



Propagation Analysis of COVID-19: An SIR Model-Based Investigation of the Pandemic

Rahul Saxena^{1,2} · Mahipal Jadeja¹ · Vikrant Bhateja³

Received: 31 December 2020 / Accepted: 17 June 2021
© King Fahd University of Petroleum & Minerals 2021

Abstract

The paper investigates the spread pattern and dynamics of Covid-19 propagation based on SIR model. Using the model dynamics, an analytical estimation has been obtained for virus span, its longevity, growing pattern, etc. Experimental simulations are carried out on the data of four regions of India over a period of two months of country-wide lockdown. The analysis illustrates the effect of lockdown on the contact rate and its implication. Simulation results illustrate that there is a cut-down in effective contact rate by a considerable factor ranging from 2 to 4 for the selected regions. Further, the estimates for the vaccines to be developed, maximum range and span of the disease can be also estimated. Results portray that the SIR model is a significant tool to cast the dynamics and predictions of Covid-19 outbreak in comparison to other epidemic models. The study demonstrates the progression of real time data in accordance with the SIR model with high accuracy.

Keywords Covid-19 · Epidemic modeling · SIR model · Trend analysis · Predictive model

1 Introduction

The novel Corona Virus (Covid-19) outbreak emerged from Wuhan city of China on 2nd January 2020 as per the reports by WHO [1] and was declared Public Health Emergency of International Concern (PHEIC) on 30th January 2020. WHO named this disease as coronavirus disease (Covid-19) in February 2020 as per the same report in [1]. Till 16th May 2020, the count of Corona positive patients raises to 4.53 Million resulting in 307 K deaths. The spread of the disease is so rapid that it has resulted in 1,484,287 Corona positive cases and 88,507 deaths in USA; 2,76,505 positive cases and 27,563 deaths in Spain; 2,24,760 positive cases and 31,763

deaths in Italy and so on in many other nations [2]. Currently, India has surpassed China with a count of above 86,000 positive patients and stands 11th in the tally of the nation's having maximum positive patients. The impact for the pandemic [3] stands high due to the highly contagious nature of the disease which spreads with a rate of an exponential factor through contractual transmission, mainly body contact either through hands or any other body part. The trauma is still on and is expected to peak more in certain regions given the current growth rate and conditions. This all leads to concern that how the virus has hit hard the economies and various sector growths of the country. As per domestic rating agency Crisis [4], GDP growth forecast for the financial year 2021 will drop from 5.7 to 5.2% owing Covid-19 outbreak. A similar disruption in the economic front of other nations like the USA, Italy, China, Spain, etc. is viable and has been presented in the BBC report on its homepage [5]. Given the fact that the medicine to Covid-19 is not known right now, people need to be cautious and smart to handle the situation. We try to understand this spread dynamics with the help of a very basic model developed by Ronald Ross [6] known as the SIR model to explain the growth pattern and malaria life cycle. Since then, this has become a base model to explain the epidemic growth of various viral diseases.

✉ Rahul Saxena
2019rcp9153@mnit.ac.in

Mahipal Jadeja
mahipaljadeja.cse@mnit.ac.in

Vikrant Bhateja
bhateja.vikrant@ieee.org

¹ Malaviya National Institute of Technology, Jaipur, India

² Manipal University Jaipur, Jaipur, India

³ Shri Ramswaroop Memorial Group of Professional Colleges,
Dr. A.P.J. Abdul Kalam Technical University, Lucknow, Uttar
Pradesh, India



Covid-19 is a fairly new viral outburst and it is drastically impacting almost every aspect of human living. As a researcher, the real motivation is to explore the untouched about Covid-19 and to contribute to the mitigation of its lethal impact. Since the facts about Covid-19 are still evolving and not much is known about various aspects of the virus, the study holds its significance. It is interesting to explore the detailed information about the disease behavior in terms of its growth and other implications. The article presents a simulative study of the SIR model of Covid-19 infected patients' for the number of days. The objective of the study is to analyze the viral spread dynamics based on the mathematical modeling for the epidemic. Based on this, we can get a prediction analysis of the trend of the disease spread as well as an understanding of other related aspects.

Some of the novel aspects of the study along with the pandemic spread analysis are effects of lockdown, vaccination and herd immunity attainment, maximum possible spread, etc. We have taken the data of few states of India to understand how this disease's spread pattern conforms to the mathematical model described by Ross. The study tries to cover an interesting aspect of the effect of lockdown which is less explored in contemporary studies. We have been able to establish a strong empirical demonstration in support of the theoretical results of the SIR model that there is a significant drop in the effective contact rate as a result of lockdown. Further, we have analyzed how Covid-19 modeling differs from other epidemic disease models (modeled using similar SIR models). Our results offer realistic views regarding near-future trends and results inclination. In summary, contributions of the manuscript can be listed as:

- Modelling Covid-19 data using SIR epidemic model. Further, justifying the need to use the model in comparison to other predictive models (logistic regression).
- Analysing the impact of lockdown based on the model and observing the cut down in the contact rate and contagion.
- Determining an estimate to some other vital information that can be deduced based on the model like span of the disease, maximum range of infection, vaccination and herd immunity.
- Comparing the disease intensity based on the aspects of spread and mortality with other diseases using the SIR model.

The flow of the paper, thus, is as follows: Sect. 1 gives a general introduction to the disease, its impact and effects on the country and the need to understand its mathematical modelling. Section 2 reviews state of the art work of various epidemic models for various outbreaks. Section 3 analyzes the basic growth pattern of three states and a union territory of India using logistic regression model. The growing trend of

the data of infection spread is well matched with the logistic curve. However, the analysis over the results bring us to the conclusion that a more concrete model is needed. Section 4 presents a brief description to SIR model and its mathematical setup. The study shows that how SIR parameters fit to the growth model and how varying the model parameters can make us understand the control over epidemic. As an implication, a small investigation is done to visualize the cut down on the contact rate during the lock down period in India. Section 5 presents result and discussion of the experimental investigation done over the real data using model postulates. The section brings out further explanation, interpretation and limitations of the analysis made. Section 6 finally concludes over the results and insights gained from Sect. 5.

