory tests, and over-the-counter treatment options for common symptoms like fever and cough (Table 6).

Discussion

In this retrospective, descriptive, ecological study of GT search terms related to the second wave of the COVID-19 pandemic in India, we observed that searches for symptoms (e.g. fever, cough), disease states (e.g. infection, pneumonia), COVID-19-related medications (e.g. remdesivir, ivermectin), testing modalities (e.g. PCR, CT Scan), healthcare utilization (e.g. oxygen cylinders, hospital), and preventive behaviors (e.g. pulse oximetry, lockdown) all demonstrated an increase over our time period from 2/12/21 - 5/9/2021, in line with increases in official case counts. While peak RSVs and doubling rates varied across search categories, we generally observed that symptoms, PCR testing, outpatient medications, and preventive behaviors peaked around April 24th, approximately two weeks prior to the peak RSV in official case counts, but exhibited a doubling date one-two weeks later than official case counts.

Contrarily, healthcare utilization factors, including searches for hospital, physicians, beds, disease states, and inpatient medications did not peak until the first week of May. There was a highly significant correlation between ‘Coronavirus Disease 2019’, ‘Fever’, ‘Pulse oximetry’, ‘Oxygen saturation’, ‘C-reactive protein’, ‘D-Dimer’, & ‘Fabiflu’ and official case counts.

GT has been used to understand various aspects of the first wave of the COVID-19 pandemic worldwide, including its use as a digital surveillance tool [10-14], its use as a psychosocial behavior informant [15], and its use as a method to characterize novel symptoms such as ageusia and agnosia [16]. GT has many advantages, including real-time, real-world analysis, large sample size, and non-healthcare-related searches that may be missed in studies looking only at patients seeking healthcare.

We found that fever, cough, headache, fatigue, and chest pain were the most prevalent symptom-related searches during our time period, suggesting that these may be the most common presenting symptoms during the current COVID-19 wave. Searches for gastrointestinal symptoms, agnosia/ageusia, body/muscle aches, and other commonly reported symptoms from the first wave were not nearly as prevalent nor did they demonstrate a significant rise during our time period. These results are in line with current literature describing prevalence of symptoms [17]. Additionally, fever and cough demonstrated the strongest correlation with official case counts, whereas headache and fatigue showed significant, but lower, correlation and did not display a second doubling. This may suggest that either headache and fatigue are specific to certain patients, that there was already a baseline high level of RSV possibly due to ‘long-haul’ Covid [18], or that the pattern of symptom searching is biased away from headache and fatigue, which can be easily overlooked. The increase in searches for chest pain is likely the result of pleuritic chest pain from COVID-19 pneumonia rather than chest pain due to cardiovascular disease; however, Ciofani et al. showed in the U.S. population that hospital admissions for myocardial infarction decreased, while searches for chest pain increased in line with COVID-19 cases. This may suggest that patients may be self-triaging themselves or avoiding seeking care for true cardiovascular chest pain [19]. Interestingly, searches for fever and cough peaked nearly two weeks earlier than case counts, suggesting that symptomatic individuals may search symptoms prior to getting tested and being included in official case counts. If this is in fact true, GT may provide a useful tool to target populations with increases in searches for symptoms, by for example, providing testing facilities earlier in disease course.

These results demonstrate that GT may be useful in not only characterizing the symptoms of a novel population-wide disease but also in understanding the relative prevalence of certain symptoms among all individuals including those who otherwise have mild infection but may still search about fever or cough.

As severe COVID-19 infection is known to cause a pro-inflammatory state with elevated inflammatory markers and susceptibility to clot formation, we explored searches related to laboratory testing and disease complications including myocardial infarction, stroke, and pulmonary embolism. While searches for C-reactive protein and D-dimer were amongst the most highly correlated searches with official case counts - likely because PCR tests for disease are done in coordination with other laboratory tests and CT scans in suspected cases - we did not see a significant rise in myocardial infarction, stroke, or pulmonary embolism, suggesting that individuals with severe complications may not search before seeking care. Future studies should include terms for signs/symptoms of complications rather than the disease complication itself.

Our search term topics included commonly-used related terms (e.g. heart attack), but may not have been comprehensive enough to capture all related searches.

In addition to clinically-relevant searches, we also explored preventive behaviors and healthcare utilization. Searches for ‘oxygen saturation’, ‘pulse oximetry’, and ‘oxygen’ (cylinders) were highly prevalent and also significantly associated with official case counts. This is in agreement with the well-publicized issue of oxygen shortages in India and difficulties of finding hospital beds [20]. Despite the increase in searches for ‘hospitals’, we did not see a concomitant increase in searches for ‘ambulance’, making it unclear whether ambulance shortages were lower than hospital shortages, whether families were more likely to transport patients, or whether searches for hospitals were the result of media publicization rather than the primary cause.

In addition to GT’s use as a potential method to characterize disease, of equal importance is its role to better understand public questions concerning COVID-19 [21]. We found that related search queries could be grouped into three main themes, administrative, symptom expectations, and symptom management. This is in line with GT analyses by Hu et al. and Springer et al. that remarked that there is a need for increased disease awareness, with respect to treatment options and disease course [22,23]. This is especially evident and necessary in India, where health literacy remains low and many search queries were concerned with symptom management. There is also a need for India to address administrative questions including cost, location, and availability of testing, vaccine, and hospital facilities. Inability to access healthcare due to administrative issues is a primary source of disease burden and disproportionately affects those of lower socioeconomic status.

The results should be interpreted with caution due to many limitations. Firstly, our ecological study design does not allow us to make inferences at the individual patient level, nor does it provide information on the direction of correlation. Increases in searches may, at least partially, be due to increased presence of related topics in the media rather than individual situations. For example, Remdesivir exports were prohibited by Indian government officials on April 11th in anticipation of the predicted spike in demand for the drug. This may have contributed to the rise in interest for this term earlier than other medications.

Furthermore, searches for symptoms may not be solely due to COVID-19. However, given the high prevalence of COVID-19 in India and the significant correlation with confirmed COVID-19 cases, it is reasonable to infer that rises in related search terms are due to COVID-19. Additionally, search terms may not capture all non-English languages and regional colloquialisms. Although we tried to choose terms that would capture the largest percentage of related terms, this may have decreased specificity of searches. For example, the search topic “mask” rather than “face mask” or “covid mask” may encompass all of the intended searches with a high sensitivity; although the decision to choose the more general search term decreases specificity when including searches intended to find other masks, such as those for skincare or fashion purposes. Despite these limitations, GT’s large sample size, real-world and real-time benefits, and ability to capture populations that may otherwise be excluded from traditional research studies provide a useful tool to understand COVID-19. In particular, our major strength is there has been no comprehensive study on various clinical and social aspects of the second wave of the COVID-19 pandemic in India.

Future Directions

Future studies should attempt to reconcile the predictive power of GT search terms, including highly correlated symptoms, testing, and healthcare utilization factors, in India, within specific regions in India, and other countries. Studies should also attempt to understand the impact of mass gatherings on GT searches to allow for prediction after upcoming events such as Eid. Future studies may also focus on understanding the association between search terms and official death counts in India, including prediction of underreporting of data based on GT models developed in other countries.

Furthermore, examination of other Internet activity besides Google Web Search (such as Youtube, Facebook, and Whatsapp) may provide a more accurate depiction of Internet-based health searches. In particular, Whatsapp is one of the largest modes of learning and spreading health-related information in South Asia [24]. Popular “chain-messages” or posts may be studied for keywords and interest based on time-course to check for correlation with COVID-19 case data and risk prediction capability.

References

1. COVID-19 Dashboard. World Health Organization. <https://covid19.who.int/region/searo/country/in>. Accessed on May 19, 2021.
2. Mallapaty S. India's massive COVID surge puzzles scientists. *Nature*. 2021 Apr 21;592(7856):667-8.
3. Kuppalli K, Gala P, Cherabuddi K, Kalantri SP, Mohanan M, Mukherjee B, Pinto L, Prakash M, Pramesh CS, Rathi S, Pai NP. India's COVID-19 crisis: a call for international action. *The Lancet*. 2021 May 14.
4. *The Lancet*. India's COVID-19 emergency. 2021 May 08; 397(10286):P1683. [E-pub ahead of print]. [https://doi.org/10.1016/S0140-6736\(21\)01052-7](https://doi.org/10.1016/S0140-6736(21)01052-7)
5. Lau H, Khosrawipour T, Kocbach P, et al. Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters. *Pulmonology*. 2021 Mar 1;27(2):110-5.
6. Gautam, V., Dileepan, S., Rustagi, N., et al. Health literacy, preventive COVID 19 behaviour and adherence to chronic disease treatment during lockdown among patients registered at primary health facility in urban Jodhpur, Rajasthan. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 2020. 15(1);205-211.
7. Quadri SA, Padala PR. An Aspect of Kumbh Mela Massive Gathering and COVID-19. *Current Tropical Medicine Reports*. 2021 Mar 29:1-6.
8. Carneiro HA, Mylonakis E. Google Trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis*. 2009;49:1557-1564.

