medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Analyses and Forecast for COVID-19 epidemic in India

Rudra Banerjee, Srijit Bhattacharjee, Pritish Kumar Varadwaj

Indian Institute of Information Technology (IIIT), Allahabad, Deoghat, Jhalwa, Uttar Pradesh, India, 211015

Abstract

COVID-19 is a highly infectious disease that is causing havoc to the entire world due to the newly discovered coronavirus SARS-CoV-2. In this study, the dynamics of COVID-19 for India and a few selected states with different demographic structures have been analyzed using a SEIRD epidemiological model. A systematic estimation of the basic reproductive ratio R_0 is made for India and for each of the selected states. The study has analysed and predicted the dynamics of the temporal progression of the disease in Indian and the selected eight states: Andhra Pradesh, Chhattisgarh, Delhi, Gujarat, Madhya Pradesh, Maharashtra, Tamil Nadu, and Uttar Pradesh. For India, the most optimistic scenario with respect to duration of the epidemic shows, the peak of infection will appear before mid September with the estimated $R_0 = 1.917$, from the SEIRD model. Further, we show, a Gaussian fit of the daily incidences also indicates the peak will appear around middle of August this year. Our analyses suggest, the earliest dates when the epidemic will start to decline in most states are between Jun-August. For India, the number of infected people at the time of peak will be around 1.6 million including asymptomatic people. If the community transmission is prohibited, then the epidemic will infect not more than 3.1 million people in India. We also compared India's position in containing the disease with two countries with higher and lower number of infections than India and show the early imposition of lockdown has reduced the number of infected cases significantly.

1. Introduction

COVID-19 was announced as a global pandemic by World Health Organization (WHO)[1] owing to its highly contagious and pathogenicity that has been rapidly spreading throughout the world since its first reported outbreak in China in December, 2019. This highly contagious virus is spreading in a much higher rate than the other earlier reported corona viruses. In India, the first case of COVID-19 was reported on 30 January, 2020 and even after 23 weeks of its existence and four phases of lock-down, the daily rise of cases is alarming. To manage an epidemic like this, the administrations and health departments across the country like India with 1.35 billion population, need a more or less clear idea of the challenges they need to tackle during the course of epidemic. Also a forecasting about the said outbreak can help in decisive planning and management of the resource allocation. Therefore a detailed analysis based on real data should prove to be very much useful to estimate the possible amount of testing kits, hospitalizations, quarantine centres etc. needed to contain the disease.

Email addresses: rudrab@iiita.ac.in (Rudra Banerjee), srijuster@gmail.com (Srijit Bhattacharjee), pritish@iiita.ac.in (Pritish Kumar Varadwaj)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The motivation behind this work is to forecast the situation with the help of SEIRD population model, a well established model to study the dynamics of infectious diseases. This may be useful for the health officials and administration to estimate how they should be prepared for coming days to provide treatment and containment measures for COVID-19. The assessment for different states will also show the trend of the epidemic. With the sufficient amount of data in our hands, we can nowcast and forecast the possible situations that a state or the whole country is going to face. India invoked complete nationwide lockdown in the early phase of the epidemic on March 25th. This has prevented rapid transmission of the virus among people and also has given time to the administration and health officials to prepare for the measures that needed to be taken to tackle the situation. However, an early lockdown for a country with over 13 billion people has slowed down the pace of the spread of the virus considerably and resulted in a delayed appearance of flattening of the temporal progression of the infected people. It is thus important to find how long should the countrymen remain alert and careful about the spread of the virus.

For prediction of the progression of epidemic, we need a well structured data. The accurate estimation of model parameters would give us most plausible situation. Although the data published by ICMR, and other agencies are quite well organised but not sufficient to compute certain parameters accurately. In this situation we have two options. Either we take help from estimates of parameters from other countries where a well organised data was used or we can take data of various states in India with different demographics, compute the parameters for each of them, and compute the values of those parameters to find an estimate the parameters. We have chosen 3 states with high population density (Delhi, Tamilnadu, Uttar Pradesh), three states with medium population density (Andhra Pradesh, Maharashrtra, Gujarat) and two states of lower population density (Madhya Pradesh, Chhatisgarh). These states are also further divided into high, moderate and low infection rates. We have employed same methodology for each of these states and projected the values of the different parameters and executed the Susceptible-Exposed-Infected-Recovered-Dead (SEIRD) model to see the trend of COVID-19 transmission. The trends are then compared with the existing data for those states, which in turn enhance predictability.

The recently reported work [2] has compared the whole genome sequence of 28 viral strains from India. The analysis has reported a novel non-synonymous mutation in NSP3 gene of SARS-CoV-2 along with other frequent and important mutations reported across the globe. This suggests the possibility of having different strains at different part of India looking into the demographic density, travel history, inflow of migrant labours, cultural practices and living style. It may be also the reason of variation of the basic reproductive ratio (R_0) values across the selected states, as reported in this study.

As already mentioned, for studying the SEIRD mode, the parameters like-incubation rate, mortality rate, R_0 and recovery rate are required to be calculated. For a given epidemic R_0 is fixed,it denotes the number of secondary infections are generated from a seed infection. The value of R_0 is calculated when an epidemic is in its growing stage or early stage. We have employed a maximum likelihood estimation method depending on 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, 5, 6, 7, 8, 9, 10]. The range is expected to become narrower if analysed with more data. Containment measures may change the effective reproductive ratio R_e and in India it is currently between 1.43-1.55 as estimated in this study, for incubation periods of 5 days and serial intervals of 8 and 10 days respectively. The incubation period (d_i) for COVID-19 has been estimated between 2-14 days (WHO). In many studies it is suggested that mean incubation period is between 5-8 days [11, 12, 6, 13, 14]. Also the average incubation period is estimated to be 5.2 in a Wuhan based study by Li *et al.* [15] as notified in Worldometer. We have taken a range of 5-7 days, for mean incubation period motivated by these studies and Indian studies [16, 17]. The

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

other important parameter for modelling the epidemic is to find the period of infectiousness. For India, without having relevant data, we have resorted to look upon other studies. The other studies suggest mean infectiousness period (d_{if}) is between $3-5$ days [7, 18]. It is to also remember that d_{if} for India will be on lower side, i.e., will be close to 3 days as lock-down, social distancing, contact tracing and isolation of the infected or exposed people are done from very early stage of the disease. However, for a comparative study we have considered the values $3, 4, 5$ days for analysing the model. The mortality rate of COVID-19 is generally low. In India it is only 3.1% of the total infected population has deceased as of current data. We have calculated the mortality rates for 8 states and for India using total number of deaths and total confirmed cases over a period.

