Article

The Determinants of COVID-19 Diffusion in India: A State-level Analysis

Journal of Health Management
1–12

DOI: 10.1177/09720634251328449
journals.sagepub.com/home/jhm
Sage

Padma Suresh Mandala[1] and Shilpa Chaudhary[2]

Abstract

The first case of coronavirus disease 2019 (COVID-19) was detected in India on 30 January 2020. Subsequently, the disease spread across all states and union territories (UTs). However, the literature has not adequately addressed the differential nature of COVID-19 diffusion across states and over time as the pandemic progressed and its determinants. The article examines the role of demographic and socio-economic factors in COVID-19 diffusion across states and assesses the relative significance of these variables in diffusion as the pandemic progressed. We specify a regression model relating cumulative cases per million to demographic and socio-economic variables for a cross-section of states and UTs of India. We use ordinary least squares and quantile regression estimation techniques for discrete time points corresponding to the two waves and for the subsequent period to identify explanatory variables' relative role in diffusion. Results confirm the differential effects of demographic and socio-economic factors on COVID-19 diffusion in the progression of the pandemic. Urbanisation, per capita income, literacy, share of the 60-plus population and proportion of vulnerable populations with health risks are identified as key predictors of spread. Results of the study confirm that demographic and socio-economic variations across states are predictors of COVID-19 diffusion. The diffusion pattern suggests these variables' significance for policymaking in future pandemic control and management.

Keywords

COVID-19, demographic factors, public health, regression, socio-economic factors

Introduction

Coronavirus disease 2019 (COVID-19) is a respiratory disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus that spreads primarily through person-to-person transmission. As of 23 November 2021, India ranked second after the United States in cumulative confirmed cases, with 34.5 million cases accounting for 13.3% of the total confirmed cases worldwide (Worldometer, www.worldometers.info/coronavirus/).

The first positive case of coronavirus was reported in Kerala on 30 January 2020 (WHO, 2020). The epidemiological nature of variants, socio-economic and demographic factors and policy responses have influenced the subsequent pattern of diffusion. A 21-day national lockdown was imposed on 25 March 2020, when the country reported around 500 cases. The lockdown was extended until 31 May 2020 and helped to 'flatten the curve' by restricting the spread of COVID-19 infections (Government of India, 2021). However, with the unlocking of the economy, daily new infections peaked in mid-September 2020 before the first wave of the pandemic subsided in India. At the peak of the first wave, daily new cases averaged 100,000, and daily deaths averaged over 1,000. A puzzling phenomenon of the pandemic is that between October 2020 and February 2021, daily new confirmed cases and daily deaths declined despite the easing of restrictions. This comfortable situation ended when the second wave of infections erupted suddenly in April 2021 with a peak in May 2021. Daily cases crossed 400,000 and daily deaths averaged 3,000 leading to an unprecedented health crisis (The Lancet, 2021).

The spatial diffusion of cases in the two waves of infections has been dissimilar. While the first wave abated by mid-September 2020, the intensity and curtailment of the second wave of the pandemic varied across states. While new epidemiological variants of the virus with higher transmissibility can explain the speed and intensity of the outbreak in the second wave (Mallapaty, 2021), demographic and socio-economic factors and the success of containment measures explain differential diffusion across states. Given the consequences of a stringent national lockdown on livelihoods, the

[1]Department of Economics, Sri Venkateswara College, University of Delhi, New Delhi, Delhi, India
[2]Department of Economics, Janki Devi Memorial College, University of Delhi, New Delhi, Delhi, India

Corresponding author:
Padma Suresh Mandala, Department of Economics, Sri Venkateswara College, University of Delhi, New Delhi, Delhi 110021, India.
E-mail: padmasuresh@svc.ac.in

decision regarding the imposition of lockdowns in the second wave was left to the discretion of states and taken at the sub-national level. The second wave of the pandemic thus saw greater variation in diffusion across states. While many states had contained the pandemic by June 2021, a few states in the Northeast and the southern state of Kerala faced a situation of 'dragging the wave' with continuing high infections.

The early diffusion of the pandemic has been explained by international connectivity and global mobility associated with international trade and travel and tourism, and subsequent local diffusion has been explained by various economic, population and settlement features (Antonietti et al., 2021; Farzanegan et al., 2020; Jeanne et al., 2022; Sigler et al., 2021). Population density and urban concentration have been identified as predictors for disease transmission, with urban cities emerging as epicentres due to a greater concentration of people, human interactions, proximity and higher mobility in cities (Andersen et al., 2021; Barak et al., 2021; Carozzi et al., 2020; Nathan, 2021). Population characteristics like age structure also determine the spread, with a higher incidence in populations with a larger elderly share (Davies et al., 2020). The prevalence of comorbid conditions is known to increase the risk of disease severity from COVID-19 (Williamson et al., 2020). Greater diffusion is observed in countries with higher per capita GDP (Antonietti et al., 2021) and in regions with higher levels of GDP as reported in studies by Ascani et al. (2021) for Italy and Paez et al. (2020) for Spain. Higher human development is associated with greater spread as more affluent and educated populations have greater mobility (Sigler et al., 2021). Studies have also examined the interaction of socio-economic inequalities and racial/ethnic features and resultant disadvantages for specific groups, such as Black, Asian and minority ethnic (Andersen et al., 2021; Bambra et al., 2020; Bryan et al., 2021; Richards-Belle et al., 2020).

For India, studies have analysed the successful containment of the pandemic in Kerala in the early stages of the pandemic (Jalan & Sen, 2020; WHO, 2020) and the robust containment measures that were adopted in Asia's largest and highly populated slum, Dharavi, in Mumbai in May 2020 (Golechha, 2020). Studies also assessed the lockdown and its consequences for livelihoods (Nayyar, 2020; Ray & Subramanian, 2020) and the need for a comprehensive evaluation of healthcare policy and infrastructure (Balakrishnan & Namboodhiry, 2021; Gupta, 2020). Using cross-section regressions, Basu and Mazumder (2021) examine the determinants of the regional spread of COVID-19 in India until February 2021 and find that COVID-19 infections are clustered in urbanised, rich states with high population density and less in agriculturally dominant states and regions with greater poverty. Their dynamic panel regression model shows that infections are persistent across states and that unlocking aggravated the infections. Pandey et al. (2021), using district-level data until 1 November 2020, showed a greater concentration of cases in districts with high population density, lack of basic amenities and a larger share of vulnerable, elderly and migrant populations.

While three major cities—Delhi, Mumbai and Chennai—accounted for more than 50% of India's confirmed cases in June 2020 (Jalan & Sen, 2020), subsequent spatial and temporal diffusion of cases across states in India has been uneven. The relationship between diffusion and the relative role of socio-economic and demographic determinants as the pandemic progressed is largely underinvestigated for India. The primary objective here is to assess the comparative role of these factors in the spatial and temporal spread of COVID-19 in India.

Methods

The methodology adopted is a quantitative research design using secondary data from 33 states and union territories (UTs). We specify a regression model relating cumulative cases per million to demographic and socio-economic variables. The dependent variable, that is, cumulative cases per million is time-varying (daily), while socio-economic and demographic variables are annual data.

A cross-sectional ordinary least squares (OLS) estimation technique is used for different time points to assess the role of explanatory variables during the two waves and in the subsequent period. The OLS cross-sectional regression model using state-level data is specified as:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots \beta_k x_{ki} + \epsilon_i \quad (1)$$

where y is log-transformed cumulative cases per million population in the state, x's refers to state-level demographic and socio-economic explanatory variables in the study and ϵ is the classical disturbance term.