9. Shaughnessy AF, Slawson DC, Duggan AP. “Alexa, Can You Be My Family Medicine Doctor?” The Future of Family Medicine in the Coming Techno-World. *The Journal of the American Board of Family Medicine*. 2021 Mar 1;34(2):430-4.
10. Ortiz-Martínez Y, Garcia-Robled JE, Vásquez-Castañeda DL, Bonilla-Aldana DK, Rodriguez-Morales AJ. Can Google® trends predict COVID-19 incidence and help preparedness? The situation in Colombia. *Travel medicine and infectious disease*. 2020 Sep 1.
11. Husnayain A, Fuad A, Su EC. Applications of Google Search Trends for risk communication in infectious disease management: A case study of the COVID-19 outbreak in Taiwan. *International Journal of Infectious Diseases*. 2020 Jun 1;95:221-3.
12. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, Kalhori SR. Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study. *JMIR public health and surveillance*. 2020;6(2):e18828.
13. Venkatesh U, Gandhi PA. Prediction of COVID-19 outbreaks using google trends in India: A retrospective analysis. *Healthcare informatics research*. 2020 Jul 31;26(3):175-84.
14. Ciaffi J, Meliconi R, Landini MP, Ursini F. Google trends and COVID-19 in Italy: could we brace for impact?. *Internal and Emergency Medicine*. 2020 May 25:1-5.
15. Brodeur A, Clark AE, Fleche S, Powdthavee N. COVID-19, lockdowns and well-being: Evidence from Google Trends. *Journal of public economics*. 2021 Jan 1;193:104346.
16. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss of smell related searches during the COVID-19 outbreak. *International forum of allergy & rhinology* 2020 Jul (Vol. 10, No. 7, pp. 839-847).
17. Grant MC, Geoghegan L, Arbyn M, Mohammed Z, McGuinness L, Clarke EL, Wade RG. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. *PloS one*. 2020 Jun 23;15(6):e0234765.
18. Halpin S, O'Connor R, Sivan M. Long COVID and chronic COVID syndromes. *Journal of medical virology*. 2021 Mar;93(3):1242-3.
19. Ciofani JL, Han D, Allahwala UK, Asress KN, Bhindi R. Internet search volume for chest pain during the COVID-19 pandemic. *American heart journal*. 2021 Jan 1;231:157-9.
20. Wise J. Covid-19: Countries rally to support India through “storm that has shaken the nation”. *BMJ: British Medical Journal (Online)*. 2021 Apr 26;373.
21. Hu D, Lou X, Xu Z, Meng N, Xie Q, Zhang M, Zou Y, Liu J, Sun G, Wang F. More effective strategies are required to strengthen public awareness of COVID-19: Evidence from Google Trends. *Journal of global health*. 2020 Jun;10(1).

22. Hu D, Lou X, Xu Z, et al. More effective strategies are required to strengthen public awareness of COVID-19: Evidence from Google Trends. *J Glob Health*. 2020;10(1):011003. doi:10.7189/jogh.10.011003
23. Springer S, Menzel LM, Zieger M. Google Trends reveals: Focus of interest in the population is on treatment options rather than theories about COVID-19 animal origin. *Brain, behavior, and immunity*. 2020 Jul;87:134.
24. Bapaye JA, Bapaye HA. Demographic Factors Influencing the Impact of Coronavirus-Related Misinformation on WhatsApp: Cross-sectional Questionnaire Study. *JMIR Public Health Surveill*. 2021;7(1):e19858. Published 2021 Jan 30. doi:10.2196/19858

Table 1. Search Terms by Category

General Terms	Symptoms	Disease Course	Mass Gatherings	Testing	Healthcare Utilization	Preventive Behaviors	Medication Use	Control Group
COVID-19	Fever	Pneumonia	Holi	RT-PCR	Bed	Quarantine	Remdesivir	Bollywood
Corona	Cough	ARDS	Easter	CT Score	Oxygen	Isolation	Ivermectin	Cricket
Coronaviruses Disease	Shortness of Breath	Myocardial Infarction	Eid	CT Scan	Hospital	Lockdown	Dexamethasone	Animals
COVID-19 Testing	Trouble breathing	Stroke	Indian New Year	X-ray	Blood	Hand sanitizer	Azithromycin	Modi
COVID-19 vaccine	Fatigue	Pulmonary Embolism	New Year	D-dimer	Physician	Face mask	Doxycycline	Election
Virus	Loss of Taste	Cardiac arrest	Hindu New Year	C-reactive protein	Ambulance	N95 mask	Fabiflu	Film
Pandemic	Loss of Smell	Oxygen desaturation	Kumbh Mela	CBC	ICU	Social distancing	Paracetamol	Bank
Cremations	Chest Pain	Low oxygen	Political Rally	Ferritin	Oxygen concentrator	Hand washing	Aspirin	
Funeral	Muscle Ache	Pain	Farmer Protests	Interleukin -6	Blood plasma	Hygiene	Inhaler	
Double mutant	Body Ache	DVT	Weddings	Swab test	Pharmacy	Public health	Hydroxychloroquine	
Triple mutant	Diarrhea	Myocarditis	Cricket matches	Blood pressure	Nurse	Research	Vitamin D	
	Nausea	Heart Failure		Blood glucose	Telehealth	Event cancellation	Steroids	
	Vomiting	Kidney injury		Oxygen saturation	Bed	Vaccine	Itolizumab	
	Headache	Kidney failure		Electrolyte	Cylinder	Donations	Oxygen therapy	
	Chills	Chronic fatigue		Antibody test	Ayurvedic medicine	Gloves	Ventilator	
	Congestion	Liver failure		Troponin	Helpline	PPE	Tocilizumab	
	Sore throat	New Diabetes		Creatinine	Covid tent	Healthy diet	Plasma therapy	
	Runny nose	Organ failure		ECG	Vaccine registration	Exercise	Favipiravir	
	Weakness	Multisystem Inflammatory Disease		Calcium	Transportation	Smoking cessation		

Table 2. Top Searched Terms by Category

Category	Top Searched Topics
General	Coronavirus Disease 2019, Coronavirus, COVID-19 Testing, COVID-19 Vaccine
Symptoms	Fever, Cough, Headache, Fatigue, Chest Pain
Disease Process	Infection, Pneumonia, Myocardial Infarction, Stroke, Pulmonary Embolism
Testing	Polymerase Chain Reaction, Computed tomography, D-dimer, C-reactive protein, Oxygen Saturation
Healthcare Utilization	Oxygen Cylinder, Physician, Hospital, Pharmacy, Bed
Medication Use	Remdesivir, Ivermectin, Azithromycin, Ivermectin, Dexamethasone
Festivals	Holi, Eid, Indian New Year, Kumbh Mela
Preventive Behaviors	Lockdown, Quarantine, Mask, Pulse Oximetry, Hand Sanitizer
Control Groups	Cricket, Bollywood, Animals, Bank, Election

Table 3. Graphical Characteristics of Top Search Queries

Search Term	Peak RSV*	Peak 3-day RSV	Initial Doubling Date [#]	2nd Doubling Date [#]	Case Lag Time ^s
Official COVID-19 Cases	5/6/2021	5/5 - 5/7	3/10/2021	3/21/2021	
General Terms					
Coronavirus Disease 2019	5/1/2021	4/29 - 5/1	2/28/2021	4/2/2021	5
Coronavirus	4/26/2021	4/26 - 4/28	3/17/2021	4/5/2021	10
COVID-19 Testing	4/27/2021	4/27 - 4/29	3/17/2021	3/30/2021	9
COVID-19 Vaccine	5/1/2021	4/29 - 5/1	2/28/2021	4/22/2021	5
Symptoms					
Fever	4/24/2021	4/24 - 4/26	3/29/2021	4/13/2021	12
Headache	4/26/2021	4/29 - 5/1	4/23/2021	N/A [%]	10
Fatigue	4/21/2021	4/19 - 4/21	4/18/2021	N/A	15
Cough	4/28/2021	4/21 - 4/23	4/10/2021	4/21/2021	8
Chest pain	5/5/2021	5/5 - 5/7	4/8/2021	5/2/2021	1
Disease Process/Complications					
Stroke	5/7/2021	3/8 - 3/10	N/A	N/A	-1
Pneumonia	5/1/2021	5/1 - 5/3	4/1/2021	4/18/2021	5
Myocardial Infarction	4/30/2021	4/29 - 5/1	4/28/2021	N/A	6
Infection	5/5/2021	5/5 - 5/7	4/15/2021	N/A	1
Healthcare Utilization					
Pharmacy	4/24/2021	4/22 - 4/24	N/A	N/A	12
Physician	5/8/2021	5/4 - 5/6	N/A	N/A	-2
Oxygen	4/24/2021	4/24 - 4/26	4/6/2021	4/13/2021	12
Hospital	5/1/2021	5/7 - 5/9	4/17/2021	N/A	5
Bed	5/5/2021	5/5 - 5/7	N/A	N/A	1
Preventive Behaviors					
Quarantine	4/19/2021	4/19 - 4/21	4/3/2021	N/A	17
Pulse Oximetry	4/24/2021	4/23 - 4/25	3/16/2021	3/30/2021	12
Mask	4/26/2021	4/26 - 4/28	4/14/2021	N/A	10
Lockdown	4/19/2021	4/18 - 4/20	2/15/2021	2/18/2021	17
Hand Sanitizer	4/27/2021	5/6 - 5/8	4/1/2021	4/14/2021	9
Testing					
Oxygen Saturation	5/5/2021	5/4 - 5/6	3/20/2021	3/30/2021	1
Computed Tomography	5/3/2021	5/2 - 5/4	3/29/2021	4/14/2021	3
Polymerase Chain Reaction	4/19/2021	4/22 - 4/24	2/23/2021	3/24/2021	17
C-reactive Protein	5/8/2021	5/6 - 5/8	3/18/2021	4/2/2021	-2
D-Dimer	5/6/2021	5/6 - 5/8	3/11/2021	3/29/2021	0
Medication Use					
Dexamethasone	5/1/2021	5/2 - 5/3	4/8/2021	4/18/2021	5

