Potential Factors of Transmission and Spread of COVID 19 in Indian States and Districts

Sayeed Unisa, Balram Rai, Shreyans Rai, Abhishek Saraswat, Swagata Mondal, Preeti Dhillon, Laxmi Kant Dwivedi, Suryakant Yadav, and Chander Shekhar

An Analytical Paper for Policy Makers

International Institute for Population Sciences, Mumbai
(www.iipsindia.ac.in)

May 27, 2020

IIPS Analytical Series on Covid 19:

Paper 15: Potential Factors of Transmission and Spread of COVID 19 in Indian States and Districts

(This work has not been peer-reviewed. If you find any errors or have clarifications, please email the authors.)

For research and updates on Covid-19, visit:
https://www.iipsindia.ac.in/content/covid-19-information

Authors from the IIPS Team of Researchers on Estimation and Projection of COVID 19 Cases
Dr. Sayeed Unisa, Professor and Head, Department of Mathematical Demography and Statistics (unisa@iips.net)
Mr. Balram Rai, Research Scholar (balramrai009@gmail.com)
Mr. Shreyans Rai, Research Scholar (shreyansrai46@gmail.com)
Mr. Abhishek Saraswat, Data Analyst (abhi.srswt@gmail.com)
Ms. Swagata Mondal, Master in Biostatistics and Demography (swagata9011@gmail.com)
Dr. Preeti Dhillon, Asst. Professor, Department of Mathematical Demography and Statistics (pdhillon@iips.net)
Dr. Laxmi Kant Dwivedi, Asst. Professor, Department of Mathematical Demography and Statistics (laxmikant@iips.net)
Dr. Suryakant Yadav, Asst. Professor, Department of Development Studies (suryakant_yadav@iips.net)
Dr. Chander Shekhar, Professor, Department of Fertility Studies(buddhab@iips.net)

Potential Factors of Transmission and Spread of COVID 19 in Indian States and Districts

Sayeed Unisa, Balram Rai, Shreyans Rai, Abhishek Saraswat, Swagata Mondal, Preeti Dhillon, Laxmi Kant Dwivedi, Suryakant Yadav, and Chander Shekhar

Background: On 30th January 2020, importation of the first case of novel coronavirus (SARS-CoV-2) popularly known as COVID 19 was reported in Kerala, India. The first case in India was from Wuhan, China. The Government of India has immediately given the travel advisory about the passengers arriving from China. Only from 2nd March, other states had reported the cases. The spread of the virus has been uneven in the states of India. Previous evidence suggests the role of urbanization and the known demographic, household level, and economic differences in the spread of this contagious disease. In this paper, an attempt is made to analyze the factors that have a significant role in different stages of the spread COVID-19. The identification of potential factors could benefit in the formulation of strategies to improve future health problems.

Methods: This study, used data of COVID cases from the data-sharing portal *covid19India.org.* This website provides data on all the aspects of the epidemic of COVID-19 in India and its states. We compiled the number of COVID19 cases for three time periods, i.e., 23rd March – 1st April, 2nd April – 15th April, and 16th April – 3rd May 2020 and estimated the doubling time of the epidemic by the exponential method. For the district-level analysis, we considered the cumulative infected case until 13th May 2020. We have also used raw data of the patient for travel history analysis from the same site. For socio-economic and other indicators, we have used different sources, namely Census of India, NFHS-4, emigration data from published sources, health facilities data from National Health Profile Report (CBHI, 2019). Further, we performed a reliability analysis of confirmed cases.

Findings: We found initial reporting of cases of the disease in the states with metropolitan cities and high emigration. In the first phase of analysis, the doubling rate is mainly the function of the importation of cases in the urban areas. In the second phase, isolation of suspected cases at home, household-level factors were examined. Large family size and a large number of persons sleeping in the room and the number of rooms (less than 2) in the households have affected the doubling time negatively and suggest the spread of the disease from the suspected cases at crowded homes. In the third phase, when tracing, testing, and treating are emphasized by the Government, health facilities play an essential role. Based on cluster analysis, made identification of districts in different zones.

Conclusions: In the case of pandemic without vaccine and known origin, immediate international travel restriction, and complete screening of passengers at the airports may reduce the transmission. In the large metropolitan cities with high density and crowding in the household, the best alternative is to strengthen and prepare the isolation and quarantine facilities in India.

1.Introduction:

With increasing trade, travel, population density, human displacement, migration, and deforestation, however, as well as climate change, a new era in the risk of epidemics has begun. The number and diversity of epidemic events have been increasing over the past 30 years, a trend that is expected to intensify (World Economic Forum, 2018). The spread of infectious disease is considered a top global economic risk (Sands et al., 2016). Recent economic study suggests that the annual global cost of moderately-severe to severe pandemics is roughly $570 billion, or 0.7 percent of global income (Fan et al., 2015). The cost of a severe pandemic like the 1918 influenza pandemic could total as much as five percent of global GDP (World Bank, 2015). Hence, it is essential to examine the factors responsible for the transmission of disease at the country level.

In Wuhan, several clusters of patients with pneumonia of unknown etiology were detected by late December 2019 (WHO, 2020). A novel coronavirus was identified as the causative pathogen and was named as "COVID - 19" by WHO on Feb 11, 2020. Coronaviruses are a group of viruses that can cause disease in both animals and humans. The virus causes coronavirus disease 19 (COVID-19). It spreads through person-to-person transmission primarily through respiratory droplets produced when an infected person coughs or sneezes in the close presence of other people and causing flu-like symptoms and, in extreme cases, deaths (WebMD, 2020). The most common symptoms include fever, dry cough, and tiredness; hence it is difficult to diagnose without proper testing.

Since mid-January 2020, it has rapidly spread throughout China and other countries. Air travel and foreign relations have resulted in a spread of pandemic worldwide (Bpgoch et al., 2020; Grais et al., 2003.) Apart from China, countries like the USA, Italy, Spain, and the UK are still the worst hit (WHO Situation Report 103, May 2020). These countries have one of the best health-care systems in the world.

On the 30th of January 2020, importation of the first case of novel coronavirus (SARS-CoV-2) was reported in Kerala, India (PIB GoI, Jan 2020). The first case was from Wuhan, China. The Government of India has immediately given the travel advisory about the passengers arriving from China, followed by restriction of visas for China and also Italy, Iran, Japan, and South

Korea. Other states in india had reported the cases only from 2nd March 2020. In a country with a billion-plus population and not a very sound health system according to the Global Health Security Index (GHS, 2019), it is vital to learn the role of urbanization, economic condition, and other factors in the spread of the disease.

The entire country has been put under lockdown by the Government of India to control the spread of the virus. All international airports and seaports were closed. Hence, this has stopped the importation of cases from different countries. Furthermore, the Government of India had mandated the screening of passengers on their arrival in the country during March 2020. Only those who had symptoms were sent for quarantine or further testing. All other passengers without symptoms, including asymptomatic or in the incubation period or exposed during travel, were advised to keep themselves isolated in the homes (MoHFW, 6 March 2020, Air, 2020).