The quantile regression model is specified as:

$$Q^{\tau}\left(y_i \,/\, x_i\right) = \beta_0^{\tau} + \beta_1^{\tau} x_{1i} + \beta_2^{\tau} x_{2i} + \ldots + \beta_k^{\tau} x_{ki}; \tau \in \left(0,1\right) \quad (2)$$

where Q^{τ} is a conditional point estimate of y given x's and τ refers to specific quantiles of the distribution of cumulative cases per million (say 25th, 50th and 75th). In the models, the β coefficients of the explanatory variables are the coefficients of interest. While the OLS regression captures the mean relationship, quantile regression enables us to explore the relationship between response and predictor variables at points along the distribution, that is, at specific quantiles of the dependent variable (Koenker & Hallock, 2001). Using cross-sectional data on y_i and x_i, OLS regression was performed for three data points corresponding to each of the time spans of the first and second waves and for the period till October 2021. An exploratory quantile regression analysis was performed for three data points—September 2020, May 2021 and September 2021 corresponding to different stages of the pandemic.

The dependent variable for the regression analysis is LCUMCASES, defined as the log of the state-level cumulative number of cases per million population and is constructed from data on the state-wise cumulative confirmed number of cases obtained from covid19india.org and the state-wise projected total population as of 1 March 2020, accessed from the Report of the Technical Group on Population Projections (2020). Data on cumulative cases are obtained for three data points, each corresponding to the first wave—(18 August 2020, 15 September 2020 and 13 October 2020) and the second wave (6 April 2021, 4 May 2021 and 1 June 2021). These discrete time points are about a month apart, and for the two waves, they correspond roughly to the date when cases (at the all-India level) show a rising trend, the peak of the wave and the date by which the wave waned, respectively, thus covering the time span of each of the two waves of the pandemic. In addition, since some states had a high incidence of cases beyond the second wave, the analysis includes three data points that are roughly a month apart for the later period, that is, 18 August 2021, 16 September 2021 and 14 October 2021 (The details of data sources and clarifications are given in the Appendix).

The study includes seven independent variables that measure demographic and socio-economic features of states, namely population density, percentage of urban population, percentage of population aged 60 and above, per capita income, education, health and household access to basic facilities.

LPOPDENSITY is the log population density in a state. Population density is population per square kilometre and is constructed from data on projected state total population as of 1 March 2021 (Report of the Technical Group on Population Projections, 2020) and area figures of the state in square kilometre (Census of India, 2011).

URBANPOP is the percentage of the urban population in a state and is estimated from data on projected state urban and projected state total population as of 1 March 2021 (Report of the Technical Group on Population Projections, 2020).

SIXTYPLUS is the percentage of the population aged 60 and above. Projected state-wise elderly population (aged 60+) for 2021 is obtained from Elderly in India (2021), and the percentage share is computed using projected state-wise population.

LPCNSDP is the log of per capita net state domestic product for 2019–2020 at constant prices (Base 2011–2012) and is accessed from the RBI Handbook of Statistics on Indian Economy (2020–2021).

Three indices relating to household access to basic facilities, literacy and health vulnerability are constructed on a scale of 0–1 and obtained as the geometric mean of standardised indicators included in each index (WHO Centre for Health Development, 2014). Data for the construction of these indices are obtained from the National Family Health Survey, India (NFHS-4, 2015–2016).

HHACCESS is a composite index of four basic services, namely the percentage of households with electricity, improved drinking water sources, improved sanitation facilities and using clean cooking fuel. The index is computed using data on the four indicators, each of which is first standardised as given in Equation (3) below.

$$I^{std} = \frac{I - \min(I)}{\max(I) - \min(I)} \tag{3}$$

where I is the value of the indicator for a particular state, min(I) and max(I) are the minimum and maximum values of the indicator over all the 33 states/UTs and I^{std} is the estimated standardised value of the indicator which lies between 0 and 1. For each state, the geometric mean of the standardised indicators across all indicators (i = 1… j) is obtained as the composite index given in Equation (4) below.

$$\text{Composite Index} = \left(\prod_{i=1}^{j} I_i^{std}\right)^{1/j} \tag{4}$$

LITERACY is the literacy index and is obtained from data on the percentage of literate men and women in the state (age 15–49).

HEALTHVUL is an index of health vulnerability that is constructed from state-level data on the percentage of women and men (age 15–49 years) who had high (>140 mg/dL) blood sugar levels and above normal blood pressure (systolic 140–159 mmHg and/or diastolic 90–99 mmHg).

The LITERACY and HEALTHVUL indexes are constructed as described in Equations (3) and (4). Higher values of HHACCESS and LITERACY indexes indicate better living conditions and higher literacy for a population, respectively, whereas higher scores on HEALTHVUL indicate greater vulnerability to infection and illness due to comorbidities associated with diabetes and hypertension.

Trends in the dependent variable-cumulative cases per million: The spatial and temporal diffusion of the pandemic is given in Figure 1, which plots trends in the dependent variable on the y-axis (cumulative cases per million population for all states/UTs as well as all-India) for nine data points included in the study. On the x-axis, the first three data points correspond to the first wave, the next three to the second wave and the last three data points correspond to the more recent period of August, September and October 2021. Since the data are not normalised, the scale of the y-axis is not uniform across graphs.

Figure 1 shows the diversity in spatial and temporal spread. At the start of the first wave, highly urbanised states/UTs (Goa, Delhi, Puducherry and Chandigarh) and states with large urban centres (Maharashtra, Karnataka and Tamil Nadu) had the highest incidence of coronavirus cases per million population. Many states witnessed an intense second wave as given by the steep rise in the curve for data points in the second wave. The all-India graph shows the sharp increase in cumulative cases per million population in the second wave and the plateauing of cases by June 2021. This trend is mirrored in states like Punjab, Haryana, Delhi, Uttar Pradesh and Rajasthan. Chhattisgarh, Himachal Pradesh and