Azithromycin	5/3/2021	5/3 - 5/5	3/25/2021	4/7/2021	3
Ivermectin	4/25/2021	4/23 - 4/25	3/20/2021	4/3/2021	11
Fabiflu	4/23/2021	4/23 - 4/25	3/15/2021	3/22/2021	13
Remdesivir	4/18/2021	4/18 - 4/20	3/20/2021	3/31/2021	18

Format of dates: Month/Day/Year; RSV = Relative Search Volume. All searches conducted from 2/12/21 to 5/9/21

*Peak RSV refers to the date at which a specific search query hits 100 (its highest point in our time period)

#Initial doubling date is defined as the date at which the three-day average RSV exceeds 2 times that of the RSV from 2/12 – 2/14/21. The second doubling date is defined as the date at which the three-day average RSV exceeds 2 times that of the RSV from the initial doubling date.

§Case lag time is defined as the number of days between the peak RSV of the official case count and the peak RSV of the search term of interest.

%N/A means that the search query did not double further or reach the doubling date.

Table 4. Correlation Between Search Terms and Official Case Count

Search Term	Spearman Rank Correlation	P-value	Time-adjusted β coefficient (95% CI)	P-value
General Terms				
Coronavirus Disease 2019	0.959	<0.0001	5924 (5542 - 6305)	<0.0001
Coronavirus	0.921	<0.0001	6339 (5653 - 7024)*	<0.0001
COVID-19 Testing	0.928	<0.0001	5987 (5120 - 6853)*	<0.0001
COVID-19 Vaccine	0.849	<0.0001	7778 (6062 - 9494)*	<0.0001
Symptoms				
Fever	0.935	<0.0001	4854 (4538 - 5170)	<0.0001
Headache	0.809	<0.0001	4275 (3178 - 5372)	<0.0001
Fatigue	0.778	<0.0001	2037 (708 - 3366)	0.0038
Cough	0.905	<0.0001	4033 (3678 - 4388)	<0.0001
Chest pain	0.757	<0.0001	3553 (2853 - 4252)	<0.0001
Disease Process/Complications				
Stroke	-0.089	0.41	-865 (-2011 - 280)	0.13
Pneumonia	0.886	<0.0001	3670 (3149 - 4192)	<0.0001
Myocardial Infarction	0.414	<0.0001	2348 (1406 - 3288)	<0.0001
Infection	0.857	<0.0001	5048 (4342 - 5752)	<0.0001
Healthcare Utilization				
Pharmacy	0.495	<0.0001	3982 (3313 - 4651)	<0.0001
Ambulance	0.746	<0.0001	2714 (2246 - 3183)	<0.0001
Physician	0.494	<0.0001	-951 (-1128 - 3031)	0.37
Oxygen	0.885	<0.0001	2706 (2323 - 3090)*	<0.0001
Hospital	0.884	<0.0001	5333 (4415 - 6250)	<0.0001
Bed	0.612	<0.0001	4187 (2933 - 5440)*	<0.0001
Preventive Behaviors				
Quarantine	0.875	<0.0001	2507 (1385 - 3629)*	<0.0001
Pulse Oximetry	0.952	<0.0001	3141 (2863 - 3420)*	<0.0001
Mask	0.885	<0.0001	3689 (3075 - 4303)*	<0.0001
Lockdown	0.891	<0.0001	1363 (212 - 2514)*	0.021
Hand Sanitizer	0.88	<0.0001	3885 (3514 - 4257)	<0.0001
Testing				
Oxygen Saturation	0.944	<0.0001	3356 (3118 - 3594)	<0.0001
Computed Tomography	0.913	<0.0001	4094 (3483 - 4706)	<0.0001
Polymerase Chain Reaction	0.915	<0.0001	2410 (1570 - 3250)*	<0.0001
C-reactive protein	0.955	<0.0001	4553 (4014 - 5093)	<0.0001
D-Dimer	0.945	<0.0001	3521 (3155 - 3888)	<0.0001
Medication Use				
Inhaler	0.677	<0.0001	3344 (2969 - 3719)	<0.0001

Paracetamol	0.911	<0.0001	4158 (3259 - 5057)	<0.0001
Doxycycline	0.876	<0.0001	3510 (3174 - 3845)	<0.0001
Dexamethasone	0.765	<0.0001	3268 (2734 - 3803)*	<0.0001
Azithromycin	0.933	<0.0001	3223 (2862 - 3584)	<0.0001
Ivermectin	0.927	<0.0001	3000 (2694 - 3306)*	<0.0001
Fabiflu	0.943	<0.0001	2318 (1821 - 2815)	<0.0001
Remdesivir	0.902	<0.0001	1522 (961 - 2083)*	<0.0001

Control Search Terms

Film	0.076	0.49	-426 (-2059 - 1205)	0.61
Election	0.0092	0.93	-400 (-788 - 1589)*	0.51
Cricket	-0.735	<0.0001	-180 (-764 - 402)*	0.54
Bank	-0.389	0.0002	1683 (-2529 - -836)	0.0002
Animal	-0.44	<0.0001	641 (-2438 - 1155)	0.48

Time-adjusted β regression coefficient is estimated using a multivariate linear regression model with the search term and number of days as covariates.

*Search terms may not comply with normality assumptions as residual plots were not random. However, we still present regression coefficients for completeness and due to sufficient sample size.

Table 5. Correlation Between Search Terms and Official COVID-19 Death Counts

Search Term	Spearman Rank Correlation	P-value
General Terms		
Coronavirus Disease 2019	0.945	<0.0001
Coronavirus	0.918	<0.0001
COVID-19 Testing	0.916	<0.0001
COVID-19 Vaccine	0.816	<0.0001
Symptoms		
Fever	0.921	<0.0001
Headache	0.808	<0.0001
Fatigue	0.764	<0.0001
Cough	0.908	<0.0001
Chest pain	0.75	<0.0001
Disease Process/Complications		
Stroke	-0.131	0.23
Pneumonia	0.89	<0.0001
Myocardial Infarction	0.417	<0.0001
Infection	0.853	<0.0001
Healthcare Utilization		
Pharmacy	0.513	<0.0001
Ambulance	0.741	<0.0001
Physician	0.508	<0.0001
Oxygen	0.878	<0.0001
Hospital	0.863	<0.0001
Bed	0.585	<0.0001
Preventive Behaviors		
Quarantine	0.869	<0.0001
Pulse Oximetry	0.954	<0.0001
Mask	0.883	<0.0001
Lockdown	0.897	<0.0001
Hand Sanitizer	0.897	<0.0001
Testing		
Oxygen Saturation	0.941	<0.0001
Computed Tomography	0.908	<0.0001
Polymerase Chain Reaction	0.918	<0.0001
C-Reactive Protein	0.952	<0.0001
D-Dimer	0.938	<0.0001
Medication Use		

Inhaler	0.709	<0.0001
Paracetamol	0.89	<0.0001
Doxycycline	0.864	<0.0001
Dexamethasone	0.785	<0.0001
Azithromycin	0.929	<0.0001
Ivermectin	0.927	<0.0001
Fabiflu	0.927	<0.0001
Remdesivir	0.891	<0.0001
Control Search Terms		
Film	0.072	0.51
Election	-0.012	0.91
Cricket	-0.729	<0.0001
Bank	-0.37	0.0004
Animal	-0.448	<0.0001

Table 6: Thematic Analysis of Common Search Term-related Queries

Search Category	Search Terms	Related Queries	Associated Themes
General	Coronavirus	coronavirus vaccine registration, coronavirus testing, coronavirus third wave;	Testing
	COVID-19 Testing	covid test near me;	Testing
	COVID-19 Vaccine	covid vaccine registration, covid-19 vaccine registration online, covid-19 vaccine near me	Vaccine
Symptoms	Fever	fever after covaxin, continuous fever in covid, 99.9 fever	Vaccine, Symptoms
	Cough	how long does covid cough last, lupituss cough syrup, wet cough in corona	Treatment Options, Symptom Expectations
	Headache	headache after covid, headache after covishield, covid headache	Vaccine, Symptom Management
	Chest pain	chest pain covid, chest pain in covid, chest pain in corona	Symptom Management
	Fatigue	bio bubble fatigue, symptoms of corona, covid symptoms	Symptom Expectations

