# Chapter 9

# Modeling and forecasting the spread of COVID-19 pandemic in India and significance of lockdown: A mathematical outlook

**Brijesh P. Singh**[*]

*Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India*
[*]*Corresponding author: e-mail: brijesh@bhu.ac.in*

## Abstract

A very special type of pneumonic disease that generated the COVID-19 was first identified in Wuhan, China in December 2019 and is spreading all over the world. The ongoing outbreak presents a challenge for data scientists to model COVID-19, when the epidemiological characteristics of the COVID-19 are yet to be fully explained. The uncertainty around the COVID-19 with no vaccine and effective medicine available till today create additional pressure on the epidemiologists and policy makers. In such a crucial situation, it is very important to predict infected cases to support prevention of the disease and aid in the preparation of healthcare service. India is fighting efficiently against COVID-19 and facing greater challenges because of its large population and high population density. Though the government of India is taking all needful steps to prevent its spread but it is not enough to control and stop spread of the disease so far, perhaps due to defiant nature of people living in India. Effective measure to control this disease, medical professionals needs to know the estimated size of this pandemic and pace. In this study, an attempt has been made to understand the spreading capability of COVID-19 in India through some simple models. Findings suggest that the lockdown strategies implemented in India are not successfully reducing the pace of the pandemic significantly after first lockdown.

**Keywords:** COVID-19, Differential equations, Logistic growth, Propagation, Joinpoint regression, Lockdown

## 1 Background

A novel corona virus is responsible for epidemic popularly known as COVID-19 is a new strain that has not been identified previously in humans. World Health Organization (WHO) declared COVID-19 a pandemic on March 11, 2020. The virus that caused the incidence of Severe Acute Respiratory Syndrome (SARS) in 2002 in China, Middle East respiratory syndrome (MERS) in 2012 in Saudi Arabia and the virus that causes COVID-19 are genetically related to each other, but the diseases they caused are quite different (WHO). These viruses, in general, are a family of viruses that target and affect mammal's respiratory systems. The SARS corona virus spread to humans via civet cats, while the MERS virus spread via dromedaries. In case of the novel corona virus, typically happens via contact with an infected animal, perhaps the common carriers are bats initial reports from seafood market in central Wuhan, China.

The Novel Corona Virus (COVID-19) started from Wuhan, China and thus, initially known as the Wuhan virus, expanded its circle in South Korea, Japan, Italy, Iran, USA, France, Spain and finally spreading in India. It is named as novel because it is never seen before mutation of animal corona virus but certain source of this pandemic is still unidentified. It is said that the virus might be connected with a wet market (with seafood and live animals) from Wuhan that was not complying with health and safety rules and regulations.

As of July 16, 2020, with the continuously increasing global risk more than 14 million confirm positive cases and more than 0.58 million of deaths have occurred in the world. As number of cases growing day by day, in most of the countries of the world, some most populous countries like China, India, Brazil, USA, etc., are badly affected by it. In this context, the crucial role of modeling, transmission dynamics and estimating development of COVID-19 are expected. The population based mathematical model especially growth model in this scenario are the most preferable techniques to understand the epidemic future trajectory. Epidemiological characteristics like propagating dynamics, severity, susceptibility, and the effects of control measures, for COVID-19 has produced a greater concern for researchers (Cowling and Leung, 2020; Lipsitch et al., 2020).

Since preventive measures like lockdown and social distancing have immense pressure on economy of the country, quantitative estimates and predictions are necessary to learn the impact of spread that will help in plan the strategies against COVID-19. Given the paucity of such quantitative measures, the predictions on the basis of different idea given in this paper become critical and to know when the COVID-19 stops. In recent past a number of studies with various technique and tools have been carry out to understand the dynamics of propagation of disease and future course of action. For COVID-19, various models which are capable of providing worth insights for health care policy making are being continuously developed and used to explain this pandemic retrospectively as well as to project the events

(Batista, 2020; Koo et al., 2020; Kucharski et al., 2020; Tuite and Fisman, 2020; Wu et al., 2020).

Wu et al. (2020) has been done to analyzing the pace of virus transmissibility through estimating the value of $R_0$ with the help of stochastic Markov Chain Monte Carlo method. Another analysis with mathematical incidence decay and exponential adjustment is performed. Further to explain growth behavior of COVID-19 a statistical exponential growth model adopting the serial interval from Severe Acute Respiratory Syndrome is applied by Zhao et al. (2020). A three-parameter logistic growth function is applied and predicted for China as well as some other countries is found very satisfying (Shen, 2020). In the context of India, an early study of COVID-19 (when it started spreading in India) done by Singh and Adhikari (2020) rightly believed that countrywide lockdown on March 24 for 21 days may be insufficient for controlling the COVID-19 pandemic. Malhotra and Kashyap (2020) tried to forecast the endpoints to explain the progression of COVID-19 in Indian States, using SIR and logistic growth models and found the endpoint of COVID-19 in India is in July 23, 2020.

India with a huge population about 1.3 billion, among majority of the people are living in poor hygienic condition and the medical facilities like number of doctors and hospitals are less in India as compared to developed countries indicates that the situation of India will become very critical but comparatively better public health system and political control in India than the above developed countries. The picture of India is not so good and has more than 1 million confirm positive cases and more than 26 thousand of deaths. Although the death rate of this pandemic is low in comparison of other pandemics and diseases but its high rate of spread and no proper cure available so far is the major concern in the present time. Right now in India only 29 districts out of 739 districts have COVID-19 case more than 4000. These districts are mainly metropolitans; if we implement preventive measures properly then spread can be under control at desired level, but due to defiant nature of people living in India, political desire and rivalry, still we India society are facing problem made by COVID-19.

The first case of COVID-19 is reported in India on January 30, 2020 when a student returned from Wuhan, China (covid19india.org). The Government of India was quick to launch various levels of travel advisories beginning from February 26, 2020, with restrictions on travel to China and nonessential travel restrictions to Singapore, South Korea, Iran and Italy. The efforts to control by the Hon'ble Prime Minister Narendra Modi Ji through Janata Curfew (public curfew) on March 22, 2020, can be seen as the beginning of wide-scale public preventive measures. India has launched several social distancing measures and personal hygiene measures during the second week of March.

Symptoms of COVID-19 are reported as cough, acute onset of fever and difficulty in breathing. Out of all the cases that have been confirmed,

up to 20% have been deemed to be severe. Cases vary from mild forms to severe ones that can lead to serious medical conditions or even death. It is believed that symptoms may appear in 2–14 days, as the incubation period for the COVID-19 has not yet been confirmed. However, in India 14 days minimum quarantine period is declared by Government for suspected cases. Since it is a new type of virus, there is a lot of research being carried out across the world to understand the nature of the virus, origins of its spreads to humans, the structure of it, possible cure/vaccine to treat COVID-19. India also became a part of these research efforts after the first two confirmed cases were reported here on January 31, 2020. Then in India screening of traveler at airport migrant was started, immediate Chinese visas was canceled, and who was found affected from COVID-19 kept in quarantine centers (Ministry of Home Affaires Government of India, Advisory).

For the spread of COVID-19, when disease dynamics are still unclear, mathematical modeling helps us to estimate the cumulative number of positive cases in the present scenarios. Now India is interring in the mid stages of the epidemic. It is important to predict how the virus is likely to grow among the population. The COVID-19 pandemic presents a challenge for data scientists to model it; however, the epidemiological characteristics of the COVID-19 are yet to be fully explained. The uncertainty around the COVID-19 with no vaccine and effective medicine available until today create additional pressure on the epidemiologists and policy makers. In such a crucial situation, it is very important to predict infected cases to support prevention of the disease and support in the preparation of healthcare service. A mathematical modeling approach is a suitable tool to understand the dynamics of epidemic. In the study some mathematical approach to understand the dynamics of novel COVID-19 in India has been discuss.

In absence of a definite treatment modality like vaccine, physical distancing has been accepted globally as the most efficient strategy for reducing the severity of disease and gaining control over it (Ferguson et al., 2020). Also in India it is reported that the country is well short of the WHO's recommendations of minimum threshold of 2.28 skilled health professionals per 1000 population (Anand and Fan, 2016). Therefore, on March 24, 2020, the Government of India under Prime Minister Narendra Modi Ji ordered a nationwide lockdown for 21 days, limiting movement of the entire 1.3 billion population of India as a preventive measure against the COVID-19 pandemic in India. It was ordered after a 14-h voluntary public curfew on 22 March. The lockdown was placed when the number of confirmed COVID-19 cases in India was approximately 500. On 14 April, Prime Minister of India extended the nationwide lockdown until 3 May, with a conditional relaxation after 20 April for some regions. On 4 May, the Government of India again extended the nationwide lockdown further by 2 weeks until 17 May. Also, the Government has divided the entire nation into three zones viz. green, red and orange with relaxations applied accordingly.

There are already various measures such as social distancing, lockdown masking and washing hand regularly has been implemented to prevent the spread of COVID-19, but in absence of particular medicine and vaccine it is very important to predict how the infection is likely to develop among the population that support prevention of the disease and aid in the preparation of healthcare service. This will also be helpful in estimating the health care requirements and sanction a measured allocation of resources. It is well known fact that COVID-19 has spread differently in different countries, any planning for increasing a fresh response has to be adaptable and situation-specific. Data obtained on COVID-19 outbreak have been studied by various researchers using different mathematical models (Chang et al., 2020; Srinivasa Rao Arni et al., 2020). Many other studies (Anastassopoulou et al., 2020; Corman et al., 2020; Gamero et al., 2020; Huang et al., 2020; Hui et al., 2020; Rothe et al., 2020) on this recent epidemic have been reported so many meaningful modeling results based on the different principles of mathematics.

