Policy Perspectives

What Can Really Explain the Inter-state Variations in COVID-19 Outcomes in India?

M. Dinesh Kumar, Saurabh Kumar, Nitin Bassi, Ajath Sanjeev and Sujit Raman[1]

Abstract

The study investigates into the explanatory factors for the variation in COVID-19 infections and deaths reported in Indian states as on March 31, 2021. The analysis considered the following state-wise data: proportion of people living in cities with population density higher than 5,000 persons per sq. km, per capita public health expenditure, health infrastructure per thousand population, per capita NSDP, and proportion of aged (above 60) people. As regards COVID-19 infections, the proportion of people living in densely populated areas (above 5,000 persons per sq. km), per capita NSDP and proportion of aged people explained the variation across states. As regards the deaths due to COVID-19, in addition to these three factors, per capita public health infrastructure was found to be a contributing factor, with its impact on death being negative. The curious situation of income increasing COVID-19 transmission and death could be probably explained by the considerable proportion of the people in some of the high-income states living in congested slums under extreme poverty with poor access to basic infrastructure, and the high mobility and exposure of some of these states to domestic and international travel footprint and large migrant population, all resulting in increased risk.

Keywords: COVID-19 Infections; COVID-19 Deaths; Net State Domestic Product; Densely Populated Area; Health Infrastructure; Public Health Expenditure; Population Density

While the COVID-19 Pandemic has taken a heavy toll on the world economy and public health, what has caught the real attention of the epidemiologists and researchers is the vast variation in the incidence of the diseases and more importantly, the dramatic variation in the deaths associated with the disease for the same size of the population. What has surprised most is the phenomenon that is taking a heavy toll on the life of people in the developed countries such as the US, UK, France, Germany, Spain and Italy, which are generally known for effective

---
[1] All authors are with the Institute for Resource Analysis and Policy, Hyderabad.
E-mail: dineshcgiar@gmail.com

governance, robust public health system, high literacy and good public awareness about the diseases. Not only were the incidence of COVID-19 (per thousand people) high, but the number of reported deaths per million population were also high in these countries.

Since the pandemic started in December 2019, several studies have been undertaken worldwide to identify the factors that led to spread of the COVID-19 infections and mortality in different countries and regions. Most of such studies either used statistical (regression) analysis or machine learning tools to predict the dynamics of the spread of COVID infections and mortality rates. The main factors that were explored include demographic indicators (population density, aging population, per capita income, etc.), environmental variables (temperature, humidity, UV radiations, etc.), and healthcare and infrastructure facilities.

In this article, we investigate into the various factors that explain the variation in the incidence of COVID-19 infections and deaths associated with the disease across Indian states and seek to find scientific explanation for the same, using the knowledge available in the field from scientific research. An extensive review of the research studies done internationally (including that in India) on COVID-19 infections and death was done to identify the most dominant demographic, socio-economic and public health infrastructure related factors that were found by these studies to be the influencing factors and were used to inform the current study vis-à-vis selection of variables for the analysis.

Review of International Research on COVID-19 Transmission

Irrespective of the regions, many studies found population density as the major socio-economic factor influencing the spread of COVID-19 infections. In the US, states with high population density and testing exhibited consistently high infections and deaths (Roy and Ghosh, 2020). Further, the spread was higher in the vulnerable groups that include African Americans, Hispanic-Latina, and older adults (Wong and Lee, 2020; Jin *et al.*, 2021). In Algeria of north Africa, a strong correlation was established between the population density and the number of COVID-19 infections, i.e., the spread of the infections was higher in cities with high population density (Kadi and Khelfaoui, 2020). Further, in the city of São Paulo, epicentre of COVID-19 in Brazil, cumulative confirmed cases were found to be positively correlated with population density, and negatively correlated with isolation rate, indicating that the physical distancing has been effective in reducing the viral transmission (Nakada and Urban, 2020). However, in the European Union, irrespective of the population density, countries with higher proportion of the population living in urban areas experienced higher peak of COVID-19 deaths (Jablonska *et al.*, 2021).

Some research studies also looked at the role of environmental variables in the spread of COVID-19 infections. Based on the data of 188 countries, air pollution (% CO_2 in the air) along with the population density was found to be main factor driving the increased viral spread. Further, the temperature or air pressure in these

countries did not have the same effects as pollution or population (Aabed and Lashin, 2021). In Brazil, an inverse relationship was observed between the COVID-19 confirmed cases and temperature and also with the UV radiation, suggesting that the sunlight might be effective in reducing the infectivity of the virus (Nakada and Urban, 2020).

Another set of analysis using global data set found that regions with low and high annual average temperature, both favour the transmission and incidence of the disease with different intensities (Magd *et al*., 2020). Nevertheless, in China, UK, Germany and Japan, the spread and decay stages of the COVID-19 pandemic were directly correlated with absolute humidity, temperature, and population density (Diao *et al*., 2021). Velasco et al. (2021) found that the temperature of 14.5°C is in the favourable range for the growth of the virus. In Russia, the seasonality of climate also had an impact on the COVID-19 transmission and infection. In the humid continental region, seasonal variation in temperature, which is the difference between the annual maximum and annual minimum temperature was the primary influencing variable for the COVID-19 transmission, with increased difference resulting in greater transmission of the virus. In the sub-artic region, the mean temperature diurnal range, which is the difference between average daily maximum and average daily minimum temperature was the primary influencing factor. Higher difference resulted in higher transmission of the disease (Pramanik et al., 2020). Thus, unlike population density, environmental factors had varying impacts in different climatic zones.