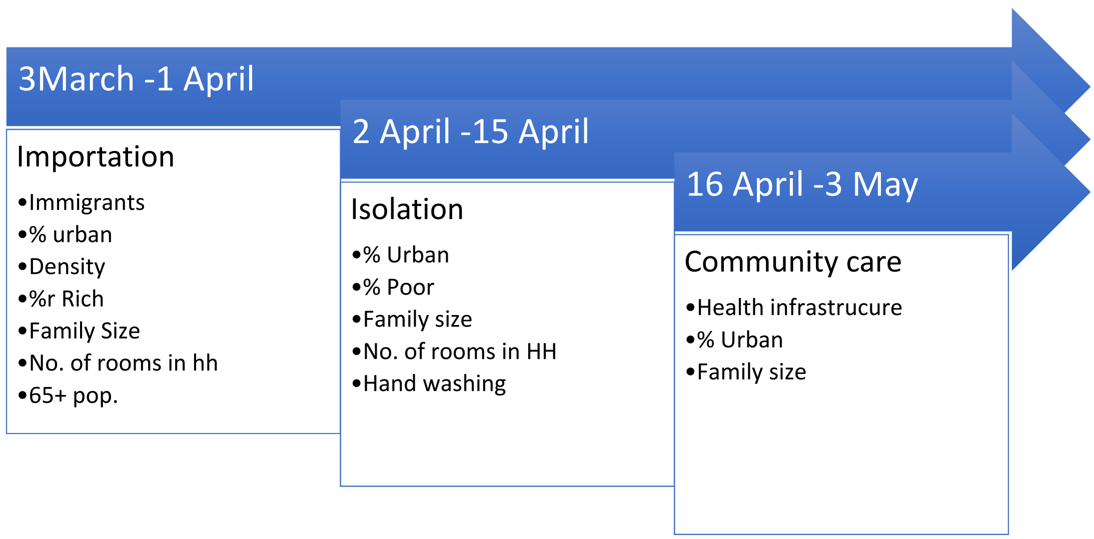
Social distancing would help in containing the spread as the virus spreads from human to human respiratory aspirates, droplets. Quarantine, isolation, and social distancing would reduce human interactions (Liu et al., 2020; Wilder-Smith & Freedman, 2020). Owing to high population density and reasonable health care, it is essential to control the spread of disease as early as possible. Since there is no vaccine or specific antiviral treatment for COVID-19, there are several suggested strategies in control like social distancing and personal hygiene practices.

1.1Research Context and Added Value of this Study:

Most of the work on COVID 19 are based on models of projecting the infection (Jung et al. 2020; Kucharshi et al.,2020; Riou, & Althaus, 2020; Wu et al., 2020). Some of the studies from Europe, China, Philippines have considered the association of socio-economic and urbanization as factors associated with the spread (Chen et al., 2020; Djalante et al., 2020; Mogi & Spijker, 2020; Nicola et al., 2020; Qiu et al., 2020; Stier et al., 2020). However, in the case of India, not much work is done to relate the spread in association with socio-economic indicators. India is a vast country with a large population, and the spread of transmission according to available confirmed cases is not uniform all over the country. Hence, it is crucial to examine and quantify the association of socio-economic factors in meditating the transmission of the virus under the different regimes of lockdown periods at the state level. It is also essential to identify the cluster of districts that have a high risk of cases based on the characteristics. This study will help to identify behavioral and socio-economic risk factors for infection.

2. Scenario and Selection of Instrumental Variables:

We created three scenarios representing the evolution of the pandemic and the response of governments. For each scenario, we have selected instrumental variables that represent the real-world possibilities in the pandemic spread. We named the three phases as "Importation," "Isolation," and "Community Care."



In the first phase, the role of health services and household level variables are not necessary as the cases were coming from foreign countries. Patient data available on COVID19India.org, till April 19, shows that only for 2691 patients, detailed notes are available about the transmission of disease. We found from this data that 21 percent confirmed cases are from international destinations. According to a travel advisory from MoHFW, international passengers were advised to stay at home in isolation (MoHFW, March 6, 2020;). Hence, in this phase, we have taken duration till the first lockdown and one week more to take account of the incubation period after the closing of the ports of arrival (Airport Industry Review, 2020; PIB, March 24, 2020). Potential factors for the spread are based on the available evidence from the earlier studies (Chinazzi et al., 2020; Li et al., 2020; Mogi & Spijker, 2020; Priyadarsini & Suresh, 2020; Qiu et al., 2020). In scenario 2, travel, offices, schools, and all recreational centers were closed. Hence, there was a possibility of the spread of disease to persons within the premises of the house and among close relatives of the international passenger who was asymptomatic or in the incubation period at the time of entry into India.For the international passengers who were tested positive and quarantined at the time of entry, separate standard operating procedures (SOPs) were issued on April 2, 2020 (MHA, April 2, 2020). Thus, this period covers 14 days and 21 days from the last arrival of international passengers. This period also covers the first lockdown duration until April 15, 2020.

In scenario 3, we fixed the period until the second lockdown date (MHA, 14th April 2020). The Government of India emphasized on tracing the cases, testing and treating. Hence, the importance is on the health infrastructure as it is capable of covering all three "Ts." Till 3rd May, all the cases that had come from international destinations with the disease must have either recovered or deceased.

Currently, we have the fourth lockdown, and we assume the factors of the third phase of our analysis will continue as the situation remains constant over the lockdowns.

2.1. Data:

For this study, we used COVID cases and patient data from the data-sharing portal *covid19India.org*. This website provides data on all the aspects of the epidemic of COVID-19 in India and its states. Recorded from the website, the number of COVID19 cases for three time periods, i.e., 23 March – 1 April, April – 15 April, and 16 April – 3 May 2020. From the recorded cases, estimated the doubling time of the epidemic by the exponential method. For this study, we have considered only those states with at least one case of COVID19. We have also taken the cumulative number of screening tests performed to detect the COVID-19 infections for all these states except Telangana for which the data is not available. For the district level data, we have taken district with at least one infected case till 13th May, and the analysis is for 500 districts with cumulative cases. We have not considered COVID cases without district name or mentioned details awaited. We have also used raw data of the patient for travel history analysis. Patient raw data is until 19th April, and it is frozen afterward on the website.

We have used multiple sources to get required socio-economic indicators. The population is projected till 31 March 2020 and used for calculation of density, urban, population above 65 years, for rates per million population. Immigration data is not available; we have considered as a proxy of it as emigration rate of the states in India. Taken the emigration rate data from a published paper, and it pertains to 2008 (Bhagat et al., 2011). Data on emigrants also available only for ten states from the Ministry of Tourism (2019). We have not used the tourism data as the analysis is for all the states with cases. Data on airport traffic data is from the World Bank, Airport Council, and Tourism. We have computed all household-level indicators for state and districts from NFHS -4, 2015-16 (IIPS 2017), and taken recent health infrastructure statistics from the National Health Profile 2019. The data sharing websites of COVID19India and ICMR does not provide information on isolation bed, quarantine beds, ventilators for all states. Data on these crucial factors would have helped in the understanding of disease control. Environmental factors such as temperature, wind, and rain are not considered in this paper as most of the states in India are experiencing summer.

3. Empirical Modelling:

In the sections below, we have discussed the quality of the confirmed COVID cases, mathematical models, and statistical methods of the study in detail.

