Determining the nutritional immunity information-seeking behaviour during the COVID-19 pandemic in India: a Google Trends data analysis

Savitesh Kushwaha, Poonam Khanna, Rachita Jain and Rachana Srivastava*
Department of Community Medicine & School of Public Health, Postgraduate Institute of Medical Education and Research, Chandigarh 160012, India

Submitted 27 January 2021: Final revision received 19 July 2021: Accepted 2 August 2021: First published online 5 August 2021

Abstract
Objective: During COVID-19, the Internet was a prime source for getting relevant updates on guidelines and desirable information. The objective of the present study was to determine the nutritional immunity information-seeking behaviour during COVID-19 in India.

Design: Google Trends (GTs) data on relevant COVID-19 and nutritional topics were systematically selected and retrieved. Data on newly reported COVID-19 cases were also examined on a daily basis. The cross-correlation method was used to determine the correlation coefficient between the selected terms and daily new COVID-19 cases, and the joinpoint regression models were utilised to measure monthly percent change (MPC) in relative search volumes (RSV).

Setting: Online.

Participants: People using Google search during the period 1 January 2020 – 31 August 2020 in India.

Results: The date of peak searches can be attributed to the COVID-19 guidelines announcement dates. All the nutritional terms showed a significant increase in average monthly percentage change. The higher than the average daily rise in COVID-19 cases leads to a higher than average increase in RSV of nutritional terms with the greatest association after 14–27 d. The highest mean relative search volume for nutritional terms was from Southern India (49·34 ± 7·43), and the lowest was from Western India (31·10 ± 6·30).

Conclusion: There was a significant rise in the Google searches of nutritional immunity topics during COVID-19 in India. The local/regional terms can be considered for better outreach of public health guidelines or recommendations. Further automation of Google Trends using programming languages can help in real-time monitoring and planning various health/nutritional events.

Worldwide 3·4 billion people have access to the Internet, and India is the second largest country with trending 749 million Internet users. Catalysts for this magnitude of Internet access includes low-cost mobile data availability and the usefulness of smartphones over desktops and tablets(1). The Internet gives instant access to an enormous amount of information, and infodemiology has been used to assess human behaviours related to the COVID-19 pandemic(2). During pandemic situations, the most feasible source of information is television, the Internet, and telephone, etc. The accessibility of the Internet has affected our social lives along with dietary and lifestyle behaviours by providing immediate access to a huge amount of information. The COVID-19 outbreak has brought drastic changes in human behaviour. People are diving into Google searches to understand their wants and needs for staying healthy in the current situation(3). In response to the increased COVID-19 cases, the lockdown was implemented across various countries, and the Internet was the only valuable tool for accessing health-related information about immunity-boosting nutrients and vaccines, enhancing their self-efficacy in fighting against the virus(4). Immune boosting was a trending topic correlated with the coronavirus pandemic due to people’s heightened concern of developing a fighting fit immune system(5). The pandemic has also affected the content explored by Internet users, and Internet services rise from 40 to 100%(6).
Nutritional surveillance during COVID-19

World Health Organization (WHO)\(^7\), FAO\(^8\), UNICEF\(^9\) as well as national institutions like National Institute of Nutrition (NIN)\(^10\), The Ministry of Ayurveda, Yoga, Naturopathy, Unani, Siddha, Sowa-Rigpa and Homeopathy (AYUSH)\(^11\), and Food Safety and Standards Authority of India (FSSAI)\(^12\) have issued guidelines such as physical distancing, facial cover, proper sanitisation, healthy eating tips to stop the spread of SARS-CoV-2. General awareness and awareness about guidelines issued during COVID-19\(^8\)\(^11\)\(^13\) requires an assessment to determine the impact of nutritional guidelines and recommendations. Determining the nutritional information-seeking behaviour was more challenging during pandemic lockdown situations. The usability of the Internet and other contactless methods was the only source to spread awareness and information during COVID-19.