Some studies also found that the hardest hit countries either had an aging population (Gardner *et al*., 2020; Upadhyaya et al., 2020) or underdeveloped healthcare systems (Tanne et al., 2020). The importance of healthcare infrastructure was obvious in Thailand where large scale infections were controlled by combination of a good healthcare system and regulation on the tourism activities (Tantrakarnapa *et al*., 2020). Further, the per capita income had a negative and statistically significant effect on COVID-19 death rate globally (Upadhyaya et al., 2020). For instance, in Mexico, high poverty and income inequality aggravated the spread of the pandemic (Benita and Gasca-Sanchez, 2020). In the European Union, lower reduction in mobility at the beginning of the pandemic and countries having more infected people when closing borders (lockdown) experienced higher mortality rate (Jablonska *et al*., 2021). High mobility (either through air or road) was also identified as one of the factors leading to spread of COVID-19 infections in the US (Roy and Ghosh, 2020) and Brazil (Nakada and Urban, 2021).

A recent study in the UK carried out by the London School of Hygiene and Tropical Medicine after the second wave found higher risks for testing positive and subsequent poor outcomes amongst minority ethnic groups. When compared with wave 1, the relative risk for testing positive, hospitalisation, ICU admission, and death were smaller in pandemic wave 2 for all minority ethnic communities compared to white people, with the exception of South Asian groups. South Asian groups remained at higher risk for testing positive, with relative risks for

hospitalisation, ICU admission, and death, which were greater in magnitude compared to the first wave. Despite the improvements seen in most minority ethnic groups in the second wave compared to the first, the disparity widened among South Asian groups (Mathur *et al.*, 2021).

After accounting for age and sex, social deprivation was the biggest potential explanatory variable for disparities in all minority ethnic groups except South Asian. In South Asian groups, health factors (e.g., body mass index, blood pressure, underlying health conditions) played the biggest role in explaining excess risks for all outcomes. Household size was an important explanatory variable for the differences in COVID-19 mortality in South Asian groups.

Thus, studies globally suggest the influence of the following factors on the spread of COVID-19 infections and COVID-19 related deaths, viz., demography, climate, economy, social diversity, mobility, and health infrastructure. However, the factors identified and their natures of influence were different for different countries.

Studies on COVID-19 Transmission in India

During the first wave, India had one of the largest numbers of COVID-19 infections (15 million), which is a little more than 1% of the population. The number of COVID-19 deaths stood at 1,50,000 people, which is nearly 1% of the total reported cases. Even among the 20 worst affected countries in terms of the number of reported cases as on 16th May 2021 when the second lethal wave was ongoing in India, it had the second lowest number of confirmed cumulative cases per million population (Figure 1). Though as a proportion of the total population, these numbers are still very small, the sharp variation in the reported cases of COVID-19 infection and deaths across Indian states had created an equal amount of curiosity among researchers working in the field from the country. Several theories were postulated by medical professionals and public intellectuals over the past few months on the factors that could probably explain the sharp variation in the COVID-19 cases and COVID-19 deaths between developed and developing countries and between regions within India. In both the cases, one frequent explanation that was provided was in the high discrepancy in reporting of cases and deaths. It was argued that the reporting system is very accurate in the developed countries, and not good in developing countries like India.

Figure 1: Confirmed cumulative cases of COVID-19 in the 20 most- affected countries as of 16 May 2021
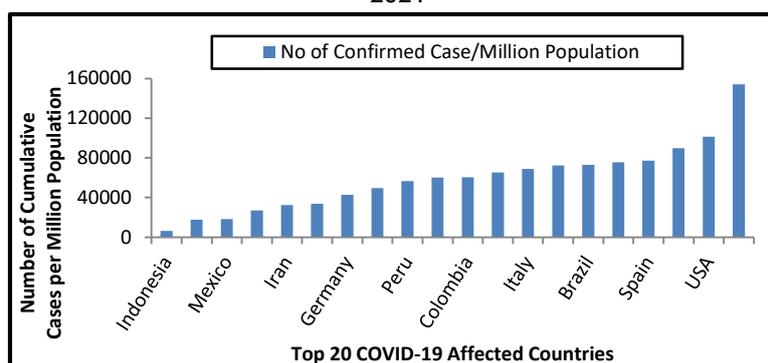
(Source: Authors' own analysis based on data accessed from the world-o-meter)

A recent paper by Balakrishnan and Namboothiri (2021) examined the factors responsible for the variation in COVID-19 cases in India, by considering the Case Fatality Ratio (CFR), using multivariate analysis. They revised the CFR estimates available from the state health departments and ran regressions considering the following factors: 1) population density; 2) the public health expenditure as a share of the state Gross Domestic Product (GDP); 3) public health infrastructure; and 4) per capita income. They concluded that the case fatality ratio is inversely proportional to the proportion of the government expenditure on health. The population density was also considered as one of the social determinants influencing the spread of COVID-19 pandemic in India in the studies by Arif and Sengupta (2020) and Pandey *et al.* (2021).

