The Peak of COVID-19 in India

Suryakant Yadav, Pawan Kumar Yadav, Neha Yadav, Chandra Kant Yadav

Submitted to: JMIR Public Health and Surveillance
on: October 05, 2020

Table of Contents

The Peak of COVID-19 in India

Suryakant Yadav[1] PhD; Pawan Kumar Yadav[1] MPhil; Neha Yadav[2] MA; Chandra Kant Yadav[3] MSW

[1]International Institute for Population Sciences; Mumbai IN
[2]Jawaharlal Nehru University New Delhi IN
[3]Tata Institute for Social Sciences Mumbai IN

Corresponding Author:
Suryakant Yadav PhD
International Institute for Population Sciences;
Academic Building
Govandi St Road, Deonar
Mumbai
IN

Abstract

Background: Following the USA, India ranks the second position in the world for COVID-19 cases with the highest number of daily confirmed cases since September 2020. The peak of daily confirmed cases is the most warranted feature for understanding the epidemiological stage of COVID-19 disease in India.

Objective: The objectives of the study are to analyse the growth rates of the confirmed cases of COVID-19 in India, and to provide an expected count of the peak of confirmed cases and a possible track of confirmed cases.

Methods: Exponential model was applied to estimate the growth rates of daily confirmed cases. The estimated growth rates were used for calculating the doubling time. The Lotka-Euler method was applied to calculate the effective reproduction rate. SARIMA model was developed for the growth rates to predict daily confirmed cases.

Results: show the best fit of the exponential model over the daily confirmed cases. The growth rates estimated from the exponential model shows an unsteady, modest decline. Doubling time shows a linear increase. The effective reproduction rate declined from 3.6 persons in the third week of March 2020 to 1.14 persons at the end of August 2020 and 1.10 at the end of September 2020. The diagnosis of the developed SARIMA model confirmed no trends in the residuals, no outliers, and nearly constant variance. The forecast suggests the peak value of daily confirmed cases wavers around 104,500 counts in the third week of September 2020. The cumulative COVID-19 cases account for approximately 105 lakhs at the end of December 2020.

Conclusions: The exponential model unravels a shift and a modest decline in the growth of daily confirmed cases. The trends in R(t) show analogous to the trends in growth rates of daily confirmed cases. The study shows that the SARIMA model is suitable for projecting daily confirmed cases. The results shed light on the understanding of the trends and epidemiological stage of COVID-19 disease, in the cognisance of the peak. This study based on moments of the distribution of the daily confirmed cases of COVID-19 disease unravels the uncertainty about the peak and curvature of COVID-19 disease. The prediction made based on our calculations is adjoining to the real-time peak value of daily confirmed cases in India and successfully explores the epidemiological stage of COVID-19 disease in India.

(JMIR Preprints 05/10/2020:24477)
DOI: https://doi.org/10.2196/preprints.24477

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
- ✓ **Please make my preprint PDF available to anyone at any time (recommended).**
- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
- Only make the preprint title and abstract visible.
- No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
- ✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

The Peak of COVID-19 in India

Research Highlights

1. The exponential model is the best fit over daily confirmed COVID-19 cases.
2. The trends in growth rates of daily confirmed cases show a swift decline during the five-and-a-half month's period since April 2020.
3. The effective reproduction rate in India declined from 3.6 persons in the third week of March 2020 to 1.14 persons at the end of August 2020 and 1.10 persons at the end of September 2020.
4. The forecast reveals that the peak of daily confirmed cases wavers at approximately 104,500 cases since the third week of September 2020.
5. The effective reproduction rate value equals to one at the end of November 2020.
6. The cumulative confirmed cases of COVID-19 in India accounts to

Abstract

Background: Following the USA, India ranks the second position in the world for COVID-19 cases with the highest number of daily confirmed cases since September 2020. The peak of daily confirmed cases is the most warranted feature for understanding the epidemiological stage of COVID-19 disease in India.

Objective: The objectives of the study are to analyse the growth rates of the confirmed cases of COVID-19 in India, and to provide an expected count of the peak of confirmed cases and a possible track of confirmed cases.

Methods: Exponential model was applied to estimate the growth rates of daily confirmed cases. The estimated growth rates were used for calculating the doubling time. The Lotka-Euler method was applied to calculate the effective reproduction rate. SARIMA model was developed for the growth rates to predict daily confirmed cases.

Results: Results show the best fit of the exponential model over the daily confirmed cases. The growth rates estimated from the exponential model shows an unsteady, modest decline. Doubling time shows a linear increase. The effective reproduction rate declined from 3.6 persons in the third week of March 2020 to 1.14 persons at the end of August 2020 and 1.10 at the end of September 2020. The diagnosis of the developed SARIMA model confirmed no trends in the residuals, no outliers, and nearly constant variance. The forecast suggests the peak value of daily confirmed cases wavers around 104,500 counts in the third week of September 2020. The cumulative COVID-19 cases account for approximately 105 lakhs at the end of December 2020.

Conclusions: The exponential model unravels a shift and a modest decline in the growth of daily confirmed cases. The trends in $R(t)$ show analogous to the trends in growth rates of daily confirmed cases. The study shows that the SARIMA model is suitable for projecting daily confirmed cases. The results shed light on the understanding of the trends and epidemiological stage of COVID-19 disease, in the cognisance of the peak. This study based on moments of the distribution of the daily confirmed cases of COVID-19 disease unravels the uncertainty about the peak and curvature of COVID-19 disease. The prediction made based on our calculations is adjoining to the real-time peak value of daily confirmed cases in India and successfully explores the epidemiological stage of COVID-19 disease in India.

KEYWORDS: COVID-19, SARS-CoV-2, Exponential model, ARIMA, Effective Reproduction rate, Growth rates, Pandemic.

Introduction

In late December 2019, for the first-time pneumonia cases with an unknown cause were reported in Wuhan (China). In January 2020, the cause of those pneumonia cases was identified as a new type of coronavirus [1]. On 12 January 2020, the official name of the coronavirus as COVID-19 (Coronavirus disease 2019) for the disease as well as SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) for the virus was given by World Health Organisation [2]. [3]

Coronavirus disease of 2019 (COVID-19) has shown an unprecedented increase worldwide. Following the United States of America (USA), India ranks the second position in the world with cumulate COVID-19 cases at 36.8 lakhs on 31 August 2020 and crossed 50 lakhs on 15 September 2020 [4]. India accumulates approximately 15 lakh cases in 15 days, a large stockpile in two weeks only. India has recently surpassed Brazil in the number of COVID-19 cases in the first week of September 2020. Also, the growth rates of COVID-19 cases for the past in India do not show a dramatic and unexceptional decline and are higher as compared to that of other low ranked countries in pandemic. It is reckoning that India has one of the largest sizes of COVID-19 cases in the world. Hence, it is foreseeable that India's may overtake the USA in a ceteris paribus environment.

