



OPEN

Spatio-temporal epidemiology and associated indicators of COVID-19 (wave-I and II) in India

Karuppusamy Balasubramani^{1,11}, Venkatesh Ravichandran^{2,11}, Kumar Arun Prasad^{1,11}, Mu. Ramkumar³, Sulochana Shekhar¹, Meenu Mariya James⁴, Naveen Kumar Kodali⁴, Sujit Kumar Behera⁴, Natarajan Gopalan⁴, Rakesh Kumar Sharma⁵, Devojit Kumar Sarma⁶, M. Santosh^{7,8}, Aditya Prasad Dash⁹ & Praveen Balabaskaran Nina¹⁰✉

The spatio-temporal distribution of COVID-19 across India's states and union territories is not uniform, and the reasons for the heterogeneous spread are unclear. Identifying the space–time trends and underlying indicators influencing COVID-19 epidemiology at micro-administrative units (districts) will help guide public health strategies. The district-wise daily COVID-19 data of cases and deaths from February 2020 to August 2021 (COVID-19 waves-I and II) for the entire country were downloaded and curated from public databases. The COVID-19 data normalized with the projected population (2020) and used for space–time trend analysis shows the states/districts in southern India are the worst hit. Coastal districts and districts adjoining large urban regions of Mumbai, Chennai, Bengaluru, Goa, and New Delhi experienced > 50,001 cases per million population. Negative binomial regression analysis with 21 independent variables (identified through multicollinearity analysis, with VIF < 10) covering demography, socio-economic status, environment, and health was carried out for wave-I, wave-II, and total (wave-I and wave-II) cases and deaths. It shows wealth index, derived from household amenities datasets, has a high positive risk ratio (RR) with COVID-19 cases (RR: 3.577; 95% CI: 2.062–6.205) and deaths (RR: 2.477; 95% CI: 1.361–4.506) across the districts. Furthermore, socio-economic factors such as literacy rate, health services, other workers' rate, alcohol use in men, tobacco use in women, overweight/obese women, and rainfall have a positive RR and are significantly associated with COVID-19 cases/deaths at the district level. These positively associated variables are highly interconnected in COVID-19 hotspot districts. Among these, the wealth index, literacy rate, and health services, the key indices of socio-economic development within a state, are some of the significant indicators associated with COVID-19 epidemiology in India. The identification of district-level space–time trends and indicators associated with COVID-19 would help policymakers devise strategies and guidelines during public health emergencies.

The ongoing coronavirus disease 2019 (COVID-19) pandemic that started in December 2019 has devastated millions of families by affecting physical, mental and economic health. India's first COVID-19 wave peaked in September 2020, and the cases gradually declined until February 2021. The first COVID-19 wave in India was relatively mild, especially due to the nationwide lockdown. However, the second COVID-19 wave (February–December, 2021) was devastating, with widespread mortality and morbidity. During the course of the second wave, the Delta variant (B.1.1617.2) of SARS-CoV-2, initially identified in the state of Maharashtra in late 2020,

¹Department of Geography, School of Earth Sciences, Central University of Tamil Nadu, Thiruvarur 610005, India. ²Department of Civil Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, India. ³Department of Geology, Periyar University, Salem, India. ⁴Department of Epidemiology and Public Health, School of Life Sciences, Central University of Tamil Nadu, Thiruvarur 610005, India. ⁵Shree Guru Gobind Singh Tricentenary University, Gurugram, New-Delhi-NCR 122505, India. ⁶ICMR- National Institute for Research in Environmental Health, Bhopal Bypass Road, Bhopal, Madhya Pradesh, India. ⁷School of Earth Sciences and Resources, China University of Geosciences, Beijing, People's Republic of China. ⁸Department of Earth Sciences, University of Adelaide, Adelaide, SA, Australia. ⁹Asian Institute of Public Health University, Phulnakhara, Cuttack, Odisha 754001, India. ¹⁰Department of Public Health and Community Medicine, Central University of Kerala, Kasaragod, Kerala 671316, India. ¹¹These authors contributed equally: Karuppusamy Balasubramani, Venkatesh Ravichandran and Kumar Arun Prasad. ✉email: praveen.nina@cukerala.ac.in

outcompeted the pre-existing variants Kappa (B.117.1), and Alpha (B.1.1.7), and became the dominant variant¹. The third COVID-19 wave in India started in late December 2021 and has been driven by the Omicron variant (B.1.1.529). India is the most affected country, second only to the USA in the number of cases, and third behind the USA and Brazil in deaths². As of July 2022, India has witnessed > 42 million cases and > 0.5 million deaths due to COVID-19. The states of Maharashtra and Kerala rank number one and two in the total cumulative number of COVID-19 cases (7,800,000 and 6,500,000, respectively) and deaths (150,000 and 70,000, respectively)³.

Socio-economic factors, race, human behavior, pollution levels, exposure density, urban–rural differences, access to medical facilities, pre-existing non-communicable disease conditions, and environmental factors are some of the major influential COVID-19 explanatory variables identified worldwide⁴. Among the 28 states and eight union territories (UT) in India, there are significant geographical differences that influence COVID-19 epidemiology. Literacy levels, state's economy, urban percent, 15–59 age group population, population density, the testing ratio at the district level, non-communicable diseases (NCDs), nutritional status, meteorological variables, and PM_{2.5} have been found to be positively associated with COVID-19 burden in India^{5–8}. These studies have used partial state/district-wise COVID-19 datasets prior to the completion of the second wave and have not included all the possible explanatory variables that could affect COVID-19. Furthermore, little is known about the effect of the wealth index, current NCD rates, and worker population rates on COVID-19 cases and deaths at the district level across India.

Given the background, this study is aimed to spatio-temporally map the spread of COVID-19 in wave-I and wave-II at the district level across India for the identification of COVID-19 hotspots. Furthermore, another major objective is to identify the environmental, demographic, socio-economic, and health factors associated with district-level COVID-19 rates to help guide public health strategies.

Methods

Study design

This is a cross-sectional study that compares COVID-19 cases/deaths and their associated indicators at different geographical locations in India at a defined time interval, with individual districts as the unit of measurement. The data on COVID-19 cases and deaths over the course of COVID-19 (waves-I and II) and important geographical variables at the district level were used for spatial and statistical analysis.

Spatial data preparation

Even though the COVID-19 wave-II in India lasted till December 2021, only the period from February 2020 to August 2021 (wave-I and wave-II) was considered because the publicly available COVID-19-related datasets from <https://data.covid19india.org/> and <https://www.mohfw.gov.in/> were not regularly updated since September 2021. The datasets were further segregated based on wave-I (April 2020 to January 2021) and wave-II (February 2021 to August 2021). The administrative unit (i.e., district) of the 2011 Census of India was considered for uniform data assimilation. The data of newer districts (demarcated after the 2011 Census) were combined with their parent district. The districts of megacities such as Mumbai and New Delhi were clubbed into a single unit to avoid inconsistencies in comparing COVID-19 cases/deaths with other variables. The COVID-19 incidence and mortality data were normalized according to the projected population (2020)⁹ and used for space–time trend analysis.

Given India's diversified geography and socio-economic status^{10,11}, 28 independent variables related to the environment, demography, socio-economic, and health aspects were selected to study their association with COVID-19 cases and deaths. The source of the environmental, demographic, socio-economic, and health datasets are listed in Table 1. The 2011 census data (the latest available) was downloaded from the Census of India portal to compute demographic, household, wealth, literacy, occupation, and infrastructure characteristics at the district level. The key health variables, such as anemia, blood sugar level, blood pressure, tobacco and alcohol consumption, and obesity were obtained from the National Family Health Survey-5 (NFHS-5), carried out from 2019 to 2021.

