

# Spatial mapping and socio-demographic determinants of COVID-19 mortality in India

Ashish W. Khobragade<sup>1</sup>, Dilip D. Kadam<sup>2</sup>

<sup>1</sup>Department of Community and Family Medicine, All India Institute of Medical Sciences, Raipur, Chhattisgarh, <sup>2</sup>Department of Community Medicine, Seth G.S. Medical College, Mumbai, Maharashtra, India

## ABSTRACT

**Background:** COVID-19 is caused by SARS-CoV-2. The first case of COVID-19 was detected in Wuhan city of China in December 2019. Geographic information system (GIS) mapping is important for the surveillance of infectious diseases. **Objectives:** The objectives of the study are to map spatially total cases and case fatality rate of COVID-19 and to build a linear regression model for mortality based on socio-demographic variables. **Methodology:** We plotted the epidemiological data of COVID-19 of Indian states as on 11<sup>th</sup> May 2021 using the Q-GIS software. We used socio-demographic variables as the predictors of COVID-19 mortality and developed a linear regression model. **Results:** Adjusted R-squared in linear regression model based on socio-demographic variables for COVID-19 deaths is 0.82. **Conclusions:** There are spatial variations in COVID-19 cases and deaths.

**Keywords:** COVID-19, GIS, mortality, regression, socio-demographic, spatial

## Introduction

COVID-19 is a viral disease caused by a coronavirus (SARS-CoV-2). Total confirmed cases of COVID-19 in the world as on 11<sup>th</sup> May 2021 are 159 million. Till date, 3.1 million deaths have been reported globally due to it. More than 14% of these cases and 7% of the total deaths of the world are from India as on 11<sup>th</sup> May 2021.<sup>[1,2]</sup>

John Snow, the father of modern epidemiology, in the 19<sup>th</sup> century, first used spot maps to plot the cases of Cholera in London. His geographical analysis helped to trace the source of infection even when no *Vibrio Cholerae* bacterium was discovered.<sup>[3]</sup> Spatial epidemiology is the core branch of infectious disease epidemiology. Real-time mapping of

infectious disease data is important to track the current situation of the disease.

India is a diverse country where geographical and socio-demographic conditions vary. Many socio-demographic factors such as poor socio-economic condition, overcrowding, rapid urbanisation and migration are associated indirectly with various infectious diseases. These factors are neglected most of the time while planning health. Evaluation of these factors is important to know the epidemiology of the disease in detail and to prevent and minimise the effects of potential future pandemics. Primary health care physicians can manage the cases of COVID-19 considering the various socio-demographic factors associated with it.

## Objectives

1. To map spatially total cases and case fatality rate (CFR) of COVID-19 in India
2. To develop the linear regression model based on socio-demographic determinants associated with COVID-19 mortality.

**Address for correspondence:** Dr. Ashish W. Khobragade, Department of Community and Family Medicine, All India Institute of Medical Sciences, Raipur, Chhattisgarh, India.  
E-mail: aw\_k2008@rediffmail.com

Received: 16-05-2021

Revised: 15-07-2021

Accepted: 26-07-2021

Published: 29-11-2021

### Access this article online

#### Quick Response Code:



Website:  
www.jfmpc.com

DOI:  
10.4103/jfmpc.jfmpc\_903\_21

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow\_reprints@wolterskluwer.com

**How to cite this article:** Khobragade AW, Kadam DD. Spatial mapping and socio-demographic determinants of COVID-19 mortality in India. J Family Med Prim Care 2021;10:4200-4.

## Methodology

More than one year has passed since the 1<sup>st</sup> case of COVID-19 was reported from India. We have compiled data regarding COVID-19 from the official website of the Government of India as on 11<sup>th</sup> May 2021.<sup>[2]</sup> Compiled data was extracted in an excel sheet. The number of total cases, total deaths due to COVID-19, current active cases and recovered cases reported till 11<sup>th</sup> 2021 were entered state-wise in the excel sheet. CFR was calculated from the number of deaths and total cases. The recovery rate was also calculated for each state. The CFR is directly related to virulence and indicates the killing power of the disease.