2 Studies on Epidemic Modelling: State of the Art Overview

The study of the area of epidemic modelling is not new and various researchers from many fields have been presenting studies related to it. The very basic models of epidemic modelling, namely SI (Susceptible–Infected) and SIR (Susceptible–Infected–Recovered) were first proposed by Sir Ronald Ross et al. [7] to simulate the transmission mechanism of malarial infection. The model consists of three simple nonlinear ordinary differential equations which do not have any explicit formula solution. Since then, there have been explorations of this basic model by relaxing certain assumptions and/or by including more parameters to be investigated. Hethcote et al. [8] discussed three epidemic models: namely SI, SIR with non-dynamic vitality and SIR model with dynamic vitality for diseases prevailing long, typically 10–20 years. The model considers the birth and regeneration of susceptibles to repeat the cycle. Andersson et al. [9] presented a study over stochastic epidemic models and their statistical analysis. Authors discussed various aspects of epidemic modelling from modelling dependency of an infectious disease based upon the infected persons in the vicinity to vaccination and herd immunity. The study opened up various areas of investigation over the proposed modelling of the epidemic under conditions of longevity, quick rate spread etc. Berge et al. [10] have tried to explain the spread pattern of the 2013 *Ebola* outbreak in Africa using the SIR model. The model was described for the endemic equilibrium, i.e. the infection dies when the susceptible die out.

Rachah and Torres [11] modelled the Ebola outbreak using the SEIR model, a variation of the basic SIR model with one more state exposed introduced in it. The authors supported the model implications by assigning values to the parameters based on real data. The results obtained were in synchrony with the theorems and lemmas of the SEIR model. Further based on the results, authors also determined

an optimized control strategy through vaccination. Sylvie et al. [12] did another investigative SEIR model-based study on Ebola. The model was based on a seven compartment deterministic model which included Susceptible, exposed, infected asymptomatic, infected symptomatic, hospitalized, recovered and dead individuals. The model treated the urban and the rural populations separately and so the epidemic control strategies differ owing to awareness through educational campaigns, vaccination, etc. Kato et al. [13], in a scientific report published in Nature, has shown that how both preventive and quarantine measures are necessary to stop the spread of the infection using the SIR model postulates.

Mkhatshwa et al. [14] used the real-time data of then, to evaluate model parameter values for SARS outbreak and then compared the results reported in the literature. Mpshe et al. [15] presented a SIR model-based study over *Zika virus fever (ZIKVF)* outbreak in November 2015 in Cape Verde. The model predicted the basic reproduction number for the virus R_0 and investigated the relative effect of the parameters on R_0 . The paper concluded that epidemic control can be attained by increasing the death rate of *Aedes* mosquitoes and the recovery rate of patients. Another study conducted by Riou et al. [16], using the model estimated values for *ZIKV* and *Chickenghunia* virus made a comparative analysis. The study took into comparison the infection spread rate apart from other factors like temperature, humidity, etc. to analyze key differences in the dynamics of the two diseases. Another popular and fatal Nipah virus outbreak in Bangladesh that prevailed every year from 2001 to 2014 with a predicted mortality rate of 76%, has been put in the study by Sultana et al. [17] and Biswas et al. [18] using the dynamics of the same SIR model. Xian-Xian Liu et al. [19] proposed SEAIRD prediction model. The epidemic control is predicted as a continuous change process, and the epidemic development and control are integrated to simulate and forecast the trends. This improved model is applied to simulate the Covid-19 epidemic in the USA, and by comparing the prediction results with the traditional SEIR model. It is found that the adaptive SEAIRD model's prediction results of the US Covid-19 epidemic data are in good agreement with the actual pandemic curve. Shinde et al. [20] presented a recent survey on the forecasting and prediction trend of Covid-19. The authors in the paper categorized the methods into four categories: (i) big data sets accessed from WHO/National agencies [21] and [22]; (ii) social media data [23] and [24]; (iii) stochastic and mathematical models [25–27]; (iv) data science and machine learning models [28–30]. However, the authors concluded that the data trend analysis depends on various factors and there cannot be a single method or technique to make the predictions in all cases. The trend analysis predictions for the infection spread were found to be predicted well-using machine learning/data science models like logistic regression, Weibull equation, etc. Further, they have analyzed the

impact of disease on people of different age groups, genders and also suggested measures for quarantine, ways for better handling of disease for people. But then, the significance of these models is limited to the prediction of the growth rate of infection. So, mathematical models simulating the epidemic dynamics are more promising.

As per the discussion above, we conclude that most of the mathematical epidemic models developed for the study of infectious outbreaks are based on the basic SIR model and its variations like SIER, SIERS, etc. There exists a model that supports the stochastic approach for mathematical model development for studying the epidemics and their effects. However, the literature predominantly witnessed the use of the basic SIR model for prediction and analysis of the outbreaks of various types of contagions like SARS, Malarial infection, EBOLA, Swine Flu, etc. Taking away the learnings from the previous models [31,32], a similar estimation-based model has been investigated for Covid-19 spread out based upon some real-time data in this article. The growth pattern of infection increase, recovery rate and the death rate have studied for five highly hit Indian regions. Based upon the information available, the model parameters have been estimated. A detailed comparison of Covid-19 with other diseases has also been drawn with the help of parameter values (taken from the literature) using the same model. This enables us to predict for the future trend. Also, the preventive measures (in the form of health care arrangements, quarantine measures, social distancing etc) can be timely plan out. Further, we also investigate the vaccination process to get an overview of futuristic action to be taken in this regard.

3 Analyzing Growth Pattern and Spread Dynamics of Covid-19

Before proceeding with the parameter evaluations based on the data for the model, let us understand that how the virus spread pattern has been and how rapid the growth for infection has been. For this, we have considered the data from few states of India where the virus anomaly is high. Based upon the data from Kaggle database repository [20], the statistics for these states till 12th July 2020 are as shown in Table 1. The statistics vary for the four states in the three measures as the rate of the infection spread and recovery differ in all four cases. Based on this data set, the infection and recovery trends are shown.

The trend follow an exponential growth pattern for the infection spread rate and recovery rate. However, the rate of progression differ depending upon various factors. Still the overall trajectory patterns follow a logistic curve given by the equation:

Table 1 State wise statistics of India as per data till 12th July 2020 [33]

State	Date of the first reported case	Total infected	Recovered	Death
Rajasthan	4/3/2020	23,748	17,869	503
Gujarat	20/3/2020	40,941	28,649	2032
Maharashtra	9/3/2020	246,600	136,985	10,116
Delhi	2/3/2020	110,921	87,692	3334

Table 2 Estimated parameter values for logistic fit defined by equation (2)

States	Parameter estimated values			
	a	b	c	d
Rajasthan	8.09916169e+04	-1.16984346e+03	3.14273130e-02	1.53266491e+02
Gujarat	1.07573457e+05	-4.42758160e+03	2.89455503e-02	1.22634955e+02
Maharashtra	3.41109262e+06	-7.58805672e+03	3.65333150e-02	1.90269963e+02
Delhi	2.31877226e+05	3.70395785e+02	6.49864588e-02	1.28021048e+02

$$y = \frac{a}{1 + e^{-c(x-d)}} + b \quad (1)$$

Here, x represents the total number of confirmed cases till 12th July 2020 and y represents the closest value fitted to the actual value based on the logistic function. Parameter a represents the value that can be reached on the basis of the fit; b defines the vertical shift from the axis; c defines the slope of the curve; d defines the horizontal shift from the axis. The initial guess values are defined as : $a=1$, $b=0$, $c=1$, $d=0$. Fitting up these values in Eq. (1), yields the equation of the most basic logistic curve :

$$y = \frac{a}{1 + e^{-x}} + b \quad (2)$$

Starting from this initial guess, parameter values are estimated for the four different state cases. This estimation is based on the curve fitting to the data set. The estimated parameter values are shown in Table 2.