Currently, limited data are available on how the COVID-19 pandemic affects our dietary and lifestyle-related behaviours. Therefore, Google Trends (GTs) search data available in the form of relative search volumes (RSV) can be used to analyse relevant keywords related to nutritional topics. Google Trends is the most popular tool to gather information on web-based behaviours, and it can be used to prevent health-related issues\(^14\). Therefore, the present study aimed to predict and monitor the nutritional immunity information-seeking behaviour amongst the Indian population during the COVID-19 pandemic using GTs.

**Methods**

The data on daily new confirmed COVID-19 cases in India was accessed from (~https://ourworldindata.org/coronavirus/country/india?country=IN-IND). This website is managed by Oxford Martin School (University of Oxford) and updates worldwide COVID-19 cases daily. The data were downloaded in .CSV format and filtered for a time frame of 1 January 2020–31 August 2020 on 3 November 2020 (DD-MM-YYYY).

**Google Trends**

Google Trends (https://trends.google.com/) is a website by Google (launched on 11 May 2006) that analyses the top search queries across various regions, languages, and time frames. It shows the real-time data points scaled according to the total search volume for a given geographical location and time frame. The relative search score was provided in terms of ‘Relative Search Volume’ (RSV), which range from 0 to 100\(^15\). The RSV was shown in graphical representation, and subregion wise data were also available on the website. Google Trends offers ‘related search queries’ and ‘related search topics’, which can be searched with the selected query. These queries and topics can be used to select useful terms for determining interest. Users can compare up to five different search queries, and search filters were provided to obtain the precise data based on region, time frame, search category, and search type\(^16\). Google Trends offers very dynamic data, and the user can modify the search trend according to the interest. So, to maintain the quality of data, the authors have modified the GTs checklist developed by Nuti et al.\(^17\) (see online Supplemental File 1 Tables 1 and 2) and produced a systematic flowchart for the present study (Fig. 1).

**Selection of search inputs**

The ‘query categories’ were used instead of ‘search term’ (Fig. 1) because it was considered as a group of terms sharing similar meanings in different languages. It was essential to minimise the linguistics variations because multiple languages are spoken and written in India. The search inputs for two groups were chosen to determine people’s interest during the COVID-19 period: 1. COVID-19 and 2. Nutrition. The first group (COVID-19) was chosen to gauge public interest in COVID-19-related issues. The data for three queries (Coronavirus), (Immunity), and (Vaccine) were obtained under the first group. The ‘virus’ for (Coronavirus) and ‘Topic’ for (Immunity) and (Vaccine) was selected as query categories from the suggestions, which appears after entering the query in the search box. Following the first query term (Coronavirus), the terms (Immunity) and (Vaccine) were chosen from the ‘related search topics’ offered by GTs based on their high interest.

For the second group, i.e., Nutrition, the (Nutrients) was selected as the initial query term, and four more queries (Multivitamin), (Chyavanprash), (Vitamin), and (Zinc) were determined by the authors based on nutritional immunity literature\(^18\)\(^20\) and ‘related search topic’. The term (Chyavanprash) is a ‘Hindi’ language term for Ayurvedic supplement, a cooked mixture of various herbs and spices consumed widely in India to maintain stamina, strength, and immunity. The nutrition-related queries were also considered as ‘query category’ instead of ‘search term’\(^21\)\(^22\). The ‘Topic’ was selected as the query category for (Chyavanprash), (Vitamin), and (Nutrient). The ‘Medication’ and ‘Chemical element’ was chosen as the query category for (Multivitamin) and (Zinc), respectively (Fig. 1).