This study shows, India has done much better job than some other countries where the number of infections per million has been quite large. The infection in India will decline around 10-12 Spetember. Other states also may see flattening of the curve between June-August. With a Gaussian fit of the real data, India will see a peak around mid August if no drastic change in the disease dynamics appears. The total number of infections would have been at least three times higher than the present scenario with containment measures in place for India.

Finally, a comparison of progress of COVID-19 with two countries with higher and lower infections than India are done. Russia has more infections than India at the moment and Italy has already moved to almost zero incidence limit. The comparison shows that due to huge population, India's total number of infections may turn out to be higher than these countries but the rate of infections per million will be quite low. This would mean that the number of cases with severe illness will be under control, which not going to pose a serious challenge to the Government based health care providers. However, there will still be challenges in states like Maharashtra and Delhi due to high rate of infections.

2. Model

We have employed a standard epidemiological model for infectious disease, namely SEIRD model. Let us briefly discuss the model.

2.1. SEIRD Model

To model and analyze the data the Susceptible(S) - Exposed(E) - Infectious(I) - Recovered(R) - Dead(D) model is used. Since COVID-19 has a latent phase during which the individual is infected but not yet infectious. This 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 letting infected (but not yet infectious) individuals move from S to E and from E to I. D is the dead population.

The dynamical equations for SEIRD model are :

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI & \frac{dI}{dt} &= \beta SI - \gamma I - \nu I \\ \frac{dE}{dt} &= \beta SI - \sigma E & \frac{dD}{dt} &= \nu I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{2.1}$$

For simulation purpose we cast the above equations in the following discrete time progression form

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

$$\begin{aligned} S(t+1) &= S(t) - R_0\gamma S(t)I(t) \\ E(t+1) &= E(t) + R_0\gamma S(t)I(t) - \sigma E(t) \\ R(t+1) &= R(t) + \gamma I(t) \end{aligned} \qquad \begin{aligned} I(t+1) &= I(t) - \gamma I(t) - \nu I(t) + \sigma E(t) \\ D(t+1) &= D(t) + \nu I(t) \end{aligned} \tag{2.2}$$

with initial conditions $S(0) > 0, E(0) \geqslant 0, I(0) \geqslant 0$ and $R(0) \geqslant 0$. Constants $\beta, \gamma, \sigma, \nu$ are defined in Section 3. Mortality rate can be calculated from Covid-India data and is related to ν. The recovery rate is inverse of mean infectiousness period, that is average time to get cured from the onset of symptoms. The recovery rate is taken as $3, 4$ and 5 days, as suggested by Bi *et al.* [6] from the studies based on Wuhan, inverse of which gives γ in equation (2.1).

2.2. Data source

The number of people infected, recovered and deceased daily from 14th March to 12th June from Indian government's website [19, 20] that are publicly available and data from public domain [21, 22] was obtained. From this available data, few pre-processing were carried out for the SEIRD based model calculations. The population density was collected from 2011 census [23].

3. Methodology

Let us now describe the definitions and method of computation of the parameters for studying the population dynamics with the help of SEIRD model.

Basic Reproductive Ratio. (R_0): The reproduction rate R_0 can be calculated with the daily cases starting from day 1 (March-14). Assuming that daily incidence obeys an approximate Poisson distribution, one employs Maximum Likelihood Estimate method. Given observation of $(N_0, N_1, ..., N)$ incident cases over consecutive time units, and a generation time distribution w, R_0 is estimated by maximizing the log-likelihood

$$LL(R_0) = \sum_{t=1}^{T} Log\left(\frac{e^{-\mu_t}\mu_t^{N_t}}{N_t!}\right), \tag{3.1}$$

where $\mu_t = R_0 \sum_{i=1}^{T} N_{t-i} w_i$. The random vector w_i, can be estimated as mean doubling time. w is also called serial interval distribution [11, 24].

R_0 calculated for data upto 23 March produced a value between $1.5 - 4$ [10]. R_0 produced for data up to 24 May, is 2.093 and $R_0 = 1.917$ for data up to 12 June.

Serial Interval. (T_g): The serial interval use to be the sum of incubation period and infectious period. Incubation period is the number of days after which a person develops symptoms after being exposed to an infected. Whereas infectious period is defined as the number of days an infected person remains infectious or can infect another person, after getting infected. Doubling time is defined as time period in which the cumulative infection doubles [25, 26]. Doubling Time is a function of both T_g and R_0 as $\frac{ln(2)\times T_g}{R_0 - 1}$ [27].

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Effective Reproductive Ratio. (R_e): This is a dynamical quantity that reflects on average during a pandemic the number of person gets infected by an infected person. R_e takes care the effect of measures taken to prevent the spread of the infectious disease. Therefore it will vary as time progresses and lockdown, social distancing, contact tracing etc. are being employed by Government and local administrations. The effective reproductive ratio can be calculated in several ways [3] using a detailed data that can trace the secondary infections generated by a primary infector. However, in absence of such data we have employed a different method depicted by Lipsitch [28]. The method uses average latent period (incubation period) d_i, average serial interval T_g, growth rate λ of the epidemic when it is in exponential phase. The reproductive number then can be estimated as

$$R_e = 1 + \lambda T_g + f(1-f)(\lambda T_g)^2, . \tag{3.2}$$

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

We have seen the effective reproductive ratio has reduced to 1.55 on the 90th day from 1.84 (on 55th day) currently with $T_g = 10$ days. For $T_g = 8$ days, the value of R_e has been calculated equal to 1.43.