While the USA has experienced the second wave of COVID-19 cases, Brazil has witnessed the peak of COVID-19 cases since the end of July 2020 and after that, has shown a swift decrease in the number of COVID-19 cases. Other countries such as Russia, Peru, and South Africa have exhibited the peak of COVID-19 cases with the cumulative number of COVID-19 cases smaller than what the top three countries have exhibited during the same time [5]. Following the unprecedented 'stay-at-home' national policies, Europe and many other countries show receding COVID-19 pandemic recently [6]. Given these trends, India shows a different predicament. Unlike other countries, India is the only country showing upsurge of COVID-19 without showing a peak until the mid-September 2020.

The peak value of daily confirmed COVID-19 cases in the USA and Brazil was recorded at 77,362 and 69,074, respectively. However, India surpassed the USA in the daily confirmed COVID-19 cases and breached the peak value of the USA at the end of August 2020. In seven months since the first case on 30 January 2020, the daily confirmed COVID-19 cases have reached to the highest count in the world. The rapid rise of daily confirmed cases in India indicates an exponential growth of COVID-19 cases. This upsurge in COVID-19 cases is an alarming situation as most of the governing body guidelines have been in the best hope to observe the peak value in a short time.

India has taken note of the rise in COVID-19 cases. In view of that, the Government of India (GOI) has implemented lockdown from 25 March 2020 in many phases. For different phases of lockdowns, GOI has issued many guidelines from time to time intending to maintain social distancing. From June 2020, the GOI has also implemented unlocks with proper guidelines. In the absence of vaccine, medicine and drugs, these non-pharmaceutical interventions are the precautionary measures in a bid to flatten the curve of COVID-19 cases [3] accompanied by low growth rates as well as low effective reproduction rates. The success of non-pharmaceutical interventions is evident in many countries while encountering the spread of SARS-CoV-2 virus. The importance of non-pharmaceutical interventions, including isolation and wearing masks, to control the disease transmissibility is considered as successful among the Italian population despite a toll of deaths [7].

Effective reproduction rate ($R(t)$) is an essential epidemiological measure of an epidemic or pandemic. For COVID-19 disease , $R(t)$ measures the spread of SARS-CoV-2 from a primary infectious person to secondary infectious persons at a time 't'. The regional variation in $R_0(t)$, the basic reproduction rate is significant and varies between 3.21 and 1.90 persons among China and other Asian countries [8]. Nonetheless, $R(t)$ below a value of one is warranted to contain the spread of disease. Having achieved that, the epidemics shrink [9]. While vaccines are still under the process of development, social distancing, wearing mask, isolation, extensive testing, and quarantining of confirmed infected cases remain the most effective measures to contain the pandemic [10, 11].

The rise of confirmed cases indicates an exponential rise in the last seven months from 30 January 2020 to August 2020. Therefore, it is important to analyse the daily confirmed cases to comprehend the peak value and size of the COVID-19 cases. Authorities and academicians are looking forward to the flattening of the curve as the best outcome of non-pharmaceutical interventions in the absence of vaccine, drugs, and medicine.

Objective:

The objectives of the study are to:

1) Analyse the growth rates of the confirmed cases of COVID-19 in India,
2) Provide an expected count of the peak of confirmed cases and a possible track of confirmed cases.

Methods

Data Source

We retrieved data from an Application Programming Interface (API) portal https://www.covid19india.org/ which is open access and publicly available [4]. Data on COVID-19 cases on this portal is updated from state bulletins, official handles, Press Bureau of India (PBI), Press Trust of India (PTI), and Asian News International (ANI) reports. This study uses data from the API portal 'covid19india.org' because of the two advantages. The first is that it provides time-series data in a portable '*.csv' files, and the second is that it provides data up to the district level of India in the same format.

This 'covid19india.org' portal provides data on confirmed, active, recovered, and death cases, on a daily basis, from 30/01/2020, the first case of COVID-19 disease in India. We retrieved data between the dates 30/01/2020 and 30/09/2020. A period of five-and-a-half months, rolling from 15/03/2020 to 31/08/2020, is used for the analytical purpose of daily confirmed. The daily confirmed cases for the period between 01/09/2020 and 30/09/2020 is used for validating the outcomes for the next 30 days.

This period of five-and-a-half months accounted for analysis are consisting of the time-intervals as lockdowns and unlocks. These are Lockdown 1.0: 25/03/2020–14/04/2020, Lockdown 2.0: 15/04/2020–03/05/2020, Lockdown 3.0: 04/05/2020–17/05/2020, Lockdown 4.0: 18/05/2020–31/05/2020, Unlock 1.0: 1/06/2020–30/06/2020, Unlock 2.0: 1/07/2020–31/07/2020 and Unlock 3.0: 1/08/2020–31/08/2020. The Unlock 4.0 starts from 01/09/2020. The results in the figures are shown with these time-intervals.

Statistical Techniques

Exponential model and doubling time

The plot of confirmed cases of COVID-19 disease over time suggests an exponential model. The exponential model was applied with an intercept and also without an intercept. However, parameter estimates for the intercept converged to zero, and other shape and scale parameters converged to a significant value. One advantage of the exponential model is that it does provides time-invariant parameter estimates for the time-series data available for India. The haphazard trends in parameter estimates by states of India do not allow projections. Data at a lower geographical level in India is still a limitation for robust analysis.

The exponential model can take the form of growth and decay model depending upon growth rate greater than zero and smaller than zero, respectively. Exponential growth (decay) model is preferred for the following two reasons. First of all, the data has not shown any peak given the trends into consideration, and second, the acceleration or deceleration of the growth of confirmed cases is not explicit.

The exponential model is expressed as

$$y_t = a * b^t \quad \text{………… (1)}$$

where t is time or date, y_t is the number of daily confirmed cases at time t, a is constant, b is the acceleration or deceleration of the fit of the daily confirmed cases, and $b = e^r$ with r as the growth rate of daily confirmed cases.