The analysis by Balakrishnan and Namboothiri (2021), however, suffered from the following problems on the conceptual and practical fronts. First of all, they have treated 'health expenditure (HE) as a share of the GDP' as a variable to represent the public health expenditure by state governments. This is conceptually and theoretically incorrect. It is not the proportion of the GDP spent on public health which matters, but the actual health expenditure per capita (Rs/capita) incurred by the state, the reason being that there is wide variation in the GDP across states and even the per capita GDP. Therefore, it is quite possible for a state with very high per capita Net State Domestic Product (NSDP) spending a small fraction of its GDP for health, but the actual expenditure could be quite sizable.

Secondly, there are other major factors which probably could be driving the COVID-19 cases and CFR, an important one being the number of poor people (and NOT per capita income). Again, it is not the average population density *per se* (also included in the analysis by Arif and Sengupta, 2020 and Pandey *et al.*, 2021)., but the proportion of the population living in very densely populated areas (like densely populated cities) that really matters. For instance, in Mumbai, the population density is above 35,000 persons per sq. km, and in its slums, it can be anywhere near 200,000 per sq. km. At the same time, the population density of major states (excluding the city state of Delhi) varies from 218 (for Chhattisgarh) to 1,122 (for West Bengal) people per sq. km, which does not matter when it comes to

transmission of a disease like COVID-19. But it does matter when it crosses a certain threshold, say 5 to 10,000 persons per sq. km. What is most important is even in less densely populated states, there are cities that have very high population density. Another factor could be the proportion of people living below poverty and without basic amenities like in slums. Once these factors are considered, the results would be different.

Another important issue is the relevance of case fatality ratio (CFR). Case fatality ratio refers to number of deaths per 100 reported cases of infection. That being so, the actual number of reported cases of infection per 1,000 population varies drastically amongst the states. Therefore, the use of this indicator (CFR) would hide the gravity of the problem in situations where the case load is very high (like in Kerala where CFR is only 0.40) and exaggerate the situation in states like Punjab where the CFR is very high (3.2). But the deaths per 1,000 people in Kerala is 0.12, against 0.19 in Punjab, indicating a minor difference. In that case, what really needs to be considered is the number of deaths per 1,000 of total population. So, the studies that try to identify the determinants of virus transmission need to look at the rate of infection and deaths in relation to the total population.

Data and Approach for the Study

For undertaking the analysis, we collected the data on COVID-19 cases in Indian states, from 2020 when the first case was reported, till early March 2021. Before proceeding with the analysis, we assumed that the data are reliable and accurate enough to show the variation in cases across the states. This assumption doesn't mean that the statistics are correct. It only means that even if there are significant reporting errors or manipulations, that applies to all states more or less uniformly. However, we discarded the data for the state of Bihar, as a close examination of the data during the analysis stage revealed them to be an 'outlier', probably due to serious problems with reporting of cases from that state.

State level data used for various analysis are as follows: 1) population; 2) COVID-19 infections and COVID-19 deaths; 3) Net State Domestic Product (NSDP), 4) average human development index (HDI), 5) proportion of people living in poverty; 6) public health expenditure for nine consecutive years; 7) public and private health infrastructure (no. of hospital beds); 8) number of persons living in cities with population density exceeding 5,000 per sq. km; and 9) total population above the age of 60. From these, the values of following variables were derived: 1) no. of COVID-19 infections per 1000 people; 2) number of COVID-19 deaths per 1,000 people; 3) CFR; 4) per capita NSDP; 5) proportion of people living in areas with population density higher than 5000 per sq. km; 6) proportion of people above the age of 60; 7) average annual per capita public health expenditure; and, 8) health infrastructure per 1000 people. All the variables are described in Table 1 and their estimated values are presented in Table 2.

Table 1: Description of study variables

S. No.	Variable	Description
1	Number of COVID-19 infections per 1000 people	Ratio of the total number of reported cases of COVID-19 and the total population in 000' of the respective state
2	Number of COVID-19 deaths per 1,000 people	Ratio of the total number of reported cases of COVID-19 related deaths and the total population in 000' of the respective state
3	Case Fatality Ratio (CFR)	Ratio of the total number of COVID-19 related deaths and the total number of COVID-19 reported cases in 00' in the respective state
4	Per capita Net State Domestic Product (NSDP)	Adjusted (to constant prices) NSDP divided by the total population of the respective state
5	Proportion of people living in areas with population density higher than 5000 per sq. km	Ratio of the total number of people residing in areas (cities/towns) having a population density higher than 5000 per sq. km and the total population of the respective state
6	Proportion of people above the age of 60	Ratio of the total number of people above the age of 60 and the total population of the respective state
7	Average annual per capita public health expenditure	Ratio of the average (of nine years) annual health expenditure and the total population of the respective state
8	Health infrastructure per 1000 people	Total number of beds in public and private hospitals divided by population in 000' of the respective state

Table 2: Population density, NSDP per capita, health expenditure per capita, health infrastructure, HDI, poverty rates, COVID-19 infection and COVID-19 death rates in different states