	Loss of Taste	loss of smell but not taste, loss of taste and smell covid, loss of taste and smell in covid	Symptom Management
	Loss of Smell	loss of smell but not taste, loss of smell covid which day, corona symptoms, loss of smell reasons, loss of smell in covid	Symptom Management, Symptom Expectations
	Diarrhea	covid diarrhea, covid symptoms, is diarrhea a symptom of covid	Symptom Expectation
	Vomiting	vomiting during covid, vomiting sensation reasons, vomiting in covid	Covid Specific, Symptom
Disease Process	Pneumonia	remdesivir, stages of covid pneumonia, progressive pneumonia, atypical viral pneumonia, covid pneumonia recovery time	Treatment Options, Disease Progression
	ARDS	ards full form in medical	Meaning of term
	Infection	vaccination after covid infection, covid lung infection recovery time, fungal infection after covid, covid infection timeline, lungs infection medication, lung infection symptoms	Covid Specific, Complications, Etiology, Disease Progression

	Heart Attack	symptoms of corona, how to prevent heart attack, reasons for heart attack, first aid for heart attack, heart attack symptoms in hindi	Etiology, Treatment Options, Meaning of term
	Heart Failure	heart failure symptoms	Symptoms Expectation
	Stroke	heat stroke, stroke symptoms	Symptoms Expectation
Testing	D-dimer	d dimer covid, d dimer in covid, crp d dimer, crp, crp test, high d dimer	Normal levels, Testing
	CT Scan	CT scan for covid, covid chest CT scan, CT severity score, chest CT scan, CT scan price, ct scan is harmful, ct scan side effects	Normal levels, Cost, Side Effects, Testing
	RT-PCR	rt pcr test near me, rt pcr report, rt pcr test report, rt-pcr test, pcr, rt pcr test government centers near me	Locations, Testing
	C Reactive Protein	c reactive protein covid, c reactive protein in covid, c reactive protein and covid, d dimer test, high c reactive protein, c reactive protein increased, c reactive protein meaning in hindi	Normal levels, Meaning of Term, Testing
	IL-6	il-6 test	Testing

	Ferritin	ferritin levels in covid, serum ferritin means, ferritin test in hindi, ferritin test, ferritin in covid	Normal levels, Meaning of Term, Testing
	Oxygen Saturation	normal saturation level of oxygen, normal pulse rate, oxygen concentrator, saturation meaning, oxygen saturation level	Normal levels,
	Blood Pressure	oximeter price, blood pressure measuring instrument, pulse oximeter, diastolic blood pressure range, yoga for high blood pressure	Normal levels, Treatment options, Cost
	Blood glucose	blood glucose normal, blood glucose level, blood glucose level normal, glucose level in blood, normal glucose level in blood	Normal levels,
Healthcare Utilization	Hospital	hospital near me, apollo, apollo hospital, best hospital, eye hospital, covid hospital near me	Location
	ICU	neonatal intensive care unit, icu full form	Meaning of Term
	Ambulance	supro ambulance, drdo bike ambulance, ambulance with ventilator, ambulance service, ambulance number.	Location, Contact information

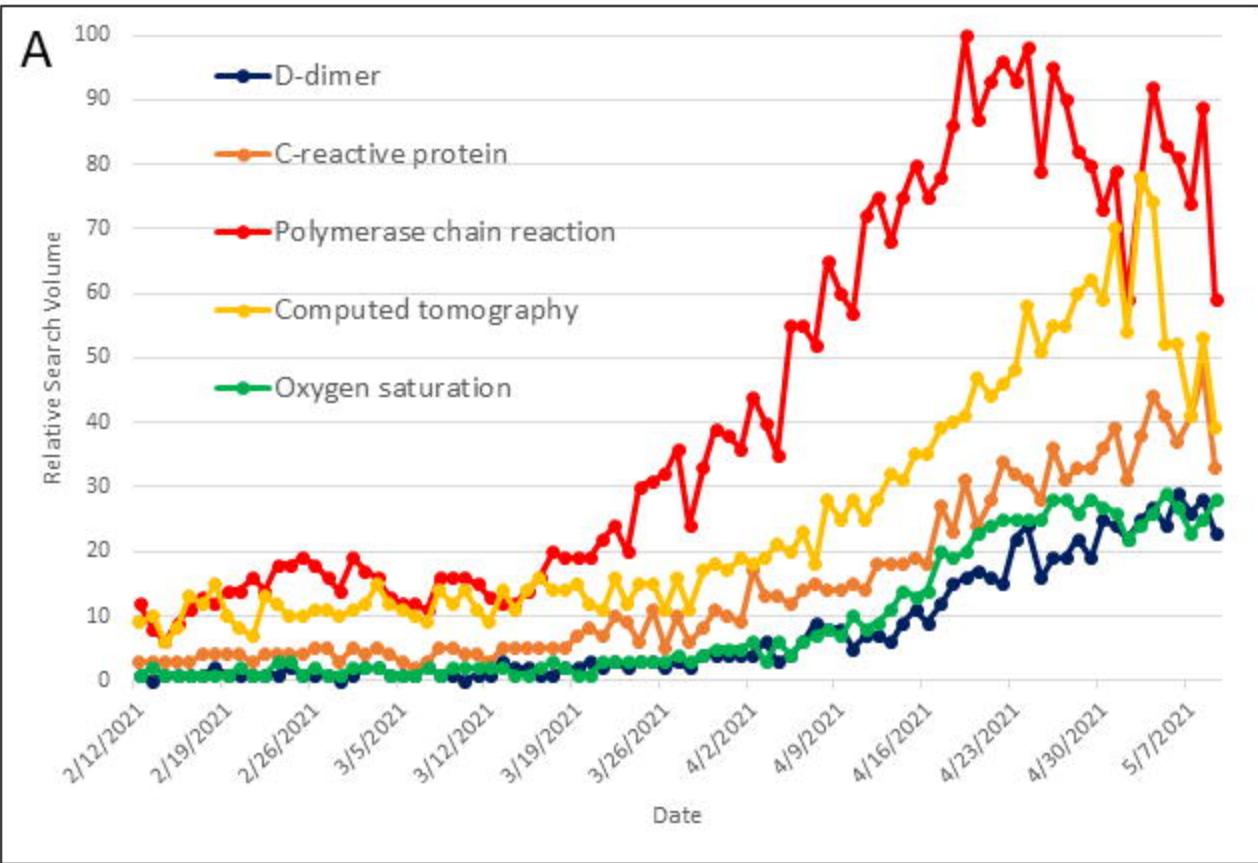
	Cylinder	oxygen concentrator price, oxygen cylinder refill near me, oxygen cylinder in lucknow, oxygen cylinder, oxygen	Cost, Location
	Doctor	doctor tamil movie cast, esanjeevani doctor login, doctor video, doctor near me, dr, doctor stranger k drama	Location
Medication Use	Remdesivir	Remdisivir online, remdesivir injection online, buy remdesivir injection, remdesivir injection buy online, remdesivir availability	Purchase options
	Paracetamol	can we take paracetamol after covid vaccine, bagged packaged goods, how many paracetamol 650 in a day, paracip 650	Medication intake, Vaccine
	Doxycycline	ivermectin use, azithromycin 500, ivermectin 12 mg, doxycycline 100 mg uses in hindi, doxycycline and lactic acid bacillus cap uses	Medication intake, Meaning of term, Medication usage
	Azithromycin	ivermectin use, doxycycline 100 mg, is azithromycin an antibiotic, azythromycin, azithromycin oral suspension ip uses in hindi	Medication usage, Meaning of term

	Ivermectin	doxycycline covid, doxy 100 uses, ivermectin 12 mg uses, tab ivermectin 12 mg hindi, ivermectin 12 mg price	Meaning of Term, Cost, Medication Usage
	Dexamethasone	remdesivir injection, dexamethasone 6 mg, dexamethasone vs remdesivir, dexamethasone brand name, dexamethasone tablet uses in tamil	Medication usage, Meaning of term
	Vitamin D	when to take vitamin d, vitamin d3 60k, vitamin d 60000 iu, vitamin c tablets, depura vitamin d3	Medication Intake
Preventative Behaviors	Lockdown	lockdown in telangana 2021, lockdown extended in delhi, raipur lockdown news, bihar lockdown guidelines, weekend lockdown in up	News, Guidelines
	Quarantine	quarantine hotels in mumbai, quarantine meaning in marathi, quarantine meaning in bengali, home quarantine schedule, bhoomi quarantine watch	Location, Meaning of term, Guidelines
	Face Masks	facemask, full face mask for men, face shield, best face mask, face mask price	Purchase options, Cost