Non-linear estimation method was applied on daily confirmed cases between 31 January 2020 and rolling from 15 March 2020 to 31 August 2020 to estimate of the value of model parameters 'b' and 'a' to get the best fit until each date. The estimated values of parameters 'b' and 'a' are time-invariant. The natural logarithm of the estimated value of parameter b from the exponential model gives the estimate of growth rate 'r' of daily confirmed cases.

The doubling time (T_d) [12] of COVID-19 cases using the growth rate of daily confirmed cases is expressed as

$$T_d = \frac{\ln(2)}{r} \quad \text{…………… (2).}$$

Effective reproduction rate

The Lotka-Euler equation [13] was applied to estimate the effective reproduction rate based on the growth rate of daily confirmed cases of COVID-19 obtained from exponential model. The Lotka-Euler equation is expressed as

$$R(t) = 1 / \int_{a=0}^{\infty} e^{-ra} g(a) da \quad \text{…………… (3)}$$

where, $R(t)$ is the basic reproduction rate, denoted by R_0, in the initial period ~5 to ~14 days of the spread of infectious disease, and is the effective reproduction rate, denoted by $R(t)$, at a time 't' of the outbreak of the disease; r is the growth rate of daily confirmed cases; $g(a)$ is the probability density function of generation (serial) interval, which is the time-span from a primary infectious person to generate secondary infectious persons in the time interval $[a, a+da]$ [14].

Mathematically, $R(t)$ depends on two statistics: the growth rate of daily cases and serial interval distribution. In the lack of duration of serial intervals in the observed duration of the infectious period, serial interval distribution applicable for China between 21 January 2020 and 8 February 2020 [15] was applied for India. The models such as Normal, Lognormal, Gamma and Weibull distribution were tested over 469 reported transmission events. It provided the shape and scale of these serial interval distributions. Of these models, the Gamma distribution shows the best fit of serial interval data in China. The Gamma distribution with the shape and scale equal to 1.46 and 0.78, respectively, is useful for India.

Projection of confirmed cases

Auto Regressive Integrated Moving Average (ARIMA) was applied to the exponential growth rate of daily confirmed COVID-19 cases [16, 17] from mid-March 2020 to 31 August 2020 to predict daily cases from 1 September 2020 to until 28 February 2021. ARIMA (p, d, q), where p is the order of autoregression, d is the degree of difference, q is the order of moving average, is expressed as

$$y_t' = \alpha + \beta_1 y_{t-1}' + \beta_2 y_{t-2}' + \ldots + \beta_p y_{t-k}' + \emptyset_1 \varepsilon_{t-1} + \emptyset_2 \varepsilon_{t-2} + \ldots \emptyset_k \varepsilon_{k-1} + \varepsilon_t \quad \ldots (4)$$

where y_t' is the differenced series of confirmed cases, y_{t-k}' is the lagged values of y_t of order k, ε_t is white noise, ε_{k-1} is the lagged errors of order k, α is an intercept term, β is an autoregressive parameter, and $\emptyset$ is a moving average parameter.

An ARIMA model is efficient in dealing at non-stationary and stationary time series. The SARIMA for time series is referred to as an ARIMA (p, d, q) (P, D, Q)[m] where (P, D, Q) represents the (p, d, q) for the seasonal part of the time series, and m refers to the number of observations per cycle. The accuracy of the SARIMA model is measured by root mean square error (RMSE) and mean absolute percentage error (MAPE).

The daily confirmed cases, adjusted R-squares, exponential growth rates, doubling time, and $R(t)$ are shown in the plot until 30 September 2020. The correlation between projected and estimated $R(t)$ values between 1 and 30 September 2020 was tested.

Assumptions:

The study assumes:

1. The trends in the future would be analogous to that of the trends in the past,
2. The seasonality in the trends in the past remains the same in the future,
3. The number of testing and the test positivity ratio does not change dramatically.

