THE EUROPEAN SOCIETY OF MEDICINE
Medical Research Archives, Volume 13 Issue 7

REVIEW ARTICLE

Factors that Caused the Inter-state Variations in COVID-19 Infections and Death in India

M. Dinesh Kumar [1] and Saurabh Kumar [2]

[1] Executive Director, Institute for Resource Analysis and Policy (IRAP), Hyderabad, INDIA
[2] Associate Counsellor at CII-ITC Centre of Excellence for Sustainable Development, New Delhi, INDIA

OPEN ACCESS

PUBLISHED
31 July 2025

CITATION
Kumar, MD., and Kumar, S., 2025. Factors that Caused the Inter-state Variations in COVID-19 Infections and Death in India. Medical Research Archives, [online] 13(7). https://doi.org/10.18103/mra.v13i7.6755

DOI
https://doi.org/10.18103/mra.v13i7.6755

ISSN
2375-1924

ABSTRACT

The study investigates into the factors responsible for the variation in COVID-19 infections and death rates reported across Indian states. 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, average climatic conditions, health infrastructure per thousand population, per capita Net State Domestic Product, and proportion of aged (above 60) people identified on the basis of an extensive review of the international scientific literature published on the pandemic, covering both COVID-19 infections and mortality. As regards COVID-19 infections, the proportion of people living in densely populated areas (above 5,000 persons per sq. km), per capita Net State Domestic Product (NSDP), proportion of aged people and climate explained the variation across states. While the first three variables had a positive influence on the infection rates, incidence of infections increased from hot & arid climate to cold & humid climate. As regards the deaths due to COVID-19, in addition to these three factors, health infrastructure per unit population was found to affect, with its impact on COVID-19 related mortality being negative. But the effect of the factors such as the proportion of aged people (above 60 years), the per capita Net State Domestic Product and the proportion of population living in densely populated areas on COVID-19 induced mortality was positive. Like in the case of COVID-19 infections, the chances of mortality increased from hot & arid regions to cold & humid regions.

1. Introduction

While the COVID-19 Pandemic has taken a heavy toll on the world economy and public health, 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 had caught the attention of the epidemiologists and researchers internationally. 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 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 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 from scientific research in this field. An extensive review of the research studies done internationally on COVID-19 infections and death was done to identify the most dominant demographic, socio-economic, environmental 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.

2. Review of International Research on COVID-19 Transmission

Studies across regions 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[22]. Further, the spread was higher in the vulnerable groups that include African Americans, Hispanic-Latina, and older adults[28-11]. In Algeria, 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[12]. Further, in the city of São Paulo, the 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[18]. However, in the European Union, irrespective of the population density of the country, those with higher proportion of the population living in urban areas experienced higher peak of COVID-19 deaths[10]. The anomalies found w. r. to the influence of population on COVID-19 infections suggests that the scale at which population density is measured influences the COVID-19 outcomes in terms of effect of population density[13].

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[1]. 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[18].

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[16]. 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[6]. Velasco et al.[27] 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[20]. Thus environmental factors had varying impacts across climatic zones.

Some studies also found that the countries that were hit hard either had an aging population[7-26] or underdeveloped healthcare systems[24]. 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[25]. Further, the per capita income had a negative and statistically significant effect on COVID-19 death rate globally[26]. For instance, in Mexico, high poverty and income inequality aggravated the spread of the pandemic[4]. 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[10]. 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[22] and Brazil[18].

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 [17].

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.

Sanfelici[23] analyzed the Italian response in different stages of the COVID-19 management. It highlighted structural inequalities that already existent in the pre-crisis stage, and has shown their consequences for the immigrants during the pandemic. A compromised socio-legal status prevailed for the immigrants due to immigration policies, forms of exploitation linked to neoliberal labour market characteristics and welfare state policies that exclude more precarious migrants. Those compromised positions became more apparent when rules pertaining to physical distancing, tracing, and the “stay home” mandate became paradoxical for those who did not have a home, or were living in overcrowded spaces, or were irregular workers with no residency and therefore no rights. The analysis of the state response showed the serious challenges faced by migrants and professionals, and a joint effort of several actors in civil society in tackling these circumstances, showing vulnerability and resilience both at the individual and at the community level. Also, due to the interventions by lawyers, some improvement have been achieved at the institutional level. However, these measures seem to be just a palliative, leaving the cultural and structural variables unchanged, within a wider socioeconomic system that creates a condition of hyperprecarity for more marginalized migrants[15].