**Data constraints**

The data for Indian states and union territories (UTs) was obtained after entering the query terms and selecting the above mentioned ‘query categories’. The time frame of 8 months (1 January 2020–31 August 2020) was selected within the category ‘Health’ and ‘Web search’ for all the terms. To avoid the impact of increased search for online food products due to lockdown, we have selected the ‘Health’ category, and only those query terms were included, which were observed to be least influenced by the lockdown period. The data based on subregions (Indian states and UTs) was also obtained from the GTs web portal.
**Statistical analysis**

The data from the Google Trends web portal was downloaded in CSV format for all the queries and opened in MS Excel. The time series graph of daily new COVID-19 cases and nutritional terms in addition to events dates were plotted separately (Figs. 2 and 3). The Joinpoint regression model was used to determine the statistically significant monthly percent change (MPC). The best-fitting points were termed as ‘Joinpoints’, which marks an increase or decrease in RSV. The Joinpoint regression programme was used for analysis (23). The software creates a graphical timeline that identifies the apparent change in the RSV. According to the software guidelines, the grid search method was used with a criterion of minimum 0 and maximum 1 Joinpoints (24). The mean search score of each month with SE was used as a dependent variable, and 1–8 coded value for each month (January–August) was selected as an independent variable. The log transformation function was used due to moderate to high skewness in the data. The nil (no data) RSV was considered as 0 and RSV showing ‘<1’ was replaced with a value of 0·5 for feasibility in data analysis. The model selection method was the permutation test and the significance level was set at P < 0·05.

Additional analysis was carried out using the cross-correlation function (CCF) in SPSS version 26.0. The correlation coefficient was calculated by comparing the explanatory (new confirmed COVID-19 cases) and dependent (Google Trends RSV) time series variables. A greater time lag between –35 and +35 d was considered to get more precise coefficient values. The natural log transformation was selected for estimating the CCF. The \( r \geq 0·3-0·5 \) (fair), \( r \geq 0·6-0·7 \) (moderate), \( r \geq 0·8-0·9 \) (very strong), and \( r = 1 \) perfect was considered in the present study (25).
Fig. 2 (colour online) Time series plot of COVID-19-related search queries in reference to various national and international events.

Fig. 3 (colour online) Time series plot of nutrition-related search queries in reference to various national and international events.
The significance CCF was demonstrated at $P < 0.01$. The increased/decreased search activity with an increase in COVID-19 cases was correlated to determine the behaviour of Indian people during COVID-19.

The mean ± SD RSV of each state and UTs was calculated, and size dot maps with colour gradients were developed using ArcGIS Pro (Esri) version 2.5, representing the high and low search areas. For estimating variations in searches in the mainland (states and UTs), the country was divided into six regions – North, South, East, West, Central, and North-East (20). The colour gradient maps were developed to represent the overall COVID-19 and nutritional searches in different regions of India. The mean RSV of Lakshadweep were not reported because the RSV was observed as nil amongst all the selected terms.

Results

Influence of COVID-19 guidelines on RSV

The first COVID-19 case in India was detected on 31 January, after which the government have taken multiple measures to control the spread of the virus. The rise in COVID-19 cases in India was controlled till 8 June 2020. The exponential rise in cases can be observed after 8 June, when the government of India announced Unlock Phase 1. Several announcements were also made by international (FAO, WHO, UNICEF (7)–(9) and national (NIN, AYUSH, FSSAI) (10–12) institutions in response to COVID-19 (Figs. 2 and 3).

The GTs time trend plots show a spike in search activities attributed to the guidelines and notifications issued by governmental institutions and development in COVID-19. The highest peak for the term (Immunity) was observed on 14 April 2020 when the Prime Minister of India urges to follow saptapadi (seven sacred vows), one of which was to increase immunity following AYUSH ministry guidelines, which was announced on 31 March 2020 (Fig. 2). The second highest peak was observed on 21 March 2020 when yoga asanas to boost immunity were shared by an Indian yogic practitioner (Baba Ramdev). The spike in searches for the term (Coronavirus) between 25 March and 1 April 2020 could be due to India’s statewide lockdown, which began on 25 March 2020. On 11 August 2020 Russian vaccine was announced; therefore, this might be the reason for the peak in term (Vaccine) on that day (Fig. 2) (see online Supplemental File 2).