Most of pandemics follow an exponential curve during the initial spread and eventually flatten out (Junling et al., 2014). SIR model is one of the best suited models for projecting the spread of infectious diseases like COVID-19 where a person once recovered is not likely to become susceptible to the infection again (Kermack and McKendrick, 1927). Susceptible-Infectious-Recovered (SIR) compartment model (Herbert, 2000) is used to include considerations for susceptible, infectious, and recovered or deceased individuals. These models have shown a significant predictive ability for the growth of COVID-19 in India on a day to day basis so far. A time dependent SIR models have been defined to observe the undetectable infected persons with COVID-19 (Chen et al., 2020). A recent study by Mandal et al. (2020) has shown that social distancing can reduce cases by up to 62%.

Further, time series models have been employed for predicting the incidence of COVID-19 disease. As compared to other prediction models, for instance support vector machine (SVM) and wavelet neural network (WNN), ARIMA model is more capable in the prediction of natural adversities (Zhang et al., 2019). Chatterjee et al. (2020) studied a stochastic mathematical model of the COVID-19 epidemic in India. The logistic growth regression model is used for the estimation of the final size and its peak time of the COVID-19 pandemic in many countries of the World and found similar result obtained by SIR model (Batista, 2020).

It is well known that the effects of social distancing become visible only after a few days from the lockdown. This is because the symptoms of the COVID-19 normally take some time to come out after getting infected from the COVID-19. An estimates indicates that, with hard lockdown and continued social distancing, the peak total infections in India will be 97 million and the number of infective by September is likely to be over 1100 million (Schueller et al., 2020).

## 2 Why mathematical modeling?

The study of infectious diseases is called epidemiology. A disease is called endemic if it persists in a population and pandemic when it occurs worldwide. The spread of an infectious disease involves not only disease related factors such as the infectious agent, mode of transmission, latent period, infectious period, susceptibility and resistance, but also social, cultural, demographic, economic and geographic factors. Mainly there are three types of models for infectious diseases that are spreading directly through person to person contact in a population. Some simple models are formulated and analyzed mathematically considering differential equations. Parameters are estimated for infectious diseases and also used to compare the vaccination levels necessary for herd immunity. The three models considered here are the simple epidemiological models and suitable for diseases which are transmitted directly from person to person. More complicated models must be used when there is transmission by insects called vectors or a reservoir of nonhuman infective. Epidemiological models are widely used to understand the pattern and policy development.

Even though vaccines are available for many infectious diseases, these diseases still cause suffering and mortality in the world, especially in developing countries. In developed countries chronic diseases such as cancer and heart disease have received more attention than infectious diseases, but infectious diseases are still a more common cause of death in the world. The transmission mechanism from an infective to susceptible is understood or nearly all infectious diseases and the spread of diseases through a chain of infections is known. However, the transmission interactions in a population are very complex so that it is difficult to comprehend the large scale dynamics of disease spread without the formal structure of a mathematical model. An epidemiological model uses a microscopic description (the role of an infectious individual) to predict the macroscopic behavior of disease spread through a population. In many sciences it is possible to conduct experiments to obtain information and test hypotheses. Experiments with infectious disease spread in human populations are often impossible, unethical or expensive. Data is sometimes available from naturally occurring epidemics or from the natural incidence of endemic; however, the data is often incomplete due to underreporting. This lack of reliable data makes accurate parameter estimation difficult so that it may only be possible to estimate a range of values for some parameters. Since repeatable experiments and accurate data are usually not available in epidemiology, mathematical models and computer simulations can be used to perform needed theoretical experiments.

Mathematical models have both limitations and capabilities that must recognized. Sometimes questions cannot be answered by using epidemiological models, but sometimes the modeler is able to find the right combination of available data, an interesting question and a mathematical model which can lead to the answer. Comparisons can lead to a better understanding of the

processes of disease spread. Modeling can often be used to compare different diseases in the same population, the same disease in different populations, or the same disease at different times. Comparisons of diseases such as measles, rubella, mumps, chickenpox, whooping cough, poliomyelitis and others are made (Hethcote, 1983; Yorke and London, 1973; Yorke et al., 1979) and in the article on rubella in this volume by Hethcote (1989). Quantitative predictions of epidemiological models are always subject to some uncertainty since the models are idealized and the parameter values can only be estimated. However, predictions of the relative merits of several control methods are often robust in the sense that the same conclusions hold over a broad range of parameter values and a variety of models. Optimal strategies for vaccination can be found theoretically by using modeling. Longini et al. (1978) use an epidemic model to decide which age groups should be vaccinated first to minimize cost or deaths in an influenza epidemic. Hethcote (1988) uses a modeling approach to estimate the optimal age of vaccination for measles.

Within a short period of time, COVID-19 has traumatized the world with a greater magnitude and coercion than older pandemics. Its eventuality is grabbed by the fact that it has infected millions and killed thousands across the globe. Global markets, accessible transportation, large scale production have largely contributed to make this pandemic spread faster. This has drastically affected the social life and health mental as well as physical of human beings worldwide. The already burdened health infrastructure across the globe is virtually exposed up to an irreparable point. The WHO declared 2019–2020 corona virus outbreak a Public Health Emergency of International Concern (PHEIC) on January 30th, 2020 and a pandemic 12 days later on February 12th, 2020. With its outbreak in Wuhan, China, the pandemic seems to occupy and include all the vitals of the world thereby affecting the mechanistic processes of any nation. The countries are trying hard to combat and contain this outbreak by following suitable set of protocols that tend to alter the transmission rate effectively. In the initial phase of spread of COVID-19; Italy, Spain, France and some other European countries are one of the worst sufferers of the pandemic and the coercive measures have resulted in the disruption of all the necessary services. On the other hand, the case is virtually less severe in South Asia. India is less affected by the COVID-19, however, China is its neighboring country having border through buffer states like Nepal and Bhutan.

Being the second most populous country of the world, India is fighting hard to minimize the damage of COVID-19. As on 15th April, the total number of infected cases in India was 12,370 with 422 deaths and most recoveries (covid19india.org). India reported its first case on 30th January and entered the countrywide lockdown on March 24th, 2020 with constantly increase in number of COVID-19 cases. Indian government as well as states government has issued early guidelines and travel advisories to limit the further damage of disease. Also, the timely precautions taken by the government have contributed greatly toward combating this pandemic. The paper attempts to devise a
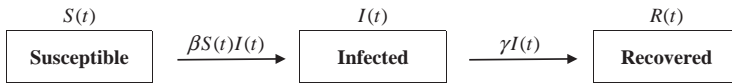
model that would conveniently help in assessing the predictability of pandemic COVID-19 in future time period. This can be achieved by evaluating the different parameters that directly or indirectly affect the ongoing rate of pandemic. Moreover, theoretical explanation, quantitative analysis and other parameters are highly required to predict the peak and size of any pandemic.

We obtained information on cumulative number of COVID-19 confirmed cases in India from covid19india.org. All cases are laboratory confirmed following the case definition by the Government of India. Some studies modeled the epidemic curve obeying the exponential growth (De Silva et al., 2009). The nonlinear least square framework is adopted for data fitting and parameter estimation for COVID-19 at this early stage. First exponential and then logistic growth curve is used to model the COVID-19 pandemic, since epidemics grow exponentially not linearly. But it is surprising that exponential growth curve always provide increasing number of daily new cases. There is no saturation point. Another deterministic model used for understanding the dynamics of epidemic is the Susceptible-Infectious-Recovered (SIR) model, which has been used to accurately predict incidence like SARS. In the SIR model, we need to know the input parameters first the stats we feed into the model (Chatterjee et al., 2020; Mandal et al., 2020; Singh and Adhikari, 2020). The first one is $R_0$ called the basic reproduction number. It is essentially the number of new cases a single infected person will cause during their infectious period. It is one of the most important parameters for assessing any epidemic. Corona virus has an $R_0 \sim 2.4$. In contrast, the swine flu virus had an $R_0 \sim 1.5$ in the 2009 swine flu epidemic (Gupta, 2020). The $R_0$ will inform us about how many people will get infected with one infected person. Other one is the case fatality rate (CFR), which is the percentage of infected people that will die due to the infection. The CFR for corona virus has been reported between 0.5% and 4%. The lower values are more appropriate in resource better settings of medical facility. But SIR model assumes that every person is moving and has equal chance of contact with each and every other person among the population irrespective of the space or distance between different people. It is assumed that the transmission rate remains constant throughout the period of pandemic. Also this model considered to have the same transmission rate for who have been diagnosed and are in quarantine or those who have not been quarantined. The harmonic analysis methods and dynamic model (Rao Srinivasa Arni et al., 2020) estimates show that the number of COVID-19 infected would be 9225 (if there were 10 infected individuals as of March 1, 2020, who was not taking any precautions to spread), 17,986 (if there were 20) and 44,265 (if there were 50).

## 2.1 SIR and SEIR model

SIR model is a theoretical epidemiological model, in which, the population is categories into three component such as: susceptible (S), which is the group of

people who are vulnerable to exposure with infectious people, infected ($I$), are those with the disease and can transmit it to the susceptible and the third component is the individuals who have recovered from the infectious disease and developed immunity and not susceptible to the same illness anymore ($R$). This framework enables us to understand the dynamics of any epidemic. Thus SIR model is a compartmental model in which individuals are separated into different compartments based on their status and follow the corresponding population sizes over the time. The diagrammatical representation of three-compartment model (Kermack and McKendrick, 1927) is given as

$$
S(t) \qquad\qquad I(t) \qquad\qquad R(t)
$$

| Susceptible | $\xrightarrow{\beta S(t)I(t)}$ | Infected | $\xrightarrow{\gamma I(t)}$ | Recovered |
|---|---|---|---|---|

where, $S(t)$ = Proportion of individual susceptible to COVID-19 at time $t$, $I(t)$ = Proportion of individual who have been infected by COVID-19 and are capable of infecting others at time $t$, and $R(t)$ = Proportion of individual who have been infected by COVID-19 and recovered at time $t$, such that $S(t) + I(t) + R(t) = 1$. Here $\beta$ is the transmission parameter controlling how much the disease can be transmitted. This is the average number of individuals that one infected individual will infect per unit time. It is determined by the chance of contact and the probability of disease transmission. While $\gamma$ is the parameter representing the rate of recovery in a particular period. The model allows us to describe the number or proportions of persons in each compartment by solving the following ordinary differential equations,

$$\frac{dS(t)}{dt} = -\beta S(t)I(t)$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t)$$

$$\frac{dR(t)}{dt} = \gamma I(t)$$

Several assumptions have been discussed with respect to the SIR model (Brauer and Castillo-Chavez, 2012; Daley and Gani, 1999). Based on the SIR model, the basic reproduction number is defined as,

$$R_0 = \frac{\beta}{\gamma}$$

Here, $R_0$ is the average number of new COVID-19 cases produced by a single COVID-19 infected case over the time. In order to fit a SIR model, the parameters were obtained by minimizing the residual sum of squares between the observed active cases and the predicted active cases.