Recovery Rate. (γ): The recovery rate is defined as the inverse of infectious period d_{if}. The WHO data and other studies [14, 29, 7, 6] indicate the infectious period is between 3-5 days.

Incubation rate. (σ): Incubation rate is calculated by taking inverse of mean incubation period d_i. WHO data and other studies [30, 11, 13, 29, 7, 6] indicate the incubation period is between 5-7 days.

Mortality Rate. (ν): The mortality rate is calculated as number of individuals died on a given period of time divided by the total infected in that period [17].

4. Results

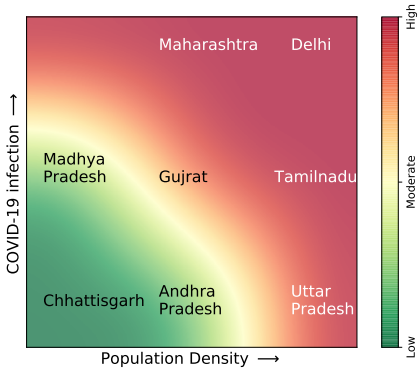
We have chosen 8 states based on their population density and COVID-19 infected per million population, as shown in Figure (1). We have omitted low population - high infection, as it does not exist in Indian scenario. We have considered population density $> 500/km^2$ as high and $< 300/km^2$ as low. On the other hand, number of infection $> 500/10^6$ people, is treated as high and $< 100/10^6$ is considered as low in our distribution.

Previous results suggest average incubation period of SARS-CoV-2 virus between 5 to 7 days [11, 31, 6]. So, we have studied the scenario for incubation period of 5(best case scenario), 6 and 7(worst case scenario) days for each state. Also for each incubation period, we have calculated the infection for γ= 3,4 & 5. Previous results suggest incubation period of SARS-CoV-2 virus between 5 to 7 days [11, 31, 6]. So, we have studied the scenario for incubation period of 5(best case scenario), 6 and 7(worst case scenario) days for each state. Also for each incubation period, we have calculated the infection for γ=3,4 & 5.

4.1. Scenario in selected states

Andhra Pradesh (AP),. a state in Deccan peninsula of India, has moderate population density and low covid infection, as shown in table (1b). The SEIRD results for $\sigma = 5,\ 6$ & 7 and $\gamma = 3,\ 4,$ & 5 is shown in Figure (2). Detailed data are shown in Table (1a). According to our calculations, $R_0(AP)$ is 2.439. Our calculations also shows the pandemic should over in AP by 89 to 134 days. *These*

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



(a) Choice of states (not to the scale)

States	Population Density($/km^2$)	Infection ($/10^6$)
Andhra Pradesh	303	84
Chhattisgarh	189	74
Delhi	11297	1489
Gujarat	308	332
Madhya Pradesh	236	119
Maharashtra	365	685
Tamilnadu	555	438
Uttar Pradesh	828	46

(b) Population density and infection/million population of the states chosen.

Figure 1: Choice of states.

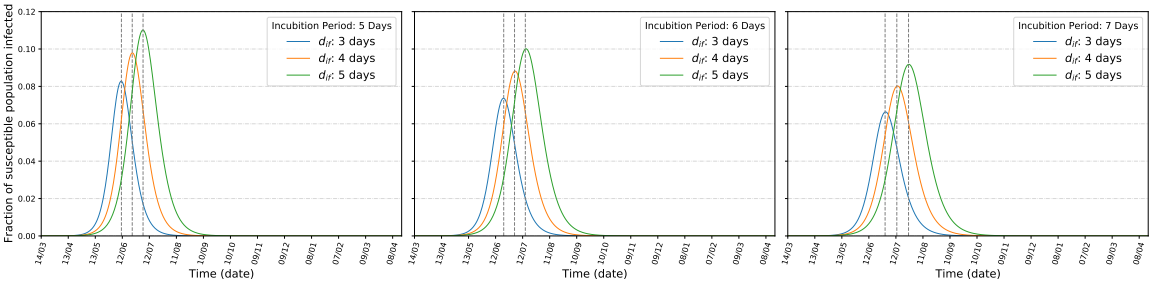


Figure 2: Andhra Pradesh: Infection with incubation period of (left) 5 days, (center) 6 days, (right) 7 days

days correspond to the best and worst scenarios in terms of temporal duration of the disease. 6.6% to 11.0% of susceptible population may be *actively* infected in AP.

Chhattisgarh (CH),. a state in central India, has low population density and low covid infection, as shown in table (1b). The SEIRD results are shown in Figure (3). Detailed data are shown in Table

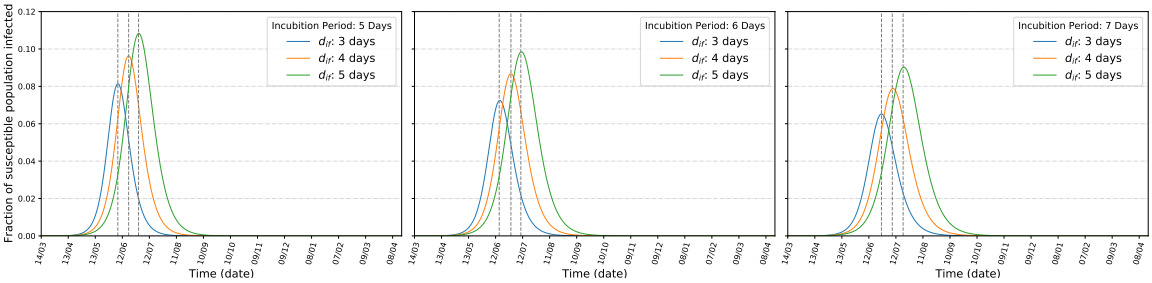


Figure 3: Chhattisgarh: Infection with incubation period of (left) 5 days, (center) 6 days, (right) 7 days

(1b). According to our calculations, $R_0(CH)$ is 2.414. Our calculations also shows the pandemic should over in CH by 85 to 128 days. 6.5% to 10.8% of susceptible population may be *actively*

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

infected in CH.