Analysis by Yimga[29], using a dataset that encompasses COVID-19 case rates, healthcare resources, and socioeconomic and transportation-related variables across multiple destination counties, showed that increased availability of hospital and ICU beds is significantly associated with a reduced risk of COVID-19 spread, after controlling for origin county infection rates, travel patterns, and socioeconomic factors.

Challen et al.[5] carried out a study in the UK to establish whether there was any change in mortality from infection with a new variant of SARS-CoV-2, which was designated a variant of concern (VOC-202012/1) in December 2020, compared with circulating SARS-CoV-2 variants. The study found that there was a high probability that the risk of mortality is increased by infection with VOC-202012/01. The study concluded that if this finding is generalisable to other populations, infection with VOC-202012/1 has the potential to cause substantial additional mortality compared with previously circulating variants.

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.

3. 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 couple of years 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.

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[2] and Pandey *et al.*[19].

Balakrishnan and Namboodhiry[3] 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.

Their analysis, however, suffered from problems both 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[13]. 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[2-19]., 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 5,000 to 10,000 persons per sq. km[13]. 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 people varies drastically amongst the states. Therefore, the use of this indicator (CFR) would hide the gravity of the problem in situations like Kerala where CFR was only 0.40 and exaggerate the situation in states like Punjab where the CFR was 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.

The statistical model developed by Kumar et al.[13] to explain the inter-state variations in COVID-19 infections considered following variables as determinants of COVID-19 infections (per 1,000 population): 1) proportion of people living in densely populated areas; 2) per capita NSDP; 3) proportion of living above the age of 60. Further, for explaining inter-state variations in COVID-19 deaths (per 1,000 people), it considered the following variables: 1) proportion of people living in densely populated areas; 2) per capita NSDP; 3) proportion of living above the age of 60; 4) average per capita public health expenditure; and 5) health infrastructure per 1000 people. The first model showed statistically very significant positive effect of per capita NSDP and proportion of people living in densely populated areas on COVID-19 infections. The second model showed statistically very significant positive effect of these two variables plus significant effect of proportion of people above the age of 60 on COVID-19 deaths. The second model did not show any effect of per capita public health expenditure and access to health infrastructure.

4 Materials and Methods

The original research that formed the basis for this article was conducted during 2022 and the analyses and the statistical models were later on refined in 2024-25 using new variables. For undertaking the analysis, we collected the data on COVID-19 cases in Indian states, since the first case was reported in 2020 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), 3) proportion of people living in poverty; 4) public health expenditure for nine consecutive years; 5) public and private health infrastructure (no. of hospital beds); 6) number of persons living in cities with population density exceeding 5,000 per sq. km; 7) total population above the age of 60; and 8) average climatic condition. 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) per capita NSDP; 4) proportion of people living in areas with population density higher than 5000 per sq. km; 5) proportion of people above the age of 60; 6) average annual per capita public health expenditure; 7) health infrastructure per 1000 people; and 8) climatic condition. All the variables are described in Table 1 and their estimated values are presented in Table 2.

Table 1: The Independent and dependent variables Considered for the Analysis

Serial Number	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	Per capita Net State Domestic Product (NSDP)	Adjusted (to constant prices) NSDP divided by the total population of the respective state

Serial Number	Variable	Description
4	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
5	Climate	A measure of the degree of cold and humidity that a region experiences
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

Note: 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.

5. Results

5.1 FACTORS EXPLAINING THE VARIATION IN COVID-19 INFECTIONS ACROSS INDIAN STATES

Following were the hypothesis to begin with. Public health research has shown that 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. But the fact remains that such variations between states and between regions are often not remarkable. Also the cases of COVID-19 were mostly reported from cities that are characterized by dense population during the first wave in India. Therefore, instead of considering the population density of the entire state, it is better to consider the proportion of the people in each state living in heavily populated areas. Hence, the proportion of the state's population which live in areas with population density more than 5,000 persons per sq. km was considered. The same approach used by Kumar[13] which found a positive effect of the variable on COVID-19 infections.