## 3 Formulation of SEIR model

The utilization of the SEIR model lies in the fact that it focuses on the basic processes that are directly related to this growing pandemic. In the preparation of this model, there is a need that the population is to be divided into some subdivisions which are susceptible subdivision $S(t)$, that denotes the population which is susceptible to catch the virus; exposed subdivision $E(t)$, that denotes the population which is infected but the symptoms are not visible yet; infected subdivision $I(t)$, that denotes the population which has been infected by the virus and are showing the symptoms; recovered subdivision $R(t)$, that denotes the population which has immunity to the infection. The basic assumption to formulate this model is that the recovered patients acquired permanent active immunity. It can be justified by the strong reason that none of the patients were re-infected by the COVID-19. There have been numerous cases where patients died after being discharged from the hospital but it was found that the patients were either discharged for having mild symptoms or the testing machine reported wrongly. Now we have normalized these components as $S+E+I+R=1$.

Furthermore, suppose that there are equal birth and death rates, i.e., $\mu$ and $\frac{1}{\alpha}$ is the mean latent period for the disease. $\frac{1}{\gamma}$ is the mean infectious period and recovered individuals are permanently immune. The contact rate $\beta$ may or may not be a function of time. Thus the SEIR model is defined as

$$\frac{dS(t)}{dt} = \mu - \beta S(t)I(t) - \mu S(t)$$

$$\frac{dE(t)}{dt} = \beta S(t)I(t) - (\mu + \alpha)E(t)$$

$$\frac{dI(t)}{dt} = \alpha E(t) - (\mu + \gamma)I(t)$$

$$R_0 = \frac{\alpha\beta}{(\mu + \alpha)(\mu + \gamma)}$$

The variable $R$ is determined from the other variables according to equation $S+E+I+R=1$.

### 3.1 Growth models

A growth curve is an empirical model of the evolution of a quantity over time. Growth curves are widely used in biology for quantities such as population size in population ecology and demography for population growth analysis, individual body height in physiology for growth analysis of individuals. Growth is also a key property of many systems such as an economic expansion, spread of an epidemic, the formation of a crystal, an adolescent's growth and the condensation of a stellar mass.

### 3.1.1 Linear growth

This is the simplest growth model, in which population grows at a constant rate over time. Linear growth is described by the equation

$$P_{t+1} = P_t + A \tag{1}$$

where $P_t$ represents the numbers or size of the system at time $t$, $P_{t+1}$ represents the system's numbers or size of the system one time unit later, and $A$ is the system's (linear) growth rate. Many times this model fails to explain natural phenomenon.

### 3.1.2 Exponential growth (unlimited population growth)

Another simple model describes exponential growth, in which population grows at a constant proportional rate over time. The relation may be expressed in either of two forms, depending on whether reproduction is assumed to be continuous or periodic (Shryock and Siegel, 1973). Exponential growth results in a continuous curve of increase or decrease, whose slope varies in direct relation to the size of the population.

$$P_t = y = P_0 e^{rt} \tag{2}$$

where $r$ is the constant rate of growth, $P_o$ is the initial population size, and the variables $t$ and $P_t$ respectively represent time and the population at time $t$ (Method 1). Another form of exponential curve is as follows

$$P_t = y = P_0 k^t \tag{3}$$

where $k = \left(\frac{P_n}{P_0}\right)^{1/n}$ and that therefore the growth rate in Eq. (3) does not a constant growth rate. David A. Swanson, University of California, USA used this type equation for prediction. We have used truncated information, i.e., only 30 days information (from March 4 to April 2, 2020) on number of COVID-19 cases for the prediction purpose. We have used two equations of exponential curve given below

  **I.** $28e^{0.14t}$ up to March 31, 2020
  **II.** $28e^{0.15t}$ from March 31, 2020 onward (adjusting faster rate of occurrence of COVID-19 cases due to Tablighi spread)

With the current incidence of the COVID-19 going on, we hear about exponential growth. In this study, an attempt has been made to understand and analyze the data through exponential growth curve. The reason for using exponential growth curve for studying the pattern of COVID-19 incidence is that epidemiologists have studied these types of happenings and it is well known that the first period of an epidemic follows exponential growth. The exponential growth function is not necessarily the perfect representation of the epidemic. I have tried to fit exponential curve first, and at the next point

to study the logistic growth curve because exponential curve is only fit the epidemic at the beginning. At some point, recovered people will not spread the virus anymore and when someone is or has been infected, the growth will stop. Logistic growth is characterized by increasing growth in the beginning period, but a decreasing growth after point of inflection. For example, in the corona virus case, the maximum limit would be the total number of exposed people in India because when everybody is infected, the growth will be stopped. After that the increasing rate of curve starts to decline and reach to the minimum.

### 3.1.3 Logistic growth (sigmoidal)

The logistic model reveals that the growth rate of the population is determined by its biotic potential and the size of the population as modified by the natural resistance, or, in other words, by all the various effects of inherent characteristics, that are density dependence Pearl and Reed, 1920. Natural resistance increases as population size gets closer to the carrying capacity. Logistic growth is similar to exponential growth except that it assumes an essential sustainable maximum point. In exponential growth curve, the rate of growth of $y$ per unit of time is directly proportional to $y$ but in practice the rate of growth cannot be in the same proportion always. The logistic curve will continue up to certain level, called the level of saturation, sometimes called the carrying capacity, after reaching carrying capacity it starts declining. A system far below its carrying capacity will at first grow almost exponentially, however, this growth gradually slows as the system expands, finally bringing it to a halt specifically at the carrying capacity (Pearl and Reed, 1920; Shryock and Siegel, 1973). The logistic relationship can be expressed as

$$P_t = y_t = \frac{k}{1 + e^{a+bt}}; \ b < 0 \tag{4}$$

where $a$, $b$ and $k$ are constant and $y_t$ is that value of the time series at the time $t$. The reciprocal of $y_t$ follows modified exponential law. Hence, the given time series observation $y_t$ will follow Logistic Law if their reciprocal $1/y_t$ follows modified exponential law. Thus in general, we may take

$$\frac{dy}{dt} = \alpha y(k - y); \alpha > 0, k > 0$$

The factor $y$ is called the momentum factor which increases with time $t$ and the factor $(k - y)$ is known as the retarding factor which decreases with time. When the process of growth approaches the saturation level $k$, the rate of growth tends to zero. Now we have

$$\frac{dy}{y(k-y)} = \alpha dt \implies \frac{1}{k}\left[\frac{1}{y} + \frac{1}{k-y}\right]dy = \alpha dt \implies \left[\frac{1}{y} - \frac{1}{k-y}\right]dy = \alpha k dt$$

Integrating, we get

$$\log\left(\frac{y}{k-y}\right) = \alpha k t + \gamma, \text{ where } \gamma \text{ is the constant of integration.}$$

$\frac{k}{y} = 1 + e^{-akt}.e^{-\gamma} \Rightarrow y = \frac{k}{1 + e^{-(\gamma + akt)}}$, this equation is same as Eq. (4) where $a = -\gamma$ and $b = -\alpha k$.

Logistic curve has a point of inflection at half of the carrying capacity $k$. This point is the critical point from where the increasing rate of curve starts to decline. The time of point of inflection can be estimate as $\frac{-a}{b}$. For the estimation of parameter of logistic curve, method of three selected point given by Pearl and Reed (1920) has been used. The estimate of the parameters can be obtained with equation given as:

$$k = \frac{y_2^2(y_1 + y_3) - 2y_1y_2y_3}{y_2^2 - y_1y_3} \tag{5}$$

$$b = \frac{1}{t_2 - t_1} \ln \left[ \frac{(k - y_2)y_1}{(k - y_1)y_2} \right] \tag{6}$$

$$a = \ln \left[ \frac{k - y_1}{y_1} \right] - bt_1 \tag{7}$$

where $y_1$ $y_2$ and $y_3$ are the cumulative number of COVID-19 cases at a given time $t_1, t_2$ and $t_3$ respectively provided that $t_2 - t_1 = t_3 - t_2$. You may also estimate the parameter $a$ and $b$ by method of least square after fixing $k$.