The time step is one, i.e. each incremental day's data is taken into consideration. Based upon the function, the curve fitting for the actual and the predicted models is shown in Fig. 1, along with their R^2 value for each fitted model. From the plot of Covid-19 data in Fig. 1 (where observations are the actual data points), it is visible that the growth rate of the infection is nonlinear. Further, fitting the data points using a logistic regression-based model brings us to the conclusion that the growth rate is exponential. The argument is justified based on the value of the coefficient of determination (R^2) which is evaluated to be close to 1, suggests that the fit exactly mimics the growth pattern of actual data. The fit is obtained based on Eq. (1) which is an equation of the exponential growth model. The closeness between the logistic curve and the real data curve depend upon the values of logistic parameters. The selected values of logistic parameters are automatically calculated and they represent the closest logistic fit for the real data set.

The fitted curve based on Eq. 2 seems to fit the growth pattern of rising cases of Covid-19 well. The R^2 estimator value (Coefficient of Determination) is found to be close to one in all the four cases which suggests that the calculated values as per the logistic relation are conforming to the real values. Again, the point to be noted is that each considered case has a different growth rate and duration. But still, each growth pattern has almost the same rising trajectory which means that the infection pattern can be modelled based on the real-valued data pattern. Hence, the same model can be generalized to many other situations as well.

The model used to make prediction although looks suitable but cannot be generalized due to following reasons:

1. The trajectory pattern is an exponential rising curve but the curvature is not guaranteed to be smooth as witnessed. The inflexions in the curvature is due to the fluctuation in the rate of people getting infected each day. It may not be a positive increase on some days owing to various factors like better social distancing measures, good immunity, climatic conditions, less population density, etc. Thus, the model cannot be taken as a general prediction model.
2. The model only takes into consideration the rate of infection over the duration. The underlying function seeks the input as a particular day number and outputs the estimated number of infected people upto that day. Other factors like the number of people recovered, initial susceptible persons, etc. all these considerations are missing.
3. Since, currently the growth trend of Covid-19 infection in India is on a rise, the trajectory is exponential. But in general, there is no process known which always has a rising pattern and follow an exponential curve. There will be a dip in future when the number of cases becomes stagnant and the recovery rate is effectively reduced. Thus, the current model may look like a good predictive model but analytically, it is not for making future predictions.

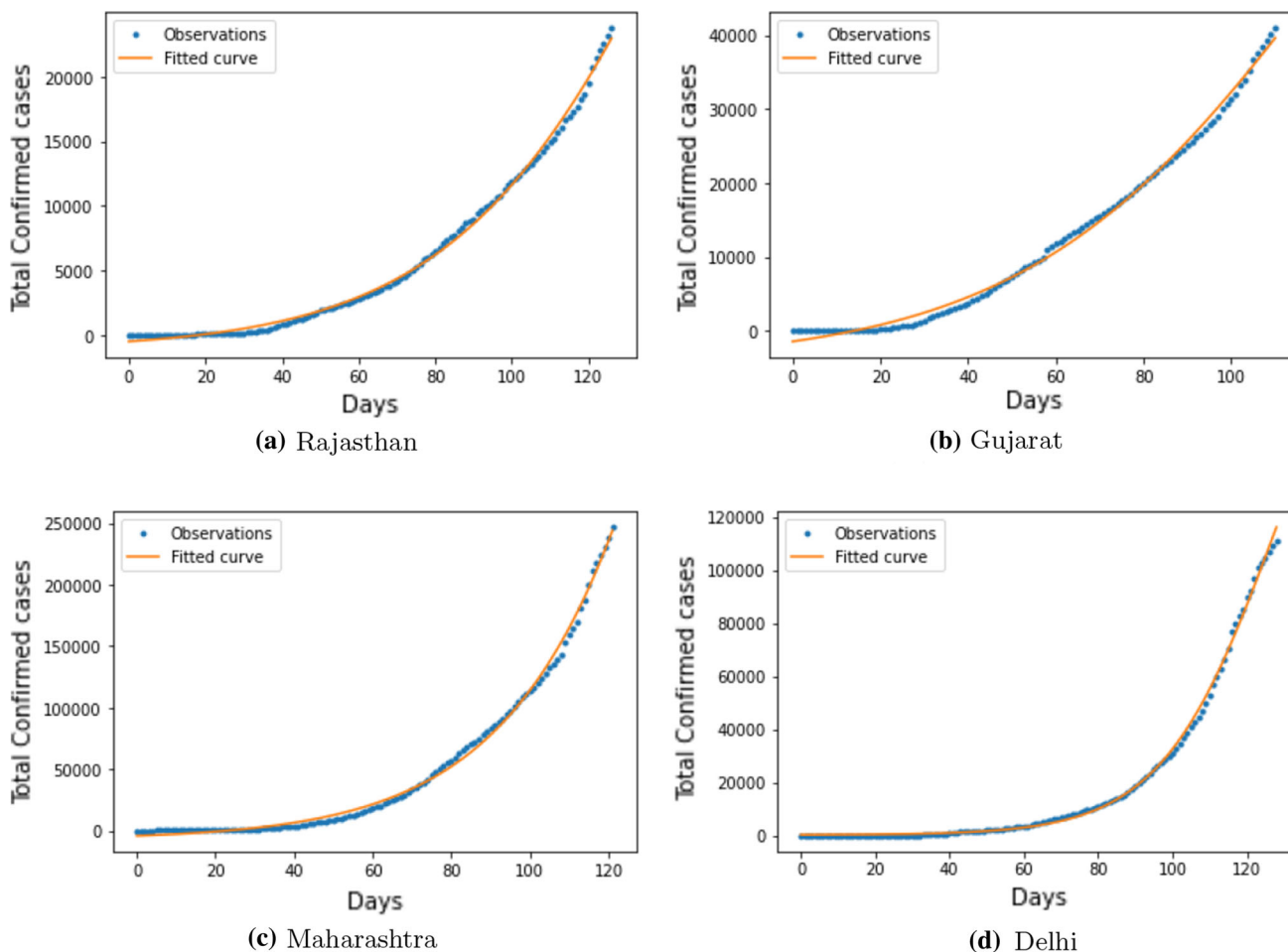


Fig. 1 Observed v/s fitted curve with R^2 values as:0.9971, 0.9975, 0.9976, 0.9981 respectively

4. Apart from growth prediction based on the trend, other information like contamination rate, how long the infection persists taking into consideration the recovered patients, the effect of the increase of susceptibles at any instance of time etc. cannot be derived or predicted based on this simple growth model.