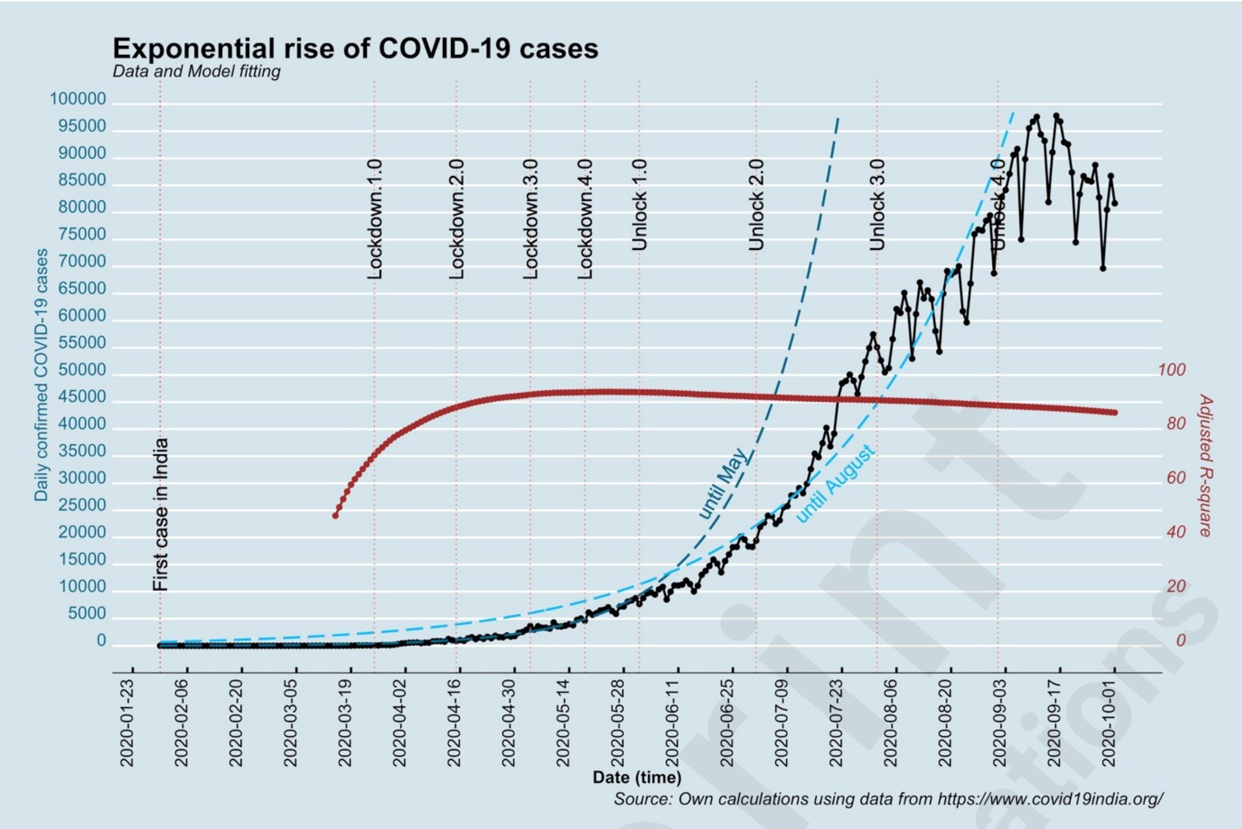
Results

Exponential model

Figure 1 shows the non-linear fits based on the Equation (1), which is an exponential function, over the daily confirmed cases until two dates at the end of May and August 2020. The daily confirmed COVID-19 cases are plotted until 30 September 2020. These non-linear fits (deep-sky-blue (dark and light) coloured long-dash line) vividly shows the exponential increase in daily confirmed cases. An interesting finding to ponder is the shift of the non-linear fit over time. The shift in the exponential fit indicates a slowdown in the increase of daily confirmed cases. Although the number of daily confirmed cases has been increasing, the cases seem lesser than expected counts of cases based on the exponential model.

Nonetheless, the burgeoning daily confirmed cases are supported by the fact that the adjusted R-square values (brown coloured connected circles) of the exponential fit is more than 85 per cent from May to August 2020. It is noteworthy that the fit of the exponential model over the data has become stronger since May when compared from mid-March to April 2020.

Figure 1: Plot of exponential fits over daily confirmed cases, India



Growth rates and doubling time

Figure 2 displays growth rates and doubling time (blue coloured connected circles) calculated from the estimate of parameter ‘b’ of the exponential model during the period of five-and-a-half months period from mid-March to August 2020. It is shown until 30 September 2020. The parameter ‘b’ or the slope of the exponential model explains the acceleration or deceleration of daily confirmed cases. The shift in the fit of the exponential model is an indication of a change in the slope, i.e. a change in the growth rates (r¿ of daily confirmed cases.

The trends in the growth rates of daily confirmed cases show high and non-steady values of growth rates with spikes in the latter half of March and April 2020. The highest value of growth rate was 28 per cent in the third week of March. The growth rates declined rapidly since the first week of April, and down the line, it declined to approximately a value of 5.5 per cent in the last week of April. However, at this end, the growth rates showed an abrupt increase. In the first week of May 2020, it increased to 6.4 per cent. This rise in growth rates was susceptible to the return of migrants from metro cities to their homeland, mostly in rural areas. The rise in growth rates has affected the trends in growth rates as its gradients ramped up. The growth rates declined, however, slowly in April and May. Also, a stagnancy in growth rates is evident, and it declined slowly to 5 per cent in the last week of May 2020.

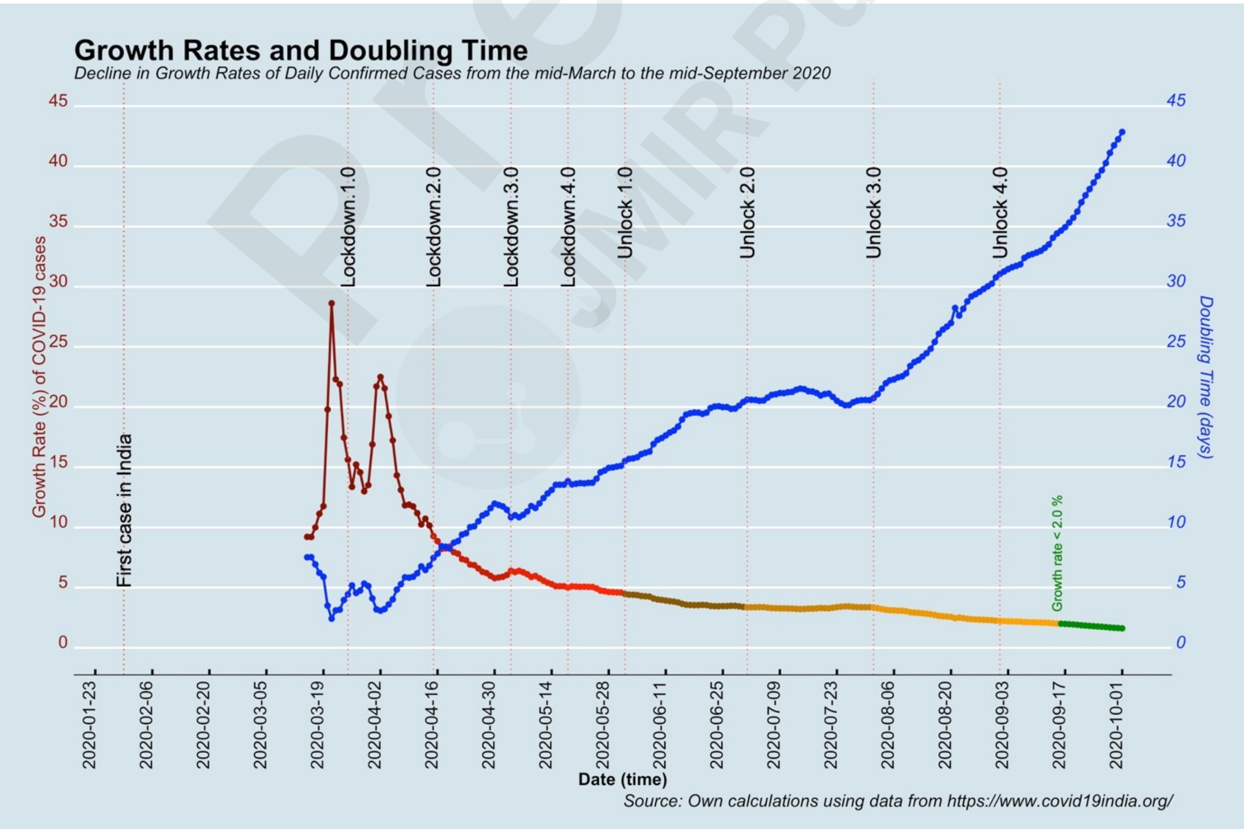
Altogether, the lockdowns were successful in lowering the growth rates; however, the pace of decline was slow. The 5 per cent growth rate at the end of May 2020 was high enough to push the daily confirmed cases to a new height.

