Quantitative Biology 2021, 9(3): 317–328
https://doi.org/10.15302/J-QB-021-0260

RESEARCH ARTICLE

A study of the COVID-19 epidemic in India using the SEIRD model

Rudra Banerjee[1], **Srijit Bhattacharjee**[2], **Pritish Kumar Varadwaj**[2,*]

[1] Department of Physics and Nanotechnology, SRM Institute of Science and Technology, Kattankulathur 603203, Tamil Nadu, India
[2] Department of Bioinformatics & Applied Sciences, Indian Institute of Information Technology, Allahabad 211015, Uttar Pradesh, India
* Correspondence: pritish@iiita.ac.in

Received December 15, 2020; Revised March 31, 2021; Accepted March 31, 2021

***Background*: The coronavirus pandemic (COVID-19) is causing a havoc globally, exacerbated by the newly discovered SARS-CoV-2 virus. Due to its high population density, India is one of the most badly effected countries from the first wave of COVID-19. Therefore, it is extremely necessary to accurately predict the state-wise and overall dynamics of COVID-19 to get the effective and efficient organization of resources across India.**
***Methods*: In this study, the dynamics of COVID-19 in India and several of its selected states with different demographic structures were analyzed using the SEIRD epidemiological model. The basic reproductive ratio R_0 was systemically estimated to predict the dynamics of the temporal progression of COVID-19 in India and eight of its states, Andhra Pradesh, Chhattisgarh, Delhi, Gujarat, Madhya Pradesh, Maharashtra, Tamil Nadu, and Uttar Pradesh.**
***Results*: For India, the SEIRD model calculations show that the peak of infection is expected to appear around the middle of October, 2020. Furthermore, we compared the model scenario to a Gaussian fit of the daily infected cases and obtained similar results. The early imposition of a nation-wide lockdown has reduced the number of infected cases but delayed the appearance of the infection peak significantly.**
***Conclusion*: After comparing our calculations using India's data to the real life dynamics observed in Italy and Russia, we can conclude that the SEIRD model can predict the dynamics of COVID-19 with sufficient accuracy.**

Keywords: COVID-19; SARS-CoV-2; epidemic; statistical analysis; SEIRD model

Author summary: The first step in combating a contagious disease, such as COVID-19, is to obtain accurate predictions of its dynamics. This is especially crucial for large countries like India, with highly diversified states. This paper would demonstrate how one can predict the spread of COVID-19 using the SEIRD model. Our calculations indicate that the first wave of COVID-19 in India would stop by the end of October of 2020, with varying dates for the different states.

INTRODUCTION

COVID-19 was announced as a global pandemic by World Health Organization (WHO) [1] due to its highly contagious and pathogenic nature. This highly contagious virus is spreading at a much higher rate than the previously reported corona viruses. In India, the first case of COVID-19 was reported on January 30, 2020. Despite the four lockdown phases during its first 30 weeks, the daily COVID-19 caseload is alarmingly rising. To manage an epidemic like this, administrations and health departments across India, a country with a population of 1.35 billion, need a fairly clear idea of the challenges they need to tackle throughout the course of the epidemic. In addition, reliable predictions regarding possible outbreaks can aid in decisive planning and resource management. Therefore, a detailed analysis following realistic data collection should prove to be very useful when estimating the number of test kits, hospital resources, and quarantine centers, needed to contain the disease.

With the help of the SEIRD population model, a well-established model used to study the dynamics of infectious diseases, we would be able to predict and assess the trend of COVID-19 in different states. After obtaining sufficient data, we can effectively nowcast and forecast the possible situations that a state or country will likely face. On March 25, 2020, India invoked a complete nationwide lockdown in the early phase of the epidemic, which prevented rapid viral transmission and bought time for the health administrations and officials to prepare and put in place effective measures needed to tackle the situation. Despite the considerable reduction in viral spread, an early lockdown for a country with over 1.3 billion people resulted in a delayed flattening of the temporal progression of COVID-19 cases. Thus, it is crucial to determine the duration in which a country should remain alert during a pandemic or epidemic.

To predict the progression of the epidemic, a well-structured data and accurate estimations of model parameters are needed. This helps to determine the most plausible situation. Although the data published by Indian Council of Medical Research (ICMR) and other agencies are fairly well organized, they are still insufficient to precisely compute specific parameters. One solution to this problem is to estimate reference parameters from other countries that utilized well organized data, and input data taken from various states of India to compute the parameter values. In this paper, we selected three states with a high population density (Delhi, Tamil Nadu, Uttar Pradesh), three states with a medium population density (Andhra Pradesh, Maharashrtra, Gujarat), and two states with a low population density (Madhya Pradesh, Chhattisgarh). We also further divided these states into high, moderate, and low infection rate categories. After using the same methodology for each of these states, we projected the values of the different parameters into the Susceptible-Exposed-Infected-Recovered-Dead (SEIRD) model to visualize the trend of COVID-19 transmission. The resulting trends are then compared to existing data to enhance predictability.

Recently published work [2] compared and analyzed the entire genome sequences of 28 viral strains from India, which showed a novel non-synonymous mutation in the *NSP3* gene of SARS-CoV-2 along with other frequent and significant mutations reported across the world. This finding suggests the possible existence of different viral strains in different parts of India, which ties into demographic densities, travel history, inflow of migrant labors, cultural practices, and living lifestyles. It may also be the reason behind the variation between the basic reproductive ratio (R_0) values across the selected states, as reported in this study.

As previously mentioned, to utilize the SEIRD model, certain parameters, such as incubation rate, mortality rate, R_0 and recovery rate, need to be calculated. For a given epidemic, the value of R_0 denotes the number of secondary infections generated from a seed infection. The value of R_0 is fixed and calculated during the early or growing stage of an epidemic. We employed a maximum likelihood estimation method that is dependent on the daily incidence and generation time outlined by Obadia *et al.* and Zhang *et al.* [3, 4]. The value of R_0 is estimated to be 1.917, which is within the ranges reported in various studies [4–10]. The range is expected to narrow if analyzed using a greater quantity of data. Containment measures may change the effective reproductive ratio, R_e, which is currently estimated to be 1.44 in India, for an incubation period of 5 days, and serial intervals of 8 and 10 d, respectively. The incubation period (d_i) for COVID-19 has been estimated to be between 2–14 d by WHO. In many studies, it is suggested that the mean incubation period is between 5–8 d [6, 11–14]. Additionally, the average incubation period is estimated to be 5.2 d in [15], as notified in Worldometer. Taking both international and India-based studies into consideration, we used an average incubation period range of 5 –7 d [16, 17]. Another crucial parameter in trend modelling is the period of infectiousness. In the case of India, we found that the mean infectiousness period (d_{if}) is between 3–5 d after referencing other studies due to the absence of relevant data [7, 18]. It is also important to keep in mind that the d_{if} value for India will be towards the lower end of the spectrum, for example, d_{if} will be close to 3 d as lockdown, social distancing, contact tracing, and isolation of the infected or exposed are done from the earlier stages of the pandemic. However, for comparative purposes, we considered 3, 4, and 5 d when using the model for analysis. The mortality rate of COVID-19 in India has gradually decreased over the past few months, from 2.49% of the total infected population in July to just 1.64% to date. The mortality rates for eight states in India were calculated by dividing the total number of deaths by the total number of confirmed cases over a specific period of time.