3.1. Reliability Analysis of Confirmed Cases

It is vital to examine the data validity of confirmed cases before proceeding for further analysis. Hence, the rate per million tests and cases were studied. We have also studied the share of positive cases with the log of it. The cumulative number of tests per million population and cumulated infected cases per million population in the three-time slots are presented in Figures 1 and 2. Public opinion is that whenever the number of tests for diagnosis increases, then the number of infected cases will also increase. In the case of Andhra Pradesh, Tamil Nadu and Jammu & Kashmir, tests per million are very high, but the rate of infected cases in these states are lower than other states. On the other hand, in the Maharashtra and Chandigarh, tests are less, but the prevalence is high (16 April -3 May). Moreover, for infectious diseases, the number of infected people usually increases first before reaching a peak and then drops, and the peak time in each state will be different.

Figure 1:Number of Tests per Million Projected Population, 2020

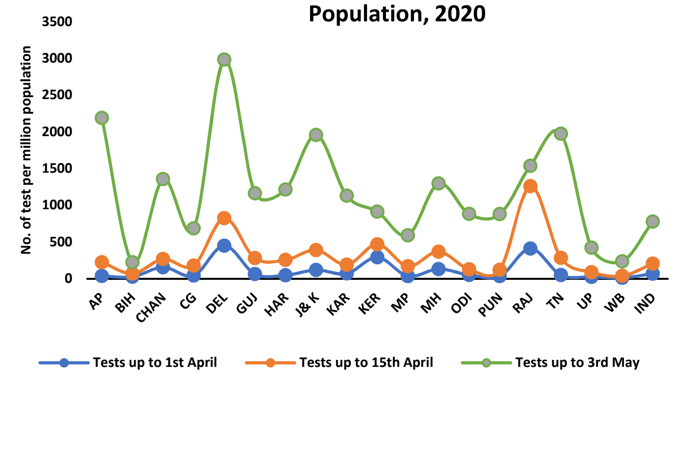


Figure 2: Number of Confirmed Cases per Million Projected Population, 2020

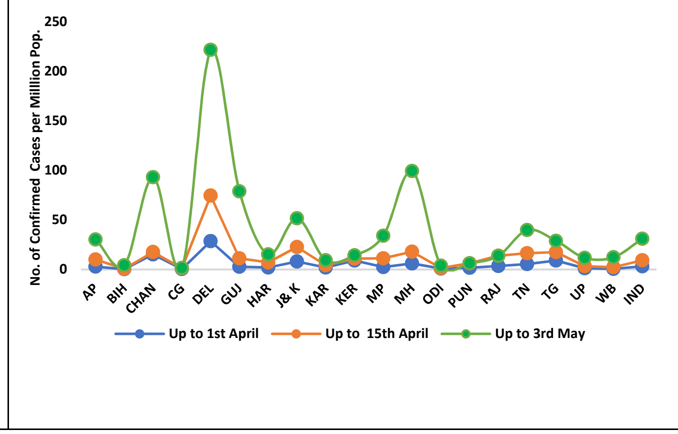


Figure 3: Number of Confirmed Cases by Share of Cases in the Tests, 3 March -1 April 2020

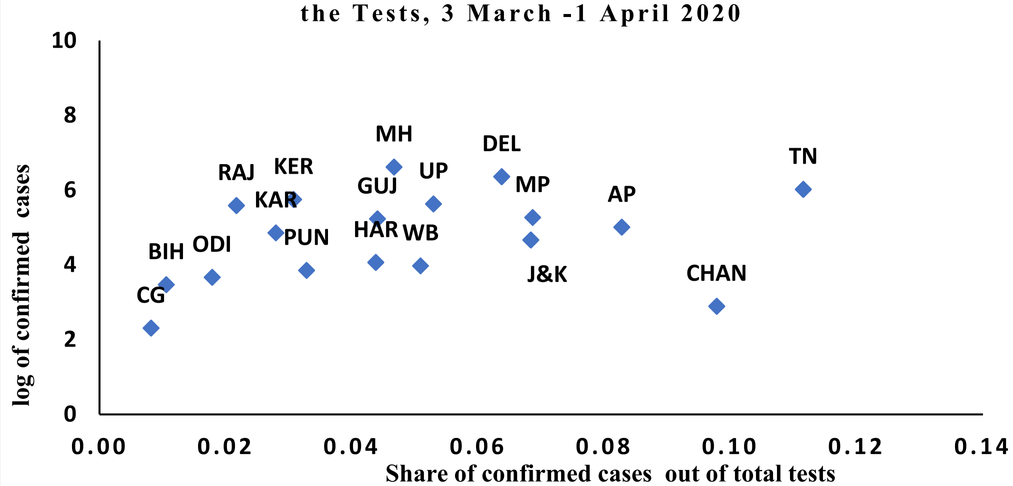


Figure 4: Number of Confirmed Cases by Share of Cases in the Tests,2 April -15 April 2020

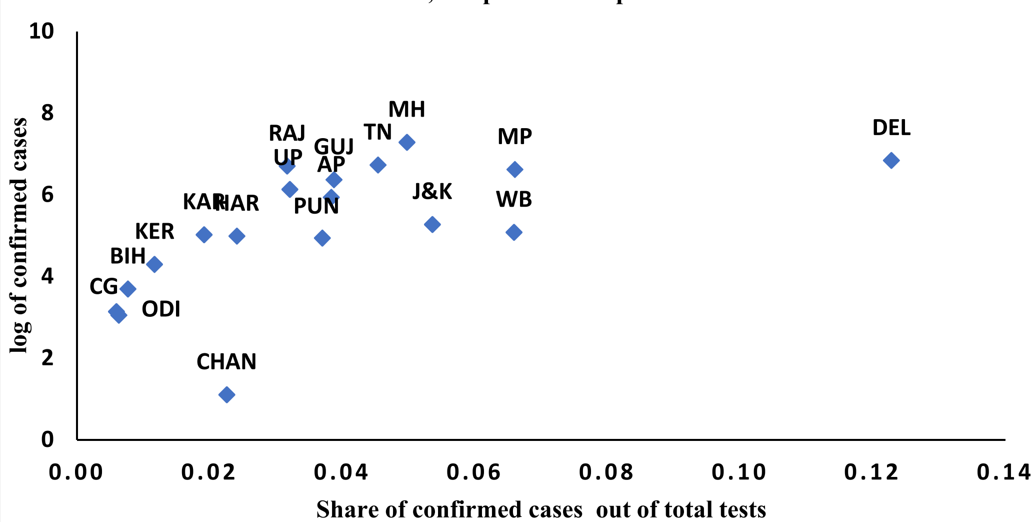
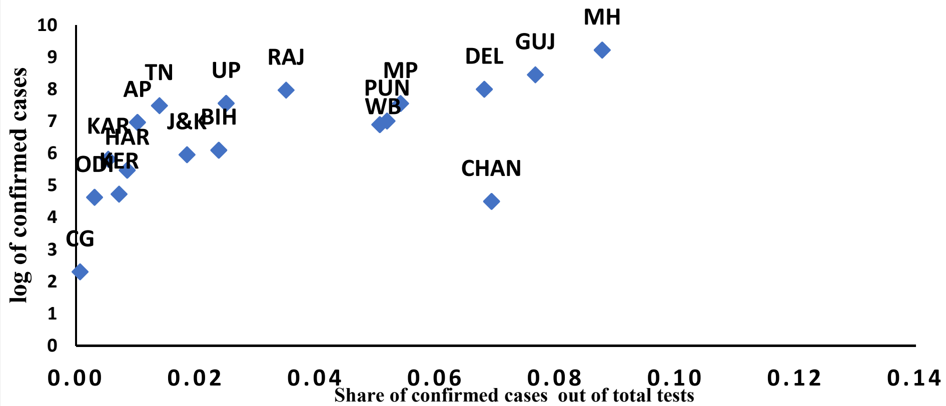


Figure 5: Number of Confirmed Cases by Share of Cases in the Tests, 16 April -3 May 2020



3.2 Estimation of Growth Rate and Doubling Time

Data from ICMR on tests provides for a certain period number and persons. It is not very clear from the data, whether it is only for the first test or it includes the test performed for patients with 48-72 hours and twice within 24 hours for those expecting to be discharged. To examine whether the number of infected cases is the result of higher testing or it is independent, we have taken an association between the share of infected cases out of tested persons and log of infected cases during these three time periods (Figures 3-5). Log of infected cases in three-time frames are more or less uniform, and it is irrespective of the share of infected cases out of tests performed. As we do not have a pattern to suggest that number of infected cases is increasing with the increase in the test, we used the number of infected cases given by the Government of India. Hence, we used the doubling rate of infected cases as the dependent variable for three time periods for the state-level analysis.