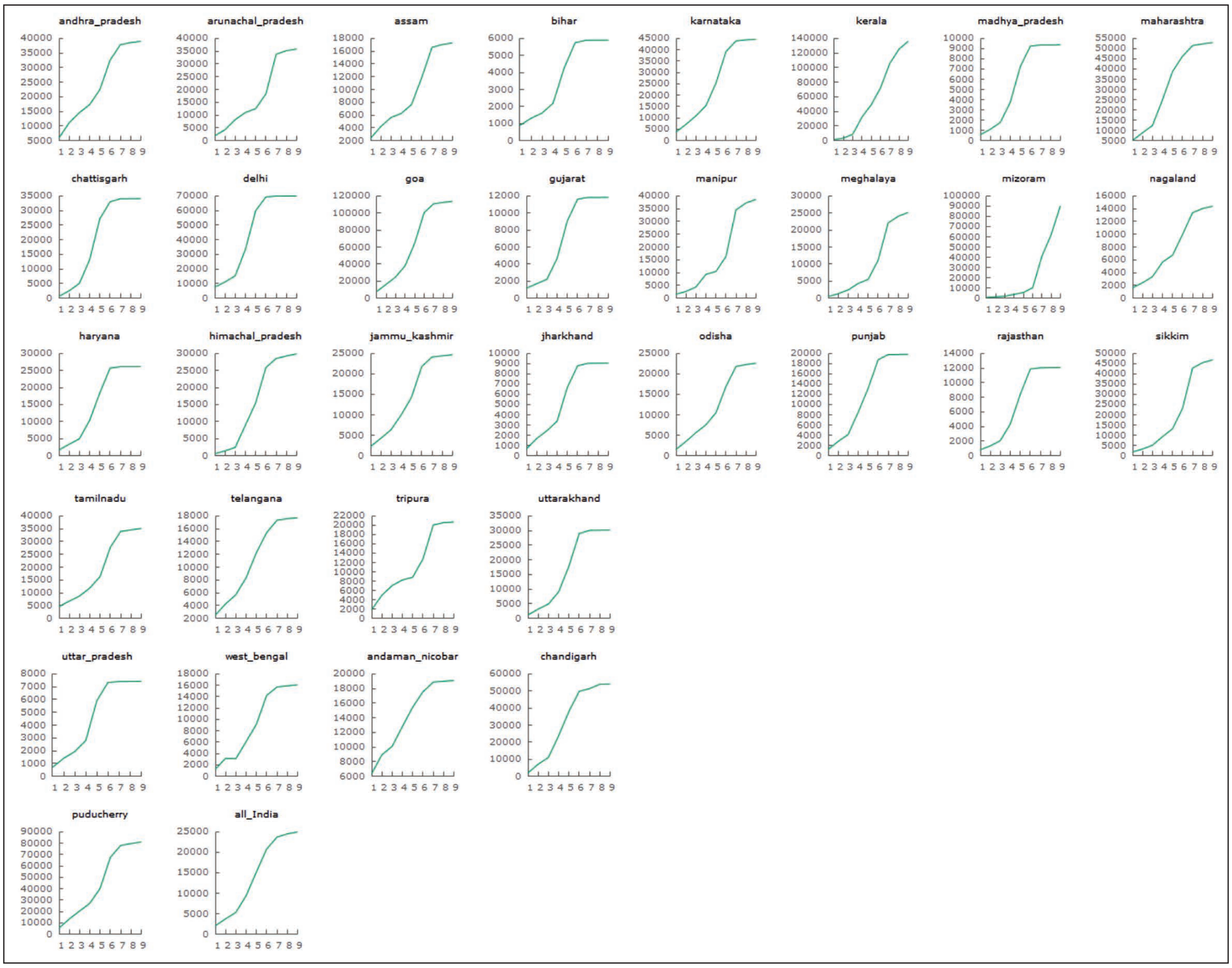

Figure 1. Trends in State-wise COVID-19 Cumulative Confirmed Cases per Million Population.

Source: Authors' construct based on secondary data.

Note: Data points 1–3 correspond to the first wave (18 August 2020, 15 September 2020 and 13 October 2020); data points 4–6 correspond to the second wave (6 April 2021, 4 May 2021 and 1 June 2021) and data points 7–9 correspond to the period after the second wave (18 August 2021, 16 September 2021 and 14 October 2021), respectively.

Uttarakhand had a smaller scale of infections in the first wave but a large jump in cases in the second wave. The Northeast states of Mizoram, Sikkim, Manipur and Arunachal Pradesh saw a sharp increase in infections after June 2021. Cumulative cases per million were low in Kerala as of August 2020, but thereafter, cases increased. Kerala and Mizoram show continuing high levels with no plateauing, and by October 2021, Kerala, Goa and Mizoram had the highest cumulative cases per million population while Bihar, Uttar Pradesh, Jharkhand, Madhya Pradesh and Rajasthan had the lowest diffusion.

Summary Statistics for Explanatory Variables

Table 1 gives summary statistics for the explanatory variables in the study. The difference in maximum and minimum values indicates the diversity of states in various dimensions and also their relevance as determinants of COVID-19 diffusion. Delhi has the highest population density, while Arunachal Pradesh has the lowest population density. Chandigarh and Delhi have a 99% urban population followed by Goa, Kerala, Puducherry and Mizoram, while Himachal Pradesh and Bihar have the lowest rates of urbanisation of 10% and 12%, respectively. Among the Northeast states, Mizoram and Sikkim had the highest urbanisation rates. Goa has the highest per capita income, while Bihar has the lowest per capita income. Kerala has the highest literacy rate followed by Mizoram, Himachal Pradesh and Goa. The literacy rate is lowest in Bihar and Jharkhand. Madhya Pradesh, Rajasthan, Uttar Pradesh and Andhra Pradesh also rank low on the literacy index. Kerala has the highest percentage of 60-plus population, while Bihar, Uttar Pradesh and some of the Northeast states have the lowest percentage of the elderly population. Chandigarh ranks highest and Bihar lowest in household access index. States

Table 1. Summary Statistics for Explanatory Variables.

Variable	Mean	Median	S.D.	Min	Max
POPDENSITY	1,194	348.5	2934	18.31	13871
PCNSDP	125,976	119,162	67,871	30,621	303,687
URBANPOP	41.11	36.97	21.76	10.27	99.70
SIXTYPLUS	9.908	9.700	2.281	6.40	16.49
HHACCESS	0.5429	0.6045	0.2550	0.004231	0.9273
LITERACY	0.4934	0.5220	0.2427	0.0003146	1.000
HEALTHVUL	0.4209	0.4443	0.1823	0.03949	0.7777

Source: Authors' estimation based on secondary data.

Note: HEALTHVUL: Index of health vulnerability; HHACCESS: Household access index; LITERACY: Literacy index; PCNSDP: Per capita NSDP; POPDENSITY: Population density; SIXTYPLUS: Percentage of population aged 60 and above; URBANPOP: Percentage of urban population.

Table 2. Results from Cross-sectional Ordinary Least Squares (OLS) Regression for the Time Period Corresponding to the First Wave of Infections (August to October 2020).

	Dependent Variable LCUMCASES		
	OLS 1	OLS 2	OLS 3
Explanatory Variables	(August 2020)	(September 2020)	(October 2020)
CONSTANT	–0.3085 (–0.068)	–0.5391 (–0.1171)	0.370 (0.079)
LPOPDENSITY	–0.0339 (–0.225)	0.0543 (0.4033)	–0.0065 (–0.0434)
LPCNSDP	0.576 (1.344)	0.6312 (1.477)	0.610 (1.426)
URBANPOP	0.0234** (2.214)	0.0184** (2.123)	0.020** (2.265)
SIXTYPLUS	0.0649 (1.014)	0.048 (1.095)	0.059 (1.490)
HHACCESS	–0.4145 (–0.413)	–0.475 (–0.579)	–0.628 (–0.795)
LITERACY	–1.6739*** (–2.802)	–1.434** (–2.572)	–0.994 (–1.563)
HEALTHVUL	1.6941** (2.253)	1.895*** (3.115)	1.385** (2.227)
R^2	0.5195	0.5908	0.5510
$F(7,25)$	5.3062***	8.6209***	8.188***

Source: Authors' estimation based on secondary data.

Notes: Figures in parentheses are *t* ratios based on heteroscedasticity-robust standard errors and * indicates significance at a 10% level, ** indicates significance at a 5% level and *** indicates significance at a 1% level. HHACCESS: Household access index; HEALTHVUL: Index of health vulnerability; LCUMCASES: Log of state-level cumulative number of cases per million population; LITERACY: Literacy index; LPCNSDP: Log per capita NSDP; LPOPDENSITY: Log population density; SIXTYPLUS: Percentage of population aged 60 and above; URBANPOP: Percentage of urban population.

with populations that have the lowest health risks are Rajasthan, Bihar and Madhya Pradesh.