The peak searches for nutritional terms were continuously increasing after March 2020 and highest during July 2020. Initial studies on vitamin D deficiency in the progression of COVID-19 came during June and July, which reported that the deficiency in vitamin D could increase the severity of COVID-19 (25,26). The Indian news reports have also mentioned the severe shortage of vitamin tablets in July (29). This could be due to the panic purchase of tablets in India in the aftermath of the COVID-19 lockdown, which saw import and manufacturing halted. National Institute of Nutrition (NIN) and the Food Safety Standard Authority of India (FSSAI have also released the guidelines on 19 June 2020 and 7 June 2020, respectively (10,12). The Chyavanprash is a consumable health product that is consumed to boost immunity (22,20). The search spike from April to July months shows the interest of buyers for boosting immunity (31), whereas the usual search spike for Chyavanprash was from November to February (winter season) (Fig. 3) (see online Supplemental File 2).

Monthly variation in RSV

The Joinpoint regression analysis shows MPC in RSV of all the selected query terms (see online Supplemental File 3 Table 1). The term (Immunity) shows one Joinpoint with a 112·34 % rise in RSV from January to April and a decline of 8·34 % after April. For term (Coronavirus), the regression analysis shows one Joinpoint with a 1234·39 % rise in RSV between January and March, which declines by 36·04 % after March. No Joinpoint was found in the selected model for the term (Vaccine), but a 46·05 % significant rise in RSV was observed. The nutritional query terms (Vitamin), (Chyavanprash), (Multivitamin), and (Zinc) have no Joinpoints but shows 8·92 %, 16·15 %, 13·17 %, and 18·26 % significant rise in RSV, respectively. The term (Nutrient) shows one Joinpoint in the selected model with a 26·61 % rise in RSV between January and May and a 3·17 % decline between May and August (Fig. 4).

Influence of daily new COVID-19 cases on RSV

The cross-correlation analysis results have shown the correlation between selected query terms with daily new COVID-19 cases in India. The highest correlation factor ± SE ($r$ ± SE) was considered for interpretation with a time lag (days). The query terms (Immunity), (Vaccine), (Vitamin), (Nutrient), (Multivitamin), and (Zinc) have shown moderate correlation ($r > 0·60$) with daily new COVID-19 cases, whereas (Chyavanprash) showed very strong correlation ($r > 0·80$) (Fig. 5).

The highest correlation between RSV and daily new COVID-19 cases for (Immunity) and (Coronavirus) was observed at the lag of 35 d. Similarly, the highest correlation for RSV of terms (Vaccine), (Vitamin), (Chyavanprash), (Nutrient), (Multivitamin), and (Zinc) occurs at the lag period of 16, 27, 22, 17, 20, and 14 d respectively (Fig. 5). At time lag 0, the RSV of selected search terms shows moderate correlation except for (Immunity) and (Coronavirus) (Fig. 6) (see online Supplemental File 3 Table 2).

The negative CCF was only observed with the RSV of the term (Coronavirus). The CCF increases from the 0 d lag and continues to increase till a time lag of 35 d. The higher than average increase in RSV of all other terms leads to a higher than average daily rise in COVID-19 cases about 0–35 d later. The strength of association here decreases with the
increase in time lag which signifies that the higher than average RSV predicting higher than the average daily rise in COVID-19 cases on the same day of increase in RSV (lag 0) with the greatest association. Similarly, the higher than the average daily rise in COVID-19 cases leads to a higher than average increase in nutritional RSV after
14–27 d with the greatest association (Figs. 5 and 7). In the present study, the strength of association of RSV with daily new cases was higher for positive time lags and lower for negative for time lags (Fig. 7).