Since we have data on the cumulative cases for all the states, it is possible to fit an exponential growth model for these cumulative cases. The exponential growth rate for the confirmed cases was calculated for three time periods, i.e., 23rd March – 1st April, 2nd April – 15th April, and 16th April – 3rd May 2020, based on which the doubling time was also calculated for these three time periods.

The basic exponential growth model is defined as

$$C(t) = C(0) * exp(rt) \qquad \text{.......................................} (1)$$

Where C(t) = number of cumulative cases at time t= t,
C (0) = number of cumulative cases at time t = 0,
r = growth rate and t = time (in days).

If we assume the population to be homogenous, i.e., each individual has an equal probability of acquiring infection through contacts, and the number of cumulative cases grows exponentially then the doubling time for the epidemic can be estimated as follows:

Considering the exponential growth model, the doubling time t_d refers to the time when the number of cases doubles, i.e., C(t_d) = 2 C(0). Substituting the value of C(t_d) = 2 C(0) in equation 1, we obtained the doubling time t(d) as

$$t(d) = ln\,(2)\,/\,r \qquad \text{.......................................(2)}$$

3.4 Multivariate analysis

For the state-level data, doubling time-based on the exponential method is used as dependent variables in the regression analyses three phases.

For within state transmissions, we consider the mediating effect with the time slot of spread. In the first phase, the objective was to examine factors responsible for the importation of cases. Hence, the potential variables as per the scenarios given in the framework are the emigration rate of the states, percentage urban, population density, percentage rich, percentage 65 and above, family size, and hygiene practices. The factors in the second phase of the disease that are family size, number of persons sharing sleeping in a room, and handwashing practice. In the third phase, the main emphasis is on the health infrastructure; hence, we selected a log of population per hospital and percentage urban. Although many health facilities are available at the community level, we required hospitals (with ventilators) for the treatment of COVID19 patients.

The Government of India has made a different zone of the districts based on the number of infected cases (The Hindu, May 2, 2020). However, district characteristics are essential for controlling the spread of disease. Thus, to identify the districts with similar characteristics, K-Mean clustering is used. A clustering method that does not require computation of all possible distances is k-means clustering. The algorithm is called k-means, where k is the number of clusters we want since a case is assigned to the cluster for which its distance to the cluster means is the smallest. We have standardized the variables to a mean of 0 and a standard deviation of 1. Based on the characteristics of housing, urban, economic condition, and the number of districts, it was decided not to have more than five clusters. After identification of cluster, used the mean number of infected cases in each cluster to give the color similar to the Government of India identification.

We have also done two district-level regression analyses, using the log of infected cases till May 13 as a dependent variable with one with the characteristic of districts and another with zones as dummy variables.

4. Results:

We have given in Appendix 1and 2, the date of the reported first case of COVID19 in the state and other indicators. From the time of the first case, it is clear that reporting of cases in the states with an international airport is much earlier than other states. In the other states, reporting of cases is with a rolling window of 1-5 days. The states with high emigration rate have early arrival of infected cases.
In Figure 6, doubling rates for three phases are presented for the states where the disease started in March and continuing until the 3rd of May. However, for all states, data is presented in Appendix 1. Some of the states, like Uttar Pradesh, Maharashtra, and Gujarat, started with a similar level of doubling rate of Kerala. However, Kerala achieved a high level of doubling time, whereas other states showed a minimal increase. Hence, this indicates that at the state level role of factors is essential to understand the transmission of COVID 19.

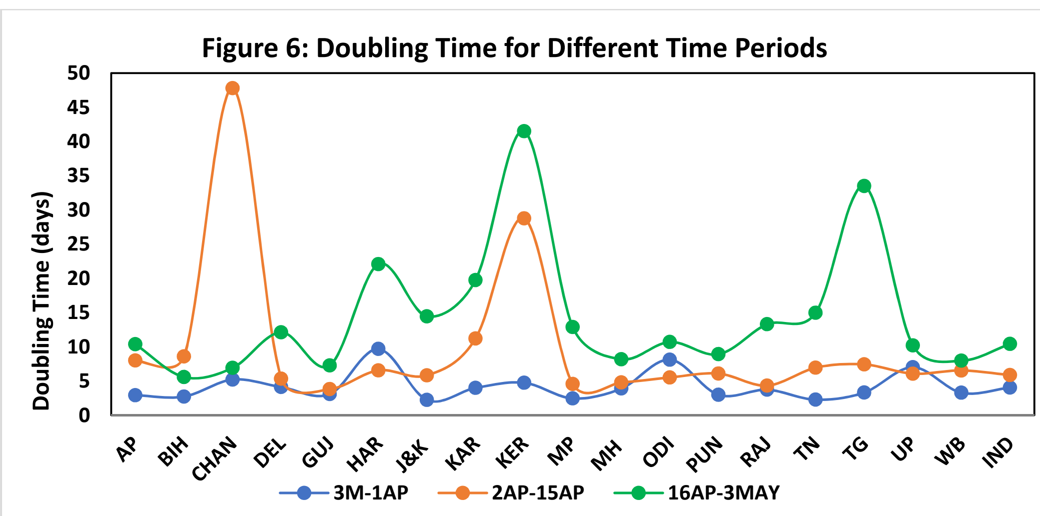


Table 1: Regression results for doubling rate for three time periods

	Unstandardized Coeff.	S.E.	Standardized Coeff.	*t*	p-value
Dependent Variable - Doubling Rate (3 March-1 April)					
Constant	-8.788	4.994		-1.76	0.102
Emigration rate per 000 population, 2008	-0.114	0.059	-0.593	-1.934	0.075
Percentage urban population	-0.154	0.056	-1.779	-2.773	0.016
Log of density	1.638	0.589	1.029	2.782	0.016
Percentage rich (4and 5 WQ)	0.127	0.048	1.469	2.634	0.021
Percentage 65+ population, 2020	0.957	0.356	0.754	2.685	0.019
Percentage of household with less than two rooms	-0.035	0.044	-0.203	-0.784	0.447
Number of observations -19					
R Square – 47.4 percent					
Dependent Variable - Doubling Rate (2 April-15April)					
Constant	106.574	44.675		2.386	0.029
Percentage urban population	0.189	0.168	0.279	1.126	0.276
Percentage poor(1 and 2 WQ)	0.703	0.356	1.04	1.972	0.065
Percentage 65+ population, 2020	-2.32	2.026	-0.228	-1.145	0.268
Mean family size	-31.764	11.094	-0.874	-2.863	0.011
Percentage of household with less than two rooms	-0.564	0.238	-0.445	-2.376	0.030
Percentage households using soap for hand washing	0.801	0.418	0.882	1.916	0.072
Number of Observations – 23					
R Square 49.1 percent					
Dependent Variable - Doubling Rate (16 April-3 May)					
Constant	390.308	110.215		3.541	0.002
Log of population per hospital	-21.609	7.791	-0.482	-2.773	0.012
Percentage urban population	-0.633	0.277	-0.433	-2.286	0.033
Mean family size	-24.36	14.139	-0.327	-1.723	0.100
Number of Observations – 23					
R Square 39.1 percent					