Another parameter considered was per capita income (the per capita net state domestic product). While it was found that the highly developed countries with very high per capita income 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[11-28]. Therefore, the research findings from outside India on the influence of per capita income on COVID-19 infections are mixed[13]. However, showed positive effect of per capita net state domestic product on COVID-19 infections.

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[7 - 26]. This is also one of the parameters used by Balakrishnan and Namboodhiry[3] in their analysis of variations in COVID-19 cases. [13] Established the positive effect of this variable on COVIDd-19 infections.

The fourth parameter considered was climate. Public health research has shown that climate has a significant influence on spread of infections[9]. Research on COVID-19 infections in other parts of the world had also shown the effect of climate in one way or the other [16-6-27]. Here, the states, which fall in cold and humid areas were given the highest score of 1.0 and which fall in hot and arid areas were given a score of 0.20. The basic premise in scoring the regions was that as a region becomes hot and arid, the chances of transmission of the virus through air would reduce[6 - 9]. Here it is important to mention that within the same state, the climatic conditions vary and accordingly, the value assigned for the state is the average of the values applicable for different climates. Accordingly, the north eastern states were assigned a value of 1.0, and the hot and arid Rajasthan was assigned a value of 0.30. The value progressively increased as an area become less hot and less arid. The gradation considered are: hot & arid; hot & semi-arid; hot and humid; and cold and humid.

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 four independent variables showed an R^2 value of 0.664. All the four parameters had a very high level of significance in explaining the variation in COVID-19 cases across states to an extent of 66.4%. The results are presented in Table 3. The regression equation is:

COVID-19 Infections per 1000 people = 16.9 + 0.000052 x Per Capita NSDP + 180.83 x Fraction of Pop. Above the age of 60 + 17.55 x Proportion of Pop.

Living in densely populated areas + 9.11 x Climate Index.

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, fraction of population above the age of 60 and climate as independent variables

Predictor	Coefficients	Standard Error	t Statistics
Constant	16.851	7.163	-2.35
Per capita NSDP (INR constant prices)	0.000052**	0.000019	2.68
Fraction of population above the age of 60	180.83**	68.91	2.62
Proportion of population living in densely populated areas	17.55**	6.88	2.55
Climate	9.11*	5.58	1.63
R-squared	0.664		
Number of observations	29		

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

5.2. FACTORS EXPLAINING THE VARIATION IN COVID-19 DEATHS

Different indicators were used by different states at different points of time to justify the actions taken to control COVID-19[1]. 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[8].

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).

COVID-19 deaths per 1000 people = - 0.222 + 0.00000115 x Per Capita NSDP + 2.04 x Fraction of Pop. above the age of 60 + 0.367 x Proportion of Pop. living in densely populated area + 0.126 x Climate Index - 0.0302 x Health Infrastructure per 1000 people

The regression analysis showed that these factors together explained the variation in COVID-19 deaths to an extent of 75.8 per cent (R-square value=0.758) (Table 4).

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

Predictor	Coefficients	Standard Error	t Statistics
Constant	-0.2218	0.10	-0.22
Per capita NSDP (INR constant price)	0.00000115**	0.00000032	3.55
Fraction of population above the age of 60	2.037*	1.009	
Proportion of population living in densely populated areas (above 5,000 per sq. km)	0.366**	0.0966	
Health infrastructure per 1000 persons	-0.03019	0.024	-1.24
Climate	0.126	0.078	1.61
R-squared	0.758		
Number of observations	29		

Note: * and ** indicates significance at 95% and 99% level, respectively

---

[1] 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 [3]. 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[13].

6. Discussion

As per the first model (presented in Table 3) that explains COVID-19 infections, 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. Colder and more humid is the climate, higher the chances of COVID 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 people living in densely populated areas and aging population on Covid-19 infections. For instance, the first trend vis-à-vis the effect of densely populated areas is in line with Jablonska et al.[10] and the second trend vis-à-vis the effect of aged population corroborates with the findings of Upadhyaya[26] and Gardner[7].

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 0.000052. If the average per capita income increases by one lac rupees, there is chance that COVID-19 cases per 1000 people would increase by around 5.20.