To predict confirmed corona cases on different day, logistic growth curve has been also used and found very exciting results. The truncated information (means not from the beginning to the present date) on confirmed cases in India has been taken from March 13 to April 2, 2020. The estimated value of the parameters are as follows $k = 18,708.28$, $a = 5.495$ and $b = -0.174$, with these estimates predicted values has been obtained and found considerably lower values than what we observed. On April 1 and 2, 2020 the number of confirmed corona cases are drastically increasing in some part of India due to some unavoidable circumstances thus there is an earnest need to increase carrying capacity of the model, thus it is increased and considered as 22,000 and the other parameters $a$ and $b$ are estimated again which are $a = 5.657$ and $b = -0.173$. The predicted cumulative number of cases is very close to the observed cumulative number of cases till date. The time of point of inflection is obtained as 32.65, i.e., 35 days after beginning. We have taken data from March 13, 2020 so that the time of point of inflection should be April 14, 2020 and by May 30, 2020 there will be no new cases found in the country. Exponential growth model and model given Swanson provided natural estimate of the total infected cases by June 30, 2020 is all most all people in India. This estimate is obtained when no preventive measure would be taken by the Government of India. The testing rate is lower in India than many western countries in the month of March and April, so our absolute numbers was low, when government initiate faster testing process then we have observed more number of cases and found this logistic model fail to provide cumulative number of corona confirm cases after April 17, 2020 thus there is a need to modify this model (Fig. 1).
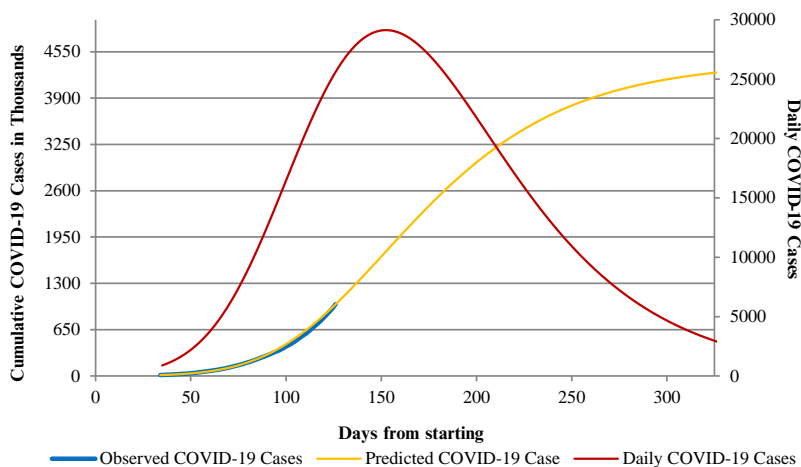
**FIG. 1** Logistic growth curve of COVID-19 cases and daily cases.

In order to the modification, I have taken natural log of cumulative number of corona confirm cases instead of cumulative number of corona confirm cases as taken in the previous model. This model provides the carrying capacity is about 80,000 cases and time of point of inflection is April 30, 2020. The present model provides reasonable estimate of the cumulative number of confirmed cases and by the end of July 2020 there will be no new cases found in the country. Further, the number of COVID-19 cases increases and the model estimate does not match to the observed number of case, therefore we need to change the data period, since the logistic curve is data-driven model that provide new estimate of point of inflection and maximum number of corona positive cases by date when disease will disappear, that helps us to plan our strategies. Finally in this study we changed the data period, i.e., we have taken data from April 15th to July 16th 2020. This provides the carrying capacity is about 45 lakh cases and time of point of inflection is August 15th, 2020 with a maximum number of new cases on a day is about 30,000 per day. The model based on this data (from April 15th to July 16th 2020) provides reasonable estimate of the cumulative number of confirmed cases, and predicted value along with 95% confidence interval provided up to August 15th, 2020 (see Table 1) and by the end of March 2021 we expect there will be no new cases in the country in absence of any effective medicine of vaccine (Fig. 2).

## 3.2 Significance of lockdown

To know the significance of lockdown we define the COVID-19 case transmission is as $c_t = \dfrac{x_t}{\sum\limits_{i=0}^{t} x_i}$, where $x_t$ is the number of confirm cases on $t^{th}$ day.

**TABLE 1** Predicted value of COVID-19 cases from logistic growth model.

| Date | Predicted | 95% Confidence limit | | Date | Predicted | 95% Confidence limit | |
| | | Lower | Upper | | | Lower | Upper |
|---|---|---|---|---|---|---|---|
| 15-Jul-2020 | 979,177 | 972,095 | 986,260 | 31-Jul-2020 | 1,414,533 | 1,404,319 | 1,424,746 |
| 16-Jul-2020 | 1,004,769 | 997,491 | 1,012,047 | 01-Aug-2020 | 1,443,118 | 1,432,717 | 1,453,519 |
| 17-Jul-2020 | 1,030,622 | 1,023,147 | 1,038,096 | 02-Aug-2020 | 1,471,804 | 1,461,217 | 1,482,391 |
| 18-Jul-2020 | 1,056,726 | 1,049,054 | 1,064,398 | 03-Aug-2020 | 1,500,579 | 1,489,809 | 1,511,350 |
| 19-Jul-2020 | 1,083,073 | 1,075,204 | 1,090,943 | 04-Aug-2020 | 1,529,434 | 1,518,482 | 1,540,386 |
| 20-Jul-2020 | 1,109,653 | 1,101,585 | 1,117,720 | 05-Aug-2020 | 1,558,357 | 1,547,226 | 1,569,488 |
| 21-Jul-2020 | 1,136,456 | 1,128,190 | 1,144,721 | 06-Aug-2020 | 1,587,339 | 1,576,032 | 1,598,647 |
| 22-Jul-2020 | 1,163,472 | 1,155,008 | 1,171,935 | 07-Aug-2020 | 1,616,371 | 1,604,889 | 1,627,852 |
| 23-Jul-2020 | 1,190,691 | 1,182,030 | 1,199,352 | 08-Aug-2020 | 1,645,441 | 1,633,788 | 1,657,093 |
| 24-Jul-2020 | 1,218,103 | 1,209,245 | 1,226,962 | 09-Aug-2020 | 1,674,539 | 1,662,718 | 1,686,360 |
| 25-Jul-2020 | 1,245,699 | 1,236,643 | 1,254,754 | 10-Aug-2020 | 1,703,657 | 1,691,671 | 1,715,644 |
| 26-Jul-2020 | 1,273,466 | 1,264,215 | 1,282,718 | 11-Aug-2020 | 1,732,784 | 1,720,636 | 1,744,933 |
| 27-Jul-2020 | 1,301,397 | 1,291,950 | 1,310,843 | 12-Aug-2020 | 1,761,912 | 1,749,604 | 1,774,219 |
| 28-Jul-2020 | 1,329,479 | 1,319,839 | 1,339,119 | 13-Aug-2020 | 1,791,030 | 1,778,567 | 1,803,492 |
| 29-Jul-2020 | 1,357,702 | 1,347,870 | 1,367,535 | 14-Aug-2020 | 1,820,129 | 1,807,514 | 1,832,743 |
| 30-Jul-2020 | 1,386,057 | 1,376,033 | 1,396,081 | 15-Aug-2020 | 1,849,200 | 1,836,437 | 1,861,963 |
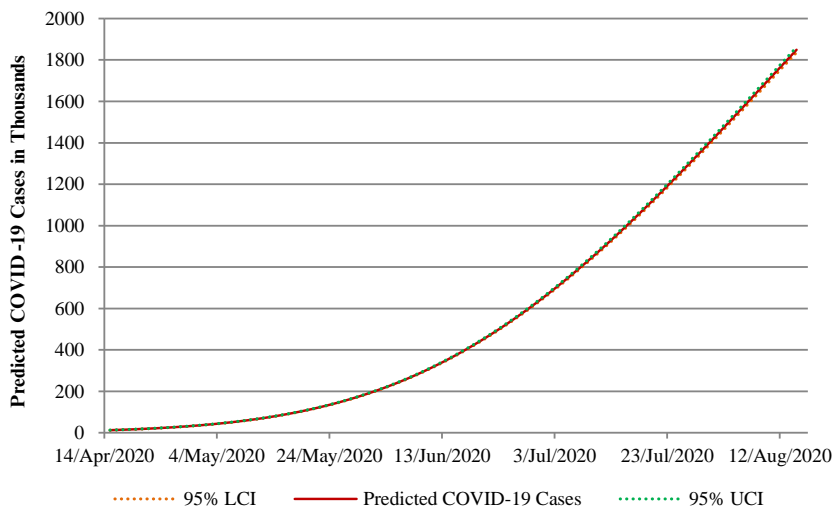
**FIG. 2** 95% Confidence interval of COVID-19 predicted cases (logistic growth model).

We have calculated $c_t$ and the doubling time of the corona case transmission in India. The doubling time is calculate as $\frac{Ln2}{c_t} = \frac{0.693}{c_t}$.

We have calculated COVID-19 case transmission $c_t$ on the basis of 5 days moving average of daily confirm cases (in the beginning the data in India is very fluctuating) and it is found gradually decreasing in India. This indicates the good sign of government attempts to combat this pandemic through implementing lockdown. These findings indicate that in future the burden of corona will be expectedly lowering down if the current status remains same. In Table 2 given below, an attempt has been made to show the summary statistics of corona case transmission $c_t$ during various lockdown periods in India. It is observed that average COVID-19 case transmission was maximum (0.16 with standard deviation 0.033) in the period prior to the lockdown. During the first lockdown period the average COVID-19 case transmission was 0.14 with standard deviation 0.032, however, in lockdown 2 it was 0.07 with standard deviation 0.009 and in lockdown 3 COVID-19 case transmission was 0.06 with standard deviation 0.007, however, in the period of fourth lockdown the average case transmission was 0.05 with standard deviation 0.005, thus it is clear that both average transmission load and standard deviation are decreasing. Table 3 reveals the result of ANOVA for average $c_t$ during various lockdown periods which is significant means that the average corona case transmission is significantly different is various lockdown periods considered. A group wise comparison of the average COVID-19 case transmission $c_t$ during various lockdown periods is shown in Table 4 which reveals that first lockdown is significantly affects the spread of corona case transmission than others but second lockdown period is not significantly different than third and fourth. Same result is observed for third and fourth lockdown period.