The number of states with confirmed cases for three phases is relatively small and does not permit us to use all covariates in multivariate analyses. However, they are used in the discussion section to explain the situation. In the first phase (3rd March to 1st April) regression analysis, the association of percentage urban with the doubling time is negative. Moreover, the emigration rate per thousand is also negatively associated with the doubling time at a 10 percent level of significance. Percentage rich, percentage 65, and above population and log of density are showing a positive relationship with doubling time of infected cases (Table 1). In the regression, we have considered the rich population as they are exposed to international travel.

In the second phase (2 April to 15 April), isolation of cases is important. In the regression analysis, it is found that the mean number of people sharing a room is the most significant negatively associated factor with the doubling time of infected cases. In the third phase, the association of log of population per hospital and percentage urban is negative with the doubling time of infected cases (Table 1).

Figure 7-8 presents the association between the log of the number of confirmed cases of COVID-19 at the district level and the percentage of urban and percentage of rich. The highest (and significant) associations are observed for the percentage of urban and rich. Hence, it is thought to have a better understanding of the characteristics of the districts by classifying them based on these two variables along with the housing characteristics.

Figure 7: Association of Log of Infected Cases by Percentage Urban, 13 May

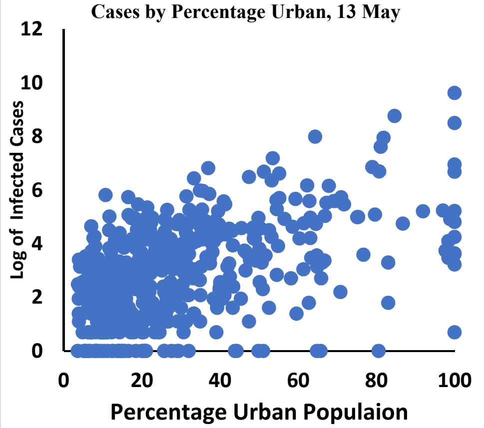
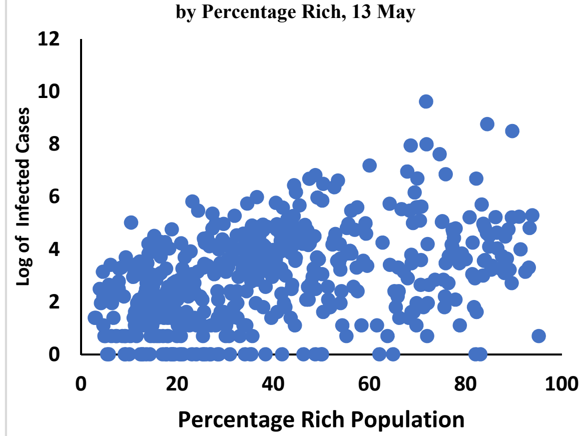


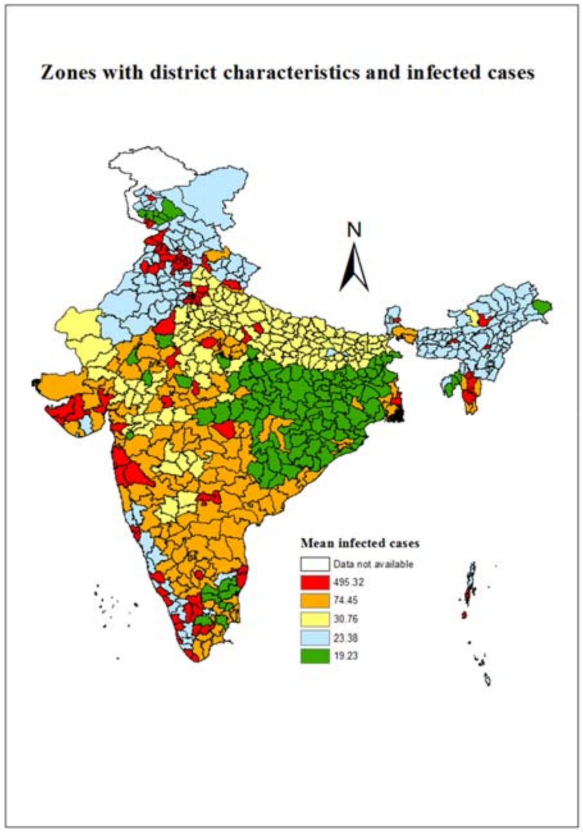
Figure 8: Association of Log of Infected Cases by Percentage Rich, 13 May



In Table 2 presented, K-Mean clustering centres, ANOVA, mean, the minimum, and the maximum number of infected cases in each cluster. All the variables selected for clustering are significant at 5 percent level of significance. Cluster 2 has the highest number of cases, followed by cluster five, four, one, and three. Districts in cluster 2, have the highest percentage of urban and rich. In the case of clusters 5 and 4, housing conditions are poor; specifically, the percentage of households with less than two rooms is highest. Based on the number of cases in each cluster, we have named these clusters as red, orange, yellow, blue, and green. The classification of districts based on our analysis is presented in Maps 1.

Table 2: Final Cluster Centers and results for ANOVA

Variables (Z_Score)	**Final Clusters Centers**					**ANOVA**		
	1	2	3	4	5	**Mean Square (Cluster)**	**Mean Square (Error)**	**F**
Percentage urban population	-0.3421	1.8909	-0.5359	-0.5101	0.1264	105.24	0.34	306.45*
Percentage rich (4and 5 WQ)	0.1811	1.6245	-0.7766	-0.6769	0.1009	96.80	0.40	244.13*
Mean family size	-0.1616	-0.5663	-0.1923	1.1096	- 0.4794	61.88	0.62	100.36*
Mean number of persons sleeping in a room	-1.0414	-0.4552	-0.1369	1.2078	0.3237	103.73	0.35	293.91*
Percentage of household with less than two rooms	-1.1513	-0.1810	-0.0438	0.6187	0.8368	88.57	0.45	197.53*
Percentage of household with water outside their premises	-0.3447	-0.5310	0.9119	-0.5186	0.5389	54.84	0.66	82.99*
% HH sharing toilet or no toilet	-0.9957	-0.9694	0.9302	0.8950	0.0160	114.55	0.29	402.32*
Percentage households using soap for hand washing	0.4294	1.0705	-1.2699	-0.2525	0.1700	84.30	0.48	177.35*
Number of districts	154	92	118	144	130			
Computed COVID cases for each cluster								
Mean number of infected cases	23.38	495.32	19.23	30.76	74.45			
Standard Deviation	79.89	1791.16	52.81	71.94	133.16			



Map: Zones classified from clustering analysis and represented by the mean number of infected case

We have carried out two district-level regression analyses with dependent variables as the log of infected cases to understand the association of potential factors. In the first regression analysis, the independent variables are percentage urban, rich, percentage of households with less than two rooms, the mean number of persons sleeping in the room (other household-level variables are not considered due to multicollinearity). All the selected variables have a significant positive association with the log of infected cases. In the second regression analysis, considered zones as dummy variables. Districts in the red zone have 2.4 times more log of infected cases in comparison to the green zone (reference category). In the case of the blue zone, it is 0.6 cases (Table 3).