Results

Multivariate Regression Results

Factors Determining Spread of COVID-19 in the Two Waves of Infections

In Tables 2 and 3, we report results from cross-sectional OLS regression for three data points corresponding to the time span of the two waves. The results support the following findings. First, during the first wave of infections, URBANPOP, that is, the proportion of the urban population in a state, is positively associated with the diffusion of COVID-19 cases and is significant at a 5% level. For all three data points in the first wave, HEALTHVUL is also a factor contributing to diffusion. Larger at-risk populations contributed more to case counts. On the other hand, LITERACY is negatively associated with the diffusion of COVID-19 cases at the start and peak of the first wave and is not significant thereafter when cases started declining after the peak of the first wave. Second, during the second wave of infections, URBANPOP continues to be a significant explanatory variable in the diffusion of cases in April and May 2021. The proportion of 60-plus population, SIXTYPLUS, is a key determinant at all three data points in the second wave. LPCNSDP emerged as an important predictor in spatial diffusion during the peak of the second wave in May 2021 and June 2021. Thus, LITERACY and HEALTHVUL, which are important predictors for diffusion in the first wave of infections, are not significant in the second wave, while LPCNSDP and SIXTYPLUS emerge as key predictors. The results show that

Table 3. Results from Cross-sectional Ordinary Least Squares (OLS) Regression for the Time Period Corresponding to the Second Wave of Infections (April to June 2021)

	Dependent Variable LCUMCASES		
	OLS 4	OLS 5	OLS 6
Explanatory Variables	(April 2021)	(May 2021)	(June 2021)
CONSTANT	1.047 (0.250)	–0.191 (–0.050)	0.658 (0.1801)
LPOPDENSITY	–0.002 (–0.016)	0.041 (0.393)	0.0434 (0.4150)
LPCNSDP	0.584 (1.552)	0.748** (2.141)	0.6944** (2.074)
URBANPOP	0.0162** (2.292)	0.0114* (1.877)	0.0086 (1.500)
SIXTYPLUS	0.074** (2.127)	0.072** (2.173)	0.0738** (2.417)
HHACCESS	–0.500 (–0.7007)	–0.472 (–0.713)	–0.354 (–0.5402)
LITERACY	–0.019 (–0.033)	0.017 (0.034)	0.3024 (0.5919)
HEALTHVUL	0.461 (1.014)	–0.245 (–0.603)	–0.1780 (–0.4721)
R^2	0.6460	0.6747	0.6846
$F(7,25)$	14.312***	15.8809***	12.6655***

Source: Authors' estimation based on secondary data.

Notes: Figures in parentheses are *t* ratios based on heteroscedasticity-robust standard errors and * indicates significance at a 10% level, ** indicates significance at a 5% level and *** indicates significance at a 1% level. HHACCESS: Household access index; HEALTHVUL: Index of health vulnerability; LCUMCASES: Log of state-level cumulative number of cases per million population; LITERACY: Literacy index; LPCNSDP: Log per capita NSDP; LPOPDENSITY: Log population density; SIXTYPLUS: Percentage of population aged 60 and above; URBANPOP: Percentage of urban population.

Table 4. Results from Cross-sectional Ordinary Least Squares (OLS) Regression for the Time Period Corresponding to the Time Period Beyond the Second Wave (August to October 2021).

	Dependent Variable LCUMCASES		
	OLS 7	OLS 8	OLS 9
Explanatory Variables	(August 2021)	(September 2021)	(October 2021)
CONSTANT	4.0699 (1.529)	5.288* (2.047)	6.2302** (2.345)
LPOPDENSITY	–0.0853 (–0.852)	–0.116 (–1.178)	–0.1413 (–1.436)
LPCNSDP	0.4492* (1.745)	0.346 (1.374)	0.268 (1.042)
URBANPOP	0.0161*** (2.808)	0.018*** (3.135)	0.0200*** (3.282)
SIXTYPLUS	0.0601 (1.598)	0.0613 (1.527)	0.0614 (1.477)
HHACCESS	–0.6333 (–1.147)	–0.591 (–1.113)	–0.5420 (–1.031)
LITERACY	0.9818* (1.752)	1.1602* (1.995)	1.2962** (2.138)
HEALTHVUL	–0.038 (–0.0799)	–0.0914 (–0.181)	–0.1446 (–0.2701)
R^2	0.7077	0.7211	0.7269
$F(7,25)$	18.928***	22.7923***	22.9551***

Source: Authors' estimation based on secondary data.

Notes: Figures in parentheses are *t* ratios based on heteroscedasticity-robust standard errors and *indicates significance at a 10% level, **indicates significance at a 5% level and ***indicates significance at a 1% level. HEALTHVUL: Index of health vulnerability; HHACCESS: Household access index; LCUMCASES: Log of state-level cumulative number of cases per million population; LITERACY: Literacy index; LPCNSDP: Log per capita NSDP; LPOPDENSITY: Log population density; SIXTYPLUS: Percentage of population aged 60 and above; URBANPOP: Percentage of urban population.

the severe second wave of infections was largely driven by diffusion in richer states with a larger urban population and states with older populations.

Beyond the Second Wave: Dragging the Wave—Kerala and Northeast India

While cases plateaued in many states by June 2021, Kerala and some Northeast states reported continuing high levels of infections. The two states of Kerala and Mizoram together accounted for a significant 75% and 66% of total daily infections in India as of 18 August 2021 and 14 October 2021, respectively. To examine the determinants of diffusion after June 2021, cross-sectional regressions were carried out for three data points in the subsequent period. The results given in Table 4 indicate that URBANPOP is highly statistically significant at a 1% level as a predictor during August, September and October 2021. LITERACY is