The trends in growth rates of daily confirmed cases show a rapid decline in June 2020. By the end of June, the growth rate values declined rapidly to approximately 3 per cent, and the doubling time increased from 10 to 15 days. Nonetheless, 3 per cent is a high growth rate as the daily confirmed cases attest an exponential increase. However, the growth rates did not decline, and it was almost stagnant in the subsequent month of July 2020. Instead, there was a slight increase in the growth rates, and that added the stacks of daily confirmed cases in July. At that nearly-constant growth rate, the number of daily confirmed cases rose from approximately 20,000 to 60,000 in July. At a doubling time of approximately 20 days, the cumulative confirmed cases were approximately 17 lakhs at the end of July. Ten lakhs more confirmed COVID-19 cases were added during July to approximately four lakhs confirmed cases during June. In the backdrop, there is a strong base for the exponential growth rate to manifold increase in daily confirmed cases. With such a big base, a small positive exponential growth rate is sufficient to pile up daily confirmed cases. It always has the potential to delay the peak of daily confirmed cases. By the end of July 2020, India's position regarding COVID-19 pandemic was entirely different what could have been just one month ago.

The growth rates rollbacked on the path of smooth decline in August 2020. The growth rates declined swiftly from 3.3 per cent to 2.2 per cent. Despite this decline in growth rates, the daily confirmed cases kept on burgeoning in this month because of a large base. The doubling time increased from 20 days to 31 days. Accordingly, the cumulative COVID-19 cases increased by twofold and added approximately 19.4 lakhs in August. India shows the doubling rate of 31 days after seven months at a stockpile of daily confirmed cases. On 29 August 2020, the count of daily confirmed cases rose to 78,480, which is the highest in the world as it crossed the peak value of 77,362 in the USA.

India has taken seven months to reach a low growth rate. The delay in reaching a low growth rate has a cost in terms of a high peak and a large cumulative confirmed case. India can circumvent this delay by having a low growth rate of value one per cent sooner rather than later; however, this delay is inevitable. At the growth rate of one per cent, the exponential model becomes a function of time (days) and the parameter estimate 'a', and then, the daily confirmed cases are not growing exponentially. Nonetheless, daily confirmed cases would be multiplicative of time, but it cannot have a snowballing effect. The peak of the daily confirmed cases would analytically appear for a growth rate between one and two per cent.

Figure 2: Trends in growth rates of daily confirmed cases and doubling time of cumulative COVID-19 cases



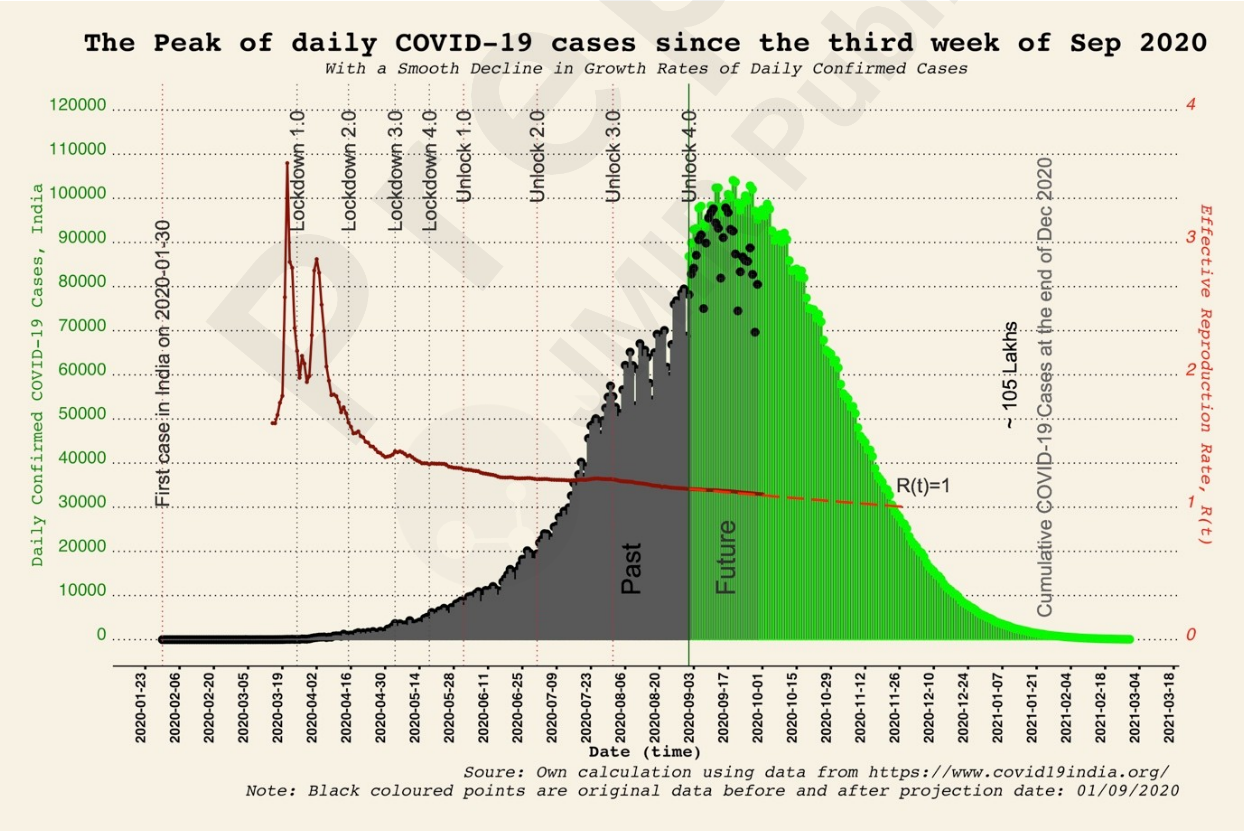
Tracking effective reproductive rates

The effective reproduction rate $R(t)$ is one of the epidemiological measures to understand the spread of COVID-19 disease. Figure 3 shows the trends in $R(t)$ (red-orange-green colored connected circles) from the latter half of March to August 2020. It is shown until 30 September 2020.

The $R(t)$ values in India declined from the highest value of 3.6 persons in the third week of March 2020 to 1.4 persons at the end of April 2020. During these two months of March and April 2020, the trends in $R(t)$ were non-steady. However, there was an increase in the $R(t)$ values in the first week of May 2020. The gradient of $R(t)$ shows an uptick. It was in resemblance to the ascent in growth rates. The growth rates have a significant effect on the $R(t)$ as compared to serial interval distribution, and appropriately, the trends in $R(t)$ is analogous to the trends in growth rates. The trends in $R(t)$ were off the track of smooth decline from achieving low $R(t)$ values in a shorter time as compared with the previous trends. The $R(t)$ value was 1.3 persons at the end of May, showing a marginal decline during these two-and-a-half months.