Owing to these reasons, we require a more concrete model considering recovered and susceptible persons which makes more sense. Thus, in the next section, we re-investigate the missing aspects of the logistic curve-based model. The model discussed so far is although good enough to have a know the general growth trend of the infection.

4 SIR Model-Based Exploration of Covid-19 Data Set

This section deals with the simulation of real data using SIR model and tuning its parameters. Further, the understanding gained from the analysis has been used to measure the impact

of lockdown as an effect of lowering R_0 value. The analysis has been presented over four prominent regions of India. The analysis helps in determining the expected population proportion which is likely catch the infection. Further, it helps in visualizing the fact that how controlling the parameters of the model actually simulates to the actual curve growth.

4.1 Understanding SIR Model and Its Tuning Parameters

We define epidemic and pandemic as per the state of art literature:

- An epidemic happens when a disease spreads to a large number of people in a very short period.
- When an epidemic goes global, it is referred to as pandemic.

Given the outbreak of Covid-19, the virus spread is a pandemic and has spread across the globe affecting around 180 countries so far. The disease has a high contagious factor

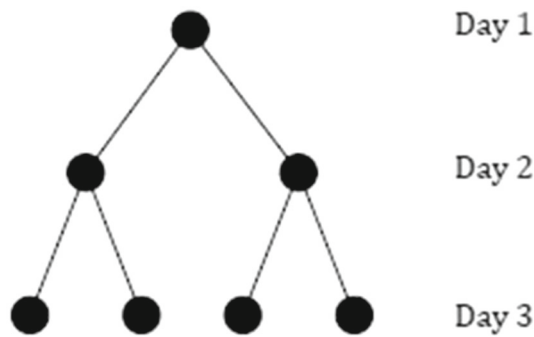


Fig. 2 Exponential growth rate of disease



Fig. 3 SIR Model Lifecycle

and thus is spreading at a very high pace. A basic simple growth model given the constant multiplication factor for the increase of infected population is shown in Fig. 2.

However, a more realistic way to model the disease is the SIR model [6], which takes into consideration a more number of factors and variables. All the other epidemic models stated (SEIR, SIS) are the derivative models of the SIR epidemic model. The SEIR model is very much similar to the SIR model with an additional stage of the *exposed* population from the susceptibles. However, in the case of Covid-19, the infection probability is so high that the transmission duration from susceptible to exposed is very much less. Thus, the exposed phase can be equated to the number of susceptibles in the population. For the SIS model, the third phase is *susceptible*, which means a person after getting infected moves to the susceptible stage. However, considering the trend of the disease to date, the possibility of cases in which a patient after a recovery gets infected is almost nil. Thus, modeling the disease based on the SIS model makes less sense in comparison to SIR. The model assumes that the people who recover from the disease will no longer be susceptible. The cycle of the disease moves as per Fig. 3.

Before discussing about the model, we briefly discuss a few related terms:

- **Effective Contact rate** It is defined as the product of the transmission rate with the contact rate. For instance, if the rate of transmission is 5% and 5 contacts per day are made, then, this implies effective contact rate = $0.05 * 5 = 0.25$.
- **Recovery rate** It is defined as the number of days required to recover from the infection. For instance, if the average

recovery period for a patient is 4 days, then, recovery rate = $\frac{1}{4} = 0.25$.

- **Reproduction Number (R_0)** It is defined as the ratio of effective contact rate to the recovery rate. The number intuitively gives an idea that how fast the epidemic spreads.

From the model dynamics, we have:

$$\frac{dI}{dt} = aSI - bI \quad (3)$$

Based upon these simple nonlinear differential equations, we estimate the model parameters. Equation (3) is a result based on the number of susceptible getting infected minus the number of infected persons recovering from the infection at time 't'. a represents transmission rate of the conversion of susceptible to infected and b represents the recovery rate. Now, investigating these equations a bit in detail, we can get few interesting facts. From Eq. (3), initial rate of change can be defined as:

$$\frac{dI}{dt_{t=0}} = aS_0b - bI_0 \quad (4)$$

The factor $\frac{aS_0}{b}$ is referred to as reproduction rate defined as R_0 . Here S_0 refers to the initial susceptible population, a refers to the transmission rate and b refers to the recovery rate. This mathematical term (R_0) is a measure to indicate how contagious the disease is. It tells the average number of people who can catch the disease from an infected person. Based on the value of R_0 , there can be three situations:

- If $R_0 < 1$, then, in this case, the infected person will infect less than one person. As a result, the virus propagation will decline and eventually die out.
- If $R_0 = 1$, then, in this case, the infected person will infect exactly one person. As a result, the disease will stay alive, but there will not be an epidemic outbreak.
- If $R_0 > 1$, then, in this case, the infected person will infect more than one person and there will be an outbreak.

The intensity of R_0 value varies depending on the viral infection strength for a disease. The R_0 value for the Spanish flu (also known as the 1918 influenza pandemic) was estimated to be in the range 1.4–2.8 [34]. Similarly, for Swine flu and H1N1 flu, the R_0 value ranges between 1.4 and 1.6. A more detailed comparative view has been presented in Sect. 5.3 of the paper. For Covid-19, R_0 value consideration is in the range 5.7–6.3 as per the study reported in [35]. This estimate has been obtained based on data from the original outbreak in Wuhan, China. In earlier studies, the value of R_0 was reported in the range 1.8–3.0, which later on was

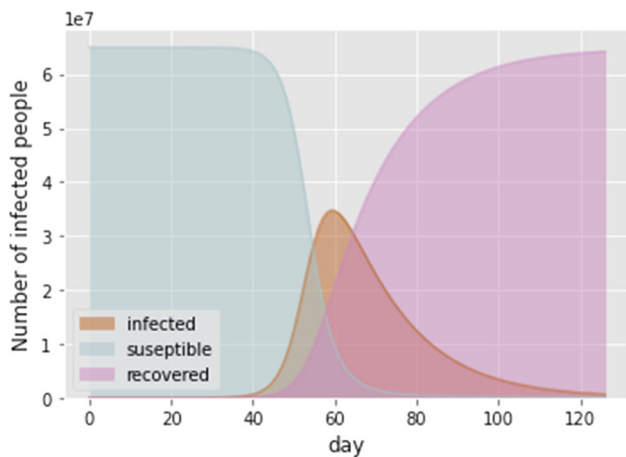


Fig. 4 Exponential growth rate of disease

discarded in [35]. So, from the perspective of being reasonable about the calculations, R_0 value consideration in this manuscript is 6.0.