Delhi,. the capital of India, is with highest COVID-19 infection and also the most densely populated state of India as shown in Figure (1). SEIRD results for Delhi is shown in Figure (4) and the detailed data is shown in Table (1c). According to our calculation, $R_0(Dl)$ is 2.053. Our calculations also

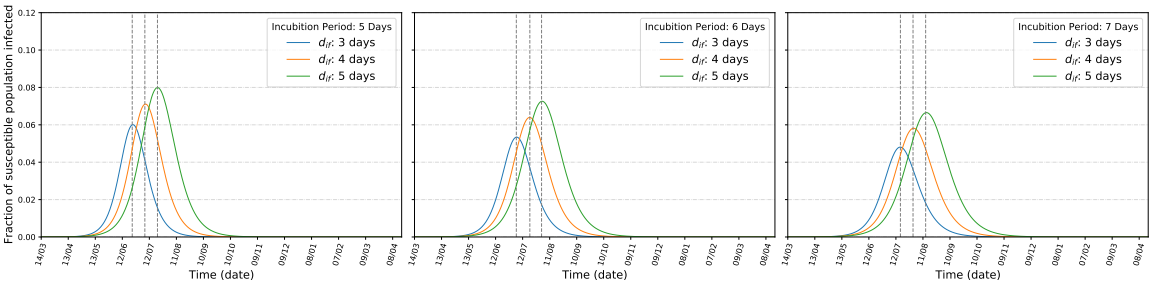


Figure 4: Delhi: Infection with incubation period of (left) 5 days, (center) 6 days, (right) 7 days

shows the pandemic should over in Delhi by 101 to 153 days. 4.8% to 7.99% of susceptible population may be *actively* infected in Delhi.

Gujarat(GJ),. the western state in India, has moderate population density and also moderate COVID-19 infection, as shown in Figure (1). SEIRD results for Gujrat is shown in Figure (5) and the detailed data is shown in Table (1d). According to our calculation, $R_0(GJ)$ is 2.339. Our

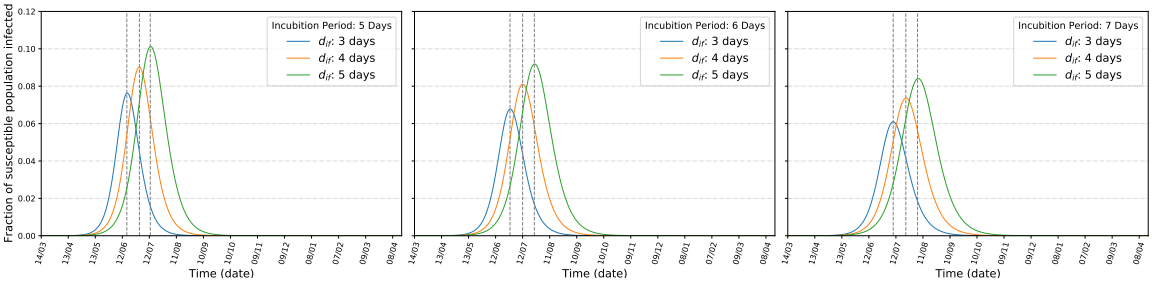


Figure 5: Gujarat: Infection with incubation period of (left) 5 days, (center) 6 days, (right) 7 days

calculations also shows the pandemic should over in Gujarat by 95 to 144 days. 6.1% to 10.1% of susceptible population may be *actively* infected in GJ.

Madhya Pradesh(MP),. the central state in India, has low population density and also moderate COVID-19 infection, as shown in Figure (1). SEIRD results for Madhya Pradesh is shown in Figure (6) and the detailed data is shown in Table (1e). According to our calculation, $R_0(MP)$ is 2.303. Our estimation also shows the pandemic should over in MP by 99 to 149 days. 5.9 to 9.9% of susceptible population may be *actively* infected in MP.

Maharashtra(MH),. another western state in India, has moderate population density and high COVID-19 infection, as shown in Figure (1). SEIRD results for Maharashtra is shown in Figure (7) and the detailed data is shown in Table (1f). According to our calculation, $R_0(MH)$ is 2.087. Our calculation also shows the pandemic should be over in MH by 116 to 176 days. 4.9% to 8.2% of susceptible population may be *actively* infected in MH.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

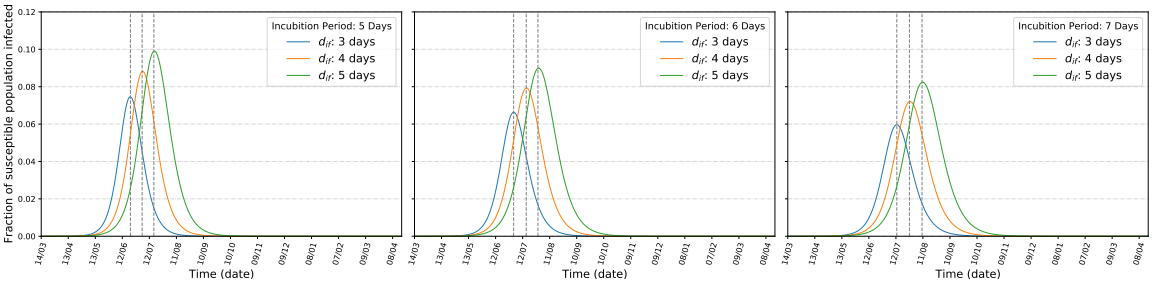


Figure 6: Madhya Pradesh: Infection with incubation period of (left) 5 days, (center) 6 days, (right) 7 days