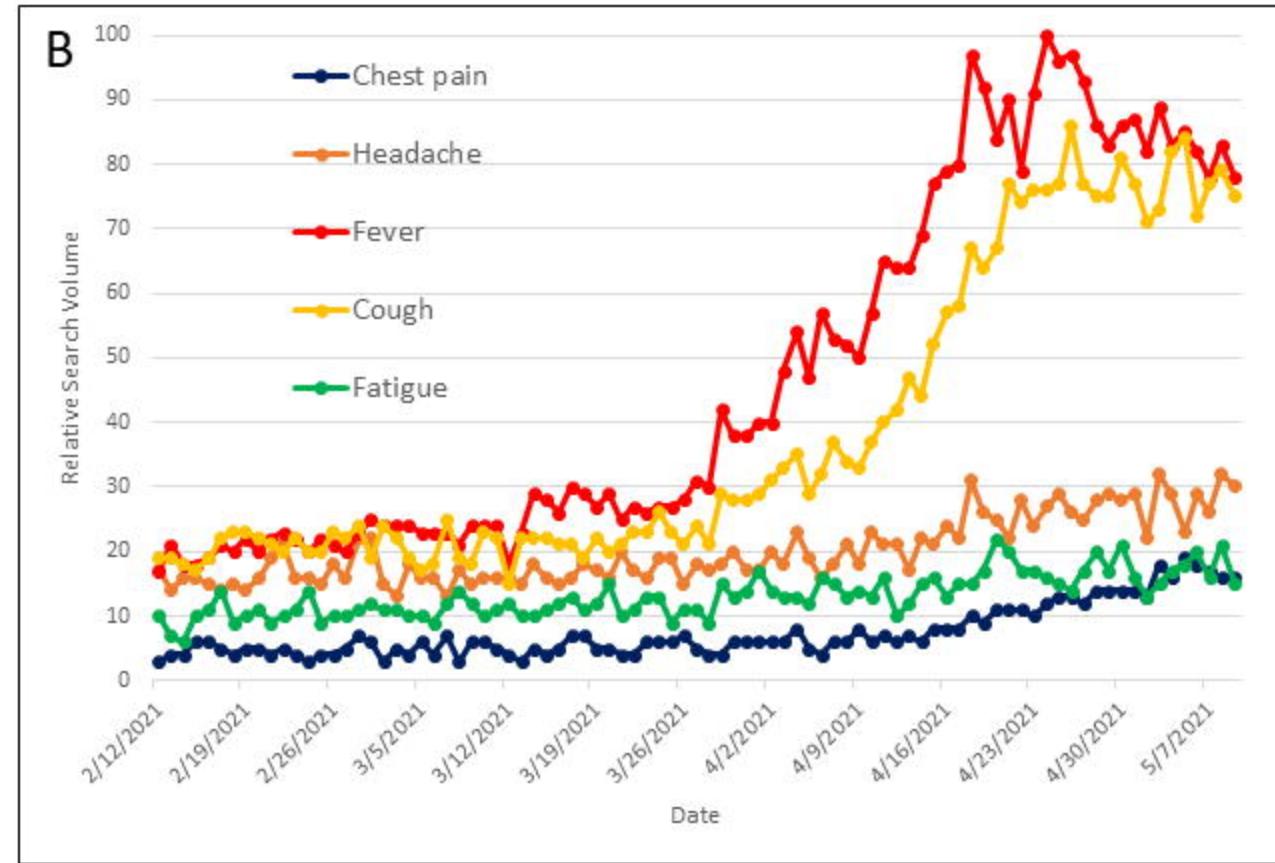
Pulse Oximeters	oximeter online india, which oximeter is best, gilma oximeter, what is pr bpm in oximeter, oximeter price bangalore	Purchase options, Cost, Meaning of Term
Handwashing	Handwashing steps, handwashing machine, second hand washing machine, step of hand washing, handwash refill	Guideline
Sanitizer	Hand sanitizer, spray sanitizer, sanitizer price, sanitizer machine	Purchase options, Cost