The TERRACLIMATE datasets with ~ 4 km and ASTER datasets with 30-m spatial resolutions¹² were used to extract elevation and climatic parameters (except rainfall) using the Google Earth Engine (GEE) platform. The PERSIANN-Cloud Classification System (PERSIANN-CCS) datasets (<https://chrsdata.eng.uci.edu/>) and Forest Survey of India (FSI) 2019 annual reports were used respectively for rainfall¹³ and forest data¹⁴. The particulate matter 2.5 (PM_{2.5}) concentration was obtained from the combined measurement of satellite (NASA MODIS, MISR, and SeaWiFS) derived Aerosol Optical Depth (AOD) and ground-derived PM_{2.5} datasets¹⁵. All these environmental datasets were aggregated at the district level to extract a mean value for each district. In order to remove the seasonal effects and short-term changes, an annual average value of 2018–2021 is considered for the climatic datasets.

As there are no readily available datasets at the district level for assessing wealth and health services, these two variables were determined based on the parameters extracted from the district census handbooks¹⁰. The health services variable was determined based on the availability of all kinds of hospital beds in a district. The methodology for deriving the wealth index is provided in Supplementary File 1.

Space–time trend analysis

The emerging hotspot analysis is a geospatial technique used to understand space–time dynamics through the identification of hot and coldspots of any spatial event¹⁶. The COVID-19 datasets were prepared as population (2020) normalized 15 days sum of cases and deaths (Supplementary Files 2, 3). The data were then converted into a space–time cube using ArcGIS Pro software (version 2.8.0; year 2021), and a total of 34-time steps with 21,760 observations/bins (640 spatial units multiplied by 34 temporal intervals) were generated, and each bin

Category	Variable	Description of the variable	Period	Source
COVID-19 (dependent variables)**	Wave-I cases	Wave-I cases (April 2020 to January 2021)	2020–2021	https://data.covid19india.org/ https://www.mohfw.gov.in/
	Wave-I deaths	Wave-I deaths (April 2020 to January 2021)		
	Wave-II cases	Wave-II cases (February 2021 to August 2021)		
	Wave-II deaths	Wave-II deaths (February 2021 to August 2021)		
	Total cases	Total cases recorded (February 2020 to August 2021)		
	Total deaths	Total deaths recorded (February 2020 to August 2021)		
Demographic and socio-economic	Population	Projected population	2020	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/RXYJR6
	Population-2011	Census population	2011	https://censusindia.gov.in/census.website
	Household density	No. of households/sq. km area		
	Population density	No. of persons/sq. km area		
	SC/ST population	Scheduled Castes/Tribes population density (no. of persons/sq. km area)		
	Literacy	Literacy rate (%)		
	Working population	Total working population rate (%)		
	Cultivators	Cultivator population rate (%)		
	Agricultural laborers	Agricultural laborers' rate (%)		
	Household industries workers	Household industries workers' rate (%)		
	Other workers	Other workers' rate (%)		
	Wealth index	Wealth index (0–1) is computed based on household indicators (Refer Supplementary File 1)		
Health services	No. of beds in all kind of hospitals			
Environmental	Forest	Forest cover (%)	2019	https://fsi.nic.in/forest-report-2019
	Elevation	Average elevation (meters above mean sea level)	2007	https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003
	Max temp	Annual average maximum temperature (°C)	2018–2021	https://developers.google.com/earth-engine/datasets/catalog/IDAHO_EPSCOR_TERRA_CLIMATE
	Min temp	Annual average minimum temperature (°C)		
	Windspeed	Annual average windspeed (km/h)		
	AET	Annual average actual evapotranspiration (mm)		
	Rainfall	Annual average rainfall (mm)		
PM _{2.5}	Annual average particulate matter 2.5 (µg/m ³)	https://chrsdata.eng.uci.edu/ https://sites.wustl.edu/acag/datasets/surface-pm2-5/		
Health	High blood glucose	Blood sugar level > 140 mg/dl or taking medicine to control blood sugar level (%)	2019–2021	http://rchiips.org/nfhs/
	High blood pressure	Elevated blood pressure (systolic ≥ 140 mmHg and/or diastolic ≥ 90 mmHg) or taking medicine to control blood pressure (%)		
	Tobacco (women)	Women age 15 years and above who use any kind of tobacco (%)		
	Tobacco (men)	Men age 15 years and above who use any kind of tobacco (%)		
	Alcohol (women)	Women age 15 years and above who consume alcohol (%)		
	Alcohol (men)	Men age 15 years and above who consume alcohol (%)		
	Obese (women)	Women who are overweight or obese (BMI ≥ 25.0 kg/m ²) (%)		
	Waist-to-hip ratio	Women who have high-risk waist-to-hip ratio (≥ 0.85) (%)		
	Anemia women	Women age 15–49 years who are anemic (%)		

Table 1. Dependent and independent variables used for spatial and statistical analysis, and their data sources.

**Since the publicly available COVID-19 related domains such as <https://data.covid19india.org/> and <https://www.mohfw.gov.in/> were not regularly updated since September 2021, the period from February 2020 to August 2021 was considered.

was considered as an observation. The output was distributed into hotspots (highest significant value over the region) and coldspots (lowest significant value over the region)¹⁷. Theoretically, the emerging hotspot analysis was computed through the Getis-Ord G_i^* formula given in Eqs. (1) and (2).

$$G_i^* = \frac{\sum_j^n w_{ij} x_j - \bar{X} \sum_j^n w_{ij}}{\sqrt{\frac{n \sum_{j=1}^n w_{ij}^2 - (\sum_{j=1}^n w_{ij})^2}{n-1}}}, \quad (1)$$

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n}, \quad S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2}, \quad (2)$$

where x_j is the attribute value for spatial feature j , w_{ij} is the spatial weight between features i and j , and n is the total number of features^{18,19}.

At each iteration, a space–time bin was generated (containing z-score and p-value) and used to determine the hotspots using the Mann–Kendall trend test^{17,20}. The metadata of space–time analysis is provided in Supplementary File 4. The hotspot bins were classified into six classes: consecutive (uninterrupted run of hot time step intervals), intensifying (at least 90% of the bins becoming hotter over time), persistent (at least 90% of the bins are hot but with no trend up or down), diminishing (at least 90% of the bins becoming less hot over time), sporadic (some of the time step intervals are hot) and oscillating (some of the time step intervals are hot, some are cold)²¹.

Negative binomial regression analysis

The linear regression model is frequently applied to count variables under the assumption that they are continuous normally distributed, but this can lead to inaccurate, inconsistent, and biased results. The Poisson Regression Model (PRM) specifically addresses count outcomes; the probability distribution of a count in the PRM is defined by a Poisson process, where the distribution's mean is a function of the independent variables. The variance and mean of the outcome are assumed equal (equal dispersion), but most often, the variance exceeds the mean (over-dispersion). To overcome this issue, the Negative Binomial Regression Model (NBRM) is used, which allows the variance to be greater than the mean^{22,23} and hence considered the best-fitting model for count data²⁴. In this case too, the six dependent variables (COVID-19 cases/deaths: Total, wave-I, and wave-II) are non-negative count data, and the variance is greater than the mean for each dependent variable. Hence, the NBRM statistical test is used to adjust for over-dispersion in the outcome variables and to assess the association between the geographic variables and COVID-19 cases and deaths (wave-I, wave-II, and total) in India at the district level from February 2020 to August 2021 (Supplementary File 5). Based on exploratory regression (spatial) analysis, all overlapping variables were removed through the iterative steps, and 21 independent variables from demographic (household density, literacy rate), socio-economic (agricultural laborers' rate, household industries workers' rate, other workers' rate, wealth index, and health services), environment (forest, minimum temperature, wind speed, actual evapotranspiration, rainfall, and $PM_{2.5}$), and health (high blood glucose, high blood pressure, tobacco women, tobacco men, alcohol women, alcohol men, anemia women, and overweight/obese women) were used for the regression models as listed in Supplementary File 6. All the independent variables were assessed with the cases and deaths for each of the waves (wave-I, wave-II, and total), with the 2020 projected population of each district as an offset variable. The independent variables with $P < 0.2$ identified through univariable analyses were used to run the multivariable analyses (full model). To obtain a reduced model through iterative multivariable fitting, the variables were removed one at a time as they were not significant in the multivariable model at the alpha level of 0.10 and, when taken out, did not change any of the remaining parameter estimates by $> 20\%$. The likelihood ratio test (Supplementary File 7) shows the reduced model to be as good as the full model in fitting the data; the reduced model was chosen for the principle of parsimony (Supplementary File 7). To ensure linearity, a few independent variables were transformed (log of household density, square root of agriculture, square root of alcohol women, log of forest, and square root of health services), and used for the analysis. Multicollinearity was assessed between independent variables, and variables with variance inflation factor (VIF) greater than 10 were eliminated. The risk ratios for COVID-19 cases (wave-I, wave-II, and total) and deaths (wave-I, wave-II, and total) were determined separately using six regression models. Each model's risk ratio (RR) value is reported with its 95% confidence interval and P-value. The results of each model, along with residual plots for reduced models, are presented in Supplementary File 7. The NBRM statistical analysis and exploratory regression (spatial) analysis were carried out in STATA software (Version 17, year 2021) and ArcGIS Pro software (2.8.0), respectively.