This study suggests that the COVID-19 safety and containment measures of India has been effective, with a lower number of infections per million than certain countries. According to our current analysis, the infection curve in India will start to flatten around mid-October to mid-November, with variations between states (some states may observe the flattening between August and October). Using a Gaussian fit on real data, our calculations suggest that India will experience a peak around mid-October, provided that no drastic changes in disease dynamics appear. This study is an example of how comparing model-based predictions and Gaussian-fitted real data allows us to estimate the limited number of parameters of the SEIRD model and to predict the total number of infected cases at the start, decline, or end of the epidemic. Without the proper preventative and containment measures in place in India, the total number of infection cases would have been at least three times higher than the present scenario.

Finally, we compared the progression of COVID-19 in two other countries with higher and lower infection rates, respectively. Currently, Russia has a greater number of infection cases, while Italy has a close to zero incidence rate. Our comparison indicates how India's infection rate per million is relatively low despite its large total number of cases due to its large population size. This suggests that the number of severe COVID-19 cases will be under control and will likely not pose a serious challenge for the government-based healthcare providers. However, this will not eliminate the challenges in states with high infection rates, such as Maharashtra.

RESULTS

The eight states were selected based on their population density and COVID-19 infection rate per million population, as shown in Fig.1. Due to India's current demographics, we omitted the low population-high infection category. In our distribution, a population density > 500 km^{-2} was considered high and < 300 km^{-2} as low. Additionally, an infection rate of > $500/10^6$ individuals was treated as high and < $100/10^6$ as low.

Since previous studies suggest that the average incubation period of the SARS-CoV-2 virus is between 5 to 7 days [6, 11, 19], we studied both the best case scenario (5 d) and the worst case scenario (6 and 7 d) of the incubation period for each state. For each incubation period category, the recovery rate for the infection was calculated as γ= 3,4 & 5.

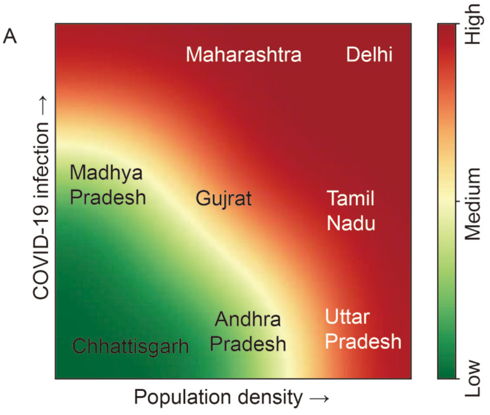
Scenario in selected states

States with high population densities

Delhi, capital of India, the most densely populated state in India, has the highest COVID-19 infection rate, as listed in Fig.1. The SEIRD results for Delhi is presented in Fig. 2A and the detailed data is shown in Table 1C. The value of $R_0(Dl)$ was calculated to be 2.053, indicating that the pandemic should decline in Delhi by 101 to 153 days. However, 4.80% to 7.99% of the susceptible population may still be actively infected in Delhi.

Tamil Nadu (TN), a southern state in India, has a high population density and a moderate COVID-19 infection rate, as reported in Fig.1. The SEIRD results for Tamilnadu is outlined in Fig. 2B and the detailed data is shown in Table 1G. According to our calculations, the value of $R_0(TN)$ is 2.356, suggesting that the pandemic should decline in TN by 95 to 143 days. However, 6.2% to 10.4% of the susceptible population may still be actively infected in TN.

Uttar Pradesh (UP), the state with the highest population in India, has a high population density, but a low COVID-19 infection rate, as reflected in Fig. 1. The SEIRD results for UP is indicated in Fig. 2C and the detailed data is listed in Table 1H. According to our calculations, the value of $R_0(UP)$ is 2.416, meaning that the pandemic should decline in MP by 99 to 150 days. However, 6.5% to 10.8% of the susceptible population may still be actively infected in UP.



B

States	Population density (km^{-2})	Infection (Per million)
Andhra Pradesh	303	84
Chhattisgarh	189	74
Delhi	11297	1489
Gujarat	308	332
Madhya Pradesh	236	119
Maharashtra	365	685
Tamil Nadu	555	438
Uttar Pradesh	828	46

Figure 1. Choice of states. (A) Choice of states (not to the scale). (B) Population density and infection/million population of the states chosen.