State	Population	Population Density	Average Public Health Expenditure (Rs)	Average Public Health Expenditure per Capita	Poverty Rate (2011-12)	Per Capita NSDP Capita	HDI	COVID-19 Deaths / 1000 People	CFR (%)	COVID-19 Cases/ 1000 People	Proportion of Population living in Density Above 5000 /sq. km	PHI per 1000	Private Health Infrastructure per 1000 People	Health Infrastructure per 1000 People	Fraction of Population above the age of 60
Andhra Pradesh	53903393	331	52371369	972	9.200	107241.0	0.650	0.133	0.805	16.513	0.019	1.115	1.115	2.230	0.090
Arunachal Pradesh	1570458	19	5154949	3282	34.670	93191.0	0.660	0.036	0.333	10.722	0.000	1.531	0.140	1.671	0.041
Assam	35607039	454	27552185	773	31.980	60695.0	0.614	0.031	0.502	6.111	0.000	0.481	0.198	0.679	0.058
Chhattisgarh	29436231	218	17893450	608	39.930	69500.0	0.613	0.131	1.228	10.634	0.000	0.320	0.272	0.592	0.068
Delhi	18710922	12617	35509529	1898	9.910	269505.0	0.746	0.583	1.706	34.188	1.000	1.303	0.806	2.109	0.061
Goa	1586250	428	5555658	3502	5.090	337745.0	0.761	0.502	1.445	34.719	0.000	1.899	0.991	2.890	0.103
Gujarat	63872399	325	54942007	860	16.630	153495.0	0.672	0.069	1.629	4.239	0.213	0.316	0.700	1.015	0.075
Haryana	28204692	638	23757545	842	11.160	169409.0	0.708	0.108	1.126	9.613	0.050	0.399	0.883	1.281	0.078
Himachal Pradesh	7451955	134	13657844	1833	8.060	139469.0	0.725	0.134	1.694	7.891	0.000	1.664	0.489	2.152	0.094
Jammu and Kashmir	13606320	61	18999975	1396	10.350	65178.0	0.688	0.144	1.547	9.304	0.000	0.536	0.052	0.588	0.068
Jharkhand	38593948	484	18221684	472	36.960	54982.0	0.599	0.028	0.909	3.116	0.028	0.279	0.407	0.687	0.061
Karnataka	67562686	352	48945556	724	20.910	148970.0	0.682	0.183	1.296	14.091	0.138	1.032	2.848	3.879	0.086
Kerala	35699443	919	39592828	1109	7.050	148078.0	0.779	0.118	0.397	29.812	0.017	1.065	1.715	2.780	0.118
Maharashtra	123144223	400	81530194	662	17.350	147450.0	0.696	0.424	2.408	17.616	0.243	0.418	1.464	1.882	0.090
Manipur	3091545	138	4568596	1478	36.890	51180.0	0.696	0.121	1.274	9.472	0.000	0.462	0.117	0.579	0.065
Meghalaya	3366710	150	5036390	1496	11.870	62458.0	0.656	0.044	1.060	4.148	0.000	1.324	0.234	1.558	0.041
Mizoram	1239244	59	3686706	2975	20.400	129609.0	0.705	0.008	0.226	3.572	0.000	1.611	0.403	2.014	0.056
Madhya Pradesh	85358965	277	44018258	516	31.650	56498.0	0.606	0.045	1.473	3.074	0.021	0.364	0.396	0.761	0.067
Nagaland	2249695	136	4285812	1905	18.880	73276.0	0.679	0.040	0.746	5.423	0.000	0.836	0.303	1.138	0.046
Odisha	46356334	298	29545457	637	32.590	75191.0	0.606	0.041	0.568	7.277	0.000	0.399	0.154	0.553	0.086
Punjab	30141373	598	24289117	806	8.260	115852.0	0.723	0.194	3.193	6.089	0.120	0.595	1.429	2.024	0.095
Rajasthan	81032689	257	58948886	727	14.710	78570.0	0.629	0.034	0.869	3.956	0.044	0.581	0.569	1.150	0.063
Sikkim	690051	97	2450316	3550	8.190	242002.0	0.716	0.196	2.196	8.905	0.000	2.260	0.568	2.828	0.059
Tamil Nadu	77841267	598.5	66853654	859	11.280	142941.0	0.708	0.161	1.467	10.951	0.094	0.996	1.000	1.996	0.097
Telangana	38510982	344	23718971	616	8.800	143618.0	0.669	0.042	0.546	7.771	0.175	0.545	2.050	2.595	0.090
Tripura	4169794	398	5235258	1256	14.050	82632.0	0.658	0.094	1.170	8.015	0.000	1.062	0.057	1.119	0.070
Uttar Pradesh	237882725	72	114920365	483	29.430	43670.0	0.596	0.037	1.446	2.538	0.040	0.321	0.862	1.183	0.065
Uttarakhand	11250858	210	12501405	1111	11.260	155151.0	0.684	0.150	1.743	8.629	0.000	0.757	1.363	2.119	0.080
West Bengal	99609303	1122	58718955	589	19.980	67300.0	0.641	0.103	1.784	5.777	0.061	0.789	0.351	1.140	0.078

Sources: Population and related variables, and public health expenditure is based on data accessed from the Ministry of Health and Family Welfare, Government of India; Poverty rates are from the RBI handbook of statistics on Indian economy 2020-21; Per capita NSDP is based on the 2018-19 data from the Ministry of Statistics and Programme Implementation; HDI is based on Global Data Lab 2019 figures of Sub-national HDI; COVID-19 infections and death related estimates are from mygov.in; and Health infrastructure estimates based on Kapoor et al., 2020.