CFR = Total number of deaths due to COVID-19/Total number of COVID-19 cases × 100.

Recovery rate = Total number of recovered cases of COVID-19/ Total number of COVID-19 cases × 100.

After calculating epidemiological parameters, such as CFR and recovery rate, we used Q-GIS software for geographical mapping of this epidemiological data.

All states and union territories were included in geographical mapping. Two such graphs were created by geographic information system (GIS) mapping of total cases and CFR.

Basic socio-demographic data were collected for each state. Data regarding the total state-wise projected population as on December 2020 was extracted.<sup>[4]</sup> The density of population per square kilometre and literacy rate according to the 2011 census was collected.<sup>[5]</sup> Percentage of urban population in the state, percentage of peoples below poverty line (BPL) according to Tendulkar Committee, percentage of persons more than 60 years residing in the state and per capita health expenditure spent on health were compiled.<sup>[5-7]</sup> Immigration rates were also obtained.<sup>[8]</sup> Other states were excluded as some data was not available for that particular state. The projected population as on December 2020 was taken into consideration to calculate COVID-19 morbidity and mortality rates. We used the ordinary least square method to develop a linear regression model using these socio-demographic variables as the predictors of COVID-19 mortality. Data of 19 states were used to develop the linear regression model.

## Statistical analysis

Statistical analysis was done by using R software version 3.6.1.

## Results

Maharashtra, Karnataka and Kerala are high-burden states at present where more than 1.9 million COVID-19 cases have been reported [Figure 1]. CFR is the highest in Punjab, followed by Sikkim, Uttarakhand, Maharashtra and Delhi [Figure 2].

Delhi has the highest number of cases and deaths of COVID-19 per 100000 general (projected) population [Table 1]. Rate of

immigration and per capita health expenditure are the significant predictor of COVID-19 deaths. The only negative predictor of COVID-19 mortality is the literacy rate. Multiple R-squared for the model of COVID-19 mortality based on various socio-demographic variables is 0.89. Adjusted R-squared in the model of COVID-19 deaths is 0.82. The model could explain 82% variations in COVID-19 deaths [Table 2].

## Discussion

We plotted point source data regarding COVID-19 total cases and CFR of Indian states and union territories using Q-GIS software. We included 19 states in the linear regression model for socio-demographic factors associated with COVID-19 mortality per 100000 general population projected for December 2020.

**Table 1: Cases and deaths of COVID-19 per 100000 general population as on 11<sup>th</sup> May 2021 in various states of India for which linear regression model is used**

States	Cases per 100000 population*	Deaths per 100000 population*
Assam	837.39	4.92
Bihar	482.09	2.69
Chhattisgarh	2932.93	36.49
Delhi	7141.38	105.09
Gujarat	1084.36	13.33
Haryana	2228.76	20.44
Himachal Pradesh	1822.10	26.01
Jharkhand	757.97	10.32
Karnataka	2921.26	28.67
Kerala	5406.57	16.47
Madhya Pradesh	798.37	7.62
Maharashtra	4173.13	62.04
Odisha	1175.40	4.74
Punjab	1495.20	35.51
Rajasthan	954.18	7.19
Tamil Nadu	1810.40	20.40
Uttarakhand	2220.40	34.63
Uttar Pradesh	640.97	6.62
West Bengal	1016.58	12.51

\*Projected December 2020 population was considered

**Table 2: Multiple linear regression model for COVID-19 deaths per 100000 general population based on socio-demographic variables**

Independent variables	Coefficient	Std error	t statistics	P (t)
Intercept	-21.32	49.11	-0.43	0.67
Population density/km <sup>2</sup>	0.001	0.002	0.67	0.52
Urban population (%)	0.67	0.36	1.87	0.09
Immigration (%)	2.67	0.78	3.42	0.005
Literacy (%)	-0.78	0.97	-0.81	0.44
Population >60 years (%)	2.75	3.09	0.89	0.39
BPL (%)	0.59	0.40	1.46	0.17
Per capita health expenditure (INR)	0.03	0.008	3.11	0.009