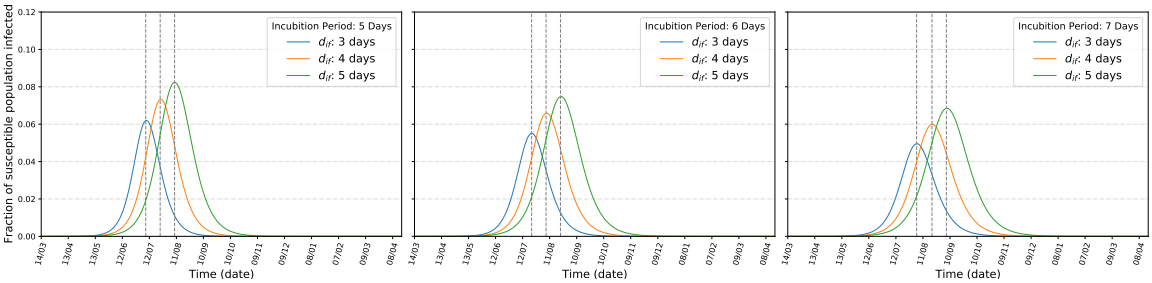


Figure 7: Maharashtra: Infection with incubation period of (left) 5 days, (center) 6 days, (right) 7 days

Tamilnadu(TN),. a southern state in India, has high population density and moderate COVID-19 infection, as shown in Figure (1). SEIRD results for Tamilnadu is shown in Figure (8) and the detailed data is shown in Table (1g). According to our calculation, $R_0(TN)$ is 2.356. Our

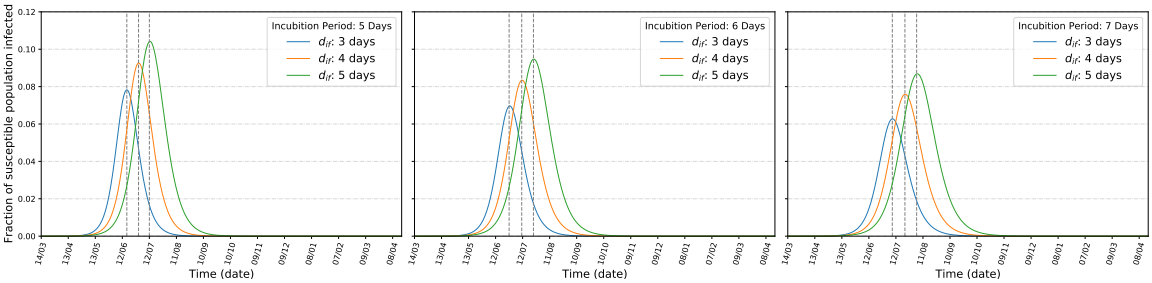


Figure 8: Tamilnadu: Infection with incubation period of (left) 5 days, (center) 6 days, (right) 7 days

computation also shows the pandemic should over in TN by 95 to 143 days. 6.2% to 10.4% of susceptible population may be *actively* infected in TN.

Uttar Pradesh(UP),. with highest population in India, has high population density but low COVID-19 infection, as shown in Figure (1). SEIRD results for UP is shown in Figure (9) and the detailed data is shown in Table (1h). According to our calculation, $R_0(UP)$ is 2.416. Our study also shows the pandemic should over in MP by 99 to 150 days. 6.5 to 10.8% of susceptible population may be *actively* infected in UP.

Finally, India as a whole, has high population density but low COVID-19 infection, with only 30 infected per million population. But, in absolute figure, with 1.3 billion population, this figure is

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

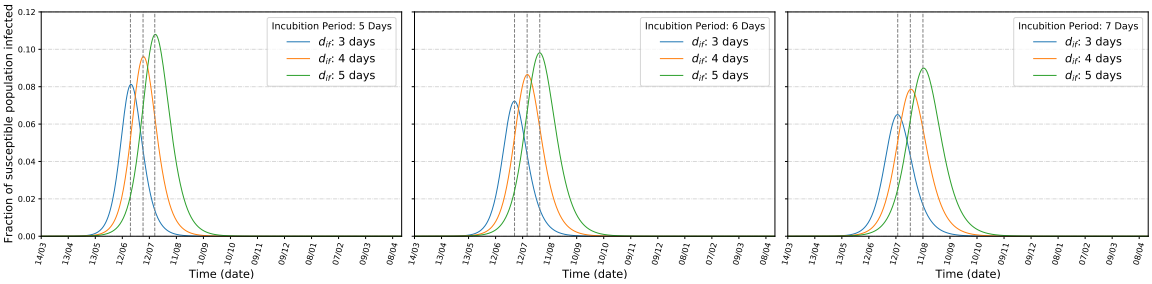


Figure 9: Uttar Pradesh: Infection with incubation period of (left) 5 days, (center) 6 days, (right) 7 days

alarmingly high and beyond the capacity of existing medical infrastructure. SEIRD results for India is shown in Figure (10) and the detailed data is shown in Table (1i). According to our calculation,