Spatial mapping

The district-wise spatio-temporal trends in COVID-19 cases and deaths were mapped using choropleth mapping techniques. Among the 21 variables used for statistical analysis, only nine (wealth index, health services, literacy rate, household density, other workers' rate, blood sugar level, forest cover, annual rainfall, and $PM_{2.5}$) were mapped based on the results from negative binomial regression analysis and spatial exploratory regression analysis of COVID-19 cases/deaths. All the maps were created using ArcGIS 10.4 software (year 2016) with online base map services²⁵.

Results

Spatio-temporal pattern of COVID-19 cases and deaths

The district-wise data of COVID-19 cases and deaths, from February 2020 to August 2021, were aggregated and used to generate the 15 days cumulative sum for the entire country. As 15 days cumulative cases and deaths were negligible until 31st March 2020, data from 01st April 2020 to 31st August 2021 were analyzed as shown in Fig. 1. During the first wave, the maximum number of cases (1,356,052) and deaths (16,787) were recorded between 01st and 15th September 2020. From then, the cases started decreasing gradually and were at their lowest in February 2021. In the second wave, the cases and deaths reached the peak in the second week of May 2021 (5,522,829 cases and 61,625 deaths were recorded between 01st and 15th May 2021).

The spatial distributions of the district-wise COVID-19 cases and deaths normalized per million population for wave-I and wave-II are shown in Fig. 2, and the cumulative sums for the same periods are presented in

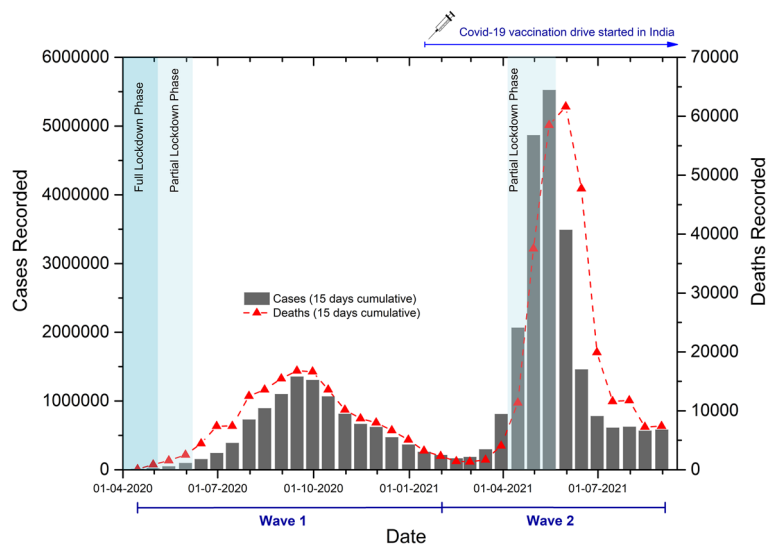


Figure 1. The cumulative COVID-19 cases and deaths in India from April 2020 to August 2021. The timeline for waves I and II is represented on the x-axis. The major lockdown periods are shaded and labeled.

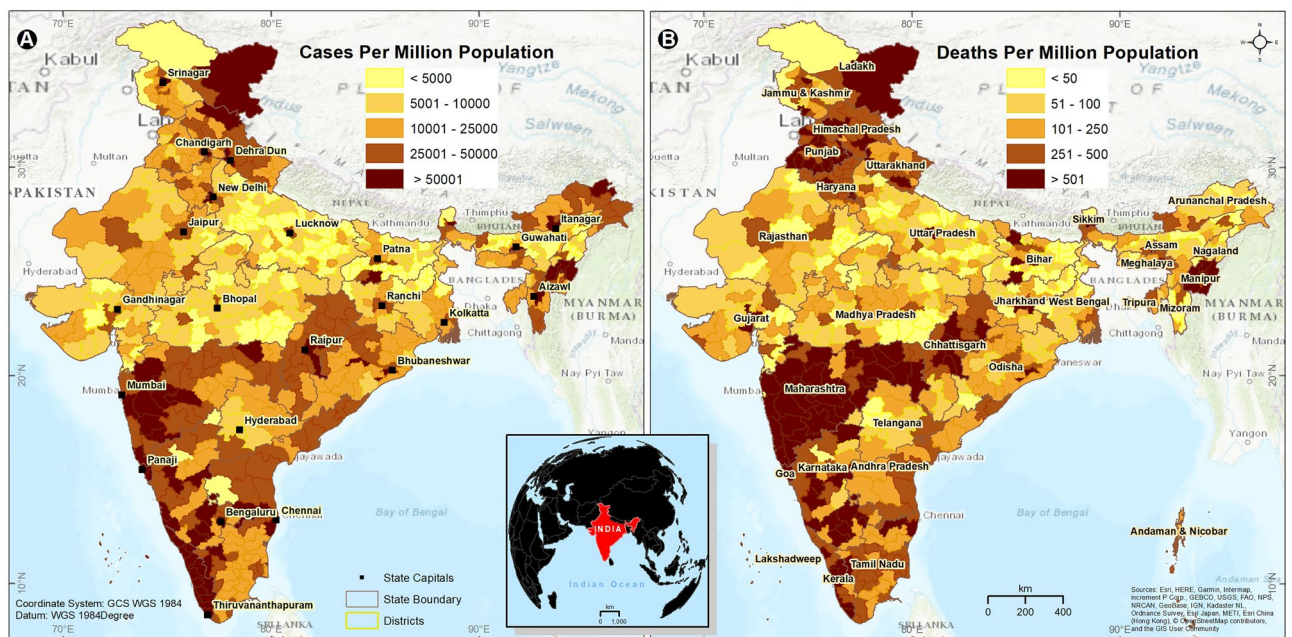


Figure 2. District-wise (A) COVID-19 cases per million population and (B) deaths per million population in India from April 2020 to August 2021. The inset map shows the location of India in Asia. The map was created using the licensed version of ArcGIS 10.4 software by Esri (www.esri.com). The background map represents the World Topographic Map (http://goto.arcgisonline.com/maps/World_Topo_Map).

Supplementary Fig. 8 (S8.1, S8.2, S8.3, S8.4). The state of Kerala is the worst hit, and is closely followed by the neighboring urban districts of Mumbai in Maharashtra. Large urban regions, such as Chennai, Bengaluru, Goa, New Delhi, and their adjoining districts, also reported very high caseloads ($> 50,001$ per million). Also, the urban districts in the North-Eastern (NE) states of Assam, Sikkim, Manipur, Mizoram, and Arunachal Pradesh reported high number of cases. Interestingly, the cases per million population are high even in the less densely populated districts of Ladakh and Dibang Valley in Arunachal Pradesh, and Lahaul and Spiti in Himachal Pradesh. On the other hand, the caseload was less (< 5000 per million) in many districts of Uttar Pradesh, Madhya Pradesh, and Bihar, states with high population density.