Residual standard error: 10.61 on 11 degrees of freedom. Multiple R<sup>2</sup>: 0.89, Adjusted R<sup>2</sup>: 0.82. F-statistic: 12.41 on 7 and 11 DF, P: 0.0002

Maharashtra is the only state where more than 5 million cumulative confirmed cases of COVID-19 have been reported as on 11<sup>th</sup> May 2021. [Figures 1 and 3]. Currently burden of the disease is found to be more in the states of Maharashtra, Karnataka, Kerala, Uttar Pradesh and Tamil Nadu. Maximum active cases are seen in Maharashtra, Karnataka, Kerala, Uttar Pradesh and Rajasthan. Punjab is the only state where CFR is more than two percentages whereas national CFR is 1.09 as on 11<sup>th</sup> May 2021. Among the states, lowest CFR is reported from Mizoram, followed by Kerala, Arunachal Pradesh and Odisha. These four states have reported CFR less than 0.5%. Recovery rate is more than 70% in all the states and union territories except Uttarakhand as on 11<sup>th</sup> May 2021. Kerala is the only state where despite having position among the top five high-burden states, it has very less CFR. Kerala has favourable socio-demographic indicators.

There are geographical disparities in morbidity, mortality, CFR as well as in the recovery rate of COVID-19 cases in different Indian states. To understand the reasons behind these differences, we developed a linear regression model taking socio-demographic factors as the predictors of mortality.

As a finding, literacy rate is the only negative predictor of COVID-19 mortality. Similar findings were obtained in a study conducted in Brazil, where deaths among hospitalised patients are highest among illiterates.<sup>[9]</sup> In our study also, those states having a high literacy rate have less COVID-19 mortality. Whereas in a study conducted in the United States, minority population and BPL population are the negative predictors of COVID-19 cases. However, only minority population is the negative predictor of mortality found in a study.<sup>[10]</sup> In a study conducted in Sweden, more hazard was observed among primary educated than secondary and tertiary educated peoples. They also noticed a higher risk of mortality among immigrants.<sup>[11]</sup> In a similar study in the United States, the odds of COVID-19 mortality is significantly higher after 60 years of age.<sup>[12]</sup> United Nations is a developed nation and India is low- and middle-income country. Hence, socio-demographic conditions vary in these

two countries. Also, socio-demographic factors associated with COVID-19 are different.

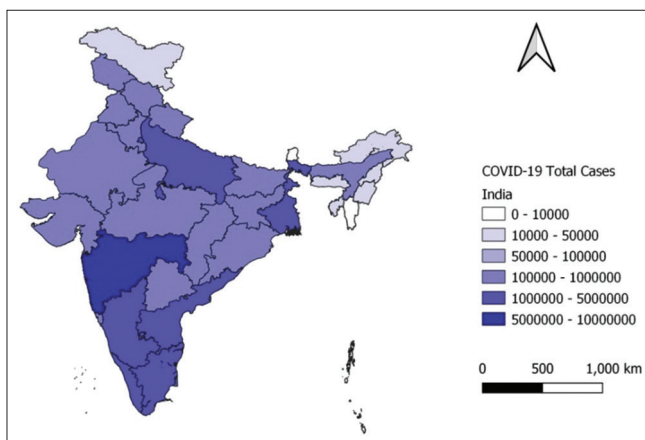
Public health expenditure and rate of immigration are significant predictors of COVID-19 case burden and deaths. In a study conducted in a tertiary hospital in Spain, more than 60% of deaths are reported in the age of >60 years.<sup>[13]</sup> Socio-economic conditions and urbanisation are also associated factors. The density of population per square kilometre is not a significant predictor.