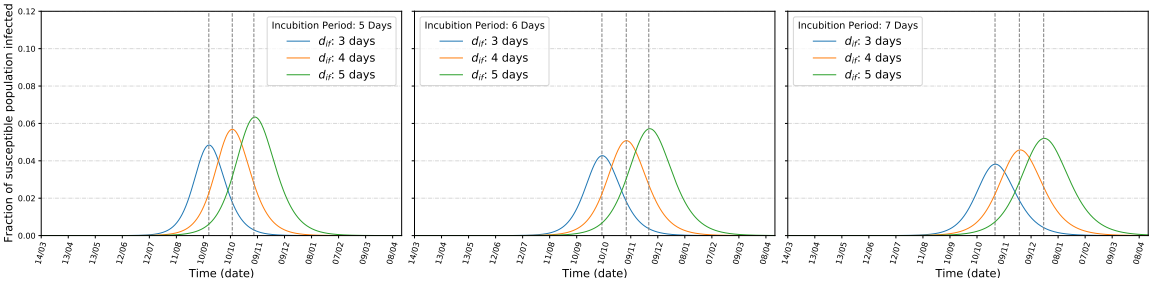


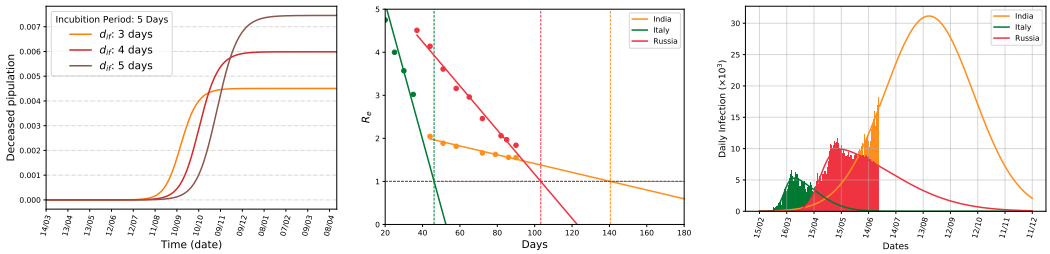
Figure 10: India: Infection with incubation period of (left) 5 days, (center) 6 days, (right) 7 days

$R_0(India)$ is 1.917. Our calculations also shows the pandemic should over in India by 186 to 284 days (depending on incubation and infectiousness periods). 4% to 6.8% of susceptible population may be active at the time of peak appears in India. To get a real estimate of total infected population when the disease will cross the $R_e \leq 1$ mark we have calculated R_e for regular periods between 44-90 days, starting from 14th May using equation (3.2). A linear fit shows R_e should go below 1.0 by first week of August (with $d_i = 5$ and $T_g = 10$ days) as shown in Figure (11b). This result matches with a Gaussian fit of daily infection, which also reaches its peak around the same time (with approx. a week's delay), as in Figure (11c). The R_e fit for $T_g = 8$ with same incubation period shows a peak at 20 days earlier than the peak at Gaussian fit. For the fitting of Figure (11c), we have got correlation functions(C) as C(amplitude, center)= 0.999, C(center, sigma)= 0.997, C(amplitude, sigma) = 0.992. These two trends support the plot with 5 days of incubation period and 3 days of infectiousness period for India (10) using the SEIRD model.

4.2. Infection scenario in India

From the Gaussian fit (11c), the total active cases is calculated to be around $5,12,700$ at the time of peak. Total number of infected people will be roughly 3.1 million at the end of transmission. With the SEIRD model, we can see the peak appears around September middle. From the ICMR data [32], it is observed that the current rate of positive cases is 6.4% of the number of tests done. If the test rate remains the same with a steady increase of 15000 tests per fortnight, the total number of tests in India will reach approximately 25.8 million around mid September. The present rate yields a total 1650560 infected cases in India around mid September. These projections are quite remarkably consistent with the predicted values from the Gaussian fit. From Gaussian fit, the total

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



(a) Death curves from SEIRD model for India.

(b) Variation in R_e, with d_i=5, T_g=10.

(c) Fitting of daily incidence.

Figure 11: Indian scenario: Comparison with Italy and Russia.

infected at the peak appears to be approximately 1.612 million. It must be noted, the numbers projected from the Gaussian fit are excluding the asymptomatic cases that are not being reported officially.

The prediction of actual number of infected people from SEIRD model will be grossly overestimated if one assumes that initially the entire population was susceptible. Therefore one needs to determine the actual number of initially susceptible population $S(0)$. This can be achieved here with the help of the Gaussian fit. From the plot (10), if we take the peak value corresponding to $d_i = 5$ and $d_{if} = 3$ (this case is the closest to the peak of the Gaussian fit), and compare with the effective active cases predicted from the Gaussian fit, then we get $S(0) \approx 10463265$. So, initial susceptible population for India is close to only 1% of the total population.

Number of deaths:. The SEIRD model predicts the peak of the Death curve, with $d_i = 5, d_{if} = 3$ will be only 0.45% of the actual susceptible population, see (11a). This projects ≈ 47085 deaths at the peak. This is quite expected as we are having total 16 million infected cases at the peak and with close to 3 percent death rate, the number turns out to be 48000, which is quite close to the value of the same predicted from the model. The total deaths at the end of the pandemic is expected to be around $90000 - 100000$.

Finally, we compare our result with others countries. For this calculation, we have taken data from Worldometers [22] from 15th February to 25th June. We have chosen Italy (3,969 cases/million population), where the pandemic is virtually over, and Russia (4,254 cases/million population), where pandemic is still in severe but in decline. As per our discussion in last paragraph, the figure (11c) shows though India has quite low infection rate (364/million population), per day. Also at the peak India will see almost 30000 case which is lower than the countries like USA, Brazil where at the peak the number exceeded well above 30k. The plots of R_e for these three countries are shown in the (3.2). As in both Italy and Russia, the peak of infection has already appeared, we can verify our analysis by comparing the plot with actual dates of peaks. The plots show that Italy has crossed the peak around fourth week of March and Russia has crossed it around third week of May. Strikingly, these are fairly accurate dates what we have observed in reality. For Italy, the peak appeared only ≈ 5 days late, and for Russia the peak appeared approximately 10-14 days earlier. Therefore, there is a high chance that India's COVID-19 infection peak will also follow the R_e curve plotted above.