The spatial distribution of the deaths (> 501 per million) shows the state of Maharashtra to be the worst affected. The densely populated urban districts in the states of Kerala, Karnataka, Tamil Nadu, Chhattisgarh, Odisha, Sikkim, Uttarakhand, Jammu and Kashmir, Himachal Pradesh, Bihar, Punjab, Haryana, Uttar Pradesh, and Gujarat also reported > 501 deaths per million population. Among the NE states, Manipur reported > 501 deaths per million population. COVID-19 deaths were low (< 50 to 250 per million) in many districts of Rajasthan, Madhya Pradesh, Andhra Pradesh, Telangana, Odisha, West Bengal, Bihar, Gujarat, and Jharkhand.

Spatial pattern of key geographical variables

The district-level spatial distributions of the key geographical variables: wealth index, literacy rate, health services, household density, other workers' rate, and high blood glucose are shown in Fig. 3. A strong spatial overlap of these indicators with high COVID-19 cases and deaths is discernible in the states of Kerala and Maharashtra.

In addition, environmental factors (forest cover, annual rainfall, minimum temperature, and $PM_{2.5}$ concentrations) mapped at the district level are shown in Supplementary Fig. 9. Forest cover, average annual rainfall, and average minimum temperature are positively correlated with COVID-19 cases, especially in Kerala and Maharashtra. The concentration of $PM_{2.5}$ is negatively associated with COVID-19; the states of Haryana, Uttar Pradesh, and Bihar have high $PM_{2.5}$ but fewer COVID-19 cases and deaths.

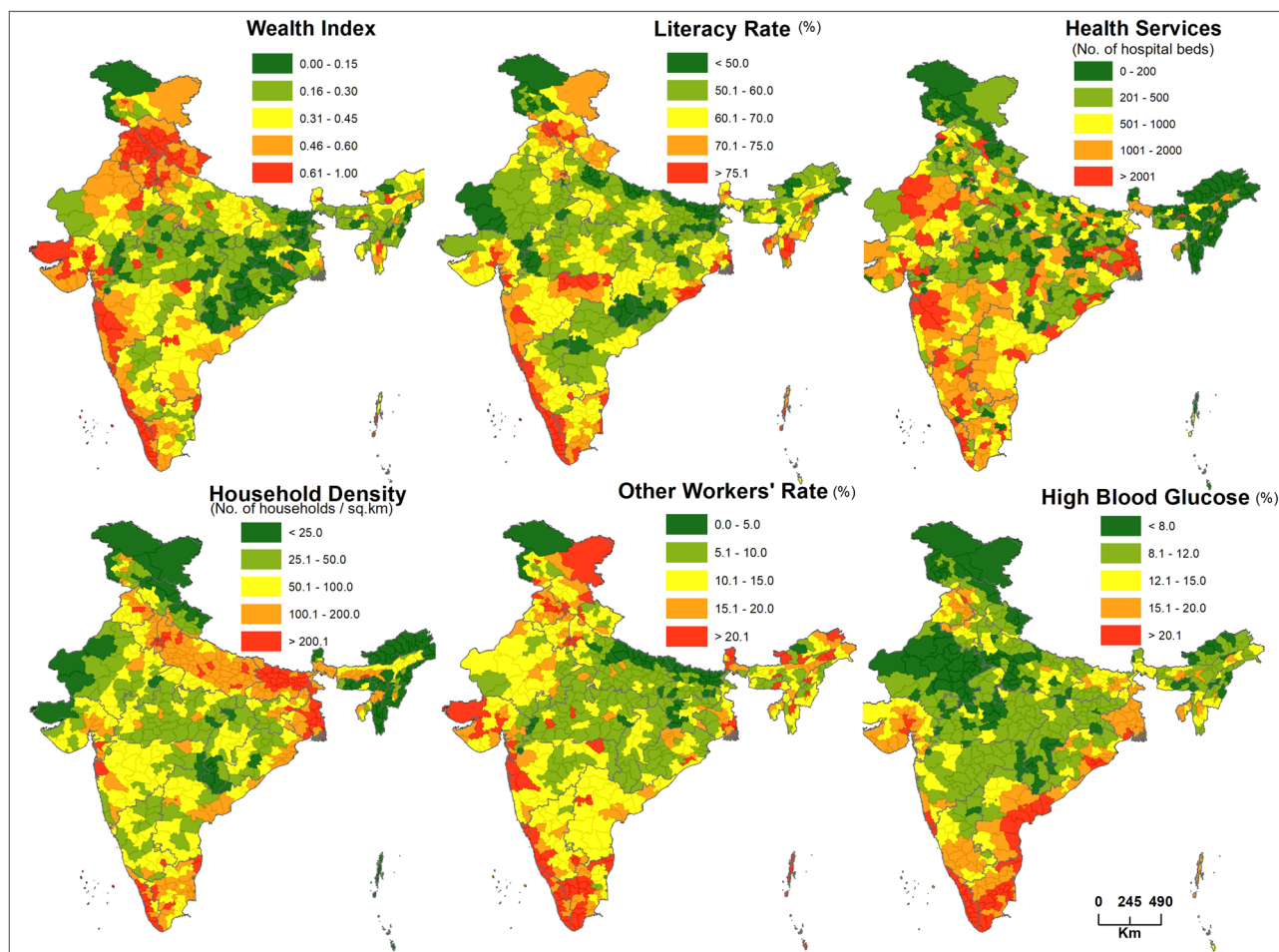


Figure 3. Spatial distributions of the influential socio-economic, demographic, and health indicators at the district level in India. The map was created using the licensed version of ArcGIS 10.4 software by Esri (www.esri.com).

Spatio-temporal trend of COVID-19 spread

The district-level space–time trends and hotspots identified by emerging hotspot analysis of population-normalized COVID-19 cases are shown in Fig. 4A,B. The trend analysis of cases over a period of 375 days (the best-fitted time period selected by the model) shows the entire state of Kerala, the western part of Tamil Nadu, districts surrounding Bengaluru, Mumbai and Hyderabad, northern districts of Chhattisgarh, most of the districts of Himachal Pradesh and NE states to have an upward trend of COVID-19 cases at 99% CI.

The entire wave's space–time hotspot analysis of the 640 census districts finds ~20% of the districts (123 districts) to have different categories of hotspots. A consecutive hotspot characteristic is observed in the entire state of Goa, 11 out of the 14 districts of Kerala, 12 districts of Karnataka, 8 districts of Maharashtra, 6 districts of Manipur, 5 districts of Tamil Nadu, and 3 districts of Himachal Pradesh. The districts of Maharashtra show an intensifying hotspot characteristic, while the New Delhi region shows a persistent hotspot characteristic. Diminishing hotspot characteristic is observed in Chennai and Mumbai suburban districts. A sporadic hotspot characteristic is observed mainly in the southern and NE states, while an oscillating hotspot trend is observed in Himachal Pradesh.

Association between the geographical variables and COVID-19

Negative binomial regression analysis results show the association between COVID-19 cases/deaths and the key demographic, socio-economic, environmental, and health-related factors, and are given in Table 2 and Supplementary Files S5, S6 and S7. Among the variables, the wealth index showed a high RR for COVID-19 cases and deaths (wave-I, wave-II, and total). The RR of COVID-19 cases increased by ~3 times for every unit increase in the wealth index in COVID-19 cases (total wave [RR: 3.577; 95% CI: 2.062–6.205]) and deaths (total wave [RR: 2.477; 95% CI: 1.361–4.506]). The literacy rate is the other significant variable that had a significant positive RR for COVID-19 cases (total wave [RR: 1.021; 95% CI: 1.011–1.032]) and deaths (total wave [RR: 1.040; 95% CI: 1.028–1.052]).

Other workers' rate, health services, tobacco use in women, alcohol use in men, overweight/obesity in women, anemia in women, and rainfall were the other positively significant factors observed in the negative binomial regression analysis. With every additional unit increase in alcohol use in men (total wave [RR: 1.018; 95% CI: 1.011–1.025]), tobacco use in women (wave-I [RR: 1.019; 95% CI: 1.008–1.031]), health services (total wave [RR: 1.005; 95% CI: 1.001–1.009]), and anemia in women (wave-I [RR: 1.007; 95% CI: 1.000–1.014]), the RR of COVID-19 cases increased. Similarly, in COVID-19 deaths, with every additional unit increase in the other workers' rate (wave-I [RR: 1.026; 95% CI: 1.001–1.051]), overweight/obesity in women (wave-I [RR: 1.020; 95% CI: 1.005–1.036]), tobacco use in women (total waves [RR: 1.015; 95% CI: 1.004–1.026]), health services (total waves [RR: 1.006; 95% CI: 1.001–1.010]), and rainfall (wave-I [RR: 1.000; 95% CI: 1.000–1.000]), the RR increased.