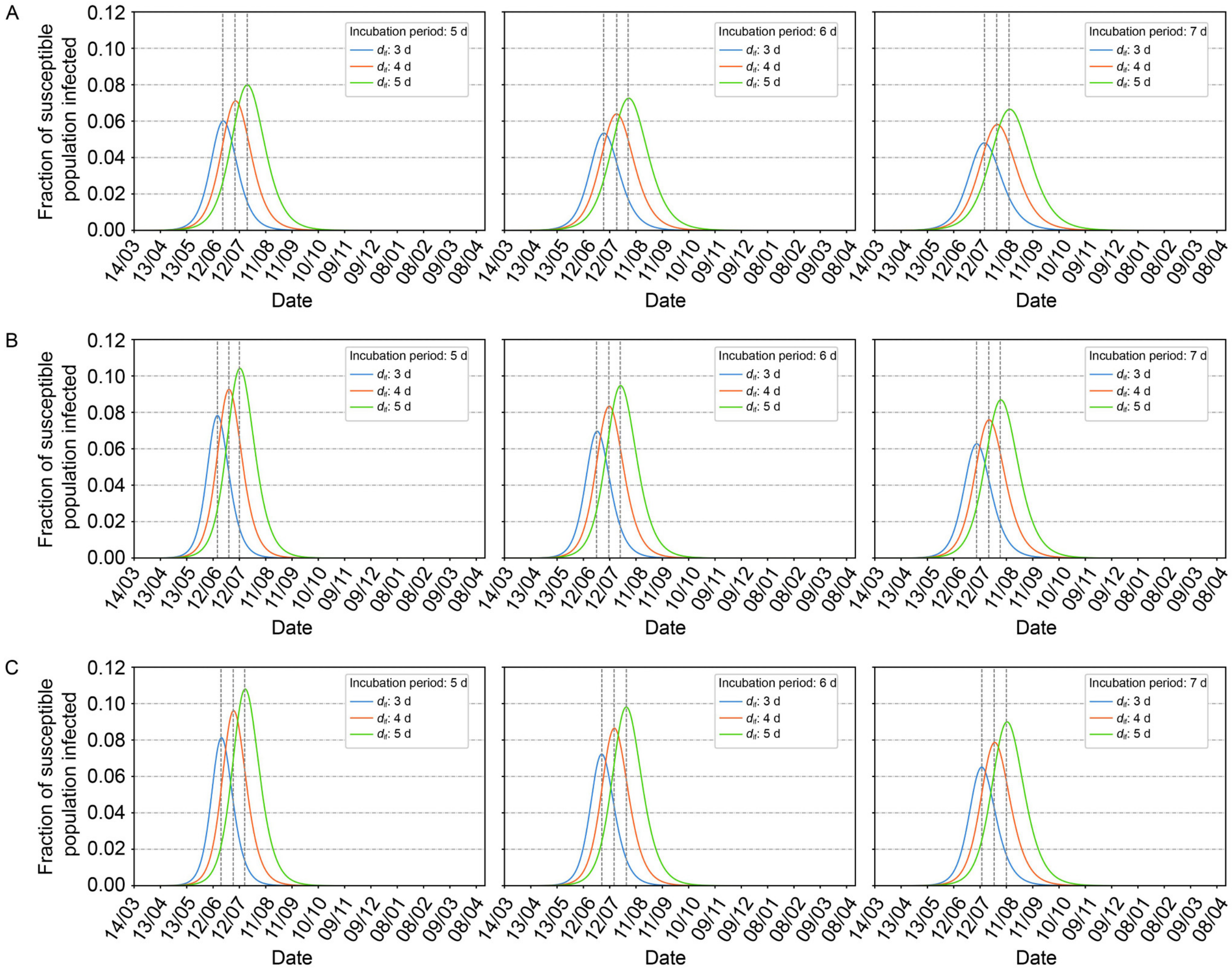


Figure 2. States with high population density. (A) Delhi, (B) Tamil Nadu, (C) Uttar Pradesh. Infection with incubation period of (left) 5 d, (center) 6 d, (right) 7 d.

States with medium population densities

Maharashtra (MH), a state located at the west of India, has a moderate population density and a high COVID-19 infection rate, as reported in Fig.1. The SEIRD results for Maharashtra is recorded in Fig. 3A and the detailed data is presented in Table 1F. According to our calculations, the value of $R_0(MH)$ is 2.087, indicating that the pandemic should decline in MH by 116 to 176 days. However, 4.9% to 8.2% of the susceptible population may still be actively infected in MH.

Gujarat (GJ), also located in western India, has a moderate population density and COVID-19 infection rate, as described in Fig.1. SEIRD results for Gujarat is presented in Fig. 3B and the detailed data is shown in Table 1D. According to our calculations, the value of $R_0(GJ)$ is 2.339, stating that the pandemic should decline in Gujarat by 95 to 144 days. However, 6.1% to 10.1% of the susceptible population may still be actively infected in GJ.

Andhra Pradesh (AP), a state located in the Deccan peninsula of India, has a moderate population density and a low COVID-19 infection rate, as shown in Table 1B. SEIRD results for $\sigma = 5$, 6 & 7 and $\gamma = 3$, 4, & 5 are summarized in Fig. 3C and the detailed data is listed in Table 1A. According to our calculations, the value of $R_0(AP)$ is 2.439, denoting that the pandemic should decline in AP by 89 to 134 days. These days correspond to the best and worst case scenarios in terms of the temporal duration of the disease. However, 6.6% to 11.0% of the susceptible population may still be actively infected in AP.