This trend is contrary to what was found globally[26]. A plausible explanation for this 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[21]. 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 kind of vulnerability further increases the risk of infection[11-26]. 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 does not reflect the socioeconomic conditions of these poor localities where most of the people are very poor.

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 [6-20-27]. 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.

As regards the second model (presented in Table 4), among these five variables identified as having influence on COVID-19 related mortality, three variables, i.e., per capita NSDP; fraction of population above the age of 60; and proportion of people living in densely populated areas had very high level of significance. Climate had a high level of significance. It was significant at 12 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 (22 per cent level).

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 footfall[21], and a substantially large migrant population[13]. Rechhi et al (2022) found that higher air passenger volume tends to coincide with more COVID-19 deaths, but this relation weakened as the pandemic proceeded. These factors increase the risk of serious COVID-19 outcomes. Incidentally, in some of the states such as Maharashtra, Telangana, West Bengal, Tamil Nadu and Delhi, the proportion of people living in highly congested slums under extreme poverty with poor access to basic health infrastructure is considerably high[8]. These factors also increase the chances of mortality as studies in United States among vulnerable groups have shown[11=28].

A study from UK substantiates this argument about the effect of extreme backwardness in boosting the death rates. 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 (source: https://www.bbc.com/news/uk-52506979). Nevertheless, it should be noted that the beta coefficient for per capita NSDP is quite low, i.e., 000000115. This means, for every 10 thousand rupee increase in per capita NSDP, the COVID deaths per one lac persons increased by around 1.15.

The negligible effect of health infrastructure, contrary to what was found by Tantrakarnapa *et al.*[25] for Thailand and Yimga[27] in US counties, 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[2].

6. Conclusion

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 COVID-19 infections and death rates are realities witnessed across Indian states[3] 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 four 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, the proportion of people above the age of 60, the climatic conditions and per capita net state domestic product. The effect of per capita NSDP was positive. States with higher per capita NSDP had higher infection rates per 1,000 people.

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 slightly less significant, was the climate. The health infrastructure from public and private sector had only a minor effect, in contrast to the findings of Tantrakarnapa *et al.*[25] and Yimga[29] . The status of health infrastructure though had negative impact on death rates, it 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. 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 this study (which also emerged from the study by Kumar and others[13] 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, UK and Mexico[4-26], 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[18-22], 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 focusing the future strategies for controlling infections like COVID-19 on regions with very high density of population and high income-inequality, or regions which have a high migrant population and have high degree of exposure to domestic and international air traffic, along with regions having cold & humid climate and region with aging population. The study shows that the risk of both infection and deaths increase as one moves from hot and arid to cold and humid regions.

2 As on May 15, 2021, the total number of cases in Kerala was 20.5 lac and that in Maharashtra was 52.7 lac. The total number of deaths in Kerala was 6150 and that in Maharashtra was 78,857. The population of Kerala and Maharashtra were 35 million and 114.2 million, respectively.