A negatively significant association was observed between COVID-19 cases/deaths and PM_{2.5} (total cases [RR: 0.987; 95% CI: 0.984–0.990]; total deaths [RR: 0.991; 95% CI: 0.987–0.994]), tobacco use in men (total cases [RR: 0.990; 95% CI: 0.982–0.998]; total deaths [RR: 0.974; 95% CI: 0.964–0.983]), and wind speed (total cases [RR:

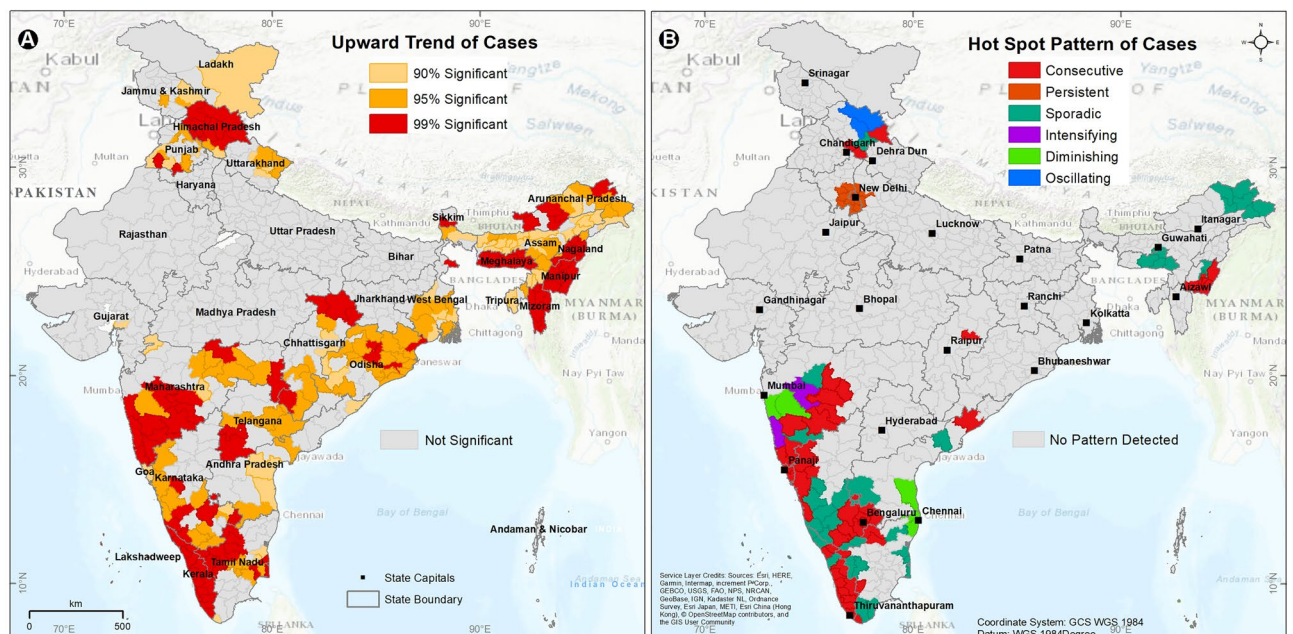


Figure 4. Distribution of (A) upward trend districts with different significance (P-value) levels and (B) distribution of different categories of hotspot districts identified by space–time pattern analysis for COVID-19 cases in India. The map was created using the licensed version of ArcGIS 10.4 software by Esri (www.esri.com). The background map represents the World Topographic Map (http://goto.arcgisonline.com/maps/World_Topo_Map).

Variables	Cases						Deaths					
	Wave-I		Wave-II		Total wave		Wave-I		Wave-II		Total wave	
	RR	P value	RR	P value	RR	P value	RR	P value	RR	P value	RR	P value
Wealth index	3.854	0.000*	3.35	0.000*	3.577	0.000*	3.731	0.001*	2.434	0.017*	2.477	0.003*
Literacy rate (%)	–	–	1.027	0.000*	1.021	0.000*	1.027	0.000*	1.037	0.000*	1.04	0.000*
Other workers' rate (%)	1.021	0.064	–	–	–	–	1.026	0.042*	–	–	–	–
Minimum temperature (°C)	–	–	–	–	–	–	0.998	0.871	–	–	–	–
Wind speed (km/h)	–	–	0.771	0.043*	0.779	0.046*	–	–	0.733	0.022*	0.78	0.054
Rainfall (mm)	–	–	–	–	–	–	1	0.049*	–	–	–	–
PM _{2.5} (µg/m ³)	0.991	0.000*	0.987	0.000*	0.987	0.000*	0.991	0.000*	0.99	0.003*	0.991	0.001*
Health services (%)	1.005	0.023*	1.004	0.051	1.005	0.021*	1.007	0.002*	1.005	0.025*	1.006	0.011*
Alcohol men (%)	1.014	0.000*	1.017	0.000*	1.018	0.000*	–	–	–	–	–	–
Alcohol women (%)	–	–	–	–	–	–	–	–	0.942	0.186	–	–
Tobacco men (%)	0.981	0.000*	0.99	0.017*	0.99	0.011*	0.984	0.016*	0.982	0.002*	0.974	0.000*
Tobacco women (%)	1.019	0.001*	–	–	–	–	1.017	0.006*	1.01	0.102	1.015	0.006*
High blood glucose (%)	–	–	–	–	–	–	0.933	0.000*	0.959	0.001*	0.963	0.001*
High blood pressure (%)	–	–	–	–	–	–	1.012	0.26	1.017	0.083	–	–
Overweight/Obesity women (%)	–	–	–	–	–	–	1.02	0.007*	–	–	–	–
Anemia women (%)	1.007	0.048*	–	–	–	–	1.006	0.178	–	–	–	–

Table 2. Estimated risk ratio (RR) for factors that significantly associated with COVID-19 cases and deaths, during wave-I and wave-II, India, February 2020 to August 2021. *Statistical significance at $P < 0.05$ “–” Variables not found significant by the model.

0.779; 95% CI: 0.609–0.995]; wave-II deaths [RR: 0.733; 95% CI: 0.562–0.956]). Similarly, a negative association was observed between COVID-19 deaths and high blood glucose (total waves [RR: 0.963; 95% CI: 0.942–0.984]).

Discussion

The COVID-19 caseload in the first and the second wave contrast themselves. The gradual increase and decrease during the first wave could be attributed to the early nationwide lockdown and its strict implementation, extensive testing, and other prevention and control strategies. However, the sudden peak and the steeper decline in the second wave point to the rapid spread of the highly transmissible Delta variant across India and the resultant natural immunity. The significant disparity observed in COVID-19 cases and deaths across the different Indian states is likely influenced by the varying physio-socio-economic-demographic factors in different parts of the country. A comprehensive analysis of the key environmental, demographic, socio-economic, and health variables shows wealth index, literacy rate, other workers' rate, health services, tobacco use in women, alcohol consumption in men, obesity/overweight in women, anemia in women, and rainfall to be positively associated with COVID-19 cases/deaths in India.

Wealth index and literacy rates are important indicators of COVID-19 epidemiology in this analysis. At the district level in India, a positive correlation ($r = 0.65$) was observed between literacy rate and wealth score²⁶. In India, the states with higher literacy rates have a relatively higher wealth index^{27,28}. Worldwide, several studies have reported a negative association between education/literacy rates and COVID-19 mortality rates^{29–31}. Literate people tend to seek diagnosis and treatment relatively earlier than people without formal education, as they are more aware of the signs and symptoms related to COVID-19³². RT-PCR of nasopharyngeal samples is the most widely used COVID-19 detection technique in India, and the COVID-19 test positivity decreases with delayed testing³³. In addition to reducing the true caseloads, false negative results also adversely affect contact tracing and other preventive and control measures. Wealth and literacy rates in a community greatly influence health-seeking behavior³⁴; this could possibly explain why the Indian states with higher wealth index and literacy rates experienced relatively high caseloads and deaths.