**TABLE 2** Summary of $c_t$ during various lockdown period.

| Lockdown period | N | Mean | Std. deviation | Std. error | 95% confidence interval for mean | |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | | *Lower bound* | *Upper bound* |
| No | 11 | 0.16 | 0.035 | 0.011 | 0.134 | 0.182 |
| 1 | 22 | 0.14 | 0.032 | 0.006 | 0.121 | 0.149 |
| 2 | 19 | 0.07 | 0.009 | 0.002 | 0.064 | 0.073 |
| 3 | 14 | 0.06 | 0.007 | 0.002 | 0.051 | 0.060 |
| 4 | 14 | 0.05 | 0.005 | 0.001 | 0.044 | 0.049 |
| Total | 80 | 0.093 | 0.048 | 0.005 | 0.082 | 0.104 |

**TABLE 3** ANOVA test for mean of $c_t$ during various lockdown period.

| Source of variations | Sum of squares | df | Mean square | F | p value |
|---|---|---|---|---|---|
| Between groups | 0.147 | 04 | 0.037 | 75.58 | 0.000 |
| Within groups | 0.036 | 75 | 0.000 | | |
| Total | 0.183 | 79 | | | |

This indicates that the COVID-19 transmission is not under control now. Fig. 3 shows corona case transmission and doubling time in India. The corona case propagation in decreasing and doubling time is increasing day by day.

### 3.3 Propagation model (based on Newton's law of cooling)

Let us define a function called tempo of disease that is the first differences in natural logarithms of the cumulative corona positive cases on a day, which is as:

$$r_t = \ln\left(\frac{p_t}{p_{t-1}}\right)$$

where $p_t$ and $p_{t-1}$ are the number of cumulative corona positive cases for period $t$ and $t-1$, respectively. When $p_t$ and $p_{t-1}$ are equal then $r_t$ will become zero. If this value of $r_t$, i.e., zero will continue a week then we can assume no new corona cases will appear further. In the initial face of the disease spread, the tempo of disease increases but after sometime when some preventive measures is being taken then it decreases.

Since $r_t$ is a function of time then the first differential is defined as

$$\frac{dr_t}{dt} = k(r_t - r_T) \tag{8}$$

where $r_t$ denotes the tempo that is the first differences in natural logarithms of the cumulative corona positive cases on a day, $r_T$ is the desired level of tempo, i.e., zero in this study, $t$ denotes the time and $k$ is a constant of proportionality.

Eq. (8) is an example of an ordinary differential equation that can be solved by the method of separating variables. The Eq. (8) can be written as

$$\frac{dr_t}{r_t} = kdt \tag{9}$$

Integrating Eq. (9), we get

$$\ln r_t = kt + C \tag{10}$$

**TABLE 4** Group wise comparison of means of $c_t$ during various lockdown periods.

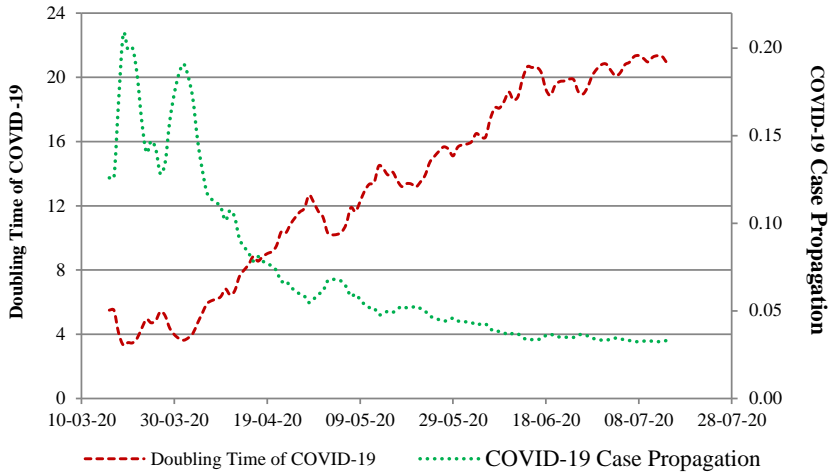| (I) group | (J) group | Mean difference (I − J) | Std. error | $p$ value | 95% Confidence interval | |
|---|---|---|---|---|---|---|
| | | | | | *Lower bound* | *Upper bound* |
| No | 1 | 0.023 | 0.008 | 0.057 | 0.000 | 0.047 |
| | 2 | 0.090 | 0.008 | 0.000 | 0.066 | 0.114 |
| | 3 | 0.102 | 0.009 | 0.000 | 0.077 | 0.128 |
| | 4 | 0.112 | 0.009 | 0.000 | 0.086 | 0.137 |
| 1 | 2 | 0.067 | 0.007 | 0.000 | 0.047 | 0.087 |
| | 3 | 0.079 | 0.008 | 0.000 | 0.058 | 0.101 |
| | 4 | 0.089 | 0.008 | 0.000 | 0.067 | 0.110 |
| 2 | 3 | 0.013 | 0.008 | 1.000 | −0.010 | 0.035 |
| | 4 | 0.022 | 0.008 | 0.059 | −0.001 | 0.044 |
| 3 | 4 | 0.009 | 0.008 | 1.000 | −0.015 | 0.033 |

**FIG. 3** Doubling time and COVID-19 propagation in India.

where $C$ is an arbitrary constant. Taking the antilogarithms of both sides of Eq. (10) we have

$$r_t = e^{kt+C} \Rightarrow e^{kt}e^C \qquad \Rightarrow \qquad r_t = Ae^{kt} \qquad (11)$$

where $A = e^C$. This Eq. (11) is the general solution of Eq. (8). If $k$ is less than zero, Eq. (11) tells us how the COVID-19 cases will decreases over the time until it reaches zero. Value of $A$ and $k$ is estimated by least square estimation procedure using the data sets.

The Government of India implemented lockdown on March 24th, 2020 and expected that the tempo of disease is decreasing. Government suggested and implemented social distancing and lockdown to control the spread of COVID-19 in the society. In Table 5, the predicted value of COVID-19 cases obtained with this method is given along with 95% confidence interval. About 21.5 lakh cases are expected by August 15th, 2020. With this model it is expected that about 45 lakh peoples will be infected in India by the end of October and after that no cases will happen since the tempo of disease $r_t$ will become zero (Fig. 4). In Table 6 an attempt has been made to show the summary statistics of tempo of COVID-19 $r_t$ during various lockdown periods in India. It is observed that average tempo is maximum (0.17 with standard deviation 0.062) in the period prior to the lockdown. During the first lockdown period the average tempo is 0.14 with standard deviation 0.044 and after that it is found decreasing in the various lockdowns. Table 7 indicates lockdown periods are significantly different in terms of tempo of disease spread. A group wise comparison of the average tempo of COVID-19 during

**TABLE 5** Predicted confirmed cases of COVID-19 till 15 August 2020, India with propagation model.

| Date | Predicted | 95% Confidence limit | | Date | Predicted | 95% Confidence limit | |
| | | Lower | Upper | | | Lower | Upper |
|---|---|---|---|---|---|---|---|
| 15-Jul-2020 | 1,140,781 | 845,770 | 1,568,823 | 31-Jul-2020 | 1,631,673 | 1,162,680 | 2,346,267 |
| 16-Jul-2020 | 1,169,628 | 864,891 | 1,613,198 | 01-Aug-2020 | 1,664,042 | 1,182,994 | 2,399,124 |
| 17-Jul-2020 | 1,198,757 | 884,133 | 1,658,183 | 02-Aug-2020 | 1,696,557 | 1,203,333 | 2,452,409 |
| 18-Jul-2020 | 1,228,163 | 903,487 | 1,703,769 | 03-Aug-2020 | 1,729,209 | 1,223,691 | 2,506,109 |
| 19-Jul-2020 | 1,257,836 | 922,949 | 1,749,945 | 04-Aug-2020 | 1,761,989 | 1,244,064 | 2,560,212 |
| 20-Jul-2020 | 1,287,768 | 942,513 | 1,796,703 | 05-Aug-2020 | 1,794,889 | 1,264,445 | 2,614,704 |
| 21-Jul-2020 | 1,317,951 | 962,171 | 1,844,032 | 06-Aug-2020 | 1,827,899 | 1,284,830 | 2,669,570 |
| 22-Jul-2020 | 1,348,376 | 981,919 | 1,891,920 | 07-Aug-2020 | 1,861,012 | 1,305,214 | 2,724,799 |
| 23-Jul-2020 | 1,379,035 | 1,001,750 | 1,940,359 | 08-Aug-2020 | 1,894,218 | 1,325,590 | 2,780,376 |
| 24-Jul-2020 | 1,409,918 | 1,021,658 | 1,989,336 | 09-Aug-2020 | 1,927,509 | 1,345,954 | 2,836,287 |
| 25-Jul-2020 | 1,441,019 | 1,041,638 | 2,038,840 | 10-Aug-2020 | 1,960,876 | 1,366,302 | 2,892,520 |
| 26-Jul-2020 | 1,472,327 | 1,061,682 | 2,088,859 | 11-Aug-2020 | 1,994,312 | 1,386,628 | 2,949,059 |
| 27-Jul-2020 | 1,503,833 | 1,081,786 | 2,139,382 | 12-Aug-2020 | 2,027,806 | 1,406,927 | 3,005,892 |
| 28-Jul-2020 | 1,535,530 | 1,101,944 | 2,190,397 | 13-Aug-2020 | 2,061,353 | 1,427,196 | 3,063,004 |
| 29-Jul-2020 | 1,567,409 | 1,122,149 | 2,241,891 | 14-Aug-2020 | 2,094,942 | 1,447,428 | 3,120,382 |
| 30-Jul-2020 | 1,599,459 | 1,142,396 | 2,293,852 | 15-Aug-2020 | 2,128,567 | 1,467,621 | 3,178,012 |
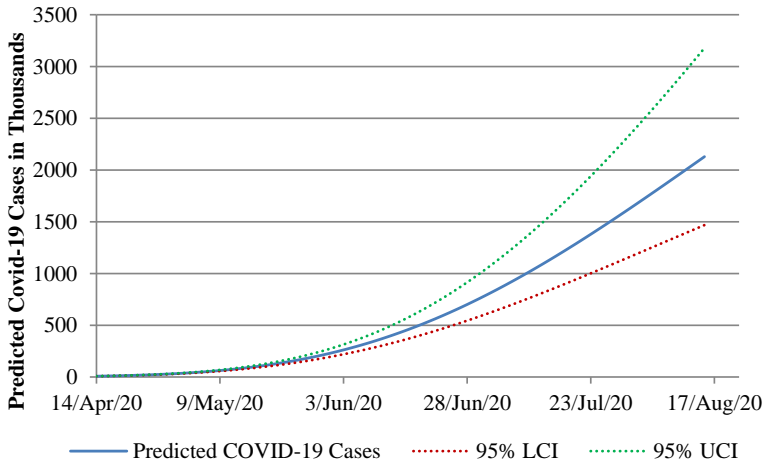
**95% Confidence Interval of COVID-19 Predicted Cases (propagation model)**

**FIG. 4** 95% Confidence interval of COVID-19 predicted cases (propagation model).

various lockdown periods is shown in Table 8 which reveals that first lockdown is significantly different than others. Consecutive mean difference shows that the decrease in disease spread has been observed but insignificant, means there is no impact of lockdown on controlling the disease spread.