Nevertheless, the decline in the $R(t)$ has been slow and timid during subsequent months from June to August 2020. It is apparent in the two months of June and July 2020 when the trends in $R(t)$ show almost near-stagnancy at the value of 1.21 persons. The warranted $R(t)$ value of one is just a few points away but, given a length of time, slowly but surely, a bulk of confirmed cases were added in these two months. Nevertheless, in August 2020, the $R(t)$ values declined swiftly. During August, it declined from 1.21 persons in the first week to 1.14 persons at the end of August. More importantly, a downward gradient of $R(t)$ was evident after three months period since May 2020. This downward gradient would provide an expected $R(t)$ values of one in near time.

Figure 3: Projection of daily confirmed cases, India



Projection, forecast and the peak of confirmed cases

For India, the projection of daily confirmed cases is the need of the hour, especially the forecast of the peak of daily confirmed COVID-19 cases. The projections for the growth rates were performed using the ARIMA model for obtaining the forecasts of daily confirmed cases. An ARIMA modelling procedure has four steps: assessment of model, estimation of parameters, diagnostic checking, and prediction. The time series plot of growth rates, autocorrelation function (ACF), and partial autocorrelation function (PACF) reveals seasonality and stationarity in growth rates of daily confirmed cases. ACF detects the degree of correlations between y_t and y_{t-k} for different values of k, and PACF detects the degree of correlations between y_t and y_{t-k} conditional on $y_{t-k+1},\dots,y_{t-1}$, i.e. partial correlations. Based on the ACF and PACF of the ARIMA model, the second difference was applied for making the time series stationary. The corrected AIC (AICc) value of the model ARIMA (2, 2, 1) was -347.3. Again, based on the seasonal part of ACF and PACF, the SARIMA (2, 2, 1) (1, 0, 1)[7] was applied that had the lowest AICc value of -357.3. The coefficients of this SARIMA model were significant at one per cent level (Table 1). The diagnosis of the developed SARIMA model confirms no trends in the residuals, no outliers, and nearly constant variance. Specifically, the ACF plot of residuals shows no significant autocorrelations. The plot of the residuals shows a standard normal variate. The p-values for the Ljung-Box statistic were above 0.05. In sum, the diagnosis of residuals confirms that the developed SARIMA (2, 2, 1) (1, 0, 1)[7] model is appropriate to the trends and seasonality in growth rates. Root mean square error (RMSE) and mean absolute percentage error (MAPE) of the SARIMA model were at 0.049 and 0.775, respectively, which is very small.

Figure 3 shows the projected daily confirmed cases (green coloured circles) based on SARIMA model and exponential model from 1 September 2020 until 28 February 2021. The peak value of COVID-19 is the most appealing statistics as so obtained from the SARIMA model. The forecasted daily confirmed cases reveal normality around a peak. Given the pace of decline is continued for September 2020, the forecasts suggest the peak of confirmed cases since the third week of September. The forecast suggests that peak value wavers around 104,500 counts of daily confirmed cases in the third week of September 2020. The forecast from the SARIMA model suggests the peak value of 104,081 on 19 September 2020. The projection reveals that the decline in the growth rates is not steep. Therefore, a significant number of confirmed cases is yet to be added. As per the forecast, the cumulative number of COVID-19 cases accounts for a total of 105 lakhs at the end of December 2020. Sooner rather than later India observes a peak, lesser the cumulative number of COVID-19 cases is witnessed. However, the timing and magnitude of the peak and total cumulative cases is contingent upon the swift decline of the growth rates.

The projected daily confirmed cases were nearly coinciding with the original data points between 1 September and 19 September 2020. It ensures an independent cross-check compared to other sensitivity analysis. The correlation between daily confirmed cases and real-time daily confirmed cases was 67 percent for the period of 19 days between 1 and 19 September 2020. However, a drop in real-time daily confirmed cases after the peak, between 20 and 31 September 2020, is achieved sharp as compared to that of in projected cases. This indicates a plausibility of smaller cumulative COVID-19 cases compared to that of the projected estimate.

Figure 3 shows the $R(t)$ values until it is equal to one (red coloured long-dash line). Based on the projected growth rates, $R(t)$ values were forecasted. The $R(t)$ values equal to one at a growth rate of value zero. However, the trends in growth rates suggest three months from September to reach a value of zero at the end of November 2020. As per the forecast, it would take more time to go further low as compared to the decline of growth rates during the previous three months, i.e. between June and August 2020. It is plausible because of the large base of confirmed cases after the peak shows up. The forecast suggests a significant lag between the peak value and the $R(t)$ value of one.

For the Indian population, learning and practising social measures in the preceding seven months have ensured deceleration in the slope of the exponential model or a decline in the growth rates of daily confirmed cases. Nevertheless, it is essential to recall the case of stagnant growth rates and $R(t)$ values in June and July 2020. India has shown a decline in the growth rates in August 2020, and down the line, it maintains the pace of decline in September 2020. Since a peak value, a modest decline in exponential growth rates is favourable, maintaining lower growth rates than 2.0. Therefore, projections suggest that the highest value of 104,081 on 19 September 2020 stands for the peak value of daily confirmed cases in the third week of September 2020.

Table 1: Parameters of SARIMA model

SARIMA	Coefficients	Std. Error
AR1	0.182	0.107
AR2	0.160	0.119
MA1	-0.956***	0.051
SAR1	0.940***	0.052
SMA1	-0.778***	0.116

Source: Own calculations using data from https://api.covid19india.org/.

Note: SARIMA model: (2, 2, 1)(1, 0, 1)[7], *** significant at 1% level; AICc=-357.3, BIC=-341.4.

Discussion

Summary of Findings

The study explores the epidemiological stage of COVID-19 disease and the curvature of the daily confirmed COVID-19 cases in India. It primarily examines the trends in the growth rates of daily confirmed cases in India, by the application of the exponential model. The $R(t)$ values were calculated using the Lotka-Euler method, and the projections were made by developing SARIMA model.