Figure 1. Clinically Relevant COVID-19 Search Terms

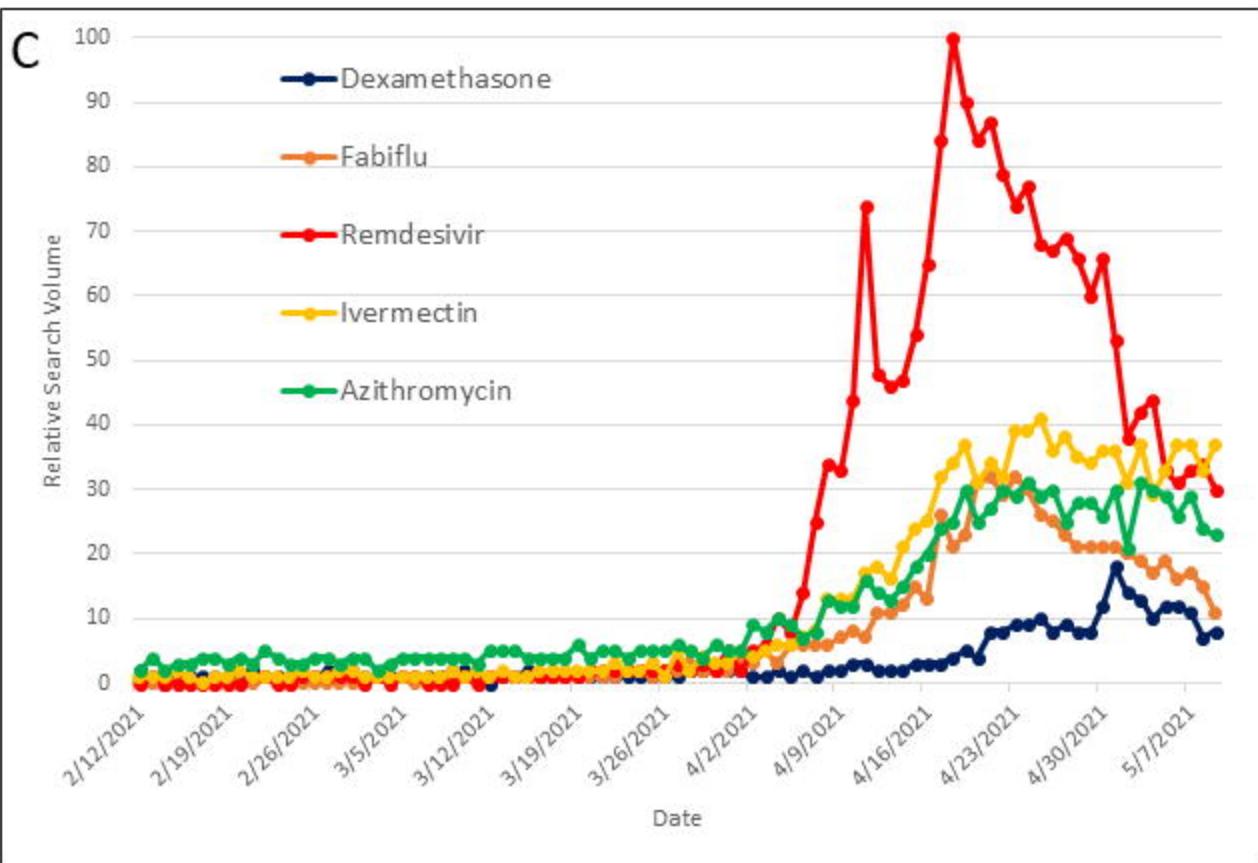
COVID-19 Testing Modalities



Most Prevalent Symptoms



Common Medications



Disease Process/Complications

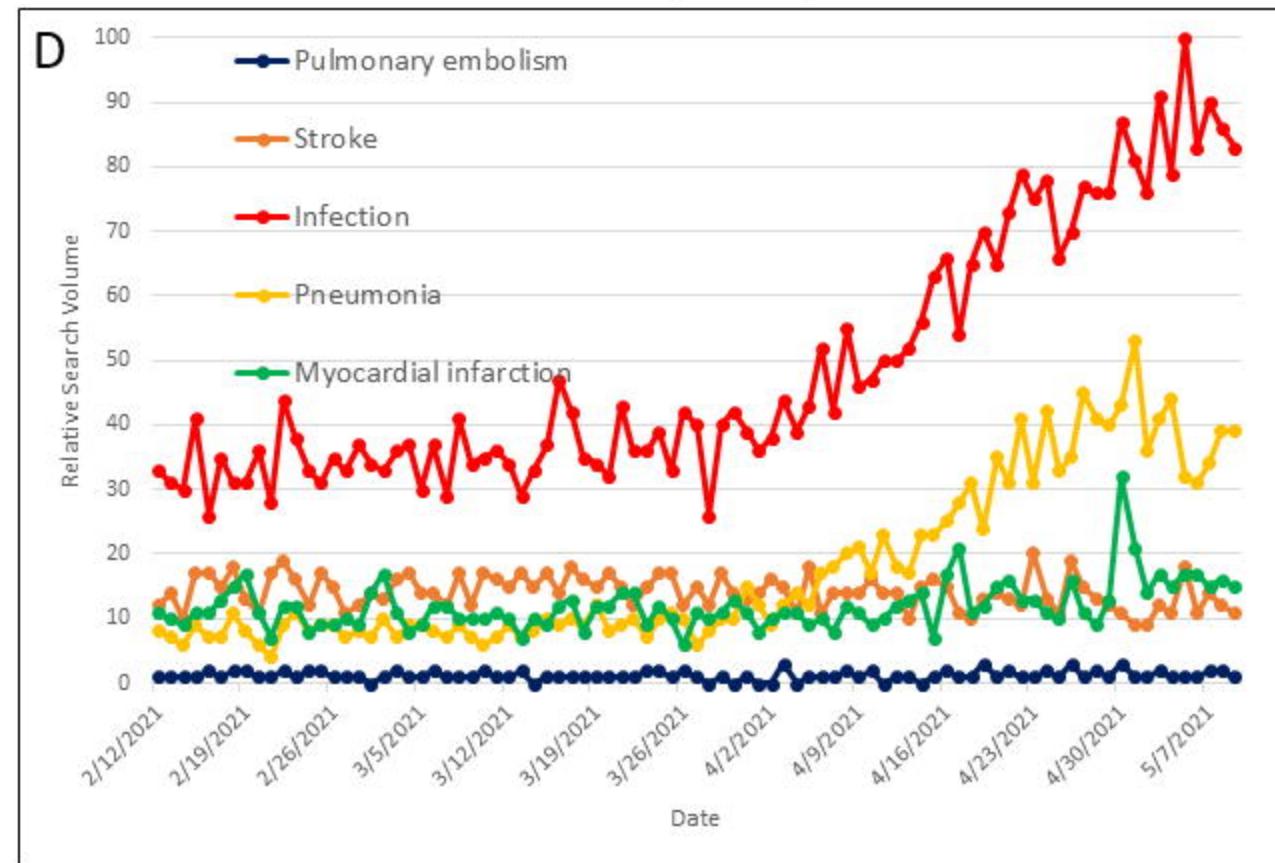
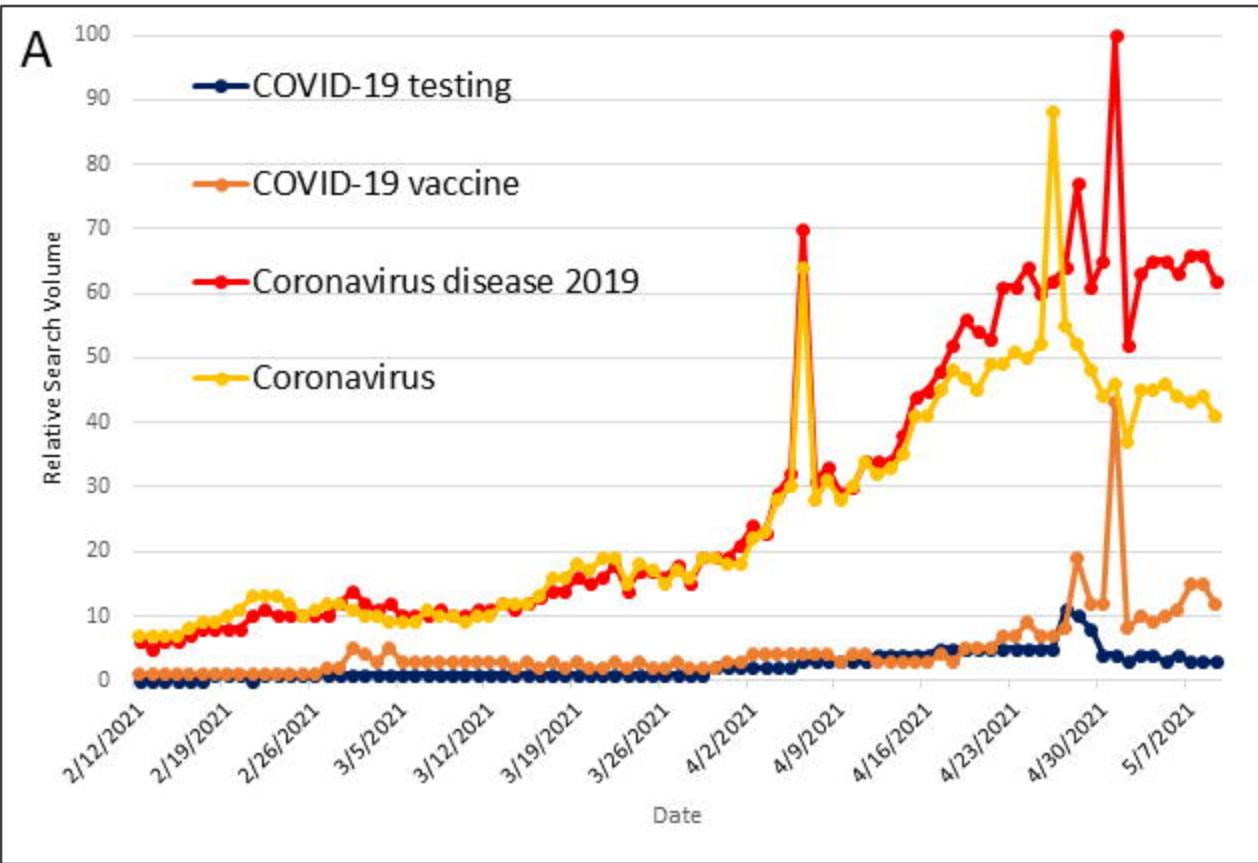
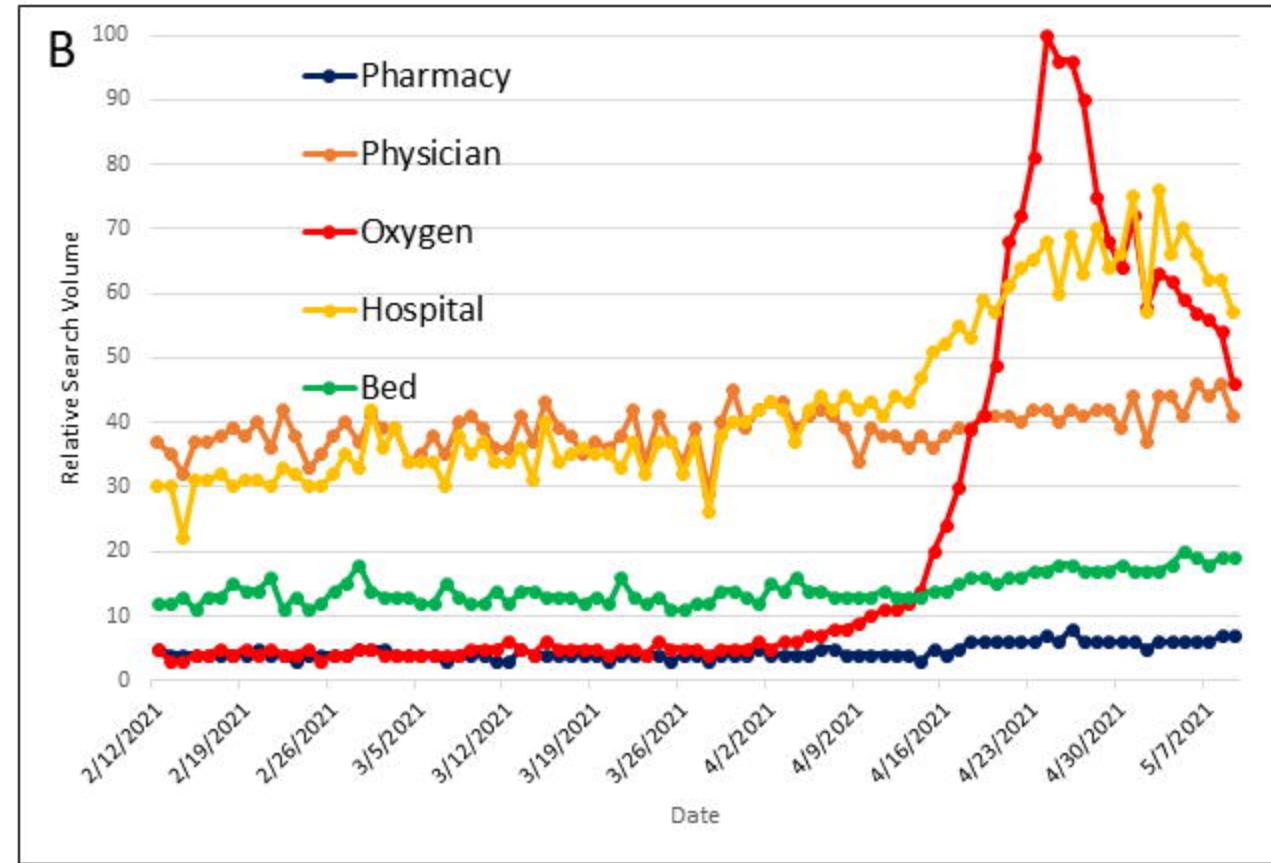