States with low population densities

Madhya Pradesh (MP) is a central state in India with a

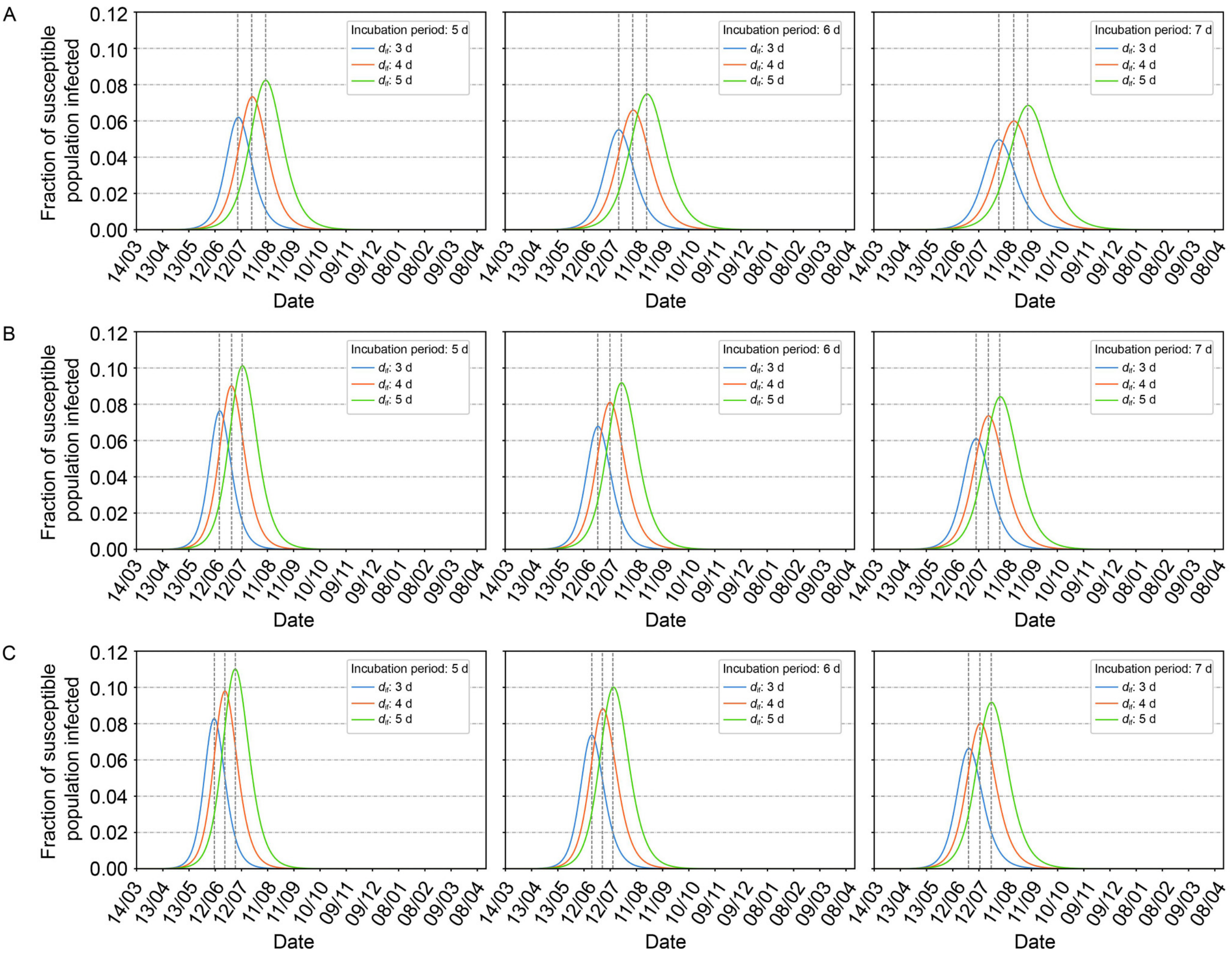


Figure 3. States with medium population density. (A) Maharashtra, (B) Gujarat, (C) Andhra Pradesh. Infection with incubation period of (left) 5 d, (center) 6 d, (right) 7 d.

low population density and a moderate COVID-19 infection rate, as outlined in Fig.1. SEIRD results for Madhya Pradesh is listed in Fig. 4A and the detailed data is shown in Table 1E. According to our calculations, the value of $R_0(MP)$ is 2.303, suggesting that the pandemic should decline in MP by 99 to 149 days. However, 5.9% to 9.9% of the susceptible population may still be *actively* infected in MP.

Chhattisgarh (CH), also a state in central India, has a low population density and COVID-19 infection rate, as illustrated in Table 1B. SEIRD results are listed in Fig. 4B and the detailed data is shown in Table 1B. According to our calculations, the value of $R_0(CH)$ is 2.414, suggesting that the pandemic should decline in CH by 85 to 128 days. However, 6.5% to 10.8% of the susceptible population may still be actively infected in CH.

As a whole, India has a high population density, but a low COVID-19 infection rate, with only 30 infected cases per million population. However, due to the 1.3 billion population, the absolute figure for India seems alarmingly high and beyond the capacity of the existing medical infrastructure. The SEIRD results for India is presented in Fig. 5 and the detailed data is shown in Table 1I. According to our calculations, the value of $R_0(India)$ is 1.917, indicating that the pandemic may come to an end in India by 180 to 240 days, depending on the incubation and infectiousness periods. During peak times, 6.5% to 8.5% of the susceptible population may be actively infected in India. To obtain a real estimation of the total infected population when the disease crosses the $R_e \leqslant 1$ mark, we calculated the R_e value for regular periods between 44–145 d. since the onset of the epidemic using Eq. (4). After applying a liner fit, we can see that the value of R_e should decrease below 1.0 by the end of October (with d_i= 5 and T_g= 10 d), as shown in Fig. 6B. This result closely corresponds with the Gaussian fit of the daily infection rate data,

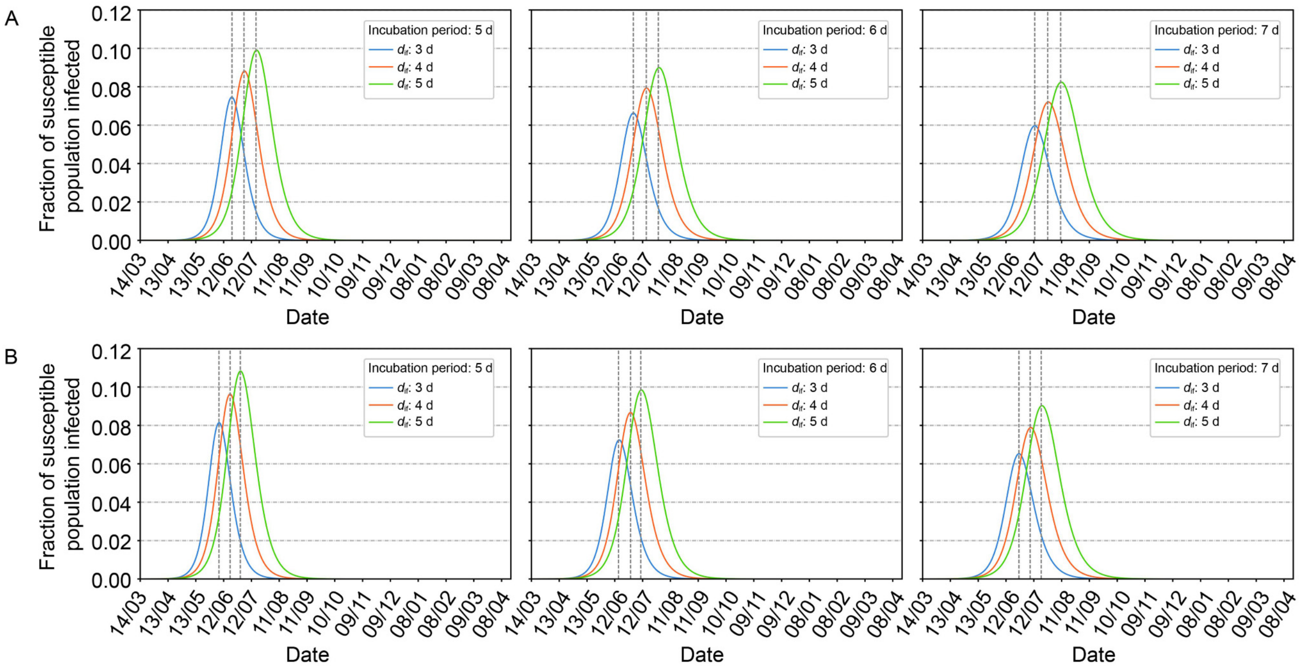


Figure 4. States with low population density. (A) Madhya Pradesh, (B) Chhattisgarh. Infection with incubation period of (left) 5 d., (center) 6 d, (right) 7 d.