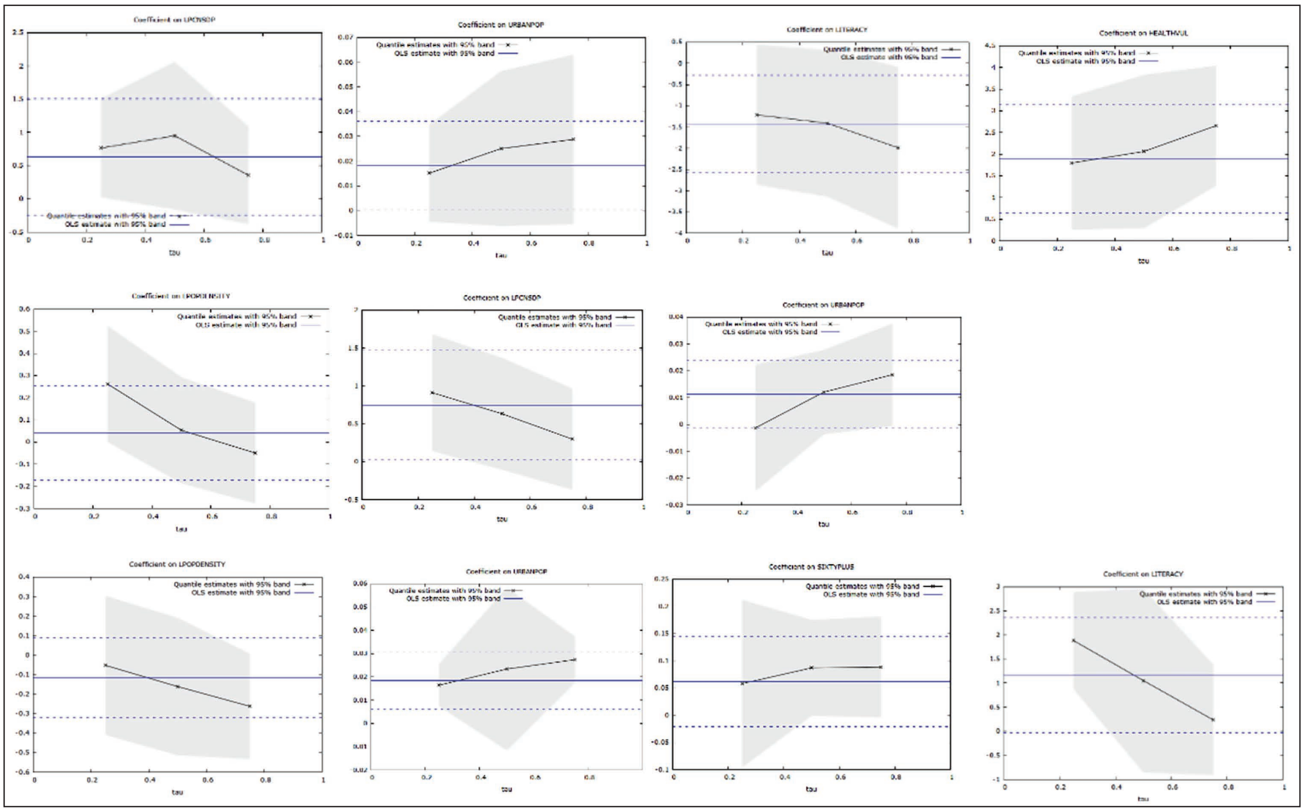

Figure 2. Quantile Regression Results for September 2020, May 2021 and September 2021(Top, Middle and Bottom Rows, Respectively).

Source: Authors' estimation based on secondary data.

Note: HEALTHVUL: Index of health vulnerability; HHACCESS: Household access index; LITERACY: Literacy index; LPCNSDP: Log per capita NSDP; LPOPDENSITY: Log population density; SIXTYPLUS: Percentage of population aged 60 and above; URBANPOP: Percentage of urban population.

significant at a 10% level in August and September 2021 and at a 5% level in October 2021. Interestingly, LITERACY variable enters with a positive sign for the regression results during these months, unlike in the first wave, when LITERACY was negatively associated with the diffusion of COVID-19 cases.

Quantile Regression Results

We performed an exploratory quantile regression exercise to examine if the effect of the predictor variables on cumulative cases per million population varies across the distribution. The analysis was performed for three data points reflecting different stages of the pandemic, that is, for September 2020, May 2021 (peak of the first and second waves, respectively) and September 2021 (with evidence of dragging the wave in a few states). The quantile regression results for estimated β coefficients for the 25th, 50th and 75th quantiles of the response variable for the three data points are given in Figure 2.

In Figure 2, we report quantile regression results of β coefficients of explanatory variables that are significant at least at a 10% level of significance. We observe that the role of explanatory variables that are significant in contributing to diffusion varies across different time points in the pandemic, a result that is similar to OLS results. There are some differences, however, in comparison with OLS results. LPOPDENSITY is not significant in the OLS estimation, but it is significant in the second wave at the 25th quantile and is significant but with a negative association in September 2021 at the 75th quantile. SIXTYPLUS is significant in September 2021 at the median and upper quantile. At the peak of the first wave, we find that LPCNSDP explains diffusion at the median and lower quantile, and HEALTHVUL is significant at all three quantiles. LITERACY has a negative sign and is significant at the higher quantile in the first wave but has a positive effect on diffusion in September 2021 in the lower quantile, with the effect decreasing at the higher quantile. Through the course of the pandemic, we find that the effect of urban population proportion is significant at higher quantiles of the distribution. We find that in all cases when the variables are significant, a clear pattern emerges from the three figures, with URBANPOP, SIXTYPLUS and HEALTHVUL having larger effects on the higher quantile than at the lower end of the distribution, while LPOPDENSITY, LPCNSDP and LITERACY have smaller effects at the higher end of the distribution as compared to the effect at the lower end of the

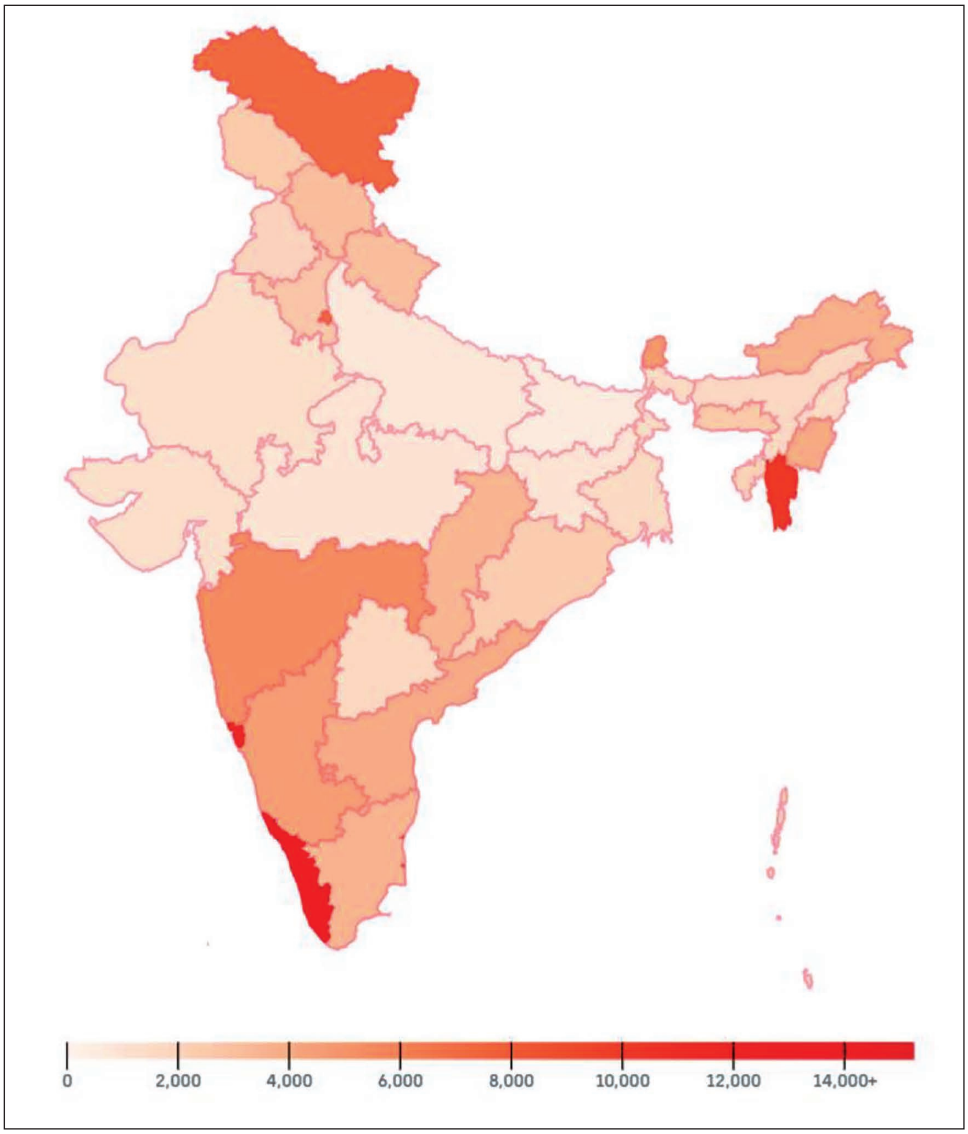

Figure 3. Map Showing Confirmed Cumulative COVID-19 Cases per 100,000 Population in States and Union Territories (UTs) of India as of 1 November 2021.