The dependent factors over which value of R_0 depends are a : transmission rate and b : recovery rate as $R_0 = \frac{aS_0}{b}$, where, S_0 being a constant which is the initial susceptible population. The value for 'a' and 'b' factors range differently for different diseases. Considering the theoretical limits: $a > 0$ and $b > 0$ holds as these values can never be negative. However, when $a \gg b$, then, value of R_0 will be high which accounts for the high rate of viral spread. On the contrary, if $b \gg a$, it states that the epidemic is near to the end. Practically, studying the range of 'a' and 'b' parameters is a complex task as these factors vary region to region, country to country. Further, they are dependent on various uncontrollable parameters like social distancing measures adopted by the people, immunity of the people in the region, environmental and climatic conditions, etc. as per [36]. However, based on the analysis in [20,37], considering the recovery rate, b to be around $1/15$. This means that if each person is infectious for 15 days then we would expect that $\frac{1}{15}$ th of the people to recover each day. Based on this, the R_0 value is estimated to be in the range 1.1–3.8 or close to 4 with effective contact rate in the range 0.073–0.2533 on the lower end. Similarly, there are regions (Wuhan in China, few regions of India, Brazil etc.) where these effective contact rate was found to be in the range 0.38–0.42 making R_0 value to lie in between 5.7 and 6.3.

For the sake of feasibility, the value for recovery rate is chosen as $1/15$ based upon which R_0 is estimated to be 6.0 as per [35]. The value of R_0 may vary region-wise as well as may go down or up. The value depends upon how infection rate and recovery rates are varying. To depict this, the graph in Fig. 4 shows how a generic SIR model growth will be.

The growth model has the following values: recovery rate = $1/15$, effective contact rate = $6/15 = 0.4$. From this,

we get the value of $R_0 = 6.0$. Using these values of equation parameters, an experimental run for the model is done over a population of 1000 people. The initial spread of the disease was quite high owing to the high R_0 value but as the days pass by, the number of susceptible reduce. Further, the growth rate improves as more infected patients complete up the span of the recovery period. So, the recovery graph starts climbing up. This causes a decline in the growth rate of infected people but still aggressive enough, and dries out after a long period from its peak attainment.

4.2 Visualizing Effect of Lockdown on Effective Contact Rate

Considering the same parameter values, we get the curve for Rajasthan state as shown in Fig. 5. The study is made over 78 days till 20th May 2020 from the day when the first Corona positive patient was found. Here in the graph, we have considered three aspects related to Covid-19 growth:

- *Infected* This indicates the number of infected people predicted by the SIR model at the given R_0 value (= 6.0) indicated by the curvature of the red line in the plots from Figs. 5, 6, 7 and 8.
- *Actual Data Trend* This refers to the trend of the number of persons getting infected indicated by the curvature of the purple line in the plots from Figs. 5, 6, 7 and 8.
- *Infected_lockdown* This represents the fact that how the infection trend curve has shifted from the original SIR model-based prediction to actual. This helps in understanding that by how much, the contact rate has been cut down till the end of lockdown, indicated by the curvature of the blue line in the plots from Figs. 5, 6, 7 and 8.

Starting with the initial population estimate of 80 million people as per [38], the curvature of the red line indicates that the total count of infected people must have raised to above 10,000 before the 25th day of the seed infection as per the SIR model growth equations. However, the actual growth rate of infected cases is depicted by the *purple growth line*. Thus, we can infer that in-between something has been done to cut down the contacts. Here, let's say that S_0 is increased by a factor 's' where $s > 0$, then, $\frac{a(S_0+S)}{b} > \frac{aS_0}{b}$. And this implies that increase in number of susceptible will cause an increase in the number of infected cases. The increase will be exponential as described by Eq. (3). This situation is a direct consequence of the fact that some migrants have turned up to their hometowns during this course of time which were not the part of original population census.

To come close to the actual data growth, we took into consideration the fact that on 24th March 2020, a complete lockdown was announced in India which continued till 20th May 2020 effectively. This is the 21st day from the day,

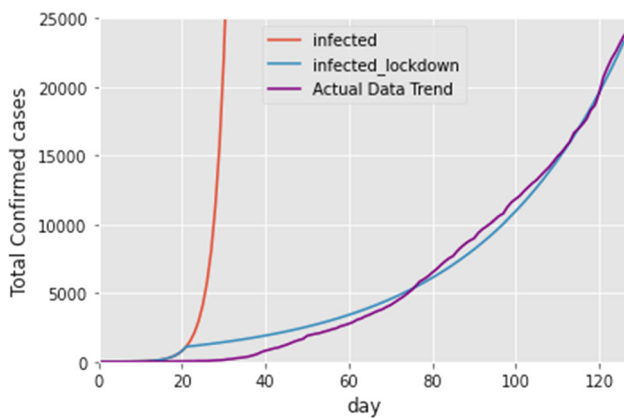


Fig. 5 Analysis curve for Rajasthan

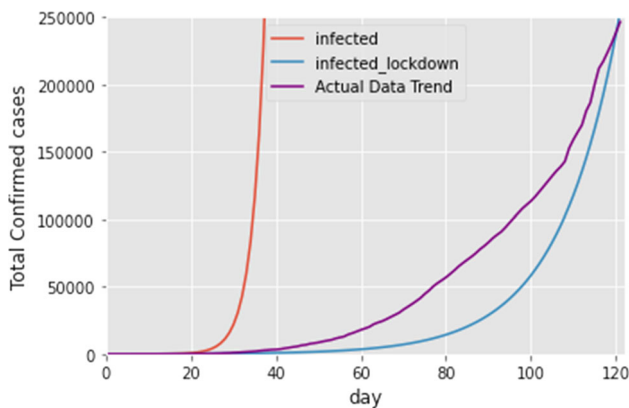


Fig. 6 Analysis curve for Maharashtra

the first case was caught in Rajasthan. This change in the trajectory is shown by the blue curve line. Till the 21st day, the blue curve followed the original model trajectory but after that the *effective contact rate* was cut down by a factor of **4.175**. Thus, the lockdown was able to suppress the growth pattern till the 78th day. This cut down was a bit closer to **4.19**, but since it is too growing exponentially at a mild pace, the overall cut down effect can be assumed to be close to 4.175. A similar kind of study is made over the datasets of Maharashtra and Gujarat shown in Figs. 6 and 7 respectively.