The variables described in Table 1 were used for developing two multivariate regression models to explain the variation in COVID-19 infections (Model 1) and COVID-19 related deaths (Model 2) across different Indian states. All the chosen independent variables for running the multivariate regression analysis were mutually exclusive. Overall, three independent variables were chosen to explain the variation in COVID-19 infections and five were chosen to explain variation in COVID-19 related deaths across different Indian states. The analysis is presented in the subsequent sections.

What Explains the Variation in COVID-19 Infections Across Indian States?

Following were the hypothesis to begin with. As public health research has shown population density would have significant influence on spread of an infection like COVID-19. However, as our review has shown, the population density figures considered by earlier researchers were at the state level or at the regional level. Given the fact that such variations are often not remarkable, and that the cases of COVID-19 were mostly reported from cities during the first wave in India, considering the population density of the entire state does not make much sense. Instead, what mattered was what proportion of the people in each state live in heavily populated areas. Hence, the proportion of the state population which live in cities with population density more than 5,000 persons per sq. km was considered.

Another parameter considered was per capita income (the per capita net state domestic product). While it was found that the highly developed countries were worse off during the first and second wave of pandemic, studies in the United States had shown that the spread of the virus was higher in the vulnerable groups that include African Americans and Hispanic-Latina (Wong and Lee, 2020; Jin *et al.*, 2021).

The value of per capita NSDP (at constant prices) for the selected states of India ranged from a lowest of Rs. 43,870 for Uttar Pradesh to a highest of Rs. 3,37,745 for Goa, which is the richest state in India in terms of per capita income. Delhi stood second with a per capita NSDP of Rs. 2,69,505.

The third parameter considered was proportion of people above the age of 60, because studies have shown that the old age people would be more susceptible to the disease, as countries having aging population were badly hit (source: based on Gardner *et al.*, 2020; Upadhyaya et al., 2020). This is also one of the parameters used by Balakrishnan and Namboodiri (2021) in their analysis of variations in COVID-19 cases.

The estimates of COVID-19 infections ranged from a lowest of 2.5 persons per 1000 population for Uttar Pradesh to 34.7 people per 1,000 population for Goa. The second highest reported COVID-19 cases was for Delhi, with 34.2 persons per 1,000 population. Kerala had the third highest reported cases with 29.8 persons per thousand.

Analysis with number of COVID-19 infection cases per 1,000 people as a dependent variable, against these three independent variables showed an R^2 value of 0.643. All the three parameters had a very high level of significance (see Table 3) in explaining the variation in COVID-19 cases across states to an extent of 64%. The regression equation is:

COVID-19 cases per 1000 people = - 7.42 + (0.000062 * Per capita NSDP in INR constant price) + (12.45 * Proportion of population living in densely populated area) + (128.49 * Fraction of population above the age of 60)

As per the model, states with high proportion of people living in very densely populated areas (like Mumbai, Delhi, Ahmedabad, Kolkata, Chennai) and higher fraction of people in the old age category (above 60 years) would have higher cases of infections. In a broad sense, these findings corroborate with findings of studies available from other countries that were reviewed in this article about the effect of population density and aging population on Covid-19 infections. For instance, the first trend is in line with Jablonska et al. (2021) and the second trend corroborates with the findings of Upadhyaya et al. (2020) and Gardner et al. (2020).

But the study clearly shows that the average population density at the aggregate level does not mean much when it comes to explaining the inter-state differences. Though the R^2 value slightly improved (to 0.64) when the multivariate analysis was carried out with (state) average population density (along with the other two variables) against Covid-19 infection rates, the same dropped to a mere 0.51 when data for 'Delhi', which is an outlier[i], was removed from the sample. More importantly, the 'p value' for 'population density' increased drastically (to 82%), indicating that the variable is not significant at all. But the original model was run by excluding 'Delhi', the R2 value remained almost the same (0.62), with the significance of each variable improving. The effect of population density becomes imperative when the density exceeds a certain threshold and therefore what matters is what proportion of the population live in such densely populated areas. In this case, the proportion of people living in areas with population density higher than 5,000 persons per sq. km was found it to be a useful criterion.

Interestingly, higher average per capita income increased the incidence of COVID-19 infections. However, it should also be pointed out that the gradient is not steep. For income to have a real effect on the disease, the rise required is quite high, as the value of the beta coefficient is 0.000018. If the average per capita income increases by one lac rupees, there is chance that COVID-19 cases would increase by around 1.8 per 1,000 people.

A plausible explanation for this trend (contrary to what was found in other countries) could be that the average per capita income considered is of a state, and not of the pockets that are badly hit by the infection. That said, some of the states having high average per capita income are also states that are high exposure to international and domestic passenger footprint by virtue of having international airports (like Delhi, Kerala, Maharashtra and Goa), and heavy influx of migrants. This increases the infection risk. On the other hand, in some of these states (Delhi and

Maharashtra), a very high proportion of the people live in slums with much greater congestion, without basic facilities of proper water supply and sanitation. This further increases the risk of infection. It should be invoked that in Mumbai, it is the slums inhabited with millions of people that was badly affected, and the average income figures for the state (in this case, Maharashtra) does not reflect the socioeconomic conditions of these poor localities where most of the people are very poor.