Even though NBRM shows a decreased risk for COVID-19 with an increase in blood glucose levels, the spatial exploratory regression analysis shows a strong overlap between elevated blood pressure/sugar levels and COVID-19 cases. NCDs are well-established risk factors of COVID-19 and are associated with increased severity of the disease^{35,36}. Diabetes and hypertension are strongly correlated with higher socio-demographic index³⁷. In a wealthy state like Kerala, which is a COVID-19 hotspot, most districts have 30–42% of people with elevated blood sugar levels (> 140 mg/dl). In contrast, in almost all the districts in states like Rajasthan, Bihar, and Jharkhand, where the wealth index and COVID-19 cases are relatively lower, less than 20% of the people have blood sugar levels > 140 mg/dl.

Overweight/obesity in women was also significantly associated with COVID-19 deaths in wave-I. A recent study has reported obesity to have a significant association with increased COVID-19 severity and mortality in India³⁸. Obesity is associated with hypertension, cardiovascular disease, and type 2 diabetes—well-documented comorbidities of COVID-19³⁹. Anemia has been reported to be a risk factor for COVID-19 cases and deaths^{40,41}; NBRM analysis shows a positive association between anemia and COVID-19 cases.

NBRM analysis shows a positive association between other (tertiary) workers' rate and COVID-19 deaths (wave-I). In India, there is a huge regional disparity in employment, especially in the industrial sector. Tamil Nadu ranks first in employment generation, while Kerala and Karnataka, the other southern states, rank among the top 10⁴². The relatively higher literacy rate and wealth index in these southern states might explain the increased tertiary employment in these states. The positive association between tertiary workers' and COVID-19 deaths suggests literacy rate and wealth index of the state to directly and indirectly affect COVID-19 deaths in India. The nature of employment in essential services, where one has to work in close proximity with others, commute to the workplace, and are continuously exposed to new contacts, makes these workers highly susceptible to COVID-19⁴³. Another reason for the high positivity rate in these workers could be because of the mandated testing and contact tracing policies in services/industrial sectors⁴⁴.

The household density and COVID-19 cases and deaths overlapped in the spatial analysis. Household density and overcrowded households are significant risk factors for COVID-19 incidence and mortality^{45,46}. In India, 37% of households have only a single room. Maharashtra, one of the hotspots, has an average household size of 4.6²⁷, and ~42% of households have only one room to dwell in⁴⁷. To reduce the airborne transmission (by droplets and aerosols) of COVID-19, the Center for Disease Control and Prevention recommends infected individuals maintain a six-foot distance between themselves and the other household members⁴⁸. It may be difficult or impossible for people to comply with this recommendation when they are in overcrowded houses⁴⁶.

The study shows tobacco smoking in women to be positively associated with COVID-19 cases and deaths. Smoking is a risk factor in increasing COVID-19 complications^{49,50}. Compared to non-smokers, smokers have a >1-time risk of increased COVID-19 severity⁵¹. In our analysis, a strong positive association was observed between tobacco use in women and COVID-19 wave-I cases/deaths. However, tobacco use in men showed a significantly decreased risk of COVID-19 cases and deaths. Hotspot analysis of tobacco consumption shows the central and eastern states of India to be the tobacco hotspots⁵², and in these states, the COVID-19 cases/deaths are relatively lower. As reported earlier⁵³, this study also finds a positive association between alcohol consumption in men and COVID-19 cases. Analysis of NFHS-4 shows the southern states, especially Kerala and Tamil Nadu, to be alcohol hotspots⁵⁴, and these are the states that have reported higher caseloads of COVID-19. The spatial mapping of COVID-19 cases shows districts in Kerala, Tamil Nadu, Telangana, Chhattisgarh, Odisha, and the NE states are following an upward trend with 99% significance; several districts in these states are hotspots of alcohol use in India⁵⁴. Kerala, followed by Maharashtra, and NE states were found to be the major COVID-19 hotspots in this analysis.

Analysis shows the health services to be positively associated with COVID-19 cases and deaths. In this pandemic, among the larger states, Kerala and Maharashtra were reported to have the best population-level health services⁵⁵. Several initiatives taken by these state governments significantly helped reduce the mortality rates, even though the incident case rates were substantially higher. Inexpensive and effective campaigns like "Break the Chain"⁵⁶ by the Kerala State Government⁵⁷, the "My family, my responsibility"⁵⁸ campaign by the Maharashtra State Government, and other similar campaigns disseminated health-related COVID-19 information to the common public. Rigorous contact tracing, testing, and home quarantine, supervised by local health authorities and the police department, helped identify more cases in Kerala⁵⁹. A similar innovative model formulated by the Municipal Corporation of Greater Mumbai helped identify more incident cases in Mumbai and Dharavi, in Maharashtra⁶⁰. The approach of "chasing the virus" with its 4Ts: (i) Tracing, (ii) Tracking, (iii) Testing, and (iv) Treating in high-risk slum clusters also served the same purpose⁶¹. The significant relationship between the number of tests, cases, and death rates⁵⁹ might explain why these states reported a high number of incident cases across the two waves.

In this study, environmental variables did not play a significant role in COVID-19 epidemiology, which is in line with similar findings from India^{62,63}. However, in the NBRM analysis, wind speed is negatively associated with COVID-19 cases and deaths. Multiple studies have reported wind speed to play a key role in COVID-19 transmission⁶⁴⁻⁶⁶. As the virus can remain active in the air for several hours, low wind speed makes the virus stay longer and concentrated in the atmosphere, which leads to increased COVID-19 cases in that area⁶⁴. Furthermore, rainfall shows a positive association with COVID-19 (wave-I deaths) in the NBRM analysis. The western slope of the Western Ghats and the NE states are among the regions receiving maximum rainfall from both monsoon seasons in India⁶⁷. The districts along the western coast (Kerala and Maharashtra) and the NE states are COVID-19 hotspots, which could explain why rainfall is positively associated with COVID-19.

Globally, exposure to ambient air pollution, especially PM_{2.5}, is widely reported as a significant risk factor in COVID-19 incidence and severity⁶⁸. A few Indian studies have also reported parallel findings linking COVID-19 and PM_{2.5}^{62,69}. We observed a relatively high PM_{2.5} (>76 µg/m³) in the highly populated great plains of India, especially in Uttar Pradesh, Bihar, and West Bengal states that lag behind in terms of wealth index²⁸, health services¹⁰, and literacy rates⁷⁰. However, the results of the regression analyses show PM_{2.5} is negatively associated with cases and deaths in waves-I and II. Hesitancy and/or lack of awareness for COVID-19 testing due to lower literacy rates in the Great Plains might explain the lack of association between COVID-19 and PM_{2.5}.

As there are no readily available datasets at the micro-administrative (district) level for assessing wealth and health services, the census data (2011) was used for computing these indices, which could be a limitation of the present study. Other limitations relate to the nature of the ecological study, which does not allow conclusions about causality, as it does not guarantee temporal relationship, and does not meet other criteria for causality. Additionally, the study is limited by the lack of temporal estimates between some independent variables (socio-economic-demographic) and the dependent variables.

Conclusion

The wealth index shows a strong positive RR for COVID-19 cases and deaths. The literacy rate, other workers' rate, alcohol consumption (men), anemia in women, health services, and rainfall are the other positively associated indicators of COVID-19 morbidity and mortality in India. These associated indicators identified through extensive spatial and statistical analysis might help guide micro-level public health emergencies in India. The space–time trends and mapping of spatial differences of COVID-19 indicators across different districts of India can assist in devising region-specific public health policies.