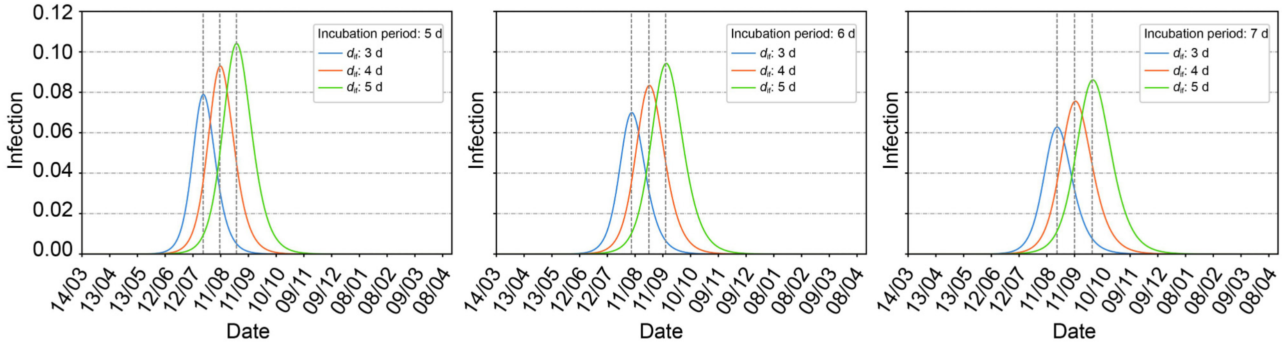


Figure 5. India: Infection with incubation period of (left) 5 d, (center) 6 d, (right) 7 d.

suggesting that the pandemic will reach its peak around the same time (with approximately a week of delay), as indicated in Fig. 6C. For the fitting of Fig. 6C, we used correlation functions(C), with C (amplitude, center) = 0.999, C (center, sigma) = 0.997, C (amplitude, sigma) = 0.992. These two trends also correspond with the SEIRD model plot of a 7 day incubation period and a 5 day infectiousness period in India (5).

The SEIRD model calculations for Italy are shown in Fig. 7.

Infection scenario in India

After applying the Gaussian fit (6c), we can see that there will be a total of approximately 2 million cases during the peak of the COVID-19 pandemic. At that point, the total number of infected individuals will be roughly 8.5 million and the daily infection rate can increase to as high as 1.2×10^5. The SEIRD model predictions will be slightly higher than the actual numbers since the model assumes that, initially, the entire population is susceptible to the disease. With that in mind, we need to determine the actual number of initially susceptible individuals $S(0)$, which can be achieved using the Gaussian fit. If we take the peak value corresponding to $d_i = 7$ and $d_{if} = 5$ (this case is the closest to the peak of the Gaussian fit) on plot (5) and compare them with the effective active cases predicted from the Gaussian fit, we can then calculate the value of $S(0) \approx 100$ million. Therefore, the initial susceptible population of India is approximately 7% of its total population.

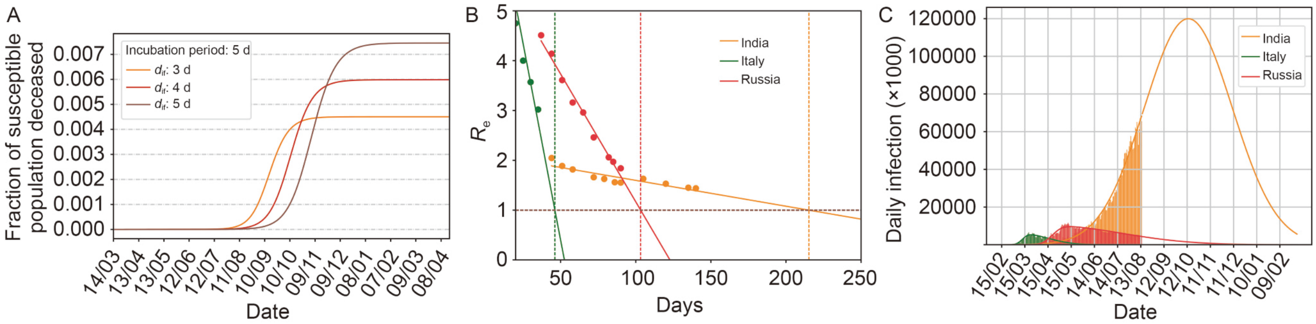


Figure 6. Indian scenario: comparison with Italy and Russia. (A) Death curves from SEIRD model for India. (B) Variation in R_e, with di=5, T_g=10. (C) Fitting of daily incidence.

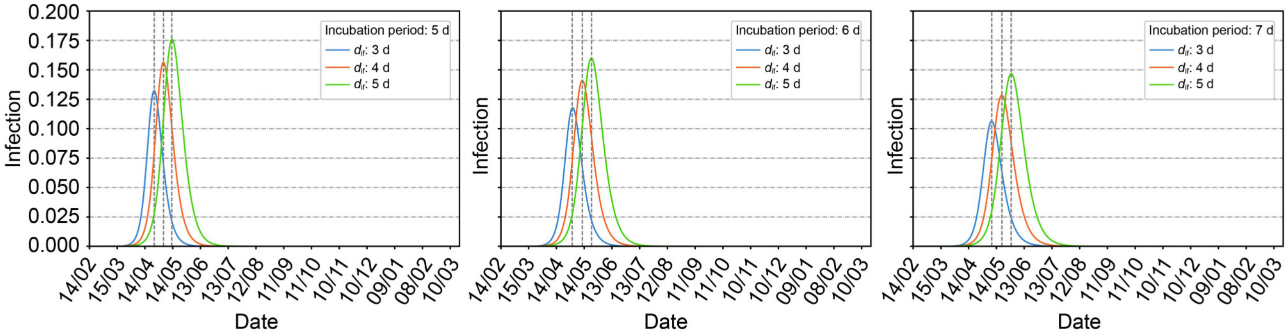


Figure 7. Infection scenario for Italy from SEIRD model.