## 3.4 Joinpoint regression model

To analyze the temporal trends and to identify important changes in the trends of the COVID-19 outbreak joinpoint regression is used in China (Al Hasan et al., 2020); here in this study we performed a joinpoint regression analysis in India to understand the pattern of COVID-19. Joinpoint regression analysis, enable us to identify time at a meaningful change in the slope of a trend is observed over the study period. The best fitting points known as joinpoints, that are chosen when the slope changes significantly in the models.

To tackle the above problem joinpoint regression analysis (Kim et al., 2000) has been employed in this study to present trend analysis. The goal of the joinpoint regression analysis is not only to provide the statistical model that best fits the time series data but also, the purpose is to provide that model which best summarizes the trend in the data (Marrot, 2010).

Let $y_i$ denotes the reported COVID-19 positive cases on day $t_i$ such that $t_1 < t_2 < \ldots < t_n$. Then the joinpoint regression model is defined as

$$\ln y_i = \alpha + \beta_1 t_1 + \delta_1 u_1 + \delta_2 u_2 + \ldots + \delta_j u_j + \varepsilon_i \tag{12}$$

**TABLE 6** Summary of $r_t$ during various lockdown period.

| Lockdown period | N | Mean | Std. deviation | Std. error | 95% Confidence interval for mean | |
|---|---|---|---|---|---|---|
| | | | | | *Lower bound* | *Upper bound* |
| No | 11 | 0.17 | 0.062 | 0.019 | 0.125 | 0.210 |
| 1 | 22 | 0.14 | 0.044 | 0.009 | 0.120 | 0.159 |
| 2 | 19 | 0.07 | 0.012 | 0.003 | 0.064 | 0.076 |
| 3 | 14 | 0.06 | 0.007 | 0.002 | 0.051 | 0.059 |
| 4 | 14 | 0.05 | 0.006 | 0.001 | 0.046 | 0.052 |
| Total | 80 | 0.10 | 0.056 | 0.006 | 0.084 | 0.109 |

**TABLE 7** ANOVA test for mean of $r_t$ during various lockdown period.

| Source of variations | Sum of squares | df | Mean square | F | p value |
|---|---|---|---|---|---|
| Between groups | 0.166 | 04 | 0.041 | 37.24 | 0.000 |
| Within groups | 0.083 | 75 | 0.001 | | |
| Total | 0.249 | 79 | | | |

where $u_j = \begin{cases} (t_j - k_j) & \text{if } t_j > k_j \\ 0 & \text{otherwise} \end{cases}$ and $k_1 < k_2 \ldots < k_j$ are joinpoints. The details of joinpoint regression analysis are given elsewhere (Kim et al., 2004).

Joinpoint regression analysis is used when the temporal trend of an amount, like incidence, prevalence and mortality is of interest (Doucet et al., 2016). However, this method has generally been applied with the calendar year as the time scale (Akinyede and Soyemi, 2016; Chatenoud et al., 2015; Missikpode et al., 2015; Mogos et al., 2016). The joinpoint regression analysis can also be applied in epidemiological studies in which the starting date can be easily established such as the day when the disease is detected for the first time as is the case in the present analysis (Rea et al., 2017). Estimated regression coefficients ($\beta$) were calculated for the trends extracted from the joinpoint regression. Additionally, the average daily percent change (ADPC), calculated as a geometric weighted average of the daily percent changes (Clegg et al., 2009). The joinpoints are selected based on the data-driven Bayesian Information Criterion (BIC) method (Zhang and Siegmund, 2007).

The equation for computing the BIC for a $k$-joinpoints regression is:

$$BIC(k) = \ln\left[\frac{SSE(k)}{n}\right] + \frac{2(k+1) \times \ln(n)}{n} \tag{13}$$

where SSE is the sum of squared errors of the $k$-joinpoints regression model and $n$ is the number of observations. The model which has the minimum value of BIC($k$) is selected as the final model. There are other methods also for identifying the joinpoints such as permutation test method and the weighted BIC methods. Relative merits and demerits of different methods of identifying the joinpoints are discussed elsewhere (National Institute Cancer, 2013). The permutation test method is regarded as the best method but it is computationally very intensive. It controls the error probability of selecting the wrong model at a certain level (i.e., 0.05). The BIC method, on the other hand, is less complex computationally.

**TABLE 8** Group wise comparison of means of $r_t$ during various lockdown periods.

| (I) group | (J) group | Mean difference (I − J) | Std. error | p value | 95% Confidence interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower bound | Upper bound |
| No | 1 | 0.027 | 0.012 | 0.306 | −0.009 | 0.063 |
| | 2 | 0.097 | 0.013 | 0.000 | 0.061 | 0.134 |
| | 3 | 0.112 | 0.013 | 0.000 | 0.073 | 0.151 |
| | 4 | 0.118 | 0.013 | 0.000 | 0.079 | 0.157 |
| 1 | 2 | 0.070 | 0.010 | 0.000 | 0.040 | 0.100 |
| | 3 | 0.085 | 0.011 | 0.000 | 0.052 | 0.118 |
| | 4 | 0.091 | 0.011 | 0.000 | 0.058 | 0.124 |
| 2 | 3 | 0.015 | 0.012 | 1.000 | −0.019 | 0.049 |
| | 4 | 0.021 | 0.012 | 0.790 | −0.013 | 0.055 |
| 3 | 4 | 0.006 | 0.013 | 1.000 | −0.030 | 0.043 |

In the present case, data on the reported confirmed cases of COVID-19 are available on a daily, thus the daily percent change ($DPC$) from day $t$ to day ($t+1$) is defined as

$$DPC = \left(\frac{y_{t+1} - y_t}{y_t}\right) \times 100 \qquad (14)$$

If the trend in the daily reported confirmed cases of COVID-19 is modeled as

$$\ln(y_t) = b_0 + b_1 t + \varepsilon \qquad (15)$$

then, it can be shown that the $DPC$ is equal to

$$DPC = \left(e^{b_1} - 1\right) \times 100 \qquad (16)$$

It is worthwhile to discuss here is that the positive value of $DPC$ indicates an increasing trend while the negative value of $DPC$ suggests a declining trend. The $DPC$ reflects the trend in the reported COVID-19 positive cases in different time segments of the reference period observed through joinpoint regression techniques. For the entire study period, it is possible to estimate average daily percent change ($ADPC$) that is the weighted average of $DPC$ of different time segments of the study period with weights equal to the length of different time segments. However, when the trend changes frequently, $ADPC$ has little meaning. It assumes that the random errors are heteroscedastic (have nonconstant variance). Heteroscedasticity is handled by joinpoint regression using weighted least squares (WLS). The weights in WLS are the reciprocal of the variance and can be specified in several ways. Thus standard error is used to control heteroscedastic in the analysis during the entire period.

To observe the trend of reported cases, the moving average method has been used in this study. The daily percent change ($DPC$) in the daily reported confirmed cases of COVID-19 during the period March 14th, 2020 through July 16th, 2020 is used for forecasting the daily reported confirmed cases of COVID-19 in the immediate future under the assumption that the trend in the daily reported confirmed cases of COVID-19 remains unchanged. The number of cases increased by the rate of 6.20% per day in India; however, the rate is different in the different segment. Also Table 9 reveals that the growth rate is positive and significant (about 19%) from 16th March to 3rd April and after that the growth rate is decreasing in comparison of first segment, i.e., for 28 days (from 3rd April to 30th April). The possible reason may be lockdown imposed in India. In the third segment, i.e., from 30th April to 4th May a high increase has been observed but it is insignificant. From 4th April to 13th May the rate is although the positive but dramatically lower than the previous segments growth rate. In the next segment, i.e., 5th segment which is of 8 days, we observe a significant increase of 6.55% in COVID-19 cases. In the last and 6th segment from 20th May to 14th July, i.e., for 56 days, the growth rate is found again positive and significant (3.03% per day) in the COVID-19 cases.