The results show that the exponential model is the best fit over the daily confirmed COVID-19 cases during the period of seven months from dates 30 January 2020 to 31 August 2020 in India. The applied exponential model reveals that initially, the growth rates of daily confirmed cases were high and non-steady positive values. However, the trends in growth rates show a linear decline during the five-and-a-half months period from the latter half of March to August 2020. Doubling time shows an increase in these five-and-a-half months period, concomitant with a steady decline in the growth rates.

$R(t)$ values rapidly declined from the highest value of 3.6 persons in the mid-March 2020 to 1.38 until April 2020 and then declined modestly to 1.14 persons until August 2020. A slowdown in $R(t)$ since May 2020 is concomitant of a modest decline in the growth rates. The trends in $R(t)$ show analogous to the trends in the growth rates of daily confirmed cases.

Based on the developed SARIMA model, the projection shows a slow, cyclic decline in the growth rates of daily confirmed COVID-19 cases since September 2020. As per the projection, the growth rates would be less than two from the mid-September and one in the third week of October and zero at the end of November 2020. The forecast suggests the peak of the daily confirmed cases since the third week of September 2020. Our estimates showed the peak value of 104,081 on 19 September 2020. The real-time peak value was 97,860 on 16 September 2020. The forecast is very close to the real-time peak value of daily confirmed cases. Projections proved significant, worthwhile for India. Most importantly, we predicted the timing of the peak and its value very close to real-time data for the first time.

As per the projections, the $R(t)$ value declines to 1.12 persons in the mid-September 2020 and 1.10 at the end of September 2020. The correlation between projected and real-time $R(t)$ values during 30 days from 1 to 30 September 2020 is 98.9 per cent. The forecast suggests the $R(t)$ value equals to one at the end of November 2020.

The total cumulative number of COVID-19 cases would be approximately 105 lakhs by the end of December 2020.

Limitations

The study is based on data of daily confirmed cases of COVID-19 disease during seven months between the dates 30 January 2020 and 31 August 2020. The data between dates 1 and 30 September 2020 were used for validating the projections in the period studied. The API portal https://www.covid19india.org/ provides data on confirmed, active, deaths, and recovery cases; nevertheless, the data on daily confirmed cases is the most appropriate for projections. The exponential model applies to daily confirmed cases in India and its states. However, the projection analysis was performed at the national level only because smaller states do have limitations for projections because of small samples and differences in the length of period for COVID-19 disease.

The study is based on a few assumptions. Apart from statistical and time series assumptions, the behavioural aspects of population and socio-economic and demographics characteristics of the population were not available for consideration in trend analysis and projections because of data constraints. The external factors such as capital and health expenditure were not possible to link with the time series. In the view of these data constraints, we do have trend and seasonality into the account for robust analysis and long-term projection.

Comparison with Prior Work

This study, for the first time, shows a close forecast of the peak value of the daily confirmed cases as well as its timing.

Numerous studies have also applied other methods ranging from non-linear methods to Susceptible-Infectious-Recovery (SIR) and its variant models in India. Despite the methods applied, most of the studies use data-driven modelling [18]. A review of mathematical and predictive models from 30 studies analysing data of India points out a large variation in predictions of COVID-19 cases [19]. This review of pieces of literature highlights regression-based models performs better than SIR, and its variant for the Indian data worked until April 2020. Another research points the variations in the evolution of the epidemic measured by its peaks and troughs is majorly because of large variations in statistical parameters sensitive to the length of time and non-pharmaceutical interventions across the countries [20]. It is acceptable as non-pharmaceutical interventions, especially the testing capacity, in each country varies.

SIR variant models comprising many components suggest a lower peak value and smaller size of COVID-19 disease when applied for countries like India, Brazil, United Kingdom, Bangladesh, and Russia. The underestimation of cases is because of the high sensitivity of the recovery rate of undetected asymptomatic carriers, rate of home quarantined or self-quarantined and transmission rate from quarantined class to susceptible class. Particularly for India, the inverse of the incubation period and self-quarantined are very sensitive parameters affecting the estimates of COVID-19 size [21]. The precautionary measures vary by the states of India, affecting the parameter estimates, and thus it adds more complexity for projections at the state level [22]. The variations in the model estimates are heavily dependent upon the assumptions being met [23]. In general, the SIR and its variant models provide better insights for planning the interventions and prepare the health systems for better clinical management [24, 25].

The peak and cumulative COVID-19 cases have been much sensitive to the effect of social distancing in India. With the same level of effective reproduction rate, the projections show large differences in cumulative cases for a minor change in the parameters representing social distancing in the model [26]. With low testing and low identification rate, the peak value is delayed in comparison to high testing and high identification rate [27, 28]. The SIR and logistic models have shown better accuracy for short-term projections. The nested exponential model reveals a sigmoid nature of the cumulative COVID-19 cases and shows flattening of the curve [29].

The projections and forecast are a crucial part of the decision-making process, and it is most required for the high-risk cases. The amount of uncertainty in the size of COVID-19 cases are risky but discarding it and acting conservatively can jeopardise the situation in a country. Most of the studies have the same assumptions which are the accuracy of real-time data, a similar pattern of precautionary and medical measures, and the number of testing does not change dramatically [30]. Other complications are the imposition of a single peak by any model. For multiple peaks, the suggested model refers to higher degree polynomial equations and yet to witness a fitting to real-time COVID-19 data.

Conclusions

The study shows that the SARIMA model is suitable for projecting daily confirmed cases. This study based on moments of the distribution of the daily confirmed cases of COVID-19 disease unravels the uncertainty about the peak and curvature of COVID-19 disease. The estimates based on our own calculations show the peak value on 19 September 2020 whereas the real-time peak value occurred on 16 September 2020. Thus, the prediction made is almost precise with the real-time peak value of daily confirmed cases for India. Virtually, for the first time, the timing of the peak and its value abutting the real-time data has been projected. The study successfully explores the epidemiological stage of COVID-19 disease in India.

The study strongly suggests keeping track of the growth rates of daily confirmed cases obtained from the exponential model to understand the flattening and size of COVID-19 cases in India.

Acknowledgements

Authors would like to thank Prof. K. S. James, the Director and Senior Professor, International Institute for Population Sciences (IIPS), Mumbai, for his recommendations and suggestions to carry out this study, initiating policy briefs on COVID–19, and for providing constructive comments which helped immensely in improving the quality of this work. We are also thankful to the members of the group “IIPS Team of Researchers on Estimation and Projection of COVID 19 Cases” for their suggestions.

Author Contributions

Conflicts of Interest

The authors declare that they have no competing interests.