Finally, to compare our results with that of other countries, we obtained data from Worldometers [20] between February 15th and June 25th. Specifically, we selected Italy with approximately 3,969 cases per million population, where the pandemic is virtually over, and Russia with 4,254 cases per million population, where the pandemic is still severe but in decline. As previously mentioned, India has a low daily infection rate (364 cases per million population) (Fig. 6C). During the peak of the pandemic, India will have approximately 1.2 million cases, which is greater than that of countries like the United States and Brazil, with over 70,000 cases at peak. The R_e plots for the above three countries are listed in Fig. 6B. Since the peak of infection had already appeared in both Italy and Russia, we can compare our calculations to the actual peak dates. Strikingly, our predictions closely correspond with the actual dates of the pandemic peaks. We calculated that Italy will peak around the fourth week of March and Russia will peak around the third week of May. In reality, Italy's peak appeared only 5 d later than predicted and Russia's peak appeared approximately 10–14 d earlier than predicted. Therefore, there is a high chance that India's COVID-19 infection peak will also follow the R_e curve plotted above[1]. Furthermore, the R_e plot reflects how increasing the testing frequency (reported by MoHFW, GoI during last 2–3 months) can positively affect the dynamics of the pandemic since the peak of infection will likely arise earlier. We can also see the effect of quarantine and social distancing on the temporal progression of the pandemic. The basic reproduction ratio will be higher if we include data up to August (see Fig. 3) compared to if we only use data up to June because the increase in testing frequency will consequently increase the number of known infected cases. Figure 8A shows a comparative plot of the total number of cases using data up to the lockdown enforced on May 4th, which indicates a saturation of cumulative incidence of around 45,000. This outcome is much lower than what would have been the case had India not enforced the lockdown as early as March 24th, 2020, which is estimated to be around 250 million infected cases. This curve is not shown in Fig. 8A due to its large deviation from the other two curves.

DISCUSSION

This study can be divided into two sections. The first

---

[1] In fact, the peak appeared roughly one week earlier than the prediction from R_e plot.

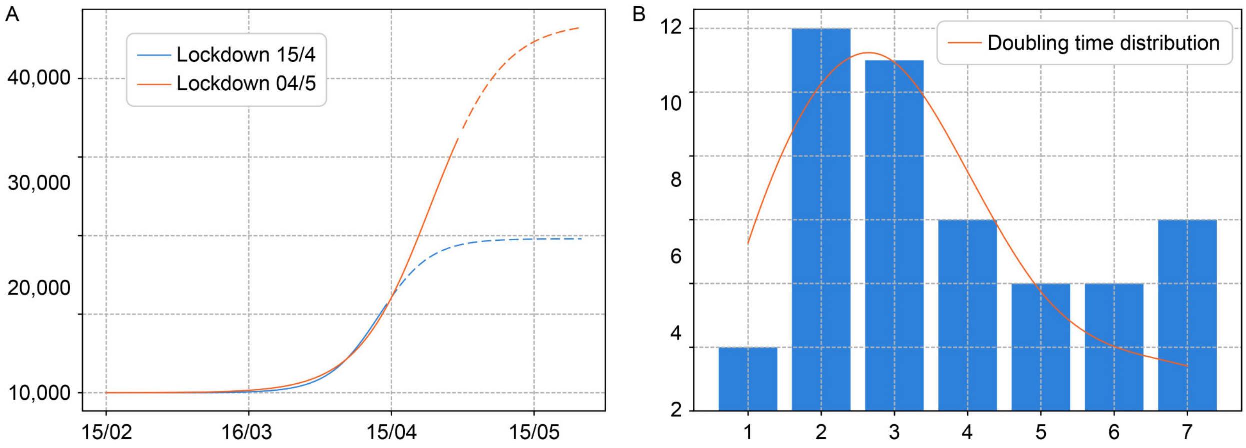


Figure 8. Variation of daily and cumulative incidence of COVID-19 cases in India. (A) Effect of lockdown on cumulative incidence. (B) Mean doubling time estimation for India.

section involves estimating several parameters used to predict the course of the COVID-19 infection in different states of India. The second section involves implementing the SEIRD model using the previously estimated parameters. More specifically, using COVID-19 data, we calculated the $R_0 s$ for India and eight of its states with varying population densities and infection rates. Overall, we implemented a simple, yet effective model to explain the growth and decline of the pandemic within the context of India.

Due to the unavailability of detailed and structured data to prominently project the incubation rate, recovery rate, etc., we considered a range of values based on the trends of the eight states and India as a whole. The general trend is that longer incubation and infectious periods would postpone the peak of infection. Our calculations show that an incubation period of 7 d and an infectious period of 5 d are the most plausible values after comparing the plots generated by the SEIRD epidemic model with real life data. Using these values, the infection would be contained within 210–240 d of onset in India (assumed to be March 14th in our study), meaning the curve would start to flatten around the middle of October. However, if the containment measures are not obeyed and community transmission occurs, the effective reproduction ratio would increase, causing a delay in the containment of the disease.

If an estimated 6.5% of the infected population require medical support, the total number of hospitalizations would be around 200,000 in the most optimistic situation, which would be well within India's healthcare capacity.

Our prediction of the peak of infection is dependent on two parameters, the incubation period (IP) and infectiousness period (IFP). Since we lacked the necessary data for an exact evaluation of these parameters, we considered a range of values (5–7 d for IP and 3–5 d for IFP) that had been reported in other studies. Figure 5 illustrates the peaks of different IP and IFP values. To predict the most likely scenario, we used a Gaussian fit (Fig. 6C), which suggests a peak of infection around October 12th. Knowing this, we selected the graph that predicted the closest date to be October 12th and found that, the combination of IP= 7 d and IPF= 5 d would give us the most accurate prediction. However, in reality, the peak of infection in India appeared around September 17th to 20th, which pointed to IP= 7 d and IPF= 4 d as the closest match. Therefore, this analysis helped us determine these two model parameters for India, which yielded in a fairly accurate prediction. Other studies, such as the one conducted by Mandal *et al.* [16], calculated a very low R_0 value of 1.03, which indicated a much lower incidence at the peak of infection. This mismatch was probably due to the adoption of an exponential growth model. On the other hand, the study conducted by Chatterjee *et al.* [17] overestimated the R_e value, which resulted in an inaccurate prediction of June, 2020 for the peak of infection.

In addition, we utilized a particular combination of methods to calculate the basic reproductive ratio in conjunction with the serial intervals, which is discussed in detail in the Methodology section. We also proposed methods to estimate the initial susceptible population, number of infected cases at peak, and number of deceased cases through comparison of the Gaussian peak and SEIRD model plots.