**TABLE 9** Results of the joinpoint regression analysis, India with sd.

| Segment | Lower endpoint | Upper endpoint | Number of days[a] | Average daily percent change (ADPC)[b] | 95% Confidence interval | | Test statistic (t) | P > \|t\| |
|---|---|---|---|---|---|---|---|---|
| | | | | | *Lower* | *Upper* | | |
| 1 | 16 March | 3 April | 19 | 19.19[c] | 17.20 | 21.20 | 20.74 | 0.00 |
| 2 | 3 April | 30 April | 28 | 5.27[c] | 4.99 | 5.56 | 37.68 | 0.00 |
| 3 | 30 April | 4 May | 5 | 11.34 | −2.48 | 27.12 | 1.61 | 0.11 |
| 4 | 4 May | 13 May | 10 | 1.78[c] | 0.44 | 3.13 | 2.64 | 0.01 |
| 5 | 13 May | 20 May | 8 | 6.55[c] | 3.63 | 9.54 | 4.54 | 0.00 |
| 6 | 20 May | 14 July | 56 | 3.03[c] | 2.98 | 3.08 | 117.45 | 0.00 |
| All | 16 March | 14 July | 121 | 6.20[c] | 5.63 | 6.78 | 21.82 | 0.00 |

[a]Number of days in a segment include both lower endpoint and upper endpoint.
[b]Average daily percent change.
[c]Statistically significant.
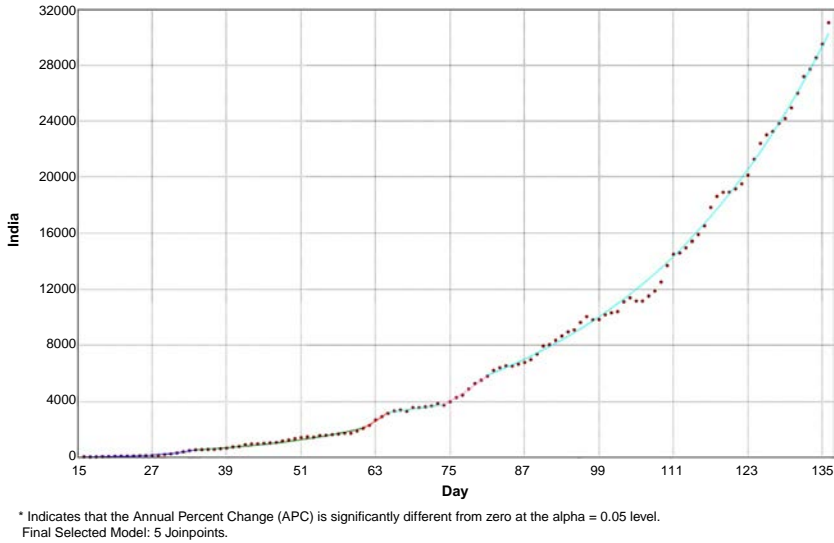
**FIG. 5**  Trend in daily reported confirmed cases of COVID-19 in India using joinpoint regression analysis with BIC criterion.

Fig. 5 shows that the trend increases in India still sharply and there is no hope of decline in COVID-19 cases. Fig. 2 shows the forecasted value of COVID-19 daily cases in India. The COVID-19 cases will increase further if the same trend prevailing.

Table 10 presents the forecast of the predicted cases of COVID-19 in India along with 95% confidence intervals. This exercise suggests that by August 15th, 2020, the confirmed cases of COVID-19 in India is likely to be 2,587,007 with a 95% confidence interval of 2,571,896–2,602,282 and daily reported cases will be 78,729 with 95% confidence interval of 77,516–79,961. This daily reported COVID-19 positive cases may change only when an appropriate set of new interventions are introduced to fight COVID-19 pandemic. It is observed that analysis indicates that in the month of August, India faces more than 50 thousand cases per day (Fig. 6).

## 4  Conclusions

India is in the comfortable zone with a lower growth rate than other countries. Logistic model shows that, the epidemic is likely to stabilize with 45 lakh cases by the end of March 2021 and peak will come in middle of the August, however, propagation model provide estimate of maximum COVID-19 case as 45 lakh but the timing is different (by end October) than the logistic model.

**TABLE 10** Forecast of daily predicted confirmed cases of COVID-19 till August 15, 2020, India.

| Date | Predicted | 95% Confidence limit | | Date | Predicted | 95% Confidence limit | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | *Lower* | *Upper* | | | *Lower* | *Upper* |
| 15-Jul-2020 | 31,208 | 31,193 | 31,223 | 31-Jul-2020 | 50,313 | 49,900 | 50,730 |
| 16-Jul-2020 | 32,153 | 32,122 | 32,184 | 01-Aug-2020 | 51,837 | 51,387 | 52,292 |
| 17-Jul-2020 | 33,127 | 33,079 | 33,176 | 02-Aug-2020 | 53,408 | 52,918 | 53,903 |
| 18-Jul-2020 | 34,131 | 34,065 | 34,198 | 03-Aug-2020 | 55,026 | 54,495 | 55,563 |
| 19-Jul-2020 | 35,165 | 35,080 | 35,251 | 04-Aug-2020 | 56,694 | 56,119 | 57,274 |
| 20-Jul-2020 | 36,231 | 36,126 | 36,337 | 05-Aug-2020 | 58,412 | 57,791 | 59,038 |
| 21-Jul-2020 | 37,329 | 37,202 | 37,456 | 06-Aug-2020 | 60,181 | 59,513 | 60,857 |
| 22-Jul-2020 | 38,460 | 38,311 | 38,609 | 07-Aug-2020 | 62,005 | 61,287 | 62,731 |
| 23-Jul-2020 | 39,625 | 39,452 | 39,799 | 08-Aug-2020 | 63,884 | 63,113 | 64,663 |
| 24-Jul-2020 | 40,826 | 40,628 | 41,024 | 09-Aug-2020 | 65,819 | 64,994 | 66,655 |
| 25-Jul-2020 | 42,063 | 41,839 | 42,288 | 10-Aug-2020 | 67,814 | 66,931 | 68,708 |
| 26-Jul-2020 | 43,337 | 43,086 | 43,590 | 11-Aug-2020 | 69,868 | 68,925 | 70,824 |
| 27-Jul-2020 | 44,650 | 44,370 | 44,933 | 12-Aug-2020 | 71,985 | 70,979 | 73,005 |
| 28-Jul-2020 | 46,003 | 45,692 | 46,317 | 13-Aug-2020 | 74,167 | 73,094 | 75,254 |
| 29-Jul-2020 | 47,397 | 47,053 | 47,743 | 14-Aug-2020 | 76,414 | 75,273 | 77,572 |
| 30-Jul-2020 | 48,833 | 48,456 | 49,214 | 15-Aug-2020 | 78,729 | 77,516 | 79,961 |
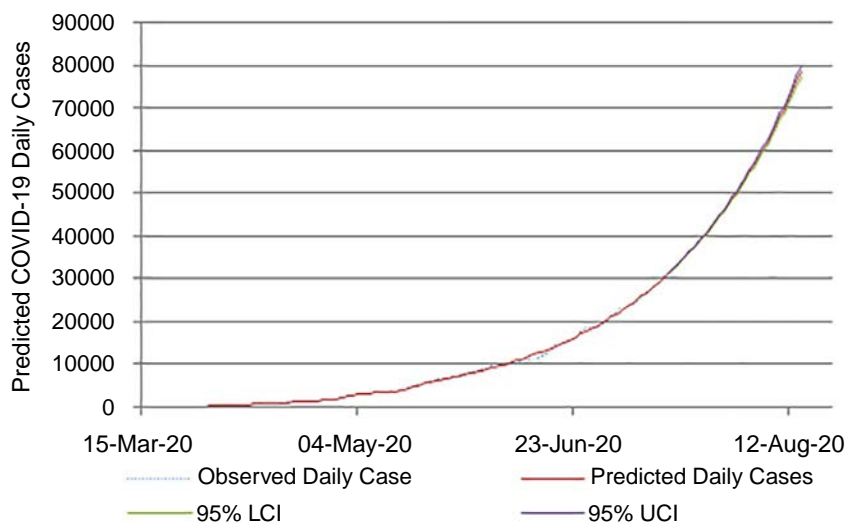
**FIG. 6** Forecast of daily reported confirmed COVID-19 cases in India up to August 15, 2020.

Logistic model need to monitor the data time to time for good long term prediction. The projections produced by the model and after their validation can be used to determine the scope and scale of measures that government need to initiate. Joinpoint regression is based on the daily reported confirmed cases of COVID-19, asserts that there has virtually been little impact of the nationwide lockdown as well as relaxations in restrictions on the progress of the COVID-19 pandemic in India. The joinpoint regression analysis provides better estimate up to 15th August for the confirmed COVID-19 cases than the other two methods. To know the better understanding of the progress of the epidemic in the country may be obtained by analyzing the progress of the epidemic at the regional level. In conclusion, if the current mathematical model results can be validated within the range provided here, then the social distancing and other prevention, treatment policies that the central and various state governments and people are currently implementing should continue until new cases are not seen. The spread from urban to rural and rich to poor populations should be monitor and control is an important point of consideration. Mathematical models have certain limitations that there are many assumptions about homogeneity of population in terms of urban/rural or rich/poor that does not capture variations in population density. If several protective measures will not be taken effectively, then this rate may be changed. However, the government of India under the leadership of Modi Ji has already taken various protective measures such as lockdown in several areas, make possible quarantine facility to reduce the rate of increase of COVID-19, thus we may hopefully conclude that, country will be successful to reduce the rate of this pandemic.

# References

Akinyede, O., Soyemi, K., 2016. Joinpoint regression analysis of pertussis crude incidence rates, Illinois, 1990-2014. Am. J. Infect. Control 44 (12), 1732–1733.

Al Hasan, S.M., Saulam, J., Kanda, K., Hirao, T., 2020. The novel coronavirus disease (COVID-19) outbreak trends in mainland China: a joinpoint regression analysis of the outbreak data from January 10 to February 11, 2020 (preprint). Bull. World Health Organ. https://doi.org/10.2471/BLT.20.253153.

Anand, S., Fan, V., 2016. The Health Workforce in India. WHO; Human resources for Health Observer, Geneva. Series No. 16 Available at: https://www.who.int/hrh/resources/hwindia_health-obs16/en/.

Anastassopoulou, C., Russo, L., Tsakris, A., Siettos, C., 2020. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS One 15 (3), e0230405. https://doi.org/10.1371/journal.pone.0230405.

Batista, M., 2020. Estimation of the Final Size of the Second Phase of the Coronavirus COVID 19 Epidemic by the Logistic Model. medRxiv. preprint https://doi.org/10.1101/2020.03.11.20024901.