5. Discussions

This study can be divided into two parts. The first part contains estimation of a few parameters that were used to predict the course of the disease in India and other states from the data. The

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	R_0=2.439	
σ(1/days)	γ(1/days)	I_{max}(days)
5	3	89
	4	101
	5	113
6	3	99
	4	111
	5	123
7	3	108
	4	121
	5	133

(a) Andhra Pradesh

	R_0=2.414	
σ(1/days)	γ(1/days)	I_{max}(days)
5	3	85
	4	97
	5	108
6	3	94
	4	107
	5	118
7	3	104
	4	116
	5	128

(b) Chhattisgarh

	R_0=2.503	
σ(1/days)	γ(1/days)	I_{max}(days)
5	3	101
	4	115
	5	129
6	3	113
	4	128
	5	141
7	3	124
	4	139
	5	153

(c) Delhi

	R_0=2.339	
σ(1/days)	γ(1/days)	I_{max}(days)
5	3	95
	4	109
	5	121
6	3	106
	4	120
	5	133
7	3	117
	4	131
	5	144

(d) Gujarat

	R_0=2.505	
σ(1/days)	γ(1/days)	I_{max}(days)
5	3	99
	4	112
	5	125
6	3	110
	4	124
	5	137
7	3	121
	4	135
	5	149

(e) Madhya Pradesh

	R_0=2.087	
σ(1/days)	γ(1/days)	I_{max}(days)
5	3	116
	4	132
	5	148
6	3	130
	4	146
	5	162
7	3	143
	4	160
	5	176

(f) Maharashtra

	R_0=2.356	
σ(1/days)	γ(1/days)	I_{max}(days)
5	3	95
	4	108
	5	120
6	3	105
	4	119
	5	132
7	3	116
	4	130
	5	143

(g) Tamilnadu

	R_0=2.416	
σ(1/days)	γ(1/days)	I_{max}(days)
5	3	99
	4	113
	5	126
6	3	111
	4	125
	5	139
7	3	122
	4	136
	5	150

(h) Uttar Pradesh

	R_0=1.917	
σ(1/days)	γ(1/days)	I_{max}(days)
5	3	186
	4	212
	5	236
6	3	208
	4	235
	5	260
7	3	230
	4	257
	5	284

(i) India

Table 1: Result for SEIRD Model

second part contains the implementation of SEIRD model with the help of the estimated parameters. In particular, using COVID-19 data we have computed the R_0s for India and 8 other states with varying population density and infection status. We have implemented a simple but effective model to explain the growth and decline on the epidemic from Indian context. Our study shows that, for states having low (Chhattisgarh) or moderate (Uttar Pradesh) infection rate the pandemic peaks as early as June, where in for a state like Maharashtra, it may take a month longer.

In view of unavailability of detailed and structured data to prominently project the incubation rate, recovery rate, etc., we have considered a range of values motivated from the trend of the disease in those 8 states and in India. The general trend is, for higher incubation and infectious period the peak appears late.

We have shown, an incubation period of 5 days and an infectious period of 3 days are the

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

most favourable values for which the infection will be under control within 186 days from the onset (assumed to be 14 March in our study) in India. This means the flattening of curve will start around middle of September. If the containment measures are not obeyed and community transmission happens then the effective reproduction ratio will increase and there will be delay in attaining the control over the disease. In the worst scenario the peak may appear around mid December.

A Gaussian plot with the available data for India indicates the flattening of the number of daily incidence will start around mid of August. Although this is earlier by almost three-four weeks than the prediction from SEIRD model but we can get a reasonable idea of expected number of infected at peak. This is calculated to be around 1.6 million. With an early estimation of R_0 as high as 4 before lockdown, the total infected population has been reduced to at least one-third of the early projected values.

As approximately 6.5% of the infected population are required hospital support, the total number of cases of hospitalisations should be around 200 thousands in the most optimistic situation. This is well within the capacity of the health care systems of India.

The states like Chhattisgarh, Karnataka, Kerala have already attained their peaks or on the verge of attaining. In particular, our analysis suggests, Chhattisgarh has already attained its peak infection. This is also consistent with the official data. For other states the peak will be attained within 1-2 months from now. So overall, the situation is improving. This is also indicated from the gradual reduction of R_e values for India in (11b).

The presentation made here have necessary caveats. The computation method for effective reproduction ratio (R_e) can be improved with a more detailed data and using a time-dependent method suggested by Wallinga *et al.* [24]. The apparent mismatch of 3-4 weeks of the Gaussian fit and model scenario may be attributed to the fact that the data was plotted from March 14 in Gaussian case, whereas the first case was detected as early as the last week of January. The incubation period, infectiousness period are needed to be calculated from more detailed data identifying an index case and then studying its subsequent transmission channels. That will make the prediction more accurate. An age-structured analysis of the disease would be useful to see the effect of the disease for different age groups.

Acknowledgements

We acknowledge the infrastructural supports provided by IIIT Allahabad where this work has been done during the lock down period.

Compliance with ethical standards

Conflict of interest We declare that we have no conflicts of interest.

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, accessed on 24th June, 2020 (2020).

[2] A. Joshi, S. Paul, Phylogenetic analysis of the novel coronavirus reveals important variants in indian strains, bioRxiv (2020). doi:10.1101/2020.04.14.041301.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[3] T. Obadia, R. Haneef, P. Y. Boëlle, The R0 package: A toolbox to estimate reproduction numbers for epidemic outbreaks, BMC Medical Informatics and Decision Making 12 (1) (2012). doi:10.1186/1472-6947-12-147.

[4] S. Zhang, M. Y. Diao, W. Yu, L. Pei, Z. Lin, D. Chen, 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, International Journal of Infectious Diseases 93 (2020) 201–204. doi:10.1016/j.ijid.2020.02.033.

[5] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the covid-19 epidemic and implementation of population-wide interventions in italy, Nature Medicine 26 (2020). doi:10.1038/s41591-020-0883-7.

[6] Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wu, Y. Wu, B. Sun, S. Huang, Y. Sun, J. Zhang, T. Ma, J. Lessler, T. Feng, Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study, The Lancet Infectious Diseases 3099 (20) (2020) 1–9. doi:10.1016/S1473-3099(20)30287-5.