References

1. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*. 2020;382(8):727-733. [doi:https://doi.org/10.1056/NEJMoa2001017]

2. World Health Organization. Naming the coronavirus disease (COVID-19) and the virus that causes it. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

3. Atangana A. Modelling the spread of COVID-19 with new fractal-fractional operators: Can the lockdown save mankind before vaccination? *Chaos, Solitons & Fractals*. 2020;136(109860) [doi:https://doi.org/10.1016/j.chaos.2020.109860]

4. COVID19-India API. Data from: A volunteer-driven, crowd-sourced database for COVID-19 stats & patient tracing in India (Online). 2020.

5. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf Dis*. 2020;20(5):533-534. [doi:https://doi.org/10.1016/S1473-3099(20)30120-1]

6. Vokó Z, Pitter JG. The effect of social distance measures on COVID-19 epidemics in Europe: an interrupted time series analysis. *GeroScience*. 2020;42(4):1075-1082. [doi:https://doi.org/10.1007/s11357-020-00205-0]

7. Chintalapudi N, Battineni G, Amenta F. COVID-19 disease outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach. *Journal of Microbiology, Immunology and Infection*. 2020;53(3):396-403. [doi:https://doi.org/10.1016/j.jmii.2020.04.004]

8. Yadav S, Yadav PK, Kishore M, et al. Basic Reproduction Rate and Case Fatality Rate of COVID-19: Application of Meta-analysis. *Demography India*. 2020;49(Special Issue):76-97.

9. Kupferschmidt K. The lockdowns worked—but what comes next? *Science* 2020;368(6488):218-219. [doi:https://doi.org/10.1126/science.368.6488.218]

10. National Guidelines for Ethics Committees Reviewing Biomedical & Health Research During Covid-19 Pandemic (Ministry of Health and Family Welfare) (2020). URL: https://ethics.ncdirindia.org//asset/pdf/EC_Guidance_COVID19.pdf

11. An Illustrative Guide on COVID Appropriate Behaviours (Ministry of Health and Family Welfare) (2020). URL: https://www.mohfw.gov.in/pdf/Illustrativeguidelineupdate.pdf

12. Galvani AP, Lei X, Jewell NP. Severe acute respiratory syndrome: temporal stability and geographic variation in death rates and doubling times. *Emerging Infectious Diseases*. 2003;9(8):991–994. [doi:https://doi.org/10.3201/eid0908.030334]

13. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*. 2007;274(1609):599-604. [doi:https://doi.org/10.1098/rspb.2006.3754]

14. Ma J. Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling*. 2020;5:129-141. [doi:https://doi.org/10.1016/j.idm.2019.12.009]

15. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerging infectious diseases*. 2020;26(6):1341-1343. [doi:https://doi.org/10.3201/eid2606.200357]

16. Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Science of the Total Environment*. 2020;729:138817. [doi:https://doi.org/10.1016/j.scitotenv.2020.138817]

17. Singh RK, Rani M, Bhagavathula AS, et al. Prediction of the COVID-19 Pandemic for the Top 15 Affected Countries: Advanced Autoregressive Integrated Moving Average (ARIMA) Model. *JMIR Public Health and Surveillance*. 2020;6(2):e19115. [doi:https://doi.org/publichealth.jmir.org/2020/2/e19115]

18. Singhal A, Singh P, Lall B, Joshi SD. Modeling and prediction of COVID-19 pandemic using Gaussian mixture model. *Chaos, solitons, and fractals*. 2020;138(110023) [doi:https://doi.org/10.1016/j.chaos.2020.110023]

19. Kotwal A, Yadav AK, Yadav J, Kotwal J, Khune S. Predictive models of COVID-19 in India: A rapid review. *Medical Journal Armed Forces India*. 2020; [doi:https://doi.org/10.1016/j.mjafi.2020.06.001]

20. Li S, Linton O. When will the Covid-19 pandemic peak? *Journal of Econometrics*. 2020; [doi:https://doi.org/10.1016/j.jeconom.2020.07.049]

21. Nabi KN. Forecasting COVID-19 pandemic: A data-driven analysis. *Chaos, Solitons and Fractals*. 2020;139(110046):1-39. [doi:https://doi.org/10.1016/j.chaos.2020.110046]

22. Sarkar K, Khajanchi S, Nieto JJ. Modeling and forecasting the COVID-19 pandemic in India. *Chaos, Solitons and Fractals*. 2020;139(110049) [doi:https://doi.org/10.1016/j.chaos.2020.110049]

23. Behl R, Mishra M. COVID-19 Lifecycle: Predictive Modelling of States in India. *Global Business Review*. 2020;21(4):883-891. [doi:https://doi.org/10.1177/0972150920934642]

24. Malavika B, Marimuthu S, Joy M, Nadaraj A, Asirvatham ES, Jeyaseelan L. Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models. *Clinical Epidemiology and Global Health*. 2020;

[doi:https://doi.org/10.1016/j.cegh.2020.06.006]

25. Sahoo BK, Sapra BK. A data driven epidemic model to analyse the lockdown effect and predict the course of COVID-19 progress in India. *Chaos, Solitons and Fractals*. 2020;139(110034) [doi:https://doi.org/10.1016/j.chaos.2020.110034]

26. Patrikar S, Poojary D, Basannar DR, Faujdar DS, Kunte R. Projections for novel coronavirus (COVID-19) and evaluation of epidemic response strategies for India. *Medical Journal Armed Forces India*. 2020;76(3):268-275. [doi:https://doi.org/10.1016/j.mjafi.2020.05.001]

27. Samui P, Mondal J, Khajanchi S. A mathematical model for COVID-19 transmission dynamics with a case study of India. *Chaos, Solitons and Fractals*. 2020;140(110173)[doi:https://doi.org/10.1016/j.chaos.2020.110173]

28. Shah PV. Prediction of the Peak, Effect of Intervention and Total Infected by the Coronavirus Disease in India. *Disaster Medicine and Public Health Preparedness*. 2020;24:1-23. [doi:https://doi.org/10.1017/dmp.2020.321]

29. Kumar A, Rani P, Kumar R, Sharma V, Purohit SR. Data-driven modelling and prediction of COVID-19 infection in India and correlation analysis of the virus transmission with socio-economic factors. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*. 2020;14(5): 1231-1240. [doi:https://doi.org/10.1016/j.dsx.2020.07.008]

30. Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. *PLoS ONE*. 2020;15(3):e0231236. [doi:https://doi.org/10.1371/journal.pone.0231236]

Supplementary Files