Figure 2. Socially Relevant COVID-19 Search Terms

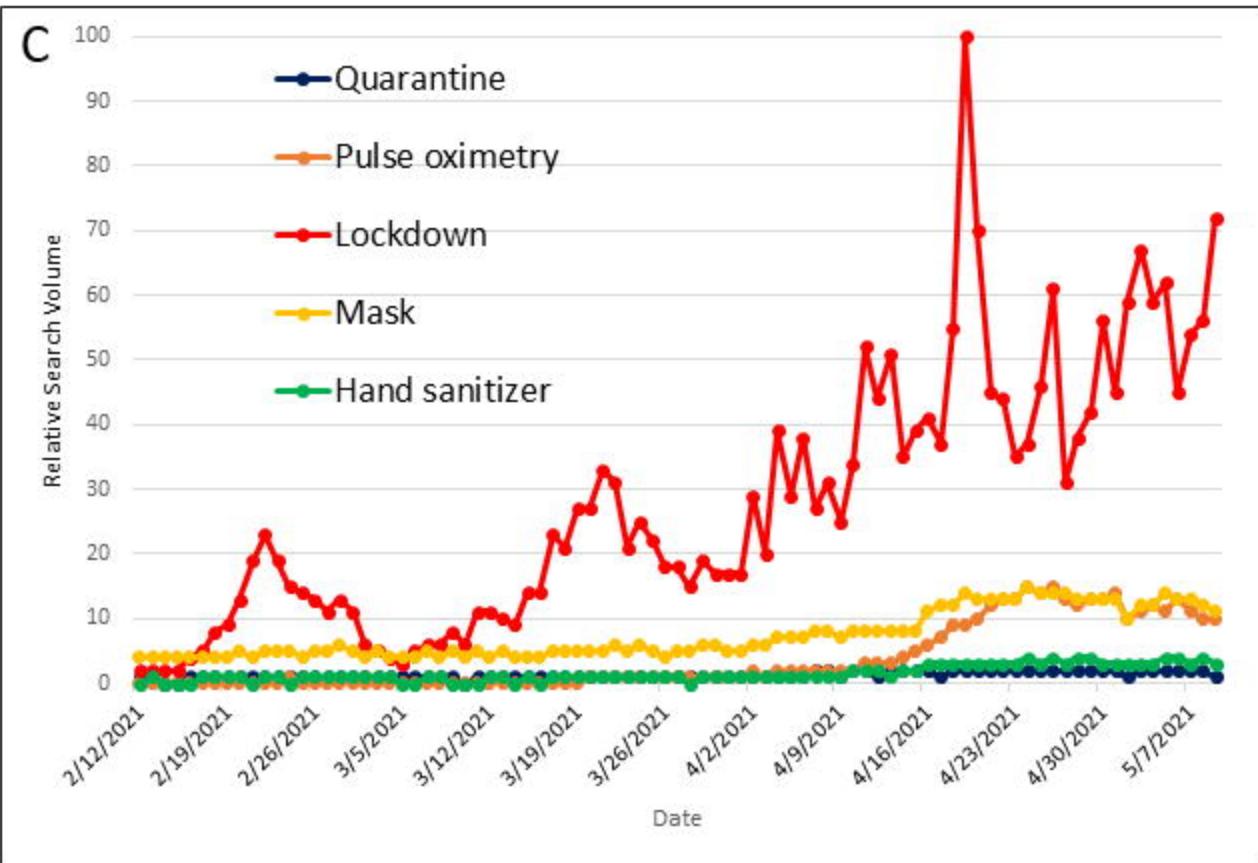
General COVID-19 Terms



Healthcare Utilization



Preventive Behaviors



Festivals/Mass Gatherings

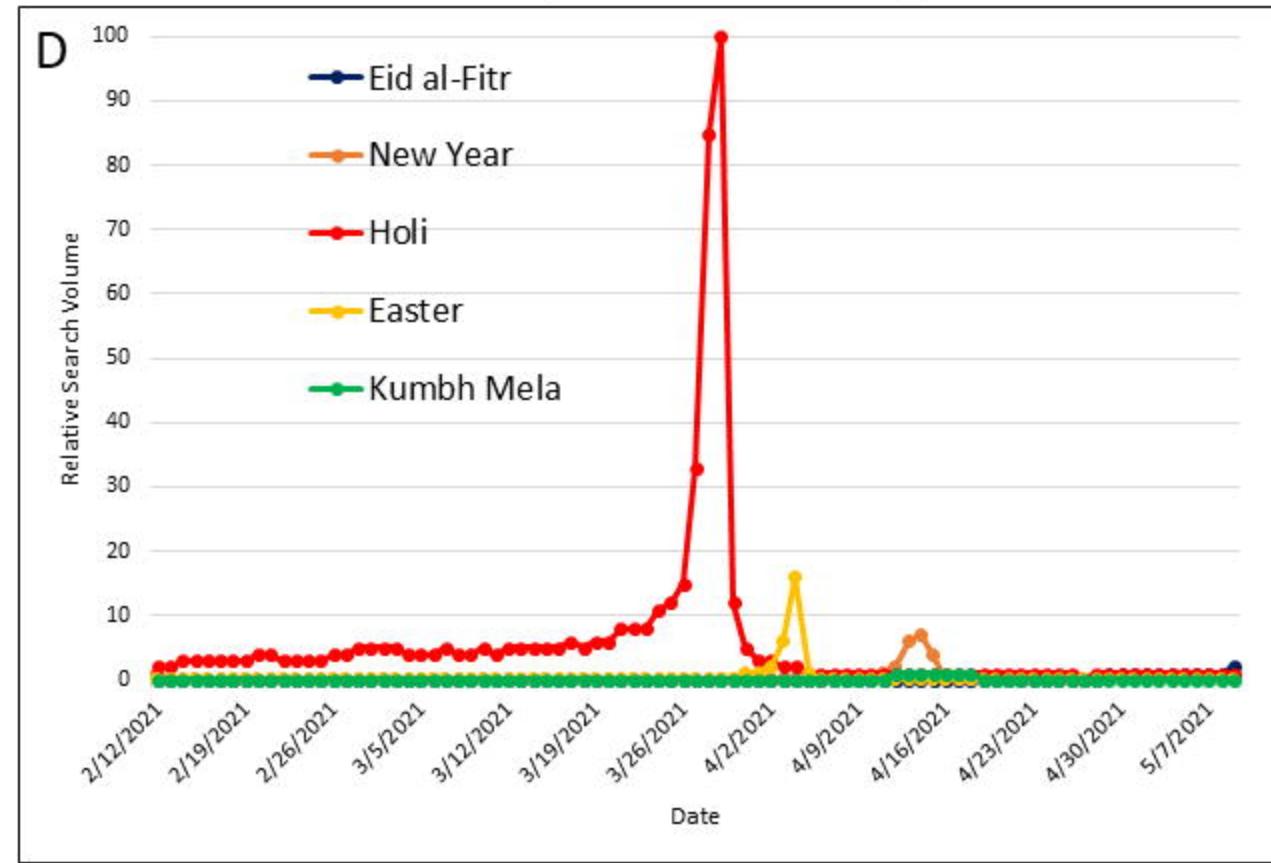
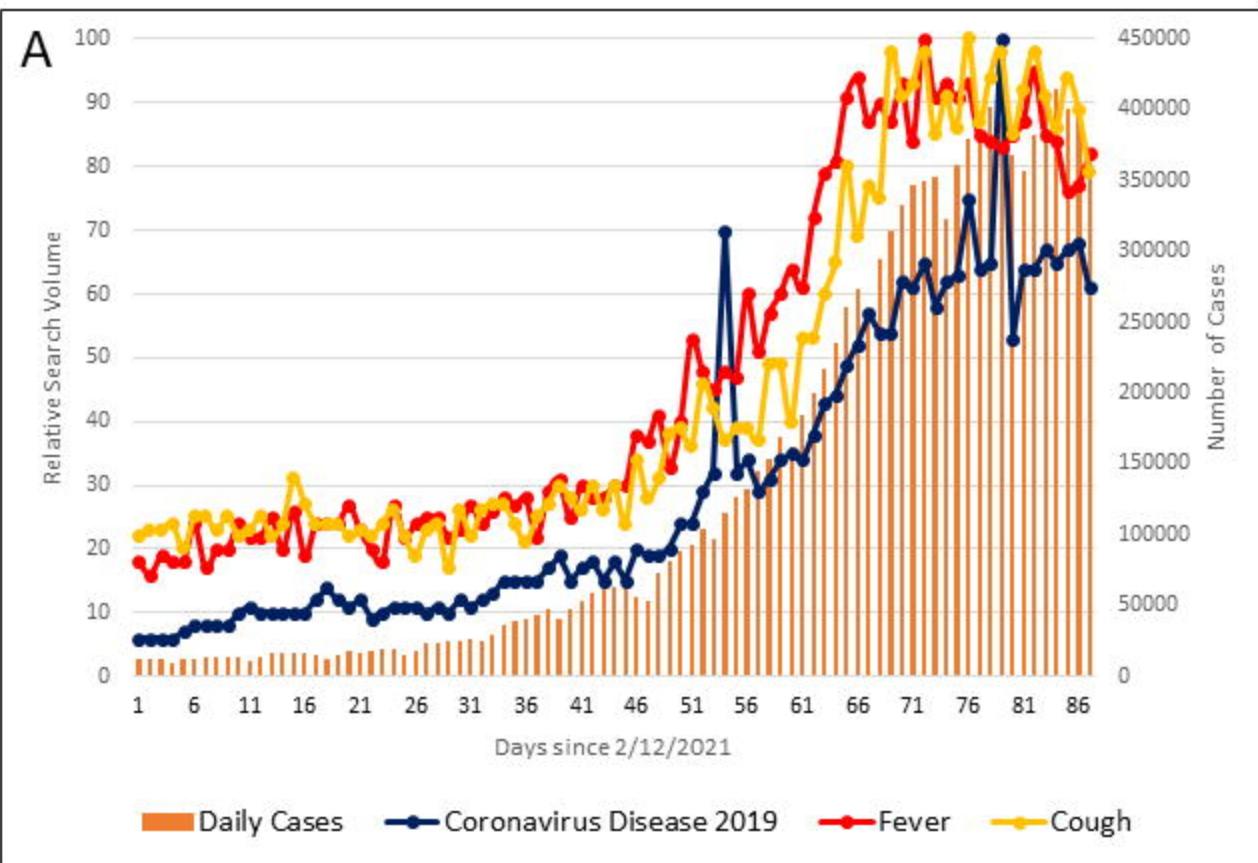