Table 3: Regression results for log of infected cases (13 May)

	Unstandardized Coeff.	S.E.	Standardized Coeff.	*t*	p-value
Dependent Variable - Log of Infected Cases (13 May)					
(Constant)	-1.172	0.542		-2.164	0.031
Percentage rich (4and 5 WQ)	0.025	0.005	0.340	5.466	0.000
Percentage urban population	0.023	0.005	0.287	5.032	0.000
Mean number of persons sleeping in a room	0.526	0.216	0.148	2.435	0.015
Percentage of household with less than two rooms	0.021	0.008	0.161	2.753	0.006
Number of Observations – 500					
R Square 30.6 percent					
Dependent Variable - Log of Infected Cases (13 May)					
(Constant)	1.937	0.185		10.476	0.000
Red zone based on characteristics	2.415	0.257	0.504	9.412	0.000
Orange Zone based on characteristics	1.397	0.237	0.334	5.888	0.000
Yellow zone based on characteristics	0.645	0.229	0.164	2.813	0.005
Blue zone based on characteristics	0.600	0.251	0.130	2.39	0.017
Number of Observations – 500					
R Square 18.4 percent					

5. Discussion:

The strength of this study is based on combining multiple data sources in the analysis of COVID-19 for India. Combining mathematical models, statistical analysis, and multiple data sets, we found that urban areas have a large number of infected cases. Moreover, the number of days for doubling is comparatively less in the states with a high percentage of the urban regions. Early reporting of COVID 19 cases noted in the states with international airports, namely Delhi, Mumbai, Bangalore, Chennai, Kolkata, Hyderabad, Ahmedabad, Cochin, and Pune. These airports have a high turnover of the enplaning and deplaning of the international and domestic passengers (Ministry of Tourism, 2019; Airport Council, 2020). From the patient data, found that the 21 percent confirmed cases arrived from international destinations, and 34 percent travel to Delhi or crossed it. Hence, there was a high chance of exposure to diseased persons among supporting staff of airports, mainly flight attendants, air hostess, security, immigration officer, counter staff, housekeeping, taxi drivers at the airport. It is difficult to identify the contact of a diseased person for the above group of people. Hence, the doubling time in the first phase was in most of the states was below four days.

From our analysis of second phase data, it is undeniable that the household level factors played a crucial role in the doubling time. In million-plus cities, most people live in the high rise building as well as the significant proportion of people live in the slums. According to the Census of India 2011, among million-plus urban agglomerations, Mumbai rank first with 18 million population followed by Delhi with 16 million, and Kolkata with 14 million population (MoHUA). Moreover, Maharashtra has 29 million population living in 6 million-plus cities, whereas Uttar Pradesh, and Kerala have seven cities, and for five cities population is less than two million. A significant proportion of people in large million-plus cities are living in flats with two or less than two rooms and a family size of 4-5 persons. In the case of Mumbai, 76 percent slum and non-slum households have less than two rooms (NFHS-4).

Furthermore, people who are living in the slums of large million-plus cities share toilets and collect water from outside. In the middle and upper-class flats/ apartment, they too share common spaces and lift for the movement. Social distancing and avoiding mass gathering, which is vital to break the chain of transmission, seems impossible in urban-slums. The recent spread of infectious diseases like Zika, Ebola, and other influenza was higher in such areas (Nguyen et al., 2016; Chen et al., 2016; Snyder et al., 2017). In the case of COVID19 also, it is found that the family cluster plays a vital role in the spread (Chan et al., 2020). It is also important to mention here that the communicable disease like Dengue has a large number of cases in the Punjab, Maharashtra, Gujarat, and Delhi (NHF, 2019). Hence, lockdown and social distancing impact in the big cities is not very impressive.

In the third phase of disease spread, the most significant role of health infrastructure is coming in the analysis. It indirectly indicates the preparedness in terms of tools, tracing, testing, and treating are required rather than lockdown that is not working in large urban areas.

District-level analysis based on comprehensive analysis gives 92 districts that are in the red zone. These districts required immediate attention to health facilities, testing, and treatment. However, in the green zones, economic and other activities may start. Uniform policy in the Indian district will not work. There should be distinct strategies at a lower level, namely districts for suppression and control of COVID-19, including progressively restoring socio-economic activity in context-appropriate and safe ways. Moreover, when we open the ports for passengers, comprehensive testing is required, like Singapore, to detect the COVID19 cases (Ministry of Health, Singapore, 2020).

We would also like to mention some limitations of our study. First, we did not consider the state and district differences in the timing of the COVID19 start date in our analyses. Another issue not covered in the study is the state-level differences in the approach to ICMR guidelines and testing for COVID-19.

6. Suggestions for Action Plan of Million Plus Cities and Red Zone Districts

Above mentioned observations have several implications for developing local responses to the spread of COVID-19. Firstly, social distancing will impact cities differently based on city size. Due to higher density, insufficient social distancing in larger cities may lead to more significant outbreaks and the creation of reservoirs for the disease. These cities need to act quickly to contain this outbreak.

- On 27 May 2020 more than 82 thousand cases are reported from eight megacities namely Mumbai, Delhi, Chennai, Ahmedabad, Kolkata, Hyderabad and Banglore. It has also been found that the states with these metro cities are also not doing well in other communicable diseases(Tuberculosis, Dengue and malaria). From the perspective of containing the transmission, larger cities require more aggressive policies in terms of isolation centres, sensitisation of community toilets & common spaces in the high rise building, supply chain of essential goods and services.
- Protocol for testing of COVID 19 should not be the same for all contacts. Test for the contacts with comorbidities or 65 and above age may be done within seven days with or without symptoms.
- Social media messages should cover not only about COVID19, but it should also include messages for improving the immune system of the population.
- There is apprehension that due to return migration of people from megacities to the place of origin, the number of COVID19 cases will increase in origin districts. People suffering in the place of destination or origin is the same, and they required health facilities. District administration with high return migrants needs to prepare the isolation centres, testing and treatment facilities.
- Looking at the trend of communicable diseases during the last thirty years it is expected that not only the COVID19 but there is also the possibility of other outbreaks. Hence, a better plan of action for control of communicable disease in the large cities are required. Thirty years back, there were specialised quarantine hospitals in the cities. However, we are not finding it now in large cities. We required rethinking about our health infrastructure and the requirements of nurses, doctors and fourth class employees in megacities.
- The cities with a high number of COVID cases also have an international airport with a large size of passenger per annum(Delhi 70 million, Mumbai 50 million, Chennai 20 million; World

Data info). Hence, when we open the ports for passengers, comprehensive testing is required, like Singapore, to detect the COVID19 cases (Ministry of Health, Singapore, 2020).

Acknowledgement:

Authors of this paper would like to thank Prof K.S. James, Director and Senior Professor, International Institute for Population Sciences, Mumbai for initiating a series related to COVID 19 at IIPS.