State and region wise distribution of RSV
Mapping of mean RSV across Indian states, UTs, and geographical regions show a differential result for overall COVID-19 and nutritional RSV of selected terms. For overall RSV, the highest mean RSV was from Sikkim (74·38 ± 17·71), followed by Andaman and Nicobar (71·75 ± 42·27). The lowest mean RSV was from Madhya Pradesh (22·63 ± 5·98), followed by Bihar (23·75 ± 7·28) (Fig. 8(a)). Amongst geographical regions, the highest mean RSV was from North-Eastern India (51·87 ± 11·61) and lowest from Western India (52·31 ± 6·43) (Fig. 8(b)) (see online Supplemental File 4).

The states and UTs wise highest mean RSV was from Andaman and Nicobar (91·33 ± 12·26) followed by Goa (90·00 ± 11·43), and the lowest mean RSV was from Madhya Pradesh (24·00 ± 6·98) (Fig. 9(a)). The mean RSV of COVID-19 group terms also show the highest and lowest search volume from North-Eastern (61·45 ± 13·96) and Western India (36·00 ± 6·67). However, the second highest mean RSV was observed from North India (52·10 ± 9·50) in the present case (Fig. 9(b)) (see online Supplemental File 4).

The mean RSV of nutritional terms was highest from Sikkim (66·40 ± 17·49), followed by Andaman and Nicobar (60·00 ± 48·99), and the lowest was from Bihar (21·60 ± 7·28) (Fig. 10(a)). Amongst geographical regions, the highest mean RSV was from South India (49·34 ± 7·43), followed by North India (47·57 ± 8·15), and the lowest was from West India (31·10 ± 6·30) (Fig. 10(b)) (see online Supplemental File 4).
Discussions

Google Trends has been widely used in the surveillance of epidemic diseases\(^{(32)}\), the impact of public health days\(^{(33)}\) and behavioural studies\(^{(34)}\). During pandemic situations and natural disasters, the online surveillance system can be the best feasible approach to monitoring population health in real-time. Sudden outbreaks or panic situations...
Fig. 8 (colour online) Map of India showing the (a) overall mean RSV across different Indian states and UTs and (b) overall mean RSV across different mainland Indian regions.

Fig. 9 (colour online) Map of India showing the (a) mean RSV from COVID-19-related terms across different Indian states and UTs and (b) mean RSV from COVID-19-related terms across different mainland Indian regions.
make people gather information for either cure or prevention. Information on Internet searches on the most popular search engine is the best tool for disease surveillance and epidemic(35).

Online or hospital-based surveys were often used to collect nutritional data from infected or normal people(36,37). Several studies predicted the COVID-19 outbreak in different parts of the world using GTs data(38,39). Google Trends provide us with the reliability of data and coverage of a huge population, making it a valuable tool to access nutritional behaviour and public health. The present study shows a spike in Google web searches near the announcement day by national (NIN, AYSUH, FSSAI) or international bodies (WHO, FAO, UNICEF). It signifies an increased interest in seeking nutritional immunity information through Internet searches. The regression analysis has shown the monthly percentage change in searches, and in the context of nutritional terms, a 9–26% rise in the search was observed from January to August. This range of MPC can be stated as good in the Indian context because several other sources of information (newspapers, television, social network sites, doctors, etc.(40,41)) are available through which people can get similar knowledge. Due to initial panic situations and fear(42), both symptomatic and asymptomatic people were aware(43) of cure or prevention from COVID-19. The public health and nutritional guidelines also influenced their searches which were evident from current results.

The daily new confirmed cases in India were found to be moderately correlated with terms (Immunity), (Vaccine), (Vitamin), (Nutrient), (Multivitamin), and (Zinc). Whereas, the term (Chyavanprash) was found to be strongly correlated with daily new confirmed cases. It confirms that the regional language-based searches were found to be more influential than the general or technical terms. After 14–27 d of the greater than average daily increase in COVID-19 cases, the people showed interest in all selected nutritional terms, which declined afterwards. It shows that the Indian population was interested in seeking nutritional information for prevention from COVID-19. The interest on the day of increase in cases was also moderately correlated for nutritional terms.