Table 3: Result of multivariate regression analysis with COVID-19 cases per 1000 people as dependent variable and per capita NSDP, proportion of population living in densely populated areas, and fraction of population above the age of 60 as independent variables

Predictor	Coefficients	Standard Error	t Statistics
Constant	-7.42	4.34	-1.71
Per capita NSDP (INR constant prices)	0.000062**	0.000019	3.30
Proportion of population living in densely populated areas (above 5000 per sq. km)	12.45*	6.32	1.97
Fraction of population above the age of 60	128.49*	62.79	2.05
R-squared	0.643		
Number of observations	29		

Note: * and ** indicates significance at 90% and 99% level, respectively

The factors that we have not considered in the analysis are the environmental conditions. There is surely a lot of variation in the climatic conditions across the country, strong enough to cause variations in the potency of the virus to spread, if we go by the studies done in Russia and other cold countries. Coastal Maharashtra, especially Mumbai is very hot and humid. So is the coastal areas and midland areas of Kerala, Chennai, the plains of West Bengal and coastal areas of Odisha. Gujarat, Karnataka, Tamil Nadu, Rajasthan, Punjab, Haryana, Andhra Pradesh and Telangana are mostly in the hot tropics. Uttar Pradesh and Bihar lie in sub-tropical, temperate zone. The north eastern states have cold and humid climate. However, the effect of these climatic variations is not captured in the model owing to inadequate information on the way they could affect infection from the virus.

What Explains the Variation in COVID-19 Deaths?

A survey by the Office for National Statistics (ONS) in the United Kingdom found in their survey that the backward areas of England and Wales, with high Index of Multiple Deprivation, had the highest mortality rates during the initial days of the pandemic. The index takes into account factors such as an area's income, employment, crime and health deprivation and disability. The ONS study, which included 20,283 deaths involving COVID-19 in England, found the mortality rate in the most deprived areas to be 55.1 deaths per 100,000 population, against 25.3 deaths in the least deprived areas (*BBC*, 2020).

Different indicators were used by different states at different points of time to justify the actions taken to control COVID-19. For instance, one of the arguments made by the government of Kerala, while the state witnessed high incidence of COVID-19 when other states were showing a steep decline in the number of cases, was the low CFR (Case Fatality Ratio). In the case of Kerala, the CFR hovered around 0.40, i.e., 4 deaths per 1,000 COVID-19 patients, during the first wave. The low CFR brought down the overall deaths per 1,000 persons to a considerably low level in Kerala and Telangana which managed to keep the CFR

below 0.5 per cent. The overall deaths per 1,000 people is the multiple of no. of cases per 1000 people and the number of deaths per 1000 cases. When the number of cases of infections increases disproportionately in some states, the low CFR in those states may not be of much relevance as the total number of deaths would increase. More importantly, from the point of view of safeguarding public health, the pressure on the health infrastructure, which is expected to protect lives, would increase with increase in number of cases. Hence, we have used the overall deaths per 1000 people for our analysis.

We ran several regression models to understand the factors that explained the variations in COVID-19 deaths per 1000 persons across the states. After several iterations, a total of five variables were included in the analysis as independent variables, while many were excluded after noticing that either they have no effect in influencing the 'COVID-19 deaths per 1000 persons', or are related to the other variables already considered for the analysis. For instance, population density was found to have no effect on the COVID-19 deaths and hence was excluded from the analysis. The poverty rate was found to be inversely proportional to the per capita net state domestic product (with a high correlation coefficient) and hence was excluded.

The final variables chosen are: proportion of people living in densely populated areas (above 5,000 persons per sq. km); average public health expenditure by the state (during the past 9 years); per capita net state domestic product; the capacity of the health infrastructure; and fraction of the population in the age group of 60 and above. The average public health expenditure per capita was found to be varying from Rs 472 per annum in Jharkhand to a highest of Rs. 3500 + per annum in Sikkim and Goa. The health infrastructure per 1,000 people was found to be varying from a lowest of 0.588 in J & K to a highest of 3.88 in Karnataka (Figure 2).

The regression analysis showed that these factors together explained the variation in COVID-19 deaths to an extent of 74.5 per cent (R-square value=0.745) (Table 4). However, among these five variables, two variables, i.e., per capita NSDP; proportion of people living in densely populated areas had very high level of significance. The fraction of aged people in the society was significant at 13 per cent level and therefore should also be considered as important. Increase in aging population certainly increases the mortality rate, probably owing to weaker immune system and the chances of co-morbidities. The health infrastructure had lower level of significance (24 per cent level), whereas public health expenditure had an inverse effect in reducing mortality.

When the regression model was run by replacing 'proportion of people living in densely populated areas with 'average state population density', though the R2 value decreased marginally (to 0.72), many of the variables except population density became insignificant with p values becoming 85% for average PHE,. 43.5% for health infrastructure and 24.4% for proportion of population above the age of 60. This analysis suggests that the average state population density is not explanatory variable for Covid-19 deaths also.