In a study conducted by Ramirez *et al.* in Colorado, it is observed that migration is negatively correlated with COVID-19 deaths. Further, the density of population is correlated significantly with the mortality of COVID-19.<sup>[14]</sup> In a GIS-based study in the United States, poor socio-economic condition is identified as a predictor of COVID-19 incidence rate.<sup>[15]</sup>

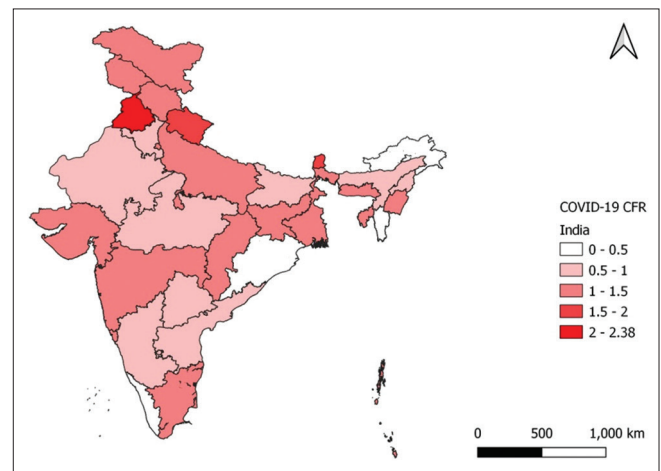
State or district authorities can delineate the areas requiring lockdown with the help of real-time GIS mapping.<sup>[16]</sup> A similar GIS-based study from China in the early stage of the pandemic has shown spatial dynamics of COVID-19.<sup>[17]</sup> They also stressed the importance of plotting recovery rate and mortality spatially to understand factors affecting it.

Manual labourers and the unemployed have the highest COVID-19 mortality rates in Ecuador.<sup>[18]</sup> Similarly, vulnerable populations, such as minorities, also have a high death rate in the United States.<sup>[19]</sup> Migrants are also one of the vulnerable populations in India. The most important reason for migration in India is employment. Hence, there is a need to provide job opportunities to these migrants in their own states and to reduce the rate of urbanisation.

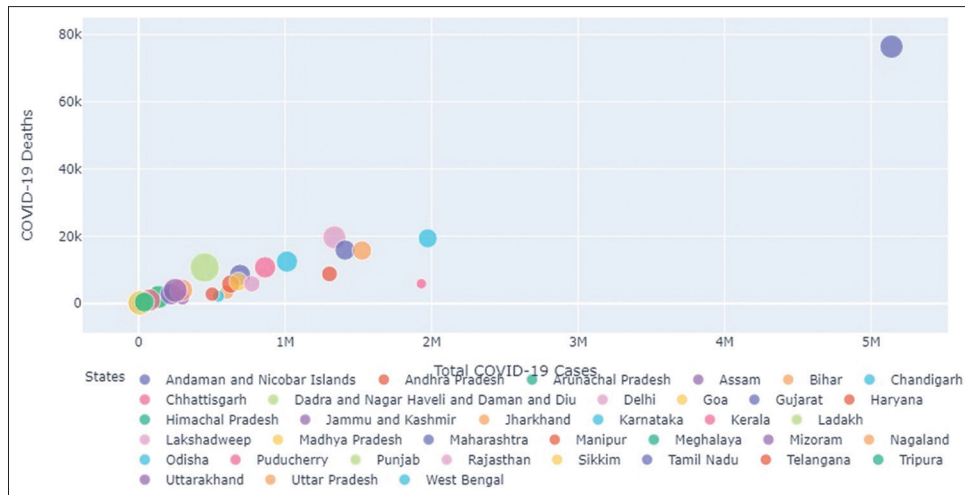
In a similar regression model study conducted in Europe, per capita income is strongly associated with COVID-19 mortality.<sup>[20]</sup> Income and health deprivation and ethnicity are also associated with COVID-19 mortality in England.<sup>[21]</sup> Non-whites have a higher risk of COVID-19 mortality in the United Kingdom.<sup>[22]</sup>



**Figure 1:** Spatial map showing state-wise total confirmed cases of COVID-19 in India as on 11<sup>th</sup> May 2021



**Figure 2:** Spatial map showing state-wise CFR of COVID-19 in India as on 11<sup>th</sup> May 2021



**Figure 3:** Scatter plot showing state-wise COVID-19 total cases and deaths in India as on 11<sup>th</sup> May 2021 (size of the bubble is proportional to the magnitude of CFR)