[7] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. B. Pearson, J. D. Munday, S. Abbott, H. Gibbs, A. Rosello, B. J. Quilty, T. Jombart, F. Sun, C. Diamond, A. Gimma, K. [van Zandvoort], S. Funk, C. I. Jarvis, W. J. Edmunds, N. I. Bosse, J. Hellewell, M. Jit, P. Klepac, The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study, The Lancet Public Health 5 (5) (2020) e261 – e270. doi:https://doi.org/10.1016/S2468-2667(20)30073-6.

[8] M. Park, A. R. Cook, J. T. Lim, Y. Sun, B. L. Dickens, A systematic review of covid-19 epidemiology based on current evidence, Journal of Clinical Medicine 9 (4) (2020) 967. doi:10.3390/jcm9040967.

[9] cddep.org, Covid-19 in india, https://cddep.org/covid-19/hospital-capacity-in-india/, accessed on 10th June, 2020 (2020).

[10] T. Bhatnagar, S. Mandal, N. Arinaminpathy, A. Agarwal, A. Chowdhury, M. Murhekar, R. Gangakhedkar, S. Sarkar, Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in india: A mathematical model-based approach, Indian Journal of Medical Research 151 (2) (2020) 190. doi:10.4103/ijmr.ijmr_504_20.

[11] S. A. Lauer, H. Grantz, Kyra, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, J. Lessler, The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application, Annals of Internal Medicine 172 (9) (2020) 577–582. doi:10.7326/M20-0504.

[12] J. F.-W. Chan, S. Yuan, K.-H. Kok, K. K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C.-Y. Yip, R. W.-S. Poon, H.-W. Tsoi, S. K.-F. Lo, K.-H. Chan, V. K.-M. Poon, W.-M. Chan, J. D. Ip, J.-P. Cai, V. C.-C. Cheng, H. Chen, C. K.-M. Hui, K.-Y. Yuen, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, The Lancet 395 (10223) (2020) 514 – 523. doi:https://doi.org/10.1016/S0140-6736(20)30154-9.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[13] J. A. Backer, D. Klinkenberg, J. Wallinga, Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from wuhan, china, 20-28 january 2020, Euro Surveill. 25 (2020) 2000062. doi:10.2807/1560-7917.ES.2020.25.5.2000062.

[14] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, C. Fraser, Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing 368 (6491) (2020). doi:10.1126/science.abb6936.

[15] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, Z. Feng, Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia, New England Journal of Medicine 382 (13) (2020) 1199–1207. doi:10.1056/NEJMoa2001316.

[16] M. Mandal, S. Mandal, Covid-19 pandemic scenario in india compared to china and rest of the world: a data driven and model analysis, medRxiv (2020). doi:10.1101/2020.04.20.20072744.

[17] S. Chatterjee, A. Sarkar, S. Chatterjee, M. Karmakar, R. Paul, Studying the progress of covid-19 outbreak in india using sird model, medRxiv (2020). doi:10.1101/2020.05.11.20098681.

[18] R. Woelfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Mueller, D. Niemeyer, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Bruenink, J. Schneider, R. Ehmann, K. Zwirglmaier, C. Drosten, C. Wendtner, Clinical presentation and virological assessment of hospitalized cases of coronavirus disease 2019 in a travel-associated transmission cluster, medRxiv (2020). doi:10.1101/2020.03.05.20030502.

[19] Goverment of India, Covid19 statewise status, https://www.mygov.in/corona-data/covid19-statewise-status, accessed on 10th June, 2020 (2020).

[20] Government of India, Covid-19 statewise status, https://www.mohfw.gov.in/, accessed on 10th June, 2020 (2020).

[21] https://www.covid19india.org/.
URL https://www.covid19india.org/

[22] Worldometers.info, Total coronavirus cases in india, https://www.worldometers.info/coronavirus/country/india/, accessed on 12th June, 2020 (2020).

[23] Office of the Registrar General & Census Commissioner, India, Provisional population totals, https://censusindia.gov.in/, accessed on 10th June, 2020 (2011).

[24] J. Wallinga, P. Teunis, Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures, American Journal of Epidemiology 160 (6) (2004) 509–516. doi:10.1093/aje/kwh255.

[25] E. Vynnycky, R. White, An Introduction to Infectious Disease Modelling, Oxford University Press, 2010.

[26] R. Anderson, M. Robert, Infectious Diseases of Humans, Oxford University Press, 1992.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.26.20141077; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[27] D. Cummings, J. Lessler, Infectious disease dynamics, http://www.interfetpthailand.net/file/June2007/Presentation/Dr.Derek/cummings.simulation.day.1.pdf (2007).

[28] M. Lipsitch, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, G. Gopalakrishna, S. K. Chew, C. C. Tan, M. H. Samore, D. Fisman, M. Murray, Transmission dynamics and control of severe acute respiratory syndrome., Science 300 (2003) 1966–1970. doi:10.1126/science.1086616.

[29] J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, The Lancet 395 (10225) (2020) 689–697. doi:10.1016/S0140-6736(20)30260-9.

[30] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, Z. Feng, Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia, New England Journal of Medicine 382 (13) (2020) 1199–1207. doi:10.1056/NEJMoa2001316.

[31] T. Yamauchi, S. Takeuchi, Y. Yamano, Y. Kuroda, T. Nakadate, Estimation of the effective reproduction number of influenza based on weekly reports in Miyazaki Prefecture, Scientific Reports 9 (1) (2019) 1–9. doi:10.1038/s41598-019-39057-w.

[32] Indian Council of Medical Research, Sars-cov-2 (covid-19) testing: Status update, https://www.icmr.gov.in/pdf/covid/update/ICMR_testing_update_25June2020_9AM_IST.pdf, accessed on 24th June, 2020.