In the case of Maharashtra, we start with the initial population of 124.9 million as per census [39]. Here, the growth rate is quite high in comparison to the curve of Rajasthan and follows the *purple line* of curvature. The model mimics the same by following the curvature of the blue line. The curvature shows an initial growth rate at a very low rate and approximately around the 30th day, curvature starts shooting up. However, the actual growth rate as per the model will be highly acute as depicted by *red line*. This is because the imposition of lockdown which was in action after 15 days of the first positive patient was found, cut down the effective contact rate by a factor of **2.92**. With a similar explanation, for a population as per census for Gujarat [40], which stands

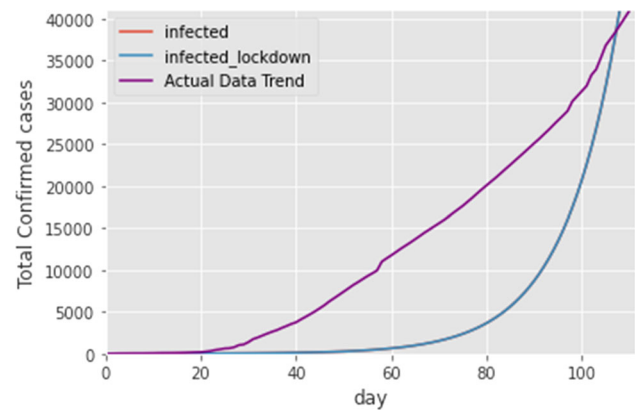


Fig. 7 Analysis curve for Gujarat

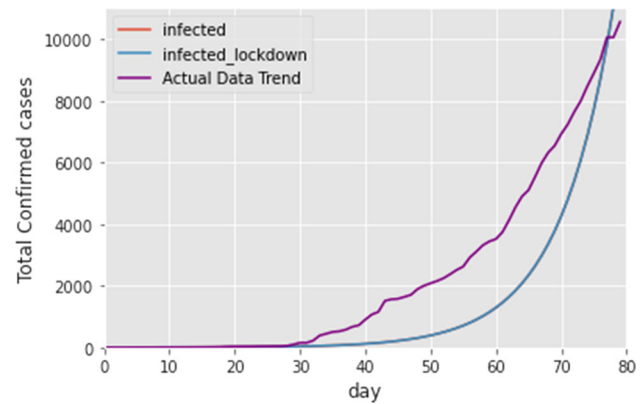


Fig. 8 Analysis curve for Delhi

out to be 65 million, the contact rate cut down post lockdown was around **2.6**. It is less than what we had for Rajasthan and Maharashtra. Due to this, total cases reported (for Gujarat) is more than 11,500 till 62nd day. The actual data growth rate denoted by the blue line is found to be above the adjusted model curve depicted by the blue line throughout the graph. This indicates that the initial reduction in contact rate was lower than the actual effective cut down which is **2.6**, which indicates the lateral effects of lockdown in Gujarat. For the given R_0 value 6.0, the effective contact rate for Delhi contracts by a factor of approximately **2.15** for a population size of 30.2 million [41]. As evident from Fig. 8, the initial outburst in Delhi was strong but soon due to lockdown, it could not spread at the rate as shown by the red line. However, post lockdown the rate at which infection is spreading is still high.

Irregular gaping for different regions depends on numerous factors. In some regions, the contagion spread started much earlier in comparison to the other regions. So the difference between the rate of infections before and after the lockdown was quite high for such regions. Another reason is that the effect of cut down in the contact rate is not the same for all the regions. This parameter is not controllable as it depends on the social distancing measures taken and fol-

lowed by the people of the underlying region. It is due to the same reason the wider gap in the latter two curves (Figs. 6, 7), the infection spread was aggressive after the initial few days of contagion. So, even after the imposition of lockdown, it took some time to cross the original prediction based on the SIR model.

The above discussion brings us to the conclusions over few points:

- The SIR model is a nonlinear differential equation-based model characterized by Eq. (4) which is again an exponential growth model. The growth trends in Figs. 5, 6, 7 and 8 based on the SIR model (purple line) merges to the actual data growth trend (blue line) owing to continuous reduction in R_0 factor due to lockdown. This merging of two curves again accounts for the nonlinearity as the actual data is again being mimicked by a nonlinear differential equation-based model.
- The growth trend does not exactly mimic the model dynamics. This is due to the fluctuating infection growth rate which is a result of vanishing susceptibility from the population at an unpredictable rate.
- As an assumption, we have kept the susceptible population constant. This metric is directly proportional to the rate of infection. We have high infected growth trend for Maharashtra due to its high population density whereas for Delhi it is low having much less population in comparison to Maharashtra. To prove it mathematically, let us consider Eq. (3), where the condition for epidemic to spread is depicted by-

$$\frac{aS_0}{b} > 1$$

Here, let's say that S_0 is increased by a factor 's' where $s > 0$, then,

$$\frac{a(S_0 + s)}{b} > \frac{aS_0}{b}$$

- This implies that an increase in the number of susceptible will cause an increase in the number of infected cases. The increase will be exponential as described by Eq. (3). This situation is a direct consequence of the fact that some migrants have turned up to their hometowns during this course of time which were not the part of original population census.
- Further, the above scenarios also bring us to the conclusion about the tuning parameter. Solving equations involve three tuning parameters a, b and S . Here, a is defined as the *effective contact rate*, b is *recovery rate* and S is the *number of Susceptibles*

- Now, here if 'S' is assumed to be constant as no vaccination has been done to the population against the viral epidemic, complete population is the contender for getting infected from virus.
- Above point implies that now the epidemic spread depends on two factors, i.e. a and b , and further equation gets modified to:

$$\frac{a}{b} > 1$$

Since the recovery rate is almost fixed i.e. 15 days, the only factor now governing the equation is *effective contact rate*. If the effective contact rate reduces, the equation inequality starts trending towards a value *less than 1*. It was due to this fact that the lockdown imposition in all the cases above helped to control the epidemic and it did not burst out at a rate as shown by the standard SIR model.