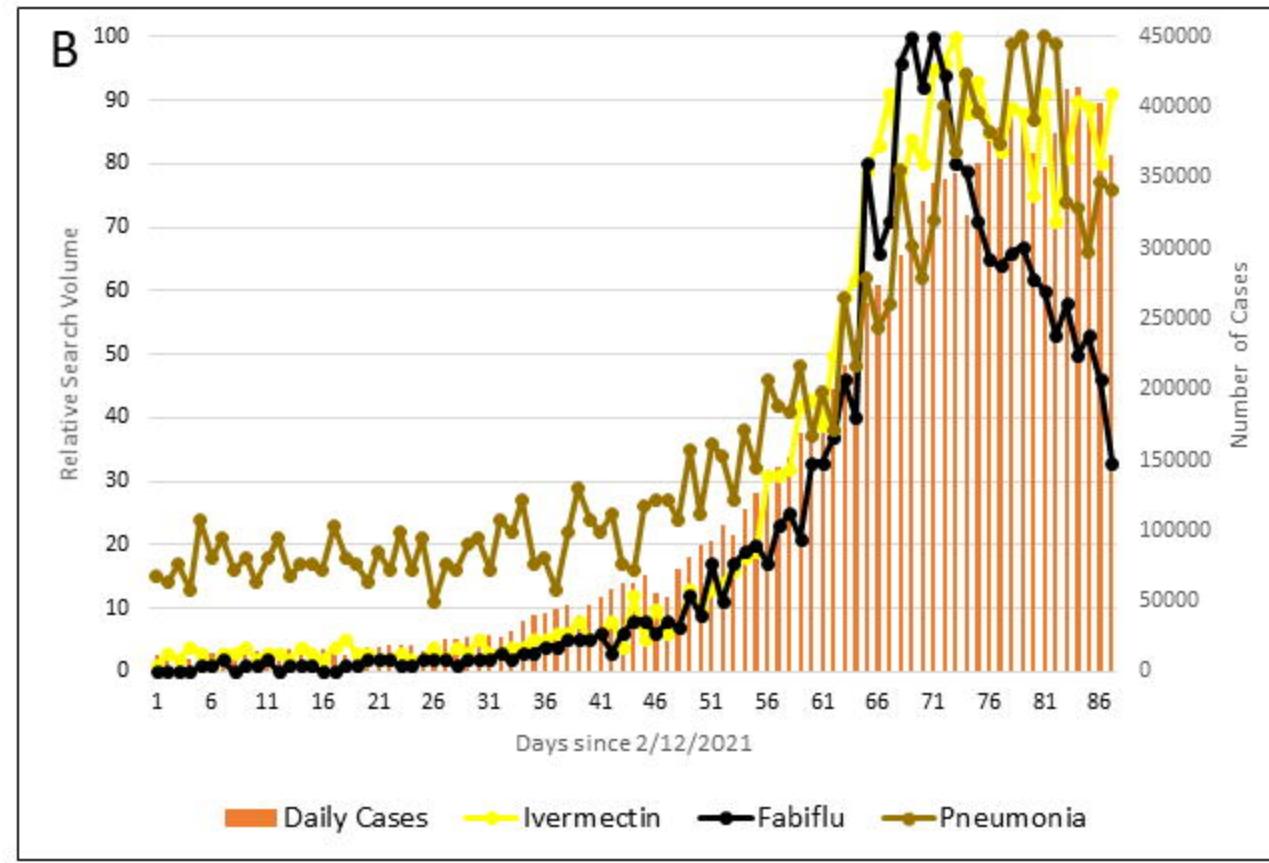


Figure 3. Correlation Between COVID-19 Cases and Search Terms

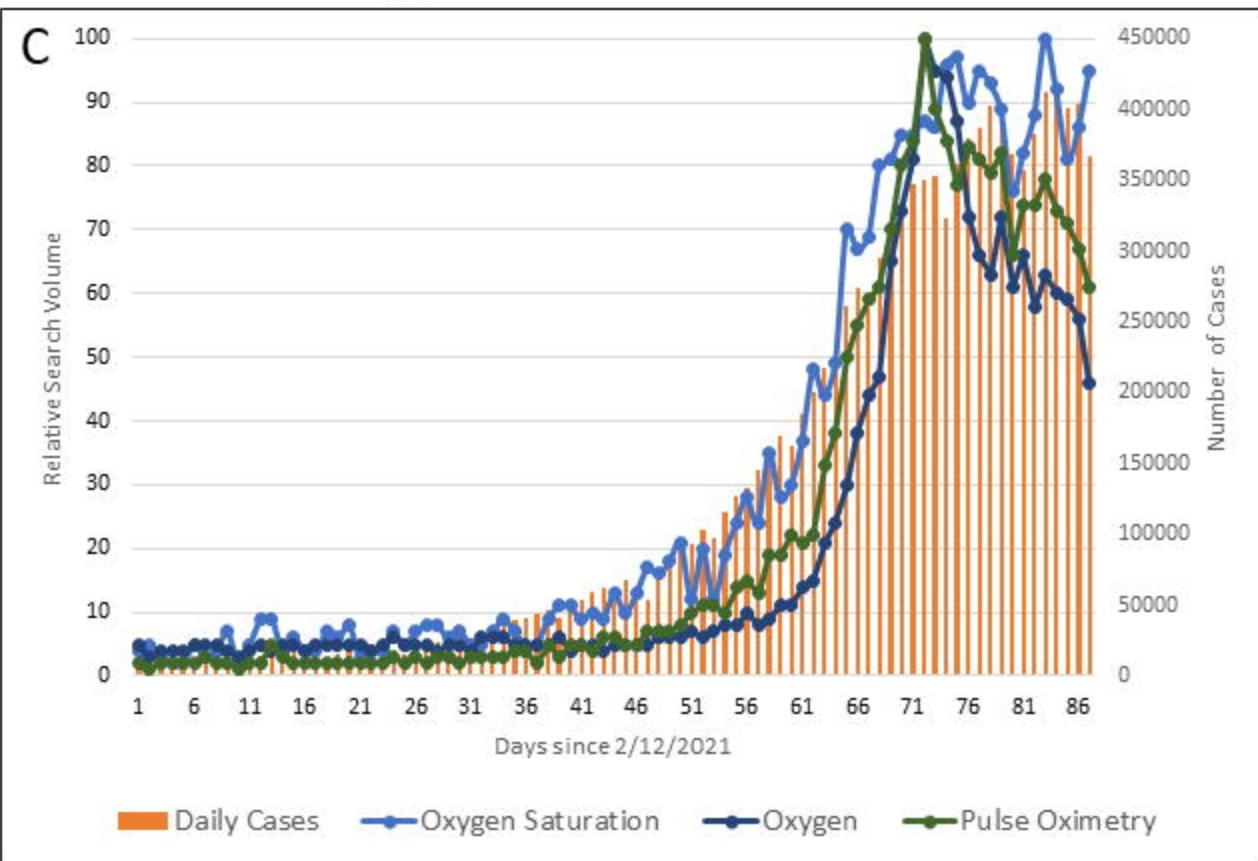
General Terms and Symptoms



Medication Use and Disease Complications



Oxygen-related Search Terms



Laboratory Testing