India is also a diverse country where people from different religions, castes and ethnicity live. Limited health infrastructure is associated with a greater risk of COVID-19 found in a study conducted in China.<sup>[23]</sup> In Indian states like Kerala, where health care indicators are good, COVID-19 mortality is more diminutive. In a spatial analysis conducted in Oman, purchasing power and population density are associated with COVID-19.<sup>[24]</sup> Illiteracy and lower socio-economic status have a substantial impact on COVID-19 mortality in India.<sup>[25]</sup>

GIS-based mapping should be a part of routine surveillance activities to understand the various spatial factors associated with the new emerging pandemic disease like COVID-19.

COVID-19 pandemic taught us a lesson to strengthen our health care system concerning infrastructure and workforce and capacity building. Health centres should be strengthened to equip them to manage the pandemics. More public health expenditure is required to strengthen our public health care system. Only then we can face future impending pandemics. Rehabilitation of the migrants should be done promptly. The holistic development of humankind is the need of the hour.

### Conclusions

There are spatial variations in COVID-19 cases as well as CFR. Education is the indirect negative predictor of COVID-19 mortality. Per capita public health expenditure and immigration are the significant independent variables.

### Recommendation

Real-time spatial mapping should be done regularly to know how the disease is progressing and to prioritise the resources. There is a need to improve indirect predictors like literacy and public health expenditure. There is also a need for occupational rehabilitation of migrants.

### Key messages

GIS mapping should be a routine part of surveillance in the containment of pandemic. Indirect predictors such as literacy, migration, urbanisation and socio-economic conditions associated with emerging infectious diseases like COVID-19 should not be ignored.

### Ethical permission

We used secondary data for analysis in this study which is freely available in the public domain. There is no direct involvement of human beings.

### Acknowledgments

The authors are thankful to the Ministry of Health and Family Welfare, Government of India for the data used in this research study.

### Financial support and sponsorship

Nil.

### Conflicts of interest

There are no conflicts of interest.

### References

1. WHO. Coronavirus (COVID-19) Dashboard. Available from: <https://covid19.who.int/>. [Last accessed on 2021 May 11].
2. MOHFW, Govt. of India. COVID-19 statewise status. Available from: <https://www.mohfw.gov.in/>. [Last accessed on 2021 May 11].
3. Shiode N, Shiode S, Rod-Thatcher E, Rana S, Vinten-Johansen P. The mortality rates and the space-time patterns of John Snow's cholera epidemic map. *Int J Health Geogr* 2015;14:21.
4. Govt. of India. Unique identification authority of India. State/UT wise Aadhaar Saturation (Overall) - All Age Groups. Available from: <https://uidai.gov.in/images/state-wise>