India has twenty-eight states, eight UTs divided into six different regions – East, West, North, South, Central, and North-East. The highest relative search volume from Sikkim (second smallest state) might be due to its highest total teledensity across all Indian states(44), resulting in many Internet users. According to regional mapping, the North-East region shows the highest relative searches for overall terms and COVID-19-related terms. At the beginning of September, North-East India accounts only for ~5% of total COVID-19 cases. A study reported that a high frequency of mutated rs2285666 allele (ACE2 gene) might accounts for fewer COVID-19 cases in the North-East region(45). But between September and October 2020, the cases got doubled in the region(46). Initial high awareness (January–July) evident in current results might be preventing the North-East region from COVID-19. However, this statement requires verification through behavioural

![Map of India showing the (a) mean RSV from nutritional terms across different Indian states and UTs and (b) mean RSV from nutritional terms across different mainland Indian regions](https://doi.org/10.1017/S1368980021003232 Published online by Cambridge University Press)
intervention studies. The highest volume in nutritional terms was from the southern region, and the lowest was from the western region. The factors influencing GTs relative search volumes could be literacy rate, socioeconomic status (SES), Internet penetrance, and information-seeking behaviour (interest) which varies widely across all geographical regions of India. In the present study, the regional comparative analyses seemed to be influenced by more than one factor, and updated data on each factor was required to grade behavioural factors according to the searches. Despite several benefits of GTs, there are still certain limitations like no defined consensus of reporting, aspects of data collection and methodological development. In both developing and developed countries, the RSV of Google Trends are primarily dependent on events (like a disaster, pandemic, national or international days), but during the complete absence of events, GTs may not accurately represent the RSV in developing countries because other factors like socioeconomic status (SES) and Internet penetrance might be influencing the RSV over time. However, this needs to be validated by a comparative study between event and non-event-based GTs data from developing and developed countries. The present study was conducted on an event (COVID-19)-based data of India (developing country), therefore, the validity and reliability of the results persist. Further enhancement of GTs results can be done using advanced programming languages like Python or Java. Automation using programming languages can help in real-time monitoring of multiple search terms and statistical analysis can be possibly embedded for a better outlook of results.

Conclusion

To improve immunity during COVID-19, the Indian population has accessed the nutritional information via performing Google searches. The search spikes were in accordance with the announcement of nutritional guidelines and daily increasing COVID-19 cases. The monthly variations in RSV in the studied time frame show increasing interest in Google searches for selected nutritional terms. This increased search can be interpreted as information-seeking behaviour depended on both awareness and fear due to COVID-19, which increases after the daily rise in COVID-19 cases. The local/regional search terms can be used in public health guidelines or recommendations for better outreach. Variation of searches across regions shows the influence of additional factors on Google searches which need further evaluation. It is difficult to access the nutritional behaviour of the larger population in the pandemic, but online surveillance using GTs data can help nutritional researchers, public health workers and government authorities to monitor the implementation of strategies or guidelines during the initial phases amongst a large population. The use of computer programming techniques in online surveillance can be utilised for real-time monitoring of desirable terms and planning event-based goals.

Acknowledgements

Acknowledgements: The author(s) would like to thank Google Inc. for making the Google Trends website freely available in the public domain and also to the University of Oxford for providing country-wise daily COVID-19 reports on the website ourworldindata.org. Financial support: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. Conflict of interest: There are no conflicts of interest. Authorship: Study design, concept, and selection of terms: S. and P.K. Data extraction, data management, and data cleaning: R.J. and R.S. Data analysis and interpretation: R.S. and S.K. Preparation of figures, maps, and supplementary files: S.K. and R.J. Literature review of the study: R.J. Writing of the manuscript: R.S. and S.K. Critical review of manuscript: P.K. Ethics of human subject participation: Not applicable.

Supplementary material

For supplementary material accompanying this paper visit https://doi.org/10.1017/S1368980021003232

References

Nutritional surveillance during COVID-19