Source: Map accessed from https://www.covid19india.org/

distribution. The quantile regression results show that the association between cumulative cases per million and the explanatory variables is not uniform across the distribution.

Discussion

The spatial and temporal nature of COVID-19 diffusion is examined in Figures 3 and 4. Figure 3 gives the pattern of diffusion at the end of nearly 2 years of the COVID-19 pandemic (as of 1 November 2021). We observe that cumulative case counts per 100,000 are highest in Kerala, Goa and Mizoram (case counts are also high for the UT of Ladakh, which is not included in the study). This is followed by Puducherry, Delhi, Maharashtra, Chandigarh, Karnataka and Andhra Pradesh. Diffusion is also high in the Northeast states of Sikkim, Manipur and Arunachal Pradesh and in Tamil Nadu, Chhattisgarh, Uttarakhand and Himachal Pradesh. On the other hand, states with the lowest COVID-19 spread are Bihar, Uttar Pradesh, Madhya Pradesh, Jharkhand, Rajasthan, Gujarat and Telangana.

While Figure 3 gives the situation at the end of nearly 2 years of the pandemic, Figure 4 illustrates the differential nature of spatial and temporal diffusion of the pandemic. Figure 4 gives the bar graph of cumulative confirmed cases per million for states/UTs of India at the end of each of the two waves (October 2020 and June 2021) and for October 2021. By the end of the first wave of infections, there was widespread

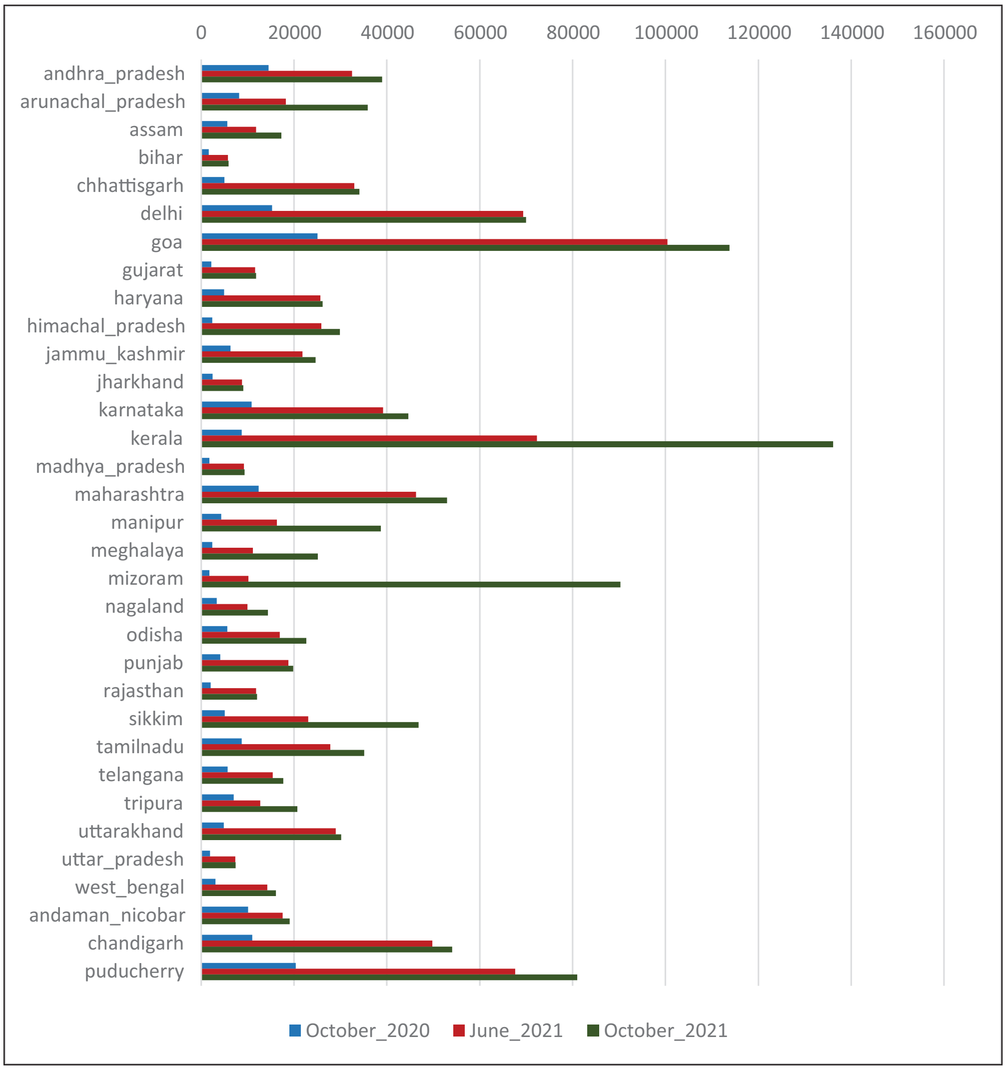

Figure 4. Cumulative Confirmed Cases per Million Population on 13 October 2020, 1 June 2021 and 14 October 2021.

Source: Authors' construction based on secondary data.

diffusion of the pandemic in Goa, Delhi, Andhra Pradesh, Puducherry, Maharashtra, Andaman and Nicobar Islands, Karnataka and Chandigarh. The second wave of infections from April to May 2021 was widespread and severe in intensity and states that witnessed a sharp spurt in cases were Goa, Kerala, Delhi, Maharashtra, Chandigarh, Puducherry, Karnataka and Andhra Pradesh. Chhattisgarh, Uttarakhand, Haryana and Himachal Pradesh also witnessed a sharp upsurge in infections in the second wave. While many states, including the northern states, had a plateauing of cases by June 2021 (Delhi, Haryana, Punjab, Uttar Pradesh, Madhya Pradesh, Rajasthan, Uttarakhand and Gujarat), the states of Kerala, Mizoram, Sikkim, Manipur, Meghalaya and Arunachal Pradesh showed continuing high infections after June 2021.

Kerala's remarkable success in containing the pandemic in the early stages was attributed to its effective public health system, prior experience in containing the Nipah virus in 2019 and ability to leverage public trust (Jalan & Sen, 2020; WHO, 2020). The subsequent puzzling high levels of infections are attributed to the state's high population density (around 900 per sq. km and the seventh highest in India), high urban population (fourth highest after Chandigarh, Delhi and Goa with over 70% urban), larger ageing population (the state has the highest 60-plus population of 17%) and also population with higher vulnerability due to prevalence of lifestyle diseases and comorbidities (FE Bureau, 2021). High case counts in Kerala are also attributed to targeted testing (Biswas & Pandey, 2021). Kerala's success in curbing the spread in the early stages of the pandemic also meant that a large proportion of the population did not have natural immunity from prior infection, making them susceptible at later stages of the pandemic (Department of Health & Family Welfare, Government of Kerala, 2021). Finally, it is argued that the infectious Delta variant, which was responsible for the massive spike in infections in the second wave, reached Kerala (and the Northeast states) late, contributing to high infection levels after June 2021.