- As an implication of the model as the contact rate starts flattening or going down, the recovery rate starts shooting up as more and more infected people gets recovered in comparison to the susceptibles converting to infected.
- The virus spread is not significantly dependent on the size of the total population. It depends on the initial infected population and the R_0 value for the disease. This fact can be illuminated in the study by looking at the infection growth rate curve for Delhi. The initial spread-out rate of infection in Delhi was much more in comparison to Rajasthan, Gujarat, and Maharashtra (depicted in Fig. 8). It exactly mimicked the SIR model estimates, given the fact that the population of Delhi is very less in comparison to other states. So, even if the population is dense, given the low contagion factor of the disease and the initial low number of infected persons, the epidemic outbreak will not be severe.

So far the discussion is how actual data fit defines SIR model and what inferences can be drawn due to the variations from the trend. In the next section, we further analyze the SIR model.

5 Results and Discussion

After gaining insights from Sect. 4, we focus on various aspects of Covid-19 pandemic in this section. We estimate the value of maximum infected people for the four selected Indian regions. We also estimate parameters related to end of the pandemic. We compare Covid-19 with other diseases. In addition to this, estimates with respect to vaccination process

Table 3 Predicting I_{\max} based upon the R_0 values

State	Population Count (Approx.)	R_0 value before lockdown	I_{\max} estimate based on original R_0 value	R_0 value after lockdown	I_{\max} estimate based on new R_0 value
Rajasthan	80 million	6.0	42.77 million	$\frac{6.0}{4.175} = 1.437$	4.146 million
Maharashtra	124 million	6.0	66.303 million	$6.0/2.92 = 2.054$	20.17 million
Gujarat	65 million	6.0	34.755 million	$6.0/2.6 = 2.30$	13.20 million
Delhi	30.2 million	6.0	16.148 million	$6.0/2.15 = 2.79$	8.26 million

and herd immunity have also been made for Covid-19. The estimated calculations are based upon the nonlinear differential equation model defined for SIR (as per Eq. (3)). Applying basic calculus concepts and by simple curve analysis, these estimates can be obtained from the defined differential equation model.

5.1 Estimating Range of Covid-19 Pandemic Spread

Considering Eq. (3) in Sect. 4, we have,

$$\frac{ds}{dt} = -aSI \text{ and } \frac{dI}{dt} = aSI - bI$$

After simplification, we obtain,

$$\frac{I_{\max}}{N} = 1 - \frac{1}{R_0} (1 + \log R_0) \quad (5)$$

This expression defines the maximum fraction of people which will be infected due to the virus spread. The whole equation is now in terms of a single variable R_0 . For the feasibility purpose, we have considered R_0 value to be constant over the entire period of investigation. Considering the fact that imposition of lockdown has brought down R_0 , we can make the estimates for maximum infected population portion as per Table 3, assuming the recovery rate is fixed.

These estimated values may slightly differ from the realistic data. This is due to the model assumptions like susceptible count equals the total number of population, effective contact rate and recovery rate are kept constant. Estimated values of I_{\max} are expected to change if any of these parameters vary in midst. This is quite likely to happen during the course of time as due to awareness better social distancing measures will cut down effective contact rate. This will lower the number of susceptibles getting converted to infected. In the meantime, if vaccine for the disease gets developed, number of susceptibles will be reduced by a significant fraction. Availability of treatment drug to cure the disease is another factor which is currently missing in the scenario which will increase the recovery rate. Thus, exact implication of the model's result gives us an idea of the aggressiveness of the growing trend. In addition, it also indicates that which regions need

more immediate attention given the current infection growing trend. This is evident from Table 3 that in Gujarat more immediate attention is required as compared to Rajasthan despite of its low population (Fig. 9).

5.2 Estimating Time Span of Covid-19 Spread

As the pandemic will proceed, the number of susceptible starts reducing from the total population. From the condition obtained in Eq. (4), monotonically increasing curve of infected population attains its peak at the condition $S = b/a$. After this point, the infected population curve declines and the trajectory follows the curve shown in Fig. 4 for the SIR growth model. Thus, at $time \rightarrow \infty$, $b \gg a$, which implies that $S \approx b$. The meaning of the statement is that almost all infected persons have recovered. At this point of time, the number of infected persons are very less and the expression for $R_0 = a/b \ll 1$, which means that epidemic is near to the end.

From this discussion, we come to an interesting conclusion that epidemic end does not depend on the absence of susceptibles in the population rather it depends on the fact that there are no more infected people. In the other words, towards the end of the pandemic, we can say that the recovery rate is significantly higher than the infection rate. This is another fact that can be illuminated mathematically and cannot be discerned from the data.

5.3 R_0 Value-Based Comparison of Covid-19 with Other Epidemics

As evident from the facts and analysis that contagion capacity of Covid-19 is high (nearly equal to 6.0), we analyze its growth as compared to some other epidemic virals. As per [42], reproduction number value is as shown in Table 4.

Figure 10 is an experimentation done on a sample of population of 5,000 people. Upper limit of R_0 value is considered for each virus, i.e. 18 for Measles, 1.08 for SARS and so on. For the simplified analysis, recovery rate is fixed as 15 days for all the diseases as it turns out to be nearly equal (15 days) for all the diseases under the consideration. Based on these parameter values, growth of various diseases as per the

Fig. 9 Graphical Plot for I_{max} estimation

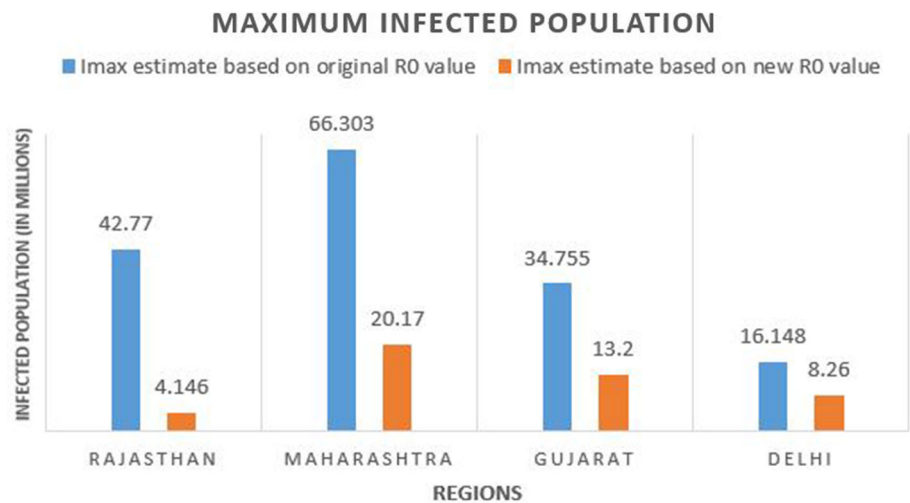


Table 4 Disease with their respective R_0 value ranges

Disease	R_0 value
Measles	12–18
SARS	0.19–1.08
Ebola	1.5–1.9
Nipah [43]	0.1–0.4
Influenza (1918 Pandemic)	1.4–2.8
MERS	0.3–0.8

SIR Model is shown in Fig. 10. Due to high R_0 value of measles, it shoots quickly and almost 90% of the population gets infected. For all other diseases, growth trend and range is not as aggressive as measles. However, the growth is still quite high for Corona in the context of a large and densely populated country like India.