References

1. Aabed K, Lashin MMA. An analytical study of the factors that influence COVID-19 spread. *Saudi J Biol Sci*. 2021;28(2):1177-1195. doi:10.1016/j.sjbs.2020.11.067
2. Arif M, Sengupta S. Nexus between population density and novel coronavirus (COVID-19) pandemic in the south Indian states: A geo-statistical approach. *Environ Dev Sustain*. 2021;23(7):10246-10274. doi:10.1007/s10668-020-01055-8
3. Balakrishnan P, Namboodhiry SK. The Interstate Variation in Mortality from COVID-19 in India. *Economic & Political Weekly*. 2021;56(6).
4. Benita F, Gasca-Sanchez F. The main factors influencing COVID-19 spread and deaths in Mexico: A comparison between Phases I and II. *bioRxiv*. Published online 2020. doi:10.1101/2020.12.22.20248716
5. Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K, Danon L. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ*. 2021;372:n579. doi:10.1136/bmj.n579
6. Diao Y, Kodera S, Anzai D, Gomez-Tames J, Rashed EA, Hirata A. 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*. 2021;12(100203):100203. doi:10.1016/j.onehlt.2020.100203
7. Gardner W, States D, Bagley N. The Coronavirus and the risks to the elderly in long-term care. *J Aging Soc Policy*. 2020;32(4-5):310-315. doi:10.1080/08959420.2020.1750543
8. Government of India. *Slums in India: A Statistical Compendium, National Buildings Organisation, Ministry of Housing and Urban Poverty Alleviation, Government of India*. GoI; 2015.
9. Hunter PR. Climate Change and Water-Borne and Vector Borne Diseases. *Journal of Applied Microbiology*. 2003;94:37-46.
10. Jabłońska K, Aballéa S, Toumi M. Factors influencing the COVID-19 daily deaths' peak across European countries. *Public Health*. 2021;194:135-142.
11. Jin J, Agarwala N, Kundu P, et al. Individual and community-level risk for COVID-19 mortality in the United States. *Nature Medicine*. 2021;27(2):264-269.
12. Kadi N, Khelfaoui M. Population density, a factor in the spread of COVID-19 in Algeria: statistic study. *Bulletin of the National Research Centre*. 2020;44(1):1-7.
13. Kumar MD, Kumar S, Bassi N, Sanjeev A, Raman S. Investigating Inter-state Variations in COVID-19 Outcomes in India. *Economic and Political Weekly*. 2023;53(38):35-41.
14. León M, Pavolini E. 'social investment' or back to 'familism': The impact of the economic crisis on family and care policies in Italy and Spain. *South Eur Soc Polit*. 2014;19(3):353-369. doi:10.1080/13608746.2014.948603
15. H Lewis and L Waite, 'Asylum, Immigration Restrictions and Exploitation: Hyper-precarity as a lens for understanding and tackling forced labour', *Anti-Trafficking Review*, issue 5, 2015, pp. 49–67, www.antitraffickingreview.org
16. Magd H, Asmi K, Karyamsetty H. COVID-19 Influencing Factors on Transmission and Incidence Rates-Validation Analysis. *Journal of Biomedical Research & Environmental Sciences*. 2020;1(7):277-291.
17. Mathur R. 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. Published online 2021:634-636.
18. Nakada LYK, Urban RC. COVID-19 pandemic: environmental and social factors influencing the spread of SARS-CoV-2 in São Paulo. *Brazil Environmental Science and Pollution Research*. Published online 2020:1-7.
19. Pandey A, Prakash A, Agur R, Maruvada G. Determinants of COVID-19 pandemic in India: An exploratory study of Indian states and districts. *Journal of Social and Economic Development*. 2021;23:248-279.
20. Pramanik M, Udmale P, Bisht P, Chowdhury K, Szabo S, Pal I. Climatic factors influence the spread of COVID-19 in Russia. *International Journal of Environmental Health Research*. Published online 2020:1-15.
21. Recchi E, Ferrara A, Rodriguez Sanchez A, et al. The impact of air travel on the precocity and severity of COVID-19 deaths in sub-national areas across 45 countries. *Sci Rep*. 2022;12(1). doi:10.1038/s41598-022-20263-y
22. Roy S, Ghosh P, eds. *Factors Affecting COVID-19 Infected and Death Rates Inform Lockdown-Related Policymaking*. Vol 15(10). PLoS ONE; 2020. doi:10.1371/journal.pone.0241165
23. Sanfelici M. The impact of the COVID-19 crisis on marginal migrant populations in Italy. *Am Behav Sci*. 2021;65(10):1323-1341. doi:10.1177/00027642211000413
24. Tanne JH, Hayasaki E, Zastrow M, Pulla P, Smith P, Rada AG. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. *BMJ*. 2020;368:m1090. doi:10.1136/bmj.m1090
25. Tantrakarnapa, K., Bhopdhornangkul, B., & Nakhaapakorn, K. Influencing factors of COVID-19 spreading: a case study of Thailand. Journal of Public Health. 2020, 1-7.
26. Upadhyaya A, Koirala S, Ressler R, Upadhyaya K. Factors affecting COVID-19 mortality: an exploratory study. *Journal of Health Research*. Published online 2020.
27. Velasco JM, Tseng WC, Chang CL. Factors Affecting the Cases and Deaths of COVID-19

Victims. *International Journal of Environmental Research and Public Health*. 2021;18(2).
28. Wong DW, Li Y. Spreading of COVID-19: Density matters. *Plos one*. 2020;15(12).
29. Yimga J. Public health infrastructure and COVID-19 spread: An air transportation network analysis. *Journal of Air Transportation Research Society*. 2024;3.