The study results confirm the differential influence of demographic and socio-economic factors in the spatial and temporal diffusion of COVID-19. Urbanisation, literacy, per capita income and share of 60-plus population emerge as important predictors for sustained spread. The diffusion pattern across states observed in India is similar to results obtained from cross-country studies wherein developed countries with high incomes, high levels of urbanisation, high socio-economic development and a larger share of elderly populations exhibit higher case counts (Antonietti et al., 2021). The study shows urbanisation is a significant predictor of diffusion throughout the pandemic. Urban spaces are places of intense economic and social interactions, crowded living proximity and greater mobility and provide opportunities for greater transmission. The literacy variable is significant, especially in the later stages of the pandemic. At the end of 2 years of the pandemic, the states of Kerala, Goa and Mizoram had the highest diffusion. These states have many features that are predictors of diffusion. All three states have high urban population proportions and the highest literacy rates in the country (the three states are amongst the top four in the literacy index). Kerala also has the highest proportion of elderly population, while Goa has the highest per capita income. Other states/UTs with high diffusion also had features like high urbanisation and high income or literacy rates like Delhi, Maharashtra, Chandigarh, Puducherry and Karnataka. In contrast, Bihar, Uttar Pradesh, Madhya Pradesh, Rajasthan and Jharkhand have the lowest spread. These states are less urban, have lower per capita incomes, lower literacy levels and have younger populations. The demographic and socio-economic features of these states are thus predictors for lower case counts.

Some policy implications follow from the study. Testing for new epidemiological variants, identifying clusters of infections, implementing local containment measures and a comprehensive vaccination programme, strengthening public health infrastructure and addressing inequities in health systems are crucial in the management of a pandemic. The diversity of states in urbanisation, income, human development and demographic features also outlines a crucial role for sub-national policy responses in pandemic management. Urbanisation is a key predictor for diffusion, and large cities are likely to emerge as epicentres in any future outbreak of infectious diseases. The efficiency of local administration and public health response in containment strategies is crucial. The success of targeted containment measures involving community and local public health systems offers encouraging lessons for the future (Golechha, 2020). Demographic differences can also impact disease burden, with states that are ahead in the demographic transition like Kerala, Tamil Nadu, Himachal Pradesh, Punjab, Andhra Pradesh, Maharashtra and Goa, and having larger ageing and at-risk populations facing the possibility of greater pressure on health systems. Being ahead in the epidemiological transition, these states also have a larger disease burden from non-communicable diseases (NCDs) in contrast to states like Bihar, Uttar Pradesh, Rajasthan and Madhya Pradesh, which have a younger age structure and a lower disease burden from NCDs (ICMR, PHFI & IHME, 2017).

Conclusion

The results highlight the differential effect of demographic and socio-economic factors in state-level COVID-19 diffusion during the two waves and subsequently. Urbanisation, per capita income, literacy, proportion of the elderly population and proportion of the vulnerable population with health risks are key predictors for diffusion. While urbanisation is a key factor in diffusion through the course of the pandemic, per capita income and proportion of the elderly population are significant in the severe second wave of infections. Significant diffusion is observed in states with high income, large urban populations, high levels of literacy and a larger share of elderly populations like Kerala, Goa and Mizoram as well as Delhi, Chandigarh, Puducherry, Maharashtra, Sikkim and Karnataka. The second wave of infections saw wider diffusion in states with high per capita income and those with larger 60-plus populations. Diffusion is lowest in states like Bihar, Uttar Pradesh, Madhya Pradesh and Rajasthan, which have low per capita income, low urban populations, low literacy levels and a young age structure. The demographic and socio-economic features of these states predict lower case counts.

The literature indicates further possible extensions of the study like the role of local policies for containment as well as behavioural attitudes of the population towards compliance with COVID-19 restrictions and protocol. The issue of pandemic fatigue (Petherick et al., 2021) and introducing some measures of compliance (for example Barak et al., 2021) are of interest but beyond the scope of the present study.

The results from the study can serve to inform strategies for containment of future waves of a pandemic. Policy suggestions include the need for public health preparedness, especially in states with large urban population clusters and in states that are at later stages in the demographic transition, with larger ageing and vulnerable populations. Additionally, states in India that are ahead in the epidemiological transition and have health systems that are geared to NCDs may need to reorient their health systems to be able to better deal with health emergencies arising from sudden outbreaks of communicable diseases.

Appendix

1. Data sources for the study are listed below:
   (a) National Family Health Survey, India, NFHS-4, accessed from https://www.nfhsiips.in/nfhsuser/nfhs4.php
   (b) RBI-Handbook of Statistics on the Indian Economy, accessed from https://dbie.rbi.org.in/DBIE/dbie.rbi?site=publications
   (c) Elderly in India (2021), Ministry of Statistics and Programme Implementation, National Statistical Office, accessed from Download Reports | Ministry of Statistics and Program Implementation | Government of India
   (d) Report of the Technical Group on Population Projections (2020), National Commission on Population, Ministry of Health & Family Welfare, accessed from https://mohfw.gov.in/?q=documents/reports-archive
   (e) Census of India, 2011, Office of the Registrar General and Census Commissioner, India, accessed from https://censusindia.gov.in/census.website/
   (f) COVID-19 data, accessed from https://www.covid19india.org/
2. Three UTs of Ladakh, Lakshadweep, Daman and Diu and Dadra and Nagar Haveli were excluded from the analysis since not all relevant data were available.
3. Rather than choosing a single time point, the study aims to evaluate the impact of state-level factors through the time span of the two waves of infections and in the subsequent period until October 2021; hence, we chose three data points for each of the time spans.
4. The elderly population data for Goa and UTs of Puducherry, Chandigarh and Andaman and Nicobar Islands are for 2011, while for Northeast states, excluding Assam, total projected figures but not state-wise figures are available. For these states, appropriate extrapolation using 2011 figures was made. Data for Delhi are for NCT Delhi. Area figures for Telangana and UT of Jammu and Kashmir are obtained from respective state profiles.
5. The NFHS-4 data relate to 2015–2016, and the figures for Jammu and Kashmir represent the combined figures for the UT of Jammu and Kashmir and the UT of Ladakh.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

References

Andersen, L. M., Harden, S. R., Sugg, M. M., Runkle, J. D., & Lundquist, T. E. (2021). Analyzing the spatial determinants of local Covid-19 transmission in the United States. *Science of the Total Environment*, *754*, 142396. https://doi.org/10.1016/j.scitotenv.2020.142396

Antonietti, R., Falbo, P., & Fontini, F. (2021). The wealth of nations and the first wave of COVID-19 diffusion. *Italian Economic Journal*, *9*(1), 61–83. https://doi.org/10.1007/s40797-021-00174-z

Ascani, A., Faggian, A., & Montresor, S. (2021). The geography of COVID-19 and the structure of local economies: The case of Italy. *Journal of Regional Science*, *61*(2), 407–441. https://doi.org/10.1111/jors.12510

Balakrishnan, P., & Namboodhiry, S. K. (2021). The importance of investing in a public health system: Evidence from COVID-19 mortality. *Indian Economic Review*, *56*(1), 233–254. https://doi.org/10.1007/s41775-021-00116-7

Bambra, C., Riordan, R., Ford, J., & Matthews, F. (2020). The COVID-19 pandemic and health inequalities. *Journal of Epidemiology and Community Health*, *74*(11), 964–968. https://doi.org/10.1136/jech-2020-214401