- aadhaar-saturation.pdf. [Last accessed on 2021 May 11].
5. Registrar General, MHA, Govt. of India. 2011 census data. Available from: <https://censusindia.gov.in/2011-common/censusdata2011.html>. [Last accessed on 2021 May 11].
  6. CBHI, MOHFW, Govt. of India. National Health Profile 2020. Available from: <http://www.cbhidghs.nic.in/showfile.php?lid=1147>. [Last accessed on 2021 May 11].
  7. Registrar General, MHA, Govt. of India. SRS report 2018. Available from: [https://censusindia.gov.in/vital\\_statistics/SRS\\_Report\\_2018/12..SRS\\_Statistical\\_Report-Detailed\\_Tables\\_2018.pdf](https://censusindia.gov.in/vital_statistics/SRS_Report_2018/12..SRS_Statistical_Report-Detailed_Tables_2018.pdf). [Last accessed on 2021 May 11].
  8. Registrar General, MHA, Govt. of India. Migration Tables. Census of India. 2001;1:1-35. Available from: [https://censusindia.gov.in/Data\\_Products/Data\\_Highlights/Data\\_Highlights\\_link/data\\_highlights\\_D1D2D3.pdf](https://censusindia.gov.in/Data_Products/Data_Highlights/Data_Highlights_link/data_highlights_D1D2D3.pdf). [Last accessed on 2021 May 11].
  9. Peres IT, Bastos LSL, Gelli JGM, Marchesi JF, Dantas LF, Antunes BBP, *et al.* Sociodemographic factors associated with COVID-19 in-hospital mortality in Brazil. *Public Health* 2021;192:15-20.
  10. Zhang CH, Schwartz GG. Spatial disparities in Coronavirus incidence and mortality in the United States: An ecological analysis as of May 2020. *J Rural Heal* 2020;36:433-45.
  11. Drefahl S, Wallace M, Mussino E, Aradhya S, Kolk M, Brandén M, *et al.* Socio-demographic risk factors of COVID-19 deaths in Sweden: A nationwide register study. *Nat Commun* 11, 5097 (2020).
  12. Wang A-L, Zhong X, Hurd Y. Comorbidity and sociodemographic determinants in COVID-19 mortality in an US urban healthcare system. medRxiv. doi: <https://doi.org/10.1101/2020.060.11.20128926>.
  13. Rivera-Izquierdo M, Del Carmen Valero-Ubierna M, R-delAmo JL, Fernández-García MÁ, Martínez-Diz S, Tahery-Mahmoud A, *et al.* Sociodemographic, clinical and laboratory factors on admission associated with COVID-19 mortality in hospitalized patients: A retrospective observational study. *PLoS One* 2020;15:e0235107.
  14. Ramírez IJ, Lee J. COVID-19 emergence and social and health determinants in Colorado: A rapid spatial analysis. *Int J Environ Res Public Health* 2020;17:3856.
  15. Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Sci Total Environ* 2020;728:138884.
  16. Sarwar S, Waheed R, Sarwar S, Khan A. COVID-19 challenges to Pakistan: Is GIS analysis useful to draw solutions? *Sci Total Environ* 2020;730:139089.
  17. Kang D, Choi H, Kim JH, Choi J. Spatial epidemic dynamics of the COVID-19 outbreak in China. *Int J Infect Dis* 2020;94:96-102.
  18. Ortiz-Prado E, Simbaña-Rivera K, Barreno LG, Diaz AM, Barreto A, Moyano C, *et al.* Epidemiological, socio-demographic and clinical features of the early phase of the COVID-19 epidemic in Ecuador. *PLoS Negl Trop Dis* 2021;15:e0008958.
  19. Karmakar M, Lantz PM, Tipirneni R. Association of social and demographic factors with COVID-19 incidence and death rates in the US. *JAMA Netw Open* 2021;4:e2036462.
  20. Sannigrahi S, Pilla F, Basu B, Basu AS, Molter A. Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustain Cities Soc* 2020;62:102418.
  21. Congdon P. COVID-19 mortality in English neighborhoods: The relative role of socioeconomic and environmental factors. *J* 2021;4:131-46.
  22. Bhaskaran K, Bacon S, Evans SJ, Bates CJ, Rentsch CT, MacKenna B, *et al.* Factors associated with deaths due to COVID-19 versus other causes: Population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. *Lancet Reg Health Eur* 2021;6:100109.
  23. Zhang H, Liu Y, Chen F, Mi B, Zeng L, Pei L. The effect of sociodemographic factors on COVID-19 incidence of 342 cities in China: A geographically weighted regression model analysis. *BMC Infect Dis* 2021;21:428.
  24. Al Kindi KM, Al-Mawali A, Akharusi A, Alshukaili D, Alnasiri N, Al-Awadhi T, *et al.* Demographic and socioeconomic determinants of COVID-19 across Oman-A geospatial modelling approach. *Geospat Health* 2021;16:145-60.
  25. Sharma AK, Gupta R, Baig VN, Singh T, Chakraborty S, Sunda JP, *et al.* Socioeconomic status and COVID-19 related outcomes in India hospital based study. medRxiv. doi: <https://doi.org/10.1101/2021.05.17.21257364>.