Data availability

The raw COVID-19 datasets used for analysis were obtained from public databases: <https://data.covid19india.org/> and <https://www.mohfw.gov.in/>. The district-wise COVID-19 cases, deaths and independent variables are provided in Supplementary Files 2, 3 and 5, respectively.

Received: 13 August 2022; Accepted: 19 December 2023

Published online: 02 January 2024

References

- Mlcochova, P. *et al.* SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* **599**, 114–119. <https://doi.org/10.1038/s41586-021-03944-y> (2021).
- Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1) (2020).
- NIC. National Informatics Centre. *Ministry of Health and Family Welfare, Government of India*. <https://www.mohfw.gov.in/> (Accessed 03 January 2022) (2022).
- Kong, J. D., Tekwa, E. W. & Gignoux-Wolfsohn, S. A. Social, economic, and environmental factors influencing the basic reproduction number of COVID-19 across countries. *PLoS ONE* **16**, e0252373. <https://doi.org/10.1371/journal.pone.0252373> (2021).
- Sebastian, I. R. *et al.* Nutritional status and COVID-19 among older persons in Kerala, India. *Inquiry* **58**, 469580211059738. <https://doi.org/10.1177/00469580211059738> (2021).
- Khobragade, A. W. & Kadam, D. D. Spatial mapping and socio-demographic determinants of COVID-19 mortality in India. *J. Fam. Med. Prim. Care* **10**, 4200–4204. https://doi.org/10.4103/jfmpc.jfmpc_903_21 (2021).
- Middya, A. I. & Roy, S. Geographically varying relationships of COVID-19 mortality with different factors in India. *Sci. Rep.* **11**, 7890. <https://doi.org/10.1038/s41598-021-86987-5> (2021).
- Sasikumar, K., Nath, D., Nath, R. & Chen, W. Impact of extreme hot climate on COVID-19 outbreak in India. *Geohealth* **4**, e2020GH000305. <https://doi.org/10.1029/2020GH000305> (2020).
- Wang, W., Kim, R. & Subramanian, S. V. (Harvard Dataverse, 2021).
- GOI. *General & Census Commissioner*. <https://censusindia.gov.in/census.website> (2022).
- Kaul, V. K. India's diversity, from conflict to innovation. *World Affairs J. Int. Issues* **19**, 10–43 (2015).
- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A. & Hegewisch, K. C. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* **5**, 170191. <https://doi.org/10.1038/sdata.2017.191> (2018).
- Nguyen, P. *et al.* The CHRS Data Portal, an easily accessible public repository for PERSIANN global satellite precipitation data. *Sci. Data* **6**, 180296. <https://doi.org/10.1038/sdata.2018.296> (2019).
- FSI. *State of Forest Report 2019*. <https://fsi.nic.in/forest-report-2019>.
- Hammer, M. S. *et al.* Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). *Environ. Sci. Technol.* **54**, 7879–7890. <https://doi.org/10.1021/acs.est.0c01764> (2020).
- Purwanto, P. *et al.* Spatiotemporal analysis of COVID-19 spread with emerging hotspot analysis and space–time cube models in East Java, Indonesia. *ISPRS Int. J. Geo-Inf.* **10**, 133. <https://doi.org/10.3390/ijgi10030133> (2021).
- Kendall, M. G. & Gibbons, J. D. In *Rank Correlation Methods* (ed. Arnold, E.) (Oxford University Press, 1990).
- Ord, J. K. & Getis, A. Local spatial autocorrelation statistics: Distributional issues and an application. *Geogr. Anal.* **27**, 286–306. <https://doi.org/10.1111/j.1538-4632.1995.tb00912.x> (1995).
- Getis, A. & Ord, J. K. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* **24**, 189–206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x> (1992).
- Mann, H. B. Nonparametric tests against trend. *J. Econometr. J. Econometr. Soc.* **13**, 245–259. <https://doi.org/10.2307/1907187> (1945).
- ESRI. *Emerging Hot Spot Analysis—Help|ArcGIS for Desktop*. <https://desktop.arcgis.com/en/arcmap/10.13/tools/space-time-pattern-mining-toolbox/emerginghotspots.html> (Accessed 03 January 2022) (2021).
- Oztig, L. I. & Askin, O. E. Human mobility and coronavirus disease 2019 (COVID-19): A negative binomial regression analysis. *Public Health* **185**, 364–367. <https://doi.org/10.1016/j.puhe.2020.07.002> (2020).
- Ibarra-Espinosa, S., de Freitas, E. D., Ropkins, K., Dominici, F. & Rehbein, A. Negative-binomial and quasi-Poisson regressions between COVID-19, mobility and environment in Sao Paulo, Brazil. *Environ. Res.* **204**, 112369. <https://doi.org/10.1016/j.envres.2021.112369> (2022).
- Chan, S., Chu, J., Zhang, Y. & Nadarajah, S. Count regression models for COVID-19. *Physica A* **563**, 125460. <https://doi.org/10.1016/j.physa.2020.125460> (2021).
- Esri. *World Topographic Map*. <http://www.arcgis.com/home/item.html?id=30e5fe3149c34df1ba922e6f5bbf808f>.
- Dehury, B. Interlinkages of literacy and household poverty in India: Inter-district and socio-demographic disparities. *Biomed. J. Sci. Tech. Res.* **1**, 1488–1495 (2017).
- Census Tables*. <https://censusindia.gov.in/census.website/data/census-tables> (2022).
- The Demographic and Health Surveys (DHS) Program*. <https://dhsprogram.com/programming/wealth%20index/India%20DHS%202015-16/IndiaNFHS4.pdf> (2015–2016).
- Hassaan, M. A., Abdelwahab, R. G., Elbarkly, T. A. & Ghazy, R. M. GIS-based analysis framework to identify the determinants of COVID-19 incidence and fatality in Africa. *J. Prim. Care Community Health* **12**, 21501327211041210. <https://doi.org/10.1177/21501327211041210> (2021).
- Bhowmik, T., Tirtha, S. D., Iraganaboina, N. C. & Eluru, N. A comprehensive analysis of COVID-19 transmission and mortality rates at the county level in the United States considering socio-demographics, health indicators, mobility trends and health care infrastructure attributes. *PLoS ONE* **16**, e0249133. <https://doi.org/10.1371/journal.pone.0249133> (2021).
- Drefahl, S. *et al.* A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. *Nat. Commun.* **11**, 5097. <https://doi.org/10.1038/s41467-020-18926-3> (2020).
- Baker, D. W. *et al.* The health care experience of patients with low literacy. *Arch. Fam. Med.* **5**, 329–334. <https://doi.org/10.1001/archfam.5.6.329> (1996).