Barak, N., Sommer, U., & Mualam, N. (2021). Urban attributes and the spread of COVID-19: The effects of density, compliance and socio-political factors in Israel. *Science of the Total Environment*, *793*, 148626. https://doi.org/10.1016/j.scitotenv.2021.148626

Basu, P., & Mazumder, R. (2021). Regional disparity of covid-19 infections: An investigation using state-level Indian data. *Indian Economic Review*, *56*(1), 215–232. https://doi.org/10.1007/s41775-021-00113-w

Biswas, S., & Pandey, V. (2021). *Covid-19: The mystery of rising infections in India's Kerala*. BBC. https://www.bbc.com/news/world-asia-india-58054124

Bryan, M. S., Sun, J., Jagai, J., Horton, D. E., Montgomery, A., Sargis, R., & Argos, M. (2021). Coronavirus disease 2019 (COVID-19) mortality and neighborhood characteristics in Chicago. *Annals of Epidemiology*, *56*, 47–54. https://doi.org/10.1016/j.annepidem.2020.10.011

Carozzi, F., Provenzano, S., & Roth, S. (2020). *Urban density and COVID-19* (IZA Discussion Paper. 2020 No. 13440). https://www.iza.org/publications/dp/13440/urban-density-and-covid-19

Davies, N. G., Klepac, P., Liu, Y., Prem, K., Jit, M., & Eggo, R. M. (2020). Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, *26*(8), 1205–1211. https://doi.org/10.1038/s41591-020-0962-9

Department of Health & Family Welfare, Government of Kerala. (2021, September). Kerala COVID-19 (SARS CoV2 IgG) seroprevalence study. https://dhs.kerala.gov.in/wp-content/uploads/2021/10/SS-Report-Aug-Sept-2021.pdf

Farzanegan, M. R., Gholipour, H. F., Feizi, M., Nunkoo, R., & Andargoli, A. E. (2020). International tourism and outbreak of coronavirus (COVID-19): A cross-country analysis. *Journal of Travel Research*, *60*(3), 687–692. https://doi.org/10.1177/0047287520931593

FE Bureau. (2021, November 21). Interview: Kerala is trying to report every Covid case, says state's Health Minister Veena George. *Financial Express*. https://www.financialexpress.com/lifestyle/health/interview-kerala-is-trying-to-report-every-covid-case-says-states-health-minister-veena-george/2373048/

Golechha, M. (2020). COVID-19 containment in Asia's largest urban slum Dharavi-Mumbai, India: Lessons for policymakers globally. *Journal of Urban Health*, *97*(6), 796–801. https://doi.org/10.1007/s11524-020-00474-2

Government of India. (2021). *Economic Survey. 2020-21*. Ministry of Finance. https://www.indiabudget.gov.in/economicsurvey/

Gupta, I. (2020). Relying on serendipity is not enough: Building a resilient health sector in India. *Indian Economic Review*, *55*(Suppl. 1), 125–147. https://doi.org/10.1007/s41775-020-00091-5

Indian Council of Medical Research, Public Health Foundation of India and Institute for Health Metrics and Evaluation. (2017).*India: Health of the nation's states. The India state-level disease burden initiative*. ICMR, PHFI and IHME. https://phfi.org/india-health-of-the-nations-states/

Jalan, J., & Sen, A. (2020). Containing a pandemic with public actions and public trust: The Kerala story. *Indian Economic Review*, *55*, 105–124. https://doi.org/10.1007/s41775-020-00087-1

Jeanne, L., Bourdin, S., Nadou, F., & Noiret, G. (2022). Economic globalization and the COVID-19 pandemic: Global spread and inequalities. *GeoJournal*, *88*(1), 1181–1188. https://doi.org/10.1007/s10708-022-10607-6

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, *15*(4), 143–156. https://doi.org/10.1257/jep.15.4.143

Mallapaty, S. (2021). India's massive COVID surge puzzles scientists. *Nature*, *592*(7856), 667–668.

Nathan, M. (2021). Critical Commentary: The city and the virus. *Urban Studies*, *60*(8), 1346–1364. https://doi.org/10.1177/00420980211058383

Nayyar, D. (2020). Lives, livelihoods and the economy: India in pandemic times. *The Indian Journal of Labour Economics*, *63*(Suppl. 1), 53–59. https://doi.org/10.1007/s41027-020-00256-z

Paez, A., Lopez, F. A., Menezes, T., Cavalcanti, R., & Pitta, M. G. D. R. (2020). A spatio-temporal analysis of the environmental correlates of COVID-19 incidence in Spain. *Geographical Analysis*, *53*(3), 397–421. https://doi.org/10.1111/gean.12241

Pandey, A., Prakash, A., Agur, R., & Maruvada, G. (2021). Determinants of COVID-19 pandemic in India: An exploratory study of Indian states and districts. *Journal of Social and Economic Development*, *23*(Suppl. 2), 248–279. https://doi.org/10.1007/s40847-021-00154-0

Petherick, A., Goldszmidt, R., Andrade, E. B., Furst, R., Hale, T., Pott, A., & Wood, A. (2021). A worldwide assessment of changes in adherence to COVID-19 protective behaviours and hypothesized pandemic fatigue. *Nature Human Behaviour*, *5*(9), 1145–1160. https://doi.org/10.1038/s41562-021-01181-x

Ray, D., & Subramanian, S. (2020). India's lockdown: An interim report. *Indian Economic Review*, *55*(Suppl. 1), 31–79. https://doi.org/10.1007/s41775-020-00094-2

Richards-Belle, A., Orzechowska, I., Gould, D. W., Thomas, K., Doidge, J. C., Mouncey, P. R., Christian, M. D., Shankar-Hari, M., Harrison, D. A., & Rowan, K. M. (2020). COVID-19 in critical care: epidemiology of the first epidemic wave across England, Wales and Northern Ireland. *Intensive Care Medicine*, *46*(11), 2035–2047. https://doi.org/10.1007/s00134-020-06267-0

Sigler, T., Mahmuda, S., Kimpton, A., Loginova, J., Wohland, P., Charles-Edwards, E., & Corcoran, J. (2021). The socio-spatial determinants of COVID-19 diffusion: The impact of globalisation, settlement characteristics and population. *Globalization and Health*, *17*(1), 56. https://doi.org/10.1186/s12992-021-00707-2

The Lancet. (2021). India's COVID-19 emergency. *Lancet (London, England)*, *397*(10286), 1683. https://doi.org/10.1016/S0140-6736(21)01052-7

WHO Centre for Health Development (Kobe, J., Weaver, S., Dai, D., Stauber, C., Luo, R., & Rothenberg, R.). (2014). *The urban health index: A handbook for its calculation and use*. World Health Organization. https://apps.who.int/iris/handle/10665/136839

WHO. (2020). *Responding to COVID-19—Lessons from Kerala*. https://www.who.int/india/news/feature-stories/detail/responding-to-covid-19---learnings-from-kerala

Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., Curtis, H. J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., McDonald, H. I., MacKenna, B., Tomlinson, L., Douglas, I. J., Rentsch, C. T., Mathur, R., Wong, A. Y. S., Grieve, R., … & Goldacre, B. (2020). Factors associated with COVID-19-related death using OpenSAFELY. *Nature*, *584*(7821), 430–436. https://doi.org/10.1038/s41586-020-2521-4