This study has several limitations, for example, the computation method for the effective reproduction ratio (R_e) can be improved using more detailed data and a time-dependent method, as suggested by Wallinga *et al.* [21]. The apparent mismatch of 3–4 weeks seen in the

Table 1 Result for SEIRD Model

(A) Andhra Pradesh

$R_0 = 2.439$		
σ (d^{-1})	γ (d^{-1})	$I_{\max}$(d)
5	3	89
	4	101
	5	113
6	3	99
	4	111
	5	123
7	3	108
	4	121
	5	133

(B) Chhattisgarh

$R_0 = 2.414$		
σ (d^{-1})	γ (d^{-1})	$I_{\max}$(d)
5	3	85
	4	97
	5	108
6	3	94
	4	107
	5	118
7	3	104
	4	116
	5	128

(C) Delhi

$R_0 = 2.503$		
σ (d^{-1})	γ (d^{-1})	$I_{\max}$(d)
5	3	101
	4	115
	5	129
6	3	113
	4	128
	5	141
7	3	124
	4	139
	5	153

(D) Gujarat

$R_0 = 2.339$		
σ (d^{-1})	γ (d^{-1})	$I_{\max}$(d)
5	3	95
	4	109
	5	121
6	3	106
	4	120
	5	133
7	3	117
	4	131
	5	144

(E) Madhya Pradesh

$R_0 = 2.505$		
σ (d^{-1})	γ (d^{-1})	$I_{\max}$(d)
5	3	99
	4	112
	5	125
6	3	110
	4	124
	5	137
7	3	121
	4	135
	5	149

(F) Maharashtra

$R_0 = 2.087$		
σ (d^{-1})	γ (d^{-1})	$I_{\max}$(d)
5	3	116
	4	132
	5	148
6	3	130
	4	146
	5	162
7	3	143
	4	160
	5	176

(G) Tamil Nadu

$R_0 = 2.356$		
σ (d^{-1})	γ (d^{-1})	$I_{\max}$(d)
5	3	95
	4	108
	5	120
6	3	105
	4	119
	5	132
7	3	116
	4	130
	5	143

(H) Uttar Pradesh

$R_0 = 2.416$		
σ (d^{-1})	γ (d^{-1})	$I_{\max}$(d)
5	3	99
	4	113
	5	126
6	3	111
	4	125
	5	139
7	3	122
	4	136
	5	150

(I) India

$R_0 = 1.917$		
σ (d^{-1})	γ (d^{-1})	$I_{\max}$(d)
5	3	126
	4	143
	5	159
6	3	140
	4	158
	5	175
7	3	155
	4	173
	5	190

Gaussian fit and model scenario may be due to the fact that the data was plotted starting on March 14th, despite the first case of COVID-19 reported as early as the last week of January. Additionally, the incubation and infectiousness periods should be calculated using the more detailed data to identify an index case and study its subsequent transmission channels, which would increase the accuracy of the final prediction. Furthermore, an age-structured analysis would be a useful way to illustrate the effect of the disease on different age groups. Finally, the methodology used in this study should yield reasonably accurate predictions when applied to other countries as well. Despite its limitations, we are still obtaining acceptable predictions due to the prolonged growth phase of the COVID-19 pandemic in India, which is likely caused by the early lockdown and containment measures.

METHODS

We employed a standard epidemiological model for infectious diseases, namely the SEIRD model, which is briefly discussed below.

SEIRD model

The Susceptible-Exposed-Infectious-Recovered-Dead (SEIRD) model is used for data modelling and analysis. Since COVID-19 has a latent phase, when the infected individual is not yet infectious, the delay between the acquisition of infection and the infectious state can be incorporated within the simpler SIR model by adding a latent/exposed population E, and moving the infected (but not yet infectious) population from S to E and from E to I. In the equation below, D refers to the deceased.

The dynamical equations for the SEIRD model are listed below:

$$\begin{aligned} &\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta S I \quad \frac{\mathrm{d}I}{\mathrm{d}t} = \beta S I - \gamma I - \nu I \\ &\frac{\mathrm{d}E}{\mathrm{d}t} = \beta S I - \sigma E \quad \frac{\mathrm{d}D}{\mathrm{d}t} = \nu I \\ &\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I \end{aligned} \tag{1}$$

For simulation purposes, we can rewrite the dynamical equations into the discrete time progression form, as listed below:

$$\begin{aligned} S(t+1) &= S(t) - R_0 \gamma S(t) I(t) \\ E(t+1) &= E(t) + R_0 \gamma S(t) I(t) - \sigma E(t) \\ I(t+1) &= I(t) - \gamma I(t) - \nu I(t) + \sigma E(t) \\ R(t+1) &= R(t) + \gamma I(t) \\ D(t+1) &= D(t) + \nu I(t) \end{aligned} \tag{2}$$

In these equations, the initial conditions are $S(0) > 0$, $E(0) > 0$, $I(0) > 0$ and $R(0) > 0$ and the constants β, γ, σ, ν are defined below. The mortality rate can be calculated using COVID-19 data from India and is related to ν. The recovery rate is the inverse of the mean infectiousness period, which is the average healing time since the onset of symptoms. The recovery rate is taken as 3, 4 and 5 d, as suggested by Bi *et al.* [6], and based on the studies, the inverse of the recovery rate gives us γ in Eq. (1).

Data source

Data of the daily COVID-19 infection, recovery, and deceased cases between March 14th and June 12th are collected from publicly available websites [20,22]. The official websites of the Indian government were also consulted periodically for an overview of the situation [23]. In addition, we also used population density data from India's 2011 census [24] and international data from the Worldometer (up to July 12th) for data pre-processing before carrying out the SEIRD-based model calculations.

Computational techniques

The definitions and computation methods of the SEIRD model parameters are described below:

Basic Reproductive Ratio (R_0): The reproduction rate R_0 can be calculated using the daily infection cases since Day 1 (March 14th). Assuming the daily incidence obeys an approximate Poisson distribution, the Maximum Likelihood Estimate method is employed. Given the observed number of incident cases (N_0, N_1, N) over consecutive time units and a generation time distribution (w), R_0 is estimated by maximizing the log-likelihood

$$LL(R_0) = \sum_{t=1}^{T} \mathrm{Log}\left(\frac{\mathrm{e}^{-\mu_t}\mu_t^{N_t}}{N_t!}\right), \tag{3}$$

where $\mu_t = R_0 \sum_{i=1}^{T} N_{t-i} w_i$. The random vector w_i, can be estimated as the mean doubling time (Fig. 8B), or the serial interval distribution, defined as the mean duration needed for the number of cases to double within a given period. [11, 21].

R_0 was calculated using data up to March 23rd, which produced a value of 1.54 [10]. Similarly, the $R_0 = 2.093$ for data up to May 24th and $R_0 = 1.917$ for data up to June 12th. The R_0 value for India is approximately 2.64 using data up to August 15, 2020. However, this should not be taken under consideration since the computation of R_0 should be conducted in the early, growing stage of epidemic.

Serial Interval (T_g): The serial interval is defined as the sum of the incubation period and infectious period. The incubation period refers to the number of days between the time of infection and symptomatic development, while the infectious period refers to the number of days an infected individual remains infectious following exposure. Doubling time is defined as the time required for the cumulative infection cases to double [25, 26], which is also a function of both T_g and R_0: $\frac{\ln(2)\times T_g}{R_0-1}$ [27].