33. Mallett, S. *et al.* At what times during infection is SARS-CoV-2 detectable and no longer detectable using RT-PCR-based tests? A systematic review of individual participant data. *BMC Med.* **18**, 346. <https://doi.org/10.1186/s12916-020-01810-8> (2020).
34. Rajan, K., Kennedy, J. & King, L. Is wealthier always healthier in poor countries? The health implications of income, inequality, poverty, and literacy in India. *Soc. Sci. Med.* **88**, 98–107. <https://doi.org/10.1016/j.socscimed.2013.04.004> (2013).
35. Nikoloski, Z. *et al.* Covid-19 and non-communicable diseases: Evidence from a systematic literature review. *BMC Public Health* **21**, 1068. <https://doi.org/10.1186/s12889-021-11116-w> (2021).
36. Williamson, E. J. *et al.* Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430–436. <https://doi.org/10.1038/s41586-020-2521-4> (2020).
37. Kassebaum, N. J. *et al.* Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1603–1658. [https://doi.org/10.1016/S0140-6736\(16\)31460-X](https://doi.org/10.1016/S0140-6736(16)31460-X) (2016).
38. Singh, R. *et al.* Association of obesity with COVID-19 severity and mortality: An updated systemic review, meta-analysis, and meta-regression. *Front. Endocrinol.* **13**, 780872. <https://doi.org/10.3389/fendo.2022.780872> (2022).
39. Russell, C. D., Lone, N. I. & Baillie, J. K. Comorbidities, multimorbidity and COVID-19. *Nat. Med.* **29**, 334–343. <https://doi.org/10.1038/s41591-022-02156-9> (2023).
40. Oh, S. M. *et al.* On-admission anaemia predicts mortality in COVID-19 patients: A single center, retrospective cohort study. *Am. J. Emerg. Med.* **48**, 140–147. <https://doi.org/10.1016/j.ajem.2021.03.083> (2021).
41. Tao, Z. *et al.* Anaemia is associated with severe illness in COVID-19: A retrospective cohort study. *J. Med. Virol.* **93**, 1478–1488. <https://doi.org/10.1002/jmv.26444> (2021).
42. Kumar, S. & Pattanaik, F. Regional disparities in employment intensity of Indian industries: A state-level analysis. *Emerg. Econ. Stud.* **6**, 23–38. <https://doi.org/10.1177/2394901520907729> (2020).
43. Hawkins, D. Differential occupational risk for COVID-19 and other infection exposure according to race and ethnicity. *Am. J. Ind. Med.* **63**, 817–820. <https://doi.org/10.1002/ajim.23145> (2020).
44. MoHFW. *Ministry of Health and Family Welfare, Government of India.* <https://www.mohfw.gov.in/>.
45. Varshney, K. & Adalbert, J. Overcrowded housing increases risk for COVID-19 mortality: An ecological study. *Health Serv. Res.* **56**, 75–76. <https://doi.org/10.1111/1475-6773.13829> (2021).
46. Varshney, K., Glodjo, T. & Adalbert, J. Overcrowded housing increases risk for COVID-19 mortality: An ecological study. *BMC Res. Notes* **15**, 126. <https://doi.org/10.1186/s13104-022-06015-1> (2022).
47. *State of Housing in India, a Statistical Compendium.* https://mohua.gov.in/upload/uploadfiles/files/Housing_in_India_Compendium_English_Version2.pdf (2013).
48. CDC. *COVID-19 Prevention Actions.* <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html> (2022).
49. Ahmed, N., Maqsood, A., Abduljabbar, T. & Vohra, F. Tobacco smoking a potential risk factor in transmission of COVID-19 infection. *Pak. J. Med. Sci.* **36**, S104–S107. <https://doi.org/10.12669/pjms.36.COVID19-S4.2739> (2020).
50. Zhao, Q. *et al.* The impact of COPD and smoking history on the severity of COVID-19: A systemic review and meta-analysis. *J. Med. Virol.* **92**, 1915–1921. <https://doi.org/10.1002/jmv.25889> (2020).
51. Baker, J., Krishnan, N., Abrams, L. C. & Berg, C. J. The impact of tobacco use on COVID-19 outcomes: A systematic review. *J. Smok. Cessat.* **2022**, 5474397. <https://doi.org/10.1155/2022/5474397> (2022).
52. Balasubramani, K., Paulson, W., Chellappan, S., Behera, S. K. & Balabaskaran Nina, P. District-level epidemiology, hot spots and sociodemographic determinants of tobacco use in Indian men and women: Analysis of national family health survey-4 (2015–16). *Public Health* **194**, 127–134. <https://doi.org/10.1016/j.puhe.2021.03.001> (2021).
53. Saurabh, S. *et al.* Tobacco, alcohol use and other risk factors for developing symptomatic COVID-19 vs asymptomatic SARS-CoV-2 infection: A case-control study from western Rajasthan, India. *Trans. R. Soc. Trop. Med. Hyg.* **115**, 820–831. <https://doi.org/10.1093/trstmh/traa172> (2021).
54. Balasubramani, K. *et al.* Epidemiology, hot spots, and sociodemographic risk factors of alcohol consumption in Indian men and women: Analysis of national family health survey-4 (2015–16), a nationally representative cross-sectional study. *Front. Public Health* **9**, 617311. <https://doi.org/10.3389/fpubh.2021.617311> (2021).
55. Mona *et al.* *Health Systems Resilience Index: A Sub-national Analysis of India's COVID-19 Response.* <https://www.orfonline.org/research/health-systems-resilience-index/> (2022).
56. Keralam, A. *National Health Mission, Government of Kerala.* <https://arogyakeralam.gov.in/2021/2004/2029/break-the-chain/> (Accessed 20 January 2022) (2022).
57. Sarkar, S. Breaking the chain: Governmental frugal innovation in Kerala to combat the COVID-19 pandemic. *Gov. Inf. Q.* **38**, 101549. <https://doi.org/10.1016/j.giq.2020.101549> (2021).
58. City, M. *My Family My Responsibility GR, Government of Maharashtra.* <https://mumbaicity.gov.in/document/my-family-my-responsibility-gr-2/> (Accessed 03 January 2022) (2022).
59. Jayesh, S. & Sreedharan, S. Analysing the covid-19 cases in Kerala: A visual exploratory data analysis approach. *SN Compr. Clin. Med.* **2**, 1337–1348. <https://doi.org/10.1007/s42399-020-00451-5> (2020).
60. Golechha, M. COVID-19 containment in Asia's largest urban slum Dharavi-Mumbai, India: Lessons for policymakers globally. *J. Urban Health* **97**, 796–801. <https://doi.org/10.1007/s11524-020-00474-2> (2020).
61. Khaparde, S., Aroskar, K., Taralekar, R. & Gomare, M. Interventions to respond to COVID 19 pandemic in Mumbai and way forward. *Indian J. Tuberc.* **69**, 234–237. <https://doi.org/10.1016/j.ijtb.2021.04.013> (2022).
62. Sahoo, M. M. Significance between air pollutants, meteorological factors, and COVID-19 infections: Probable evidences in India. *Environ. Sci. Pollut. Res. Int.* **28**, 40474–40495. <https://doi.org/10.1007/s11356-021-12709-z> (2021).
63. Kolluru, S. S. R., Patra, A. K., Nazneen, & Shiva Nagendra, S. M. Association of air pollution and meteorological variables with COVID-19 incidence: Evidence from five megacities in India. *Environ. Res.* **195**, 110854. <https://doi.org/10.1016/j.envres.2021.110854> (2021).
64. Rendana, M. Impact of the wind conditions on COVID-19 pandemic: A new insight for direction of the spread of the virus. *Urban Clim.* **34**, 100680. <https://doi.org/10.1016/j.uclim.2020.100680> (2020).
65. Şahin, M. Impact of weather on COVID-19 pandemic in Turkey. *Sci. Total Environ.* **728**, 138810. <https://doi.org/10.1016/j.scitotenv.2020.138810> (2020).
66. Bashir, M. F. *et al.* Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Sci. Total Environ.* **728**, 138835. <https://doi.org/10.1016/j.scitotenv.2020.138835> (2020).
67. Praveen, B. *et al.* Analyzing trend and forecasting of rainfall changes in India using non-parametrical and machine learning approaches. *Sci. Rep.* **10**, 10342. <https://doi.org/10.1038/s41598-020-67228-7> (2020).
68. Ali, N. & Islam, F. The effects of air pollution on COVID-19 infection and mortality—A review on recent evidence. *Front. Public Health* **8**, 580057. <https://doi.org/10.3389/fpubh.2020.580057> (2020).
69. Sahu, S. K. *et al.* Establishing a link between fine particulate matter (PM_{2.5}) zones and COVID-19 over India based on anthropogenic emission sources and air quality data. *Urban Clim.* **38**, 100883. <https://doi.org/10.1016/j.uclim.2021.100883> (2021).
70. Reserve Bank of India. *State-Wise Literacy Rate.* <https://www.rbi.org.in/Scripts/PublicationsView.aspx?id=19995>.

Author contributions

K.B. and P.B.N. conceptualized, designed, and supervised the study; K.B., V.R., K.A.P., and N.K.K. performed data extraction, data curation, software handling, data analysis, and wrote the results section; P.B.N. drafted the original manuscript; P.B.N., K.B., M.R., S.S., M.M.J., S.K.B., N.G., R.K.S., D.K.S., M.S. and A.P.D. contributed to the literature search, review and editing of the manuscript; All authors have read and agreed to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-50363-2>.

Correspondence and requests for materials should be addressed to P.B.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024