Table 2: Population density, Net State Domestic Product 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 per Capita	Per Capita NSDP Capita	COVID-19 Deaths/ 1000 People	COVID-19 Cases/ 1000 People	Proportion of Population living in Density Above 5000 /sq. km	Climatic Condition	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	972	107241.0	0.133	16.513	0.019	0.5	1.115	1.115	2.230	0.090
Arunachal Pradesh	1570458	19	3282	93191.0	0.036	10.722	0.000	1	1.531	0.140	1.671	0.041
Assam	35607039	454	773	60695.0	0.031	6.111	0.000	0.7	0.481	0.198	0.679	0.058
Chhattisgarh	29436231	218	608	69500.0	0.131	10.634	0.000	0.5	0.320	0.272	0.592	0.068
Delhi	18710922	12617	1898	269505.0	0.583	34.188	1.000	0.5	1.303	0.806	2.109	0.061
Goa	1586250	428	3502	337745.0	0.502	34.719	0.000	0.8	1.899	0.991	2.890	0.103
Gujarat	63872399	325	860	153495.0	0.069	4.239	0.213	0.6	0.316	0.700	1.015	0.075
Haryana	28204692	638	842	169409.0	0.108	9.613	0.050	0.4	0.399	0.883	1.281	0.078
Himachal Pradesh	7451955	134	1833	139469.0	0.134	7.891	0.000	0.9	1.664	0.489	2.152	0.094
Jammu and Kashmir	13606320	61	1396	65178.0	0.144	9.304	0.000	1	0.536	0.052	0.588	0.068
Jharkhand	38593948	484	472	54982.0	0.028	3.116	0.028	0.5	0.279	0.407	0.687	0.061
Karnataka	67562686	352	724	148970.0	0.183	14.091	0.138	0.5	1.032	2.848	3.879	0.086
Kerala	35699443	919	1109	148078.0	0.118	29.812	0.017	0.8	1.065	1.715	2.780	0.118
Maharashtra	123144223	400	662	147450.0	0.424	17.616	0.243	0.65	0.418	1.464	1.882	0.090
Manipur	3091545	138	1478	51180.0	0.121	9.472	0.000	1	0.462	0.117	0.579	0.065
Meghalaya	3366710	150	1496	62458.0	0.044	4.148	0.000	1	1.324	0.234	1.558	0.041
Mizoram	1239244	59	2975	129609.0	0.008	3.572	0.000	1	1.611	0.403	2.014	0.056
Madhya Pradesh	85358965	277	516	56498.0	0.045	3.074	0.021	0.6	0.364	0.396	0.761	0.067
Nagaland	2249695	136	1905	73276.0	0.040	5.423	0.000	1	0.836	0.303	1.138	0.046
Odisha	46356334	298	637	75191.0	0.041	7.277	0.000	0.5	0.399	0.154	0.553	0.086
Punjab	30141373	598	806	115882.0	0.194	6.089	0.120	0.5	0.595	1.429	2.024	0.095
Rajasthan	81032689	237	727	78570.0	0.034	3.956	0.044	0.3	0.581	0.569	1.150	0.063
Sikkim	690251	97	3550	242002.0	0.196	8.905	0.000	1	2.260	0.568	2.828	0.059
Tamil Nadu	77841267	598.5	859	142941.0	0.161	10.951	0.094	0.6	0.996	1.000	1.996	0.097
Telangana	38510982	344	616	143618.0	0.042	7.771	0.175	0.4	0.545	2.050	2.595	0.090
Tripura	4169794	398	1256	82632.0	0.094	8.015	0.000	0.8	1.062	0.057	1.119	0.070
Uttar Pradesh	237882725	72	483	43670.0	0.037	2.538	0.040	0.7	0.321	0.862	1.183	0.065
Uttarakhand	11250858	210	1111	155151.0	0.150	8.629	0.000	0.8	0.757	1.363	2.119	0.080
West Bengal	99609303	1122	589	67300.0	0.103	5.777	0.061	0.8	0.789	0.351	1.140	0.078

Note: the figures presented in the table are based on the authors' own estimates derived from the secondary data obtained from various government sources