References

Bhagat, R. B., Keshri, K., & Ali, I. (2013). Emigration and flow of remittances in India. *Migration and Development*, *2*(1), 93-105.

Bogoch, I. I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M. U., & Khan, K. (2020). Potential for global spread of a novel coronavirus from China. *Journal of Travel Medicine*, *27*(2), taaa011.

Chen, J., Chu, S., Chungbaek, Y., Khan, M., Kuhlman, C., Marathe, A., ... & Xie, D. (2016). Effect of modelling slum populations on influenza spread in Delhi. *BMJ open*, *6*(9), e011699.

Chen, T., Wu, D., Chen, H., Yan, W., Yang, D., Chen, G., ... & Wang, T. (2020). Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *BMJ*, *368*

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., ... & Viboud, C. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 368, no. 6489: 395-400

Djalante, R., Shaw, R., & DeWit, A. (2020). Building resilience against biological hazards and pandemics: COVID-19 and its implications for the Sendai Framework. *Progress in Disaster Science*, 100080.

Grais, R. F., Ellis, J. H., & Glass, G. E. (2003). Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European Journal of Epidemiology*, *18*(11), 1065-1072.

Fan, V. Y., Jamison, D. T., & Summers, L. H. (2016). *The inclusive cost of pandemic influenza risk* (No. w22137). National Bureau of Economic Research.

IIPS and ICF(2017). NAtiona Family Health Survey (NFHS-4), 2015-16, India. Mumbai: International Institute for Population Sciencs.

Jung, S. M., Akhmetzhanov, A. R., Hayashi, K., Linton, N. M., Yang, Y., Yuan, B., ... & Nishiura, H. (2020). Real-time estimation of the risk of death from novel coronavirus

(COVID-19) infection: inference using exported cases. *Journal of clinical medicine*, *9*(2), 523.

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., ... & Davies, N. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The lancet infectious diseases*.

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., ... & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, *172*(9), 577-582.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., ... & Xing, X. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*.

Liu, Y., Gayle, A. A., Wilder-Smith, A., & Rocklöv, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of travel medicine*.

Mogi, R., & Spijker, J. (2020). The influence of social and economic ties to the spread of COVID-19 in Europe(Electronic Copy).

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., ... & Agha, R. (2020). The socio-economic implications of the coronavirus and COVID-19 pandemic: A review. *International Journal of Surgery*.

Nguyen, J. L., Yang, W., Ito, K., Matte, T. D., Shaman, J., & Kinney, P. L. (2016). Seasonal influenza infections and cardiovascular disease mortality. *JAMA cardiology*, *1*(3), 274-281.

Priyadarsini, L., & Suresh, M. (2020). Factors influencing the epidemiological characteristics of pandemic COVID 19: A TISM approach. *International Journal of Healthcare Management*, 1-10.

Qiu, Y., Chen, X., & Shi, W. (2020). Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. *Journal of Population Economics*, 1.

Riou, J., & Althaus, C. L. (2020). Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance*, *25*(4), 2000058.

Sands, P., El Turabi, A., Saynisch, P. A., & Dzau, V. J. (2016). Assessment of economic vulnerability to infectious disease crises. *The Lancet*, *388*(10058), 2443-2448.

Stier, A., Berman, M., & Bettencourt, L. (2020). COVID-19 attack rate increases with city size. *Mansueto Institute for Urban Innovation Research Paper Forthcoming(Electronic copy)*.

Snyder, R. E., Boone, C. E., Cardoso, C. A. A., Aguiar-Alves, F., Neves, F. P., & Riley, L. W. (2017). Zika: A scourge in urban slums. *PLoS neglected tropical diseases*, *11*(3).

Wilder-Smith, A., & Freedman, D. O. (2020). Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. *Journal of travel medicine*, *27*(2), taaa020.

Web Based Sources

Airport Industry Review(AIR) (2020). Coronavirus screening at Airports: The problem with Thermal Detection, Available at https://airport.nridigital.com/air_apr20/coronavirus_screening_at_airports_the_problem_with_thermal_detection, accessed on May 11, 2020.

API (2020) Confirmed Cases and Raw Patient data of COVID 19 https://www.covid19india.org/ accessed on daily basis till May 27, 2020.

CBHI (2019). National Health Profile 2019, 14th Issue, Available at http://www.cbhidghs.nic.in/showfile.php?lid=1147

Global Health Security Index (2020). 2019 Global Health Security Index, Available at https://www.ghsindex.org/, accessed on May 3, 2020.

Government of India (2019), Ministry of Tourism, India Tourism Statistics at a Glance (2019), *Incredible India*, Available at http://tourism.gov.in/sites/default/files/Other/India%20Tourism%20Statistics%20at%20a%20Glance%202019.pdf , accessed on April 11, 2020.

Government of India (2020a), Ministry of Health and Family Welfare, Consolidated Travel Advisory for COVID-19, March 6, 2020 https://www.mohfw.gov.in/pdf/DraftTraveladvisorytotravelersUpdated06032020.pdf, accessed on April 15, 2020)

Government of India (2020b) Press Information Bureau, New Delhi, March 24, 2020, Available at https://www.mha.gov.in/sites/default/files/PR_NationalLockdown_26032020_0.pdf, accessed on April 15, 2020.

Government of India (2020c). Ministry of Home Affairs, Order No. 40-3/2020-DM-I(A), New Delhi, April 2, 2020, https://www.mha.gov.in/sites/default/files/3rd%20Addendum%20dated%202.4.2020%20to%20Lockdown%20measures_0.pdf, accessed on May 2, 2020.

Government of India (2020d). Ministry of Home Affairs, Order No. 40-3/2020-DM-I (A) dated April15, 2020 https://www.mha.gov.in/sites/default/files/PR_15042020RevisedConsolidatedGuidelines.pdf , accessed on April 17, 2020.

Government of India 2020e). Ministry of Health and Family Welfare, Containment Plan Novel Coronavirus Disease 2019 (COVID 19 Version 2 , updated 16.05.2020, Available at https://www.mohfw.gov.in/pdf/Containmentplan16052020.pdf , accessed on May19, 2020

The Hindu (2020). COVID-19|Under which zone does your district lie? May 2, 2020 Available at https://www.thehindu.com/data/covid-19-under-which-zone-does-your-district-lie/article31487764.ece, accessed on May 4, 2020.

ICAO(2020). Uniting Aviation, A UN specialized Agency, Available at https://www.icao.int/MID/Documents/2017/Aviation%20Data%20and%20Analysis%20Seminar/PPT11%20-%20India%20-%20Introduction%20to% , Accessed on 25 May 2020.

IIPS (2017) National Family Health Survey-4 Data, Available at https://www.dhsprogram.com/data/dataset/India_Standard-DHS_2015.cfm?flag=0

Ministry of Health Singapore. Updates on Covid-19 Local Situation. *Updates on Covid-19 Local Situation*, Available at https://www.moh.gov.sg/covid-19, accessed on May 18, 2020.

Ministry of Housing and Urban Affairs, GOI, Cities with Million Plus Population in 2011 , http://mohua.gov.in/cms/number-of-cities--towns-by-city-size-class.php

Press Information Bureau, Government of India (2020). Update on Novel Coronavirus :one positive case reported in Kerala, 30 January 2020, https://pib.gov.in/PressReleseDetail.aspx?PRID=1601095 , accessed on April 9, 2020.