Brauer, F., Castillo-Chavez, C., 2012. Mathematical Models in Population Biology and Epidemiology, second ed. Springer, New York, NY.

Chang, S.L., Harding, N., Zachreson, C., Cliff, O.M., Prokopenko, M., 2020. Modelling Transmission and Control of the Covid-19 Pandemic in Australia. arXiv. preprint arXiv:2003.10218.

Chatenoud, L., Garavello, W., Pagan, E., Bertuccio, P., Gallus, S., La Vecchia, C., et al., 2015. Laryngeal cancer mortality trends in European countries. Int. J. Cancer 842, 833–842.

Chatterjee, K., Chatterjee, K., Kumar, A., Shankar, S., 2020. Healthcare impact of COVID-19 epidemic in India: a stochastic mathematical model. Med. J. Armed Forces India 76 (2), 147–155. https://doi.org/10.1016/j.mjafi.2020.03.022.

Chen, Y.C., Lu, P.E., Chang, C.S., Liu, T.H., 2020. A Time-Dependent SIR Model for COVID-19 With Undetectable Infected Persons. http://gibbs1.ee.nthu.edu.tw/A_Time_Dependent_SIR_Model_For_Covid 19.pdf.

Clegg, L.X., Hankey, B.F., Tiwari, R., Feuer, E.J., Edwards, B.K., 2009. Estimating average annual per cent change in trend analysis. Stat. Med. 28 (29), 3670–3682.

Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K., Bleicker, T., Brunink, S., Schneider, J., Schmidt, M.L., 2020. Detection of 2019 novel coronavirus (2019-ncov) by realtime RT-PCR. Euro Surveill. 25 (3), 2000045.

Cowling, B.J., Leung, G.M., 2020. Epidemiological research priorities for public health control of the ongoing global novel coronavirus (2019-nCoV) outbreak. Euro Surveill. 25 (6), 2000110.

Daley, D.J., Gani, J., 1999. Epidemic Modelling: An Introduction. Cambridge University Press, Cambridge.

De Silva, U., Warachit, J., Waicharoen, S., Chittaganpitch, M., 2009. A preliminary analysis of the epidemiology of influenza A (H1N1) v virus infection in Thailand from early outbreak data, June-July 2009. Euro Surveill. 14 (31), 19292.

Doucet, M., Rochette, L., Hamel, D., 2016. Prevalence and mortality trends in chronic obstructive pulmonary disease over 2001 to 2011: a public health point of view of the burden. Can. Respir. J., 1–10. https://doi.org/10.1155/2016/7518287.

Ferguson, N.M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., Dighe, A., 2020. Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand. Imperial College COVID-19 Response Team.

Gamero, J., Tamayo, J.A., Martinez-Roman, J.A., 2020. Forecast of the Evolution of the Contagious Disease Caused by Novel Corona Virus (2019-ncov) in China. arXiv. preprint arXiv: 2002. 04739.

Gupta, M., 2020. Corona Virus in India: Make or Break. Available at https://medium.com/@mohakgupta_55841/coronavirus-in-india-make-or-break-5a13dfb9646d.

Herbert, W.H., 2000. The mathematics of infectious diseases. SIAM Rev. 42 (4), 599–653.

Hethcote, H.W., 1983. Measles and rubella in the United States. Am. J. Epidemiol. 117, 2–13.

Hethcote, H.W., 1988. Optimal ages or vaccination for measles. Math. Biosci. 89, 29–52.

Hethcote, H.W., 1989. Rubella. In: Levin, S.A., Hallam, T.G., Gross, L. (Eds.), Applied Mathematical Ecology. Biomathematics. vol. 18. Springer, Berlin, Heidelberg, New York.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395 (10223), 497–506.

Hui, D.S., Azhar, E.I., Madani, T.A., Ntoumi, F., Kock, R., Dar, O., Ippolito, G., Mchugh, T.D., Memish, Z.A., Drosten, C., 2020. The continuing 2019-ncov epidemic threat of novel coronaviruses to global health-the latest 2019 novel coronavirus outbreak in Wuhan, China. Int. J. Infect. Dis. 91, 264.

Junling, M., Dushoff, J., Bolker, B.M., Earn, D.J., 2014. Estimating initial epidemic growth rates. Bull. Math. Biol. 76 (1), 245–260.

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proc. Roy. Soc. Lond. A 115 (700–721), 1927.

Kim, H.J., Fay, M.P., Feuer, E.J., Midthune, D.N., 2000. Permutation tests for joinpoint regression with applications to cancer rates. Stat. Med. 19, 335–351.

Kim, H.J., Fay, M.P., Yu, B., Barrett, M.J., Feuer, E.J., 2004. Comparability of segmented line regression models. Biometrics 60 (4), 1005–1014.

Koo, J.R., Cook, A.R., Park, M., Sun, Y., Sun, H., Lim, J.T., Tam, C., Dickens, B.L., 2020. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. Lancet Infect. Dis. 20, 678–688. https://doi.org/10.1016/S1473-3099(20)30162-6.

Kucharski, A.J., Russell, T.W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Davies, N., 2020. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect. Dis. 20 (5), 553–558.

Lipsitch, M., Swerdlow, D.L., Finelli, L., 2020. Defining the epidemiology of Covid-19-studies needed. N. Engl. J. Med. 382 (13), 1194–1196.

Longini Jr., I.M., Ackerman, E., Elveback, L.R., 1978. An optimization model for influenza A epidemics. Math. Biosci. 38, 141–157.

Malhotra, B., Kashyap, V., 2020. Progression of COVID-19 in Indian States—Forecasting Endpoints Using SIR and Logistic Growth Models. https://doi.org/10.1101/2020.05.15.20103028. PPR:PPR163964.

Mandal, S., Bhatnagar, T., Arinaminpathy, N., Agarwal, A., et al., 2020. Prudent public health intervention strategies to control the corona virus disease 2019 transmission in India: a mathematical model-based approach. Indian J. Med. Res. 151, 190–199. Epub ahead of print https://doi.org/10.4103/ijmr.IJMR_504_20.

Marrot, L.D., 2010. Colorectal Cancer Network (CRCNet) User Documentation for Surveillance Analytic Software: Joinpoint. Cancer Care Ontario, pp. 1–28.

Missikpode, C., Peek-Asa, C., Young, T., Swanton, A., Leinenkugel, K., Torner, J., 2015. Trends in non-fatal agricultural injuries requiring trauma care. Inj. Epidemiol. 2 (1), 30.

Mogos, M.F., Salemi, J.L., Spooner, K.K., McFarlin, B.L., Salihu, H.M., 2016. Differences in mortality between pregnant and nonpregnant women after cardiopulmonary resuscitation. Obstet. Gynecol. 128 (4), 880–888.

National Institute Cancer, 2013. Joinpoint Regression Program. National Institutes of Health, United States Department of Health and Human Services, Bethesda, MD.

Pearl, R., Reed, L.J., 1920. On the rate of growth of the population of the United States since 1790 and its mathematical representation. Proc. Natl. Acad. Sci. U. S. A. 6 (6), 275.

Rea, F., Pagan, E., Compagnoni, M.M., Cantarutti, A., Pigni, P., Bagnardi, V., et al., 2017. Joinpoint regression analysis with time-on-study as time-scale. Application to three Italian population-based cohort studies. Epidemiol. Biostat. Public Health 14 (3), e12616.

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., 2020. Transmission of 2019-ncov infection from an asymptomatic contact in Germany. N. Engl. J. Med. 382 (10), 970–971.

Schueller, E., Klein, E., Tseng, K., Kapoor, G., Joshi, J., Sriram, A., Laxminarayan, R., 2020. COVID-19 in India: Potential Impact of the Lockdown and Other Longer Term Policies. Retrieved from Washington D.C. https://cddep.org/publications/covid-19-india-potential-impact-of-the-lockdown-and-other-longer-term-policies/.

Shen, C.Y., 2020. A logistic growth model for COVID-19 proliferation: experiences from China and international implications in infectious diseases. Int. J. Infect. Dis. 96, 582–589.

Shryock, H.S., Siegel, J.S., 1973. Methods and Materials of Demography. U.S. Dept. of Commerce, Bureau of the Census, Washington.

Singh, R., Adhikari, R., 2020. Age-Structured Impact of Social Distancing on the COVID-19 Epidemic in India. arXiv. 2003.12055v1 [q-bio.PE].

Srinivasa Rao Arni, S.R., Krantz, S.G., Thomas, K., Ramesh, B., 2020. Model-based retrospective estimates for COVID-19 or coronavirus in India: continued efforts required to contain the virus spread. Curr. Sci. 118 (7), 1023–1025.

Tuite, A.R., Fisman, D.N., 2020. Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-nCoV) epidemic. Ann. Intern. Med. 172 (8), 567–568.

Wu, J.T., Leung, K., Leung, G.M., 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 395 (10225), 689–697.

Yorke, J.A., London, W.P., 1973. Recurrent outbreaks of measles, chickenpox and mumps II. Am. J. Epidemiol. 98, 469–482.

Yorke, J.A., Nathanson, N., Pianigiani, G., Martin, J., 1979. Seasonality and the requirements for perpetuation and eradication of viruses in populations. Am J Epidemiol 109, 103–123.

Zhang, N.R., Siegmund, D.O., 2007. A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. Biometrics 63 (1), 22–32.

Zhang, Y., Yang, H., Cui, H., Chen, Q., 2019. Comparison of the ability of ARIMA, WNN and SVM models for drought forecasting in the Sanjiang Plain, China. Nat. Resour. Res. 29, 1447.

Zhao, S., Lin, Q., Ran, J., Musa, S.S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, D.He, Wang, M.H., 2020. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. Int. J. Infect. Dis. 92, 214–217.

## Further reading

COVID-19, 2020. https://www.covid19india.org.

World Health Organization, 2020. World Situation Report. (Online) Available from https://www.who.int/emergencies/diseases/novelcoronavirus-2019/situation-reports.