The adverse impact of economic conditions on death rates can be explained by the phenomenon of high degree of mobility of the people and exposure of the population in some of the high-income states (Delhi, Goa, Maharashtra and Kerala) to heavy domestic and international passenger footprint, and a substantially large migrant population. These factors increase the risk of serious COVID-19 outcomes. Incidentally, in some of the states such as Maharashtra and Delhi, the proportion of people living in highly congested slums under extreme poverty with poor access to basic health infrastructure is considerably high. These factors also increase the chances of mortality.

Table 4: Result of multivariate regression analysis with
COVID-19 deaths per 1000 people as dependent variable and
average public health expenditure, per capita NSDP,
proportion of population living in densely populated areas,
health infrastructure per 1000 persons, and fraction of
population above the age of 60 as independent variables

Predictor	Coefficients	Standard Error	t Statistics
Constant	-0.15	0.09	-17
Average public health expenditure (INR per capita)	0.000028	0.000030	0.93
Per capita NSDP (INR constant price)	0.00000094*	0.00000049	1.91
Proportion of population living in densely populated areas (above 5,000 per sq. km)	0.35**	0.11	3.20
Health infrastructure per 1000 persons	-0.029	0.025	-1.20
Fraction of population above the age of 60	2.12	1.34	1.58
R-squared	0.745		
Number of observations	29		

Note: * and ** indicates significance at 90% and 99% level, respectively

The negligible effect of health infrastructure, contrary to what was found by Tantrakarnapa *et al.*, (2020) for Thailand, in reducing the mortality rate needs explanation. As regards the effect of health infrastructure, one reason could be that the actual effect of such factors would be visible when the total number of cases crosses a threshold wherein the public health infrastructure collapses. This is what is seen during the second wave of COVID-19 infections. In spite of large number of cases per 1,000 persons, Kerala, which has one of the best public health infrastructures in the country, has been able to control the death rates to nearly 3 persons per 1,000 cases (CFR=0.3 per cent), and 1.75 persons per 10,000 population whereas the total number of deaths per 1000 infected persons in Maharashtra is 14.96 and the number of deaths per 10,000 population is 6.9. The death rate in Maharashtra today is nearly 4 times that of Kerala[ii].

As regards the negative effect of public health expenditure (PHE) on reducing COVID-19 deaths, the above phenomenon could partly explain, as many of the states incurring moderate to high expenditure on public health are also those having relatively higher NSDP (Delhi and Goa). This factor nullifies the benefit from high public health expenditure. An additional factor could be that not all people living in all areas get equal benefits of government expenditure on public health, especially in vulnerable areas in terms of public

health centres and community health centres and the number of staff equipped to handle the cases.

Concluding Remarks

Questions were raised during the past one year or so about the reliability of data relating to incidence of COVID-19 infections and deaths occurring in India by medical professionals, clinical scientists and social scientists around the world. Newspapers have occasionally reported incidence of underreporting of COVID-19 cases, especially deaths, by the state governments. Epidemiological studies on COVID-19 would require high quality data on infections and deaths. The issue relating to data reliability notwithstanding, the sharp variations in incidence of infections and death rates is a reality witnessed across Indian states. That being the case, analysis involving state-level data on COVID-19 and the range of factors that have potential bearing on the pandemic transmission and deaths can bring out certain key determinants of the variations in infection rates and death rates, if we assume that the extent of false reporting of infections and deaths is more or less same across the states.

Multivariate analysis involving state-wise data show that two important variables influence the infection rate, i.e., proportion of people living in areas with population density higher than 5,000 persons per sq. km, and the proportion of people above the age of 60. The effect of per capita NSDP was adverse. States with higher per capita NSDP had higher infection rates per 1,000 people. This probably explains the low occurrence of the infection in some of the states that are predominantly rural such as Uttar Pradesh, Jharkhand, Chhattisgarh and Odisha during the first wave of COVID-19 infections.

As regards COVID-19 related deaths, three main factors are found to influence. They are: 1) proportion of people living in areas with state's population density higher than 5,000 persons per sq. km; 2) fraction of population above the age of 60 in the state; and, 3) the per capita NSDP of the state. The fourth factor, which was less significant, was the health infrastructure from public and private sector. The status of health infrastructure though had negative impact on death rates, they had lower levels of significance. This could be due to the fact that the number of cases of infections in the states had not become high enough in the badly affected states to start showing the impact of poor health infrastructure and inadequate public health expenditure. As the data from the second wave of COVID-19 show, the states with poor health infrastructure and low level of state expenditure on public health do witness high death rates.

The scientific explanation for these factors to be important determinants of COVID-19 infection and COVID-19 death can be had from the knowledge available from past research. The most intriguing trend that emerged from the study, for which scientific explanation is hard to obtain, is the positive impact of the average economic conditions on COVID-19 infection rates and deaths. Unlike what was found in the studies elsewhere in the US and UK, high income increased both infection and mortality rates in the respective states. This was probably due to the increased mobility and associated problem of increased exposure of the people living there to domestic and international passenger footprint, and a substantially large migrant population. Incidentally, some of these high incomes states also had large and congested slums, and with poor access to water supply and sanitation and inadequate public health infrastructure.

The study highlights the need for shifting the focus of the COVID-19 control strategy to regions characterized by very high density of population and high income-inequality, or which have a high migrant population and have high degree of exposure to domestic and international air traffic, along with regions with aging population.

References

Aabed, K., & M.M. Lashin. 2021. An analytical study of the factors that influence COVID-19 spread. *Saudi Journal of Biological Sciences*, *28*(2): 1177-1195.