Effective Reproductive Ratio (R_e): This is a dynamical quantity that reflects the average number of individuals infected by other infected individuals during a pandemic. The R_e value also takes into consideration the effectiveness of the preventative measures put in place to contain the spread of the infectious disease. Therefore, as time progresses and preventative measures (*i.e.*, lockdown, social distancing, contact tracing) are employed by the government and local administrations, this value will vary. The effective reproductive ratio can be calculated in several ways [3] using detailed data that can trace secondary infections generated by a primary infection. However, without such data, we resorted to a different method described by Lipsitch [28] that uses the average latent/incubation period (d_i), average serial interval (T_g), growth rate (λ) of the epidemic in its exponential phase. The reproductive number then can be estimated as:

$$R_e = 1 + \lambda T_g + f(1-f)(\lambda T_g)^2, \tag{4}$$

where $\lambda = \ln[Y(t)]/t$, and f is the ratio of the mean latent period (time between infection and onset of symptoms), to the serial interval, or doubling time in our case. T_g is taken as the sum of the mean latent period and the mean duration of infectiousness.

Our calculations show that the effective reproductive ratio has reduced from 1.84 on the 55th day to 1.44 on the 140th day with $T_g = 10$ d. For $T_g = 8$ d, the calculated value of $R_e = 1.34$.

Recovery rate (γ): The recovery rate is defined as the inverse of the infectious period d_{if}. Data obtained from the WHO and other studies [6, 7, 14] suggest that the infectious period for COVID-19 is between 3–5 d.

Incubation rate (σ): The incubation rate is calculated by taking the inverse of the mean incubation periodd_i. Data obtained from the WHO and other studies [6, 7, 13, 11, 15] suggest that the incubation period is between 5–7 d.

Mortality Rate (ν): The mortality rate is calculated by dividing the number of deceased individuals by the total number of infected individuals within a certain time period [17].

ACKNOWLEDGEMENTS

We acknowledge the infrastructural supports provided by IIIT Allahabad, where this study was conducted during the lockdown period.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Rudra Banerjee, Srijit Bhattacharjee and Pritish Kumar Varadwaj declare that they have no conflict of interests.

All procedures performed in studies were in accordance with the ethical standards of the institution or practice at which the studies were conducted.

OPEN ACCESS

REFERENCES

1. WHO director-general's remarks at the media briefing on covid-19. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020 (2020). Accessed: 24 June, 2020
2. Joshi, A. and Paul, S. (2020) Phylogenetic analysis of the novel coronavirus reveals important variants in Indian strains. bioRxiv, 2020.04.14.041301

3. Obadia, T., Haneef, R. and Boëlle, P. Y. (2012) The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med. Inform. Decis. Mak., 12, 147
4. Zhang, S., Diao, M., Yu, W., Pei, L., Lin, Z. and Chen, D. (2020) Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. Int. J. Infect. Dis., 93, 201−204
5. Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A. and Colaneri, M. (2020) Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat. Med., 26, 855−860
6. Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S. A., Zhang, T., *et al*. (2020) Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect. Dis., 20, 911−919
7. Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., Jit, M. and Klepac, P., and the Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group. (2020) The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health, 5, e261−e270
8. Park, M., Cook, A. R., Lim, J. T., Sun, Y. and Dickens, B. L. (2020) A systematic review of COVID-19 epidemiology based on current evidence. J. Clin. Med., 9, 967
9. Covid-19 in india. https://cddep.org/covid-19/hospital-capacity-in-india/. Accessed: 10 June, 2020
10. Bhatnagar, T., Mandal, S., Arinaminpathy, N., Agarwal, A., Chowdhury, A., Murhekar, M., Gangakhedkar, R. and Sarkar, S. (2020) Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: A mathematical model-based approach. Indian J. Med. Res., 151, 190−199
11. Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G. and Lessler, J. (2020) The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann. Intern. Med., 172, 577−582
12. Chan, J. F. W., Yuan, S., Kok, K. H., To, K. K. W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C. Y., Poon, R. W. S., *et al.* (2020) A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet, 395, 514−523
13. Backer, J. A., Klinkenberg, D. and Wallinga, J. (2020) Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20−28 January 2020. Euro Surveill., 25, 2000062d
14. Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D., Fraser, C.. (2020) Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science, 368, eabb6936
15. Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., *et al*. (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. N. Engl. J. Med., 382, 1199−1207
16. Mandal, M. and Mandal, S. (2020) COVID-19 early pandemic scenario in India compared to China and rest of the world: a data driven and model analysis. World J. Advan. Res. Rev., 7, 131−140
17. Chatterjee, S., Sarkar, A., Chatterjee, S., Karmakar, M. and Paul, R. (2020) Studying the progress of COVID-19 outbreak in India using SIRD model. Indian J. Phys., https://doi.org/10.1007/s12648-020-01766-8
18. Woelfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Mueller, M. A., Niemeyer, D., Vollmar, P., Rothe, C., Hoelscher, M., *et al*. (2020) Virological assessment of hospitalized cases of coronavirus disease 2019. medRxiv, 2020. 03.05.20030502
19. Yamauchi, T., Takeuchi, S., Yamano, Y. and Nakadate, T. (2019) Estimation of the effective reproduction number of influenza based on weekly reports in Miyazaki Prefecture. Sci. Rep., 9, 2539
20. Worldometers.info. Total coronavirus cases in india. https://www.worldometers.info/coronavirus/country/india/. Accessed: 12 June, 2020
21. Wallinga, J. and Teunis, P. (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am. J. Epidemiol., 160, 509−516
22. https://www.covid19india.org/. Accessed: 12 June, 2020
23. Government of India. Covid-19 statewise status. https://www.mohfw.gov.in/. Accessed: 10 June, 2020
24. Office of the Registrar General & Census Commissioner. India. Provisional population totals. https://censusindia.gov.in/. Accessed: 10 June, 2020
25. Vynnycky, E. and White, R. (2010) An Introduction to Infectious Disease Modelling. London: Oxford University Press
26. Anderson, R. and Robert, M. (1992) Infectious Diseases of Humans. London: Oxford University Press
27. Cummings, D. and Lessler, J. (2007) Infectious disease dynamics. http://oncohemakey.com/infectious-disease-dynamics/. Accessed: 10 June, 2020
28. Lipsitch, M., Cohen, T., Cooper, B., Robins, J. M., Ma, S., James, L., Gopalakrishna, G., Chew, S. K., Tan, C. C., Samore, M. H., *et al*. (2003) Transmission dynamics and control of severe acute respiratory syndrome. Science, 300, 1966−1970