WebMD Coronavirus transmission overview, 2020 https://www.webmd.com/lung/coronavirus-transmission-overview#1, accessed May 4, 2020.

World Bank (2015). Pandemics Overview. Available at http://www.worldbank.org/en/topic/pandemics/overview, accessed on May 3, 2020

WorldData.info, Available at https://www.worlddata.info/asia/india/airports.php, Accessed on 18 May 2020.

World Economic Forum (March 15, 2020) A Visual History of Pandemics. Available at https://www.weforum.org/agenda/2020/03/a-visual-history-of-pandemics/, accessed May 3, 2020.

WHO (2020). Rolling updates on coronavirus disease https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen, accessed on May19, 2020.

WHO (2020). Novel coronavirus – China January 12,2020, Available at https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/, accessed on March 22, 2020.

Appendix 1: Growth rate and Doubling Time in Days of COVID19 Cases

States	Date of First Case	Growth Rate (exponential)			Doubling Rate		
		3March-1April	2April-15April	16April-3May	3March -1April	2April-15April	16April-3May
A&N Island	26-Mar	0.22	NA*	0.07	3.12	NA*	9.71
Andhra Pradesh	12-Mar	0.24	0.09	0.07	2.92	7.99	10.36
Arunachal Pradesh	02-Apr	NA*	NA*	NA*	NA*	NA*	NA*
Assam	31-Mar	NA*	0.04	0.01	NA*	18.73	51.73
Bihar	22-Mar	0.26	0.08	0.12	2.71	8.59	5.58
Chandigarh	19-Mar	0.13	0.01	0.10	5.20	47.80	6.92
Chhattisgarh	19-Mar	0.22	0.12	0.01	3.13	5.61	65.39
Delhi	02-Mar	0.17	0.13	0.06	4.11	5.34	12.12
Goa	25-Mar	0.51	0.17	NA*	1.36	4.12	NA*
Gujarat	19-Mar	0.23	0.18	0.10	3.08	3.82	7.27
Haryana	04-Mar	0.07	0.11	0.03	9.68	6.55	22.07
Himachal Pradesh	20-Mar	NA*	0.13	0.00	NA*	5.38	161.20
Jammu & Kashmir	09-Mar	0.31	0.12	0.05	2.24	5.82	14.44
Jharkhand	31-Mar	NA*	0.23	0.09	NA*	3.00	7.61
Karnataka	09-Mar	0.17	0.06	0.04	3.98	11.20	19.75
Kerala	30-Jan	0.15	0.02	0.02	4.74	28.76	41.51
Madhya Pradesh	20-Mar	0.28	0.15	0.05	2.45	4.56	12.88
Maharashtra	09-Mar	0.18	0.14	0.08	3.90	4.78	8.17
Manipur	24-Mar	NA*	NA*	NA*	NA*	NA*	NA*
Meghalaya	13-Apr	NA*	0.97	0.01	NA*	0.71	55.45
Mizoram	25-Mar	NA*	NA*	NA*	NA*	NA*	NA*
Odisha	16-Mar	0.09	0.13	0.06	8.12	5.51	10.70
Puducherry	15-Mar	NA*	NA*	NA*	NA*	NA*	NA*
Punjab	09-Mar	0.23	0.11	0.08	2.97	6.08	8.91
Rajasthan	03-Mar	0.19	0.16	0.05	3.72	4.32	13.30
Tamil Nadu	07-Mar	0.31	0.10	0.05	2.27	6.94	14.97
Telangana	02-Mar	0.21	0.09	0.02	3.31	7.41	33.49
Tripura	07-Apr	NA*	NA*	0.05	NA*	NA*	14.69
Uttar Pradesh	04-Mar	0.10	0.11	0.07	7.00	6.07	10.19
Uttarakhand	16-Mar	0.13	0.07	0.03	5.47	9.43	27.62
West Bengal	17-Mar	0.21	0.11	0.09	3.29	6.53	7.97
India	30-Jan	0.17	0.12	0.07	4.04	5.87	10.41

NA* -Not Applicable

Note: If the detection of the first case is after 3rd March then that time point is taken as the starting point

Appendix 2: Number of infected cases, tests, and share of infected cases of COVID-19.

States	Number of Infected cases			Number of Tests			Share of infected cases out of total tests		
	3March-1April	2April-15April	16April-3May	3March-1April	2April-15April	16April-3May	3March-1April	2April-15April	16April-3May
A&N Island	10	1	22	NA	NA	2351	NA*	NA*	0.0094
Andhra Pradesh	149	376	1058	1800	9813	103324	0.0828	0.0383	0.0102
Arunachal Pradesh	0	1	0	NA	363	447	NA*	0.0028	0.0000
Assam	1	7	10	NA	2651	9162	NA*	0.0026	0.0011
Bihar	32	40	445	3037	5226	18688	0.0105	0.0077	0.0238
Chandigarh	18	3	90	184	133	1299	0.0978	0.0226	0.0693
Chhattisgarh	10	23	10	1232	3890	14780	0.0081	0.0059	0.0007
Delhi	576	929	2971	9041	7564	43641	0.0637	0.1228	0.0681
Goa	5	2	0	220	391	1937	0.0227	0.0051	0.0000
Gujarat	186	580	4662	4224	14973	60863	0.0440	0.0387	0.0766
Haryana	58	146	238	1325	6063	27890	0.0438	0.0241	0.0085
Himachal Pradesh	NA	35	5	NA	1426	5759	NA*	0.0245	0.0009
Jammu & Kashmir	106	194	386	1551	3620	20867	0.0683	0.0536	0.0185
Jharkhand	1	25	87	NA	1803	9912	NA*	0.0139	0.0088
Karnataka	128	151	335	4587	7896	62415	0.0279	0.0191	0.0054
Kerala	314	73	113	10221	6254	15742	0.0307	0.0117	0.0072
Madhya Pradesh	193	745	1899	2812	11284	35090	0.0686	0.0660	0.0541
Maharashtra	747	1449	10058	16008	29134	114612	0.0467	0.0497	0.0878
Manipur	1	1	0	NA	NA	NA	NA*	NA*	NA*
Meghalaya	0	7	5	NA	617	1267	NA*	0.0113	0.0039
Mizoram	1	0	0	NA	NA	NA	NA*	NA*	NA*
Odisha	39	21	102	2190	3347	33121	0.0178	0.0063	0.0031
Puducherry	3	4	2	NA	NA	1948	NA*	NA*	0.0010
Punjab	47	139	1102	1434	3759	21246	0.0328	0.0370	0.0519
Rajasthan	266	810	2886	12279	25581	82380	0.0217	0.0317	0.0350
Tamil Nadu	411	831	1781	3684	18310	128113	0.1116	0.0454	0.0139
Telangana	334	316	432	NA	NA	NA	NA*	NA*	NA*
Tripura	0	2	14	NA	762	4400	NA*	0.0026	0.0032
Uttar Pradesh	278	457	1910	5255	14251	76335	0.0529	0.0321	0.0250
Uttarakhand	26	11	23	1289	1124	5165	0.0202	0.0098	0.0045
West Bengal	53	160	985	1042	2428	19445	0.0509	0.0659	0.0507
India	4293	8078	29346	89534	185065	771851	0.0479	0.0436	0.0380

NA-Not Available NA* -Not Applicable
Note: If the detection of the first case is after 3rd March then that time point is taken as the starting point