Arif, M., & S. Sengupta. 2021. Nexus between population density and novel coronavirus (COVID-19) pandemic in the south Indian states: a geo-statistical approach. *Environment, Development and Sustainability*, *23*(7): 10246-10274.

Balakrishnan, P., & S.K. Namboodhiry. 2021. The interstate variation in mortality from COVID-19 in India. *Economic & Political Weekly*, *56*(6): 36-43.

BBC. 2020. Coronavirus: Higher death rate in poorer areas, ONS figures suggest. 1 May, 2020.

Benita, F., & F. Gasca-Sanchez. 2020. On the main factors influencing COVID-19 spread and deaths in Mexico: a comparison between Phase I and II. *medRxiv*. Doi: https://doi.org/10.1101/2020.12.22.20248716

Diao, Y., S. Kodera D. Anzai J. Gomez-Tames E.A. Rashed & A. Hirata 2021. Influence of population density, temperature, and absolute humidity on spread and decay durations of COVID-19: A comparative study of scenarios in China, England, Germany, and Japan. *One Health*, *12*(2021): 1-9.

Gardner, W., D. States, & N. Bagley. 2020. The coronavirus and the risks to the elderly in long-term care. *Journal of Aging & Social Policy*, *32*(4-5): 310-315.

Jabłońska, K., S. Aballéa & M. Toumi. 2021. Factors influencing the COVID-19 daily deaths' peak across European countries. *Public Health*, *194*(2021): 135-142.

Jin, J., N. Agarwala, P. Kundu, B. Harvey Y. Zhang, E. Wallace, & N. Chatterjee. 2021. Individual and community-level risk for COVID-19 mortality in the United States. *Nature Medicine*, *27*(2): 264-269.

Kadi, N., & M. Khelfaoui. 2020. Population density, a factor in the spread of COVID-19 in Algeria: statistic study. *Bulletin of the National Research Centre*, *44*(1): 1-7.

Kapoor, G., Sriram, A., Joshi, J., Nandi, A., Laxminarayan, R. 2020. COVID-19 in India: State-wise estimates of current hospital beds, intensive care unit (ICU) beds and ventilators. The Centre for Disease Dynamics, Economics and Policy (CDDEP), Washington D.C., and Princeton University, April 2020.

Magd, H., K. Asmi & K. Henry. 2020. COVID-19 influencing factors on transmission and incidence rates-validation analysis. *Journal of Biomedical Research & Environmental Sciences, 1*(7): 277-291.

Mathur, R., C.T. Rentsch, C.E. Morton, W.J. Hulme, A. Schultze, B. MacKenna, R.M. Eggo, K. Bhaskaran, A.Y. Wong, E.J. Williamson, & H. Forbes. 2021. *Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the Open SAFELY platform*. **The Lancet**. DoI: 10.1016/S0140-6736(21)00634-6

Nakada, L.Y.K., & R.C. Urban. 2020. COVID-19 pandemic: environmental and social factors influencing the spread of SARS-CoV-2 in São Paulo, Brazil. *Environmental Science and Pollution Research, 28* (30): 40322-40328.

Pandey, A., A. Prakash, R. Agur & G. Maruvada. 2021. Determinants of COVID-19 pandemic in India: an exploratory study of Indian states and districts. *Journal of Social and Economic Development*, *23*(2021): 248–279.

Pramanik, M., P. Udmale, P. Bisht, K. Chowdhury, S. Szabo, & I. Pal. 2020. Climatic factors influence the spread of COVID-19 in Russia. *International Journal of Environmental Health Research*: 1-16.

Roy, S., & P. Ghosh. 2020. Factors affecting COVID-19 infected and death rates inform lockdown-related policymaking. *PloS one*, *15*(10): e0241165.

Tanne, J.H., E. Hayasaki, M. Zastrow, P. Pulla, P. Smith, & A.G. Rada. 2020. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. *BMJ*, 368.

Tantrakarnapa, K., B. Bhopdhornangkul & K. Nakhaapakorn. 2020. Influencing factors of COVID-19 spreading: a case study of Thailand. *Journal of Public Health*. Doi: https://doi.org/10.1007/s10389-020-01329-5

Upadhyaya, A., S. Koirala, R. Ressler, & K. Upadhyaya. 2020. Factors affecting COVID-19 mortality: an exploratory study. *Journal of Health Research*. Doi: 10.1108/JHR-09-2020-0448

Velasco, J.M., W.C. Tseng & C.L. Chang. 2021. Factors affecting the cases and deaths of COVID-19 victims. *International Journal of Environmental Research and Public Health*, *18*(2): 674.

Wong, D.W., & Y. Li. 2020. Spreading of COVID-19: density matters. *Plos one*, *15*(12): e0242398.

---

[i] The reason for considering Delhi as an outlier is that it is the only state whose average population density figure is representative of the actual population density across its entire geographical area, unlike many other states which while having low average

population density, had large proportion of the people living in very densely populated areas.

[ii] As on May 15, 2021, the total number of cases in Kerala is 20.5 lac and that in Maharashtra is 52.7 lac. The total number of deaths in Kerala is 6150 and that in Maharashtra is 78,857. The population of Kerala and Maharashtra are 35 million and 114.2 million, respectively.