The main reason of concern while comparing the Covid-19 pandemic with previous epidemics and pandemic Influenza is that they have measures for cutting down the number of susceptibles (in the form of herd immunity and vaccination). So the peak attainment for disease like measles will not be the same as shown in the figure and through vaccination, it can be tackled efficiently. Facts from [44] suggests that how herd immunity and vaccination will not allow a wild-fire spread-out of the disease. Containment is possible and soon the disease will die out owing to these handling tools accompanied with safety measures. Death rate in the case of Covid-19 is though not as aggressive as Ebola [45] and Nipah [46], but the fact that these diseases do not have such high contamination rate, makes them less severe as compared to Covid-19. The absence of known treatment measures and herd immunity attainment is still far for Covid-19. Therefore, the overall impact of Covid-19 is still more severe in comparison to the diseases like Nipah and Ebola. This discussion highlights the importance of vaccination and herd

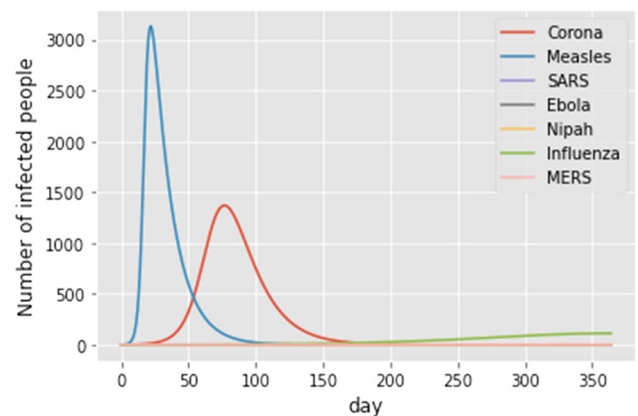


Fig. 10 SIR Model-based simulation of diseases

immunity more, no matter how aggressive the death rate and contamination is. Through vaccination, growing rate of infected individuals can be contained by saving a portion of population from the strike of the disease. Further, with the progression of time, herd immunity also helps (anti body development against the disease).

5.4 Vaccination and Herd Immunity

Considering the fact that even if the vaccine is 100% effective, it is very expensive to vaccinate the entire population. Moreover, there may be a portion of population that has a compromised immune system and cannot be vaccinated. So, is there a way by which we can ensure the pandemic spread control by vaccination of some percentage of the population? The answer to this question is yes and it can be easily inferred from SIR model analysis. The relevant concept is referred to as herd immunity. Let us assume that ' p ' be the fraction of population that gets vaccinated. Assuming that vaccine is 100% effective, the number of susceptibles vaccinated will

Table 5 Predicting population to be vaccinated

State	Population count	Current R_0 value after lockdown	Population to be vaccinated $(1 - \frac{1}{R_0}) * \text{Population count}$
Rajasthan	80 million	1.437	24.32 million
Maharashtra	124 million	2.054	63.63 million
Gujarat	65 million	2.30	36.73 million
Delhi	30.2 million	2.79	19.37 million

be $p \cdot S_0$ will be removed from S_0 . Now, remaining number of susceptibles will be $(1 - p) \cdot S_0$. To prevent epidemic spread, we now have the equation of R_0 , defined as:

$$(1 - p) \cdot S_0 \frac{a}{b} \leq 1$$

This further implies that,

$$(1 - p) \frac{a}{b} \leq 1$$

From mathematical investigation in Section 4, $R_0 = \frac{aS_0}{b}$, thus, at equilibrium, we have,

$$(1 - p) = \frac{1}{R_0}$$

and, finally, we have,

$$p = 1 - \frac{1}{R_0} \quad (6)$$

Equation (6) finally gives the expression for the fraction of population that needs to be vaccinated (p). The answer p directly depends upon the R_0 value. With the fluctuation in R_0 value(s), the corresponding p value(s) will change accordingly. Considering the current R_0 values that sustains to the rate of growth of infection as discussed above, we have the estimates for the states as per Table 5.

One more consideration that is to be taken into account is vaccine effectiveness. For example, if vaccine is 90% effective, then, the fraction of population that needs to be vaccine is increased by the factor $\frac{1}{0.9}$. For example, if R_0 value is 2.0 and total susceptible population is 1000, then, if the vaccine is fully effective, then, fraction of population needs to be vaccinated will be $1000/2.0 = 500$. But if the vaccine is 90% effective, then, we need to vaccine $500/0.9 = 556$ people. Thus, herd immunity and vaccination also depend upon how effective the vaccine is.

5.5 Limitations and Scope

We have made several assumptions to analyze evolving Covid-19 data for the feasibility purpose. The basic assumptions of the SIR model also hold, i.e. three stages S , I , and R

sums up to the total population, R_0 parameter is kept fixed, the initial susceptible population is equated to the total population. The analytical results of the model are dependent on R_0 value of the disease. The R_0 value consideration for the disease is based on the transmission rate which eventually depends upon the contacts made and its frequency. The reported R_0 value for Covid-19 (in the range 5.7–6.3) as per [35] considers various biological, socio-behavioral and environmental factors which govern the transmission of virus [36]. However, there cannot be a direct measure for estimation of R_0 value and is reported based on some mathematical observations. With respect to our reported simulations and analytical results, the impact of the R_0 value makes more sense instead of how the R_0 value is being determined (based on the contact, multiple exposures, etc.).

Further, a few facts can only be explained mathematically and can not be noticed through data. The investigation may involve the inclusion of some more model parameters like susceptibles may only be considered as working officials, some more parameters like temperature conditions may be included in the model. In future, the study can further be extended by inclusion of some more model parameters like temperature conditions or by doing precise calculations of existing parameters like number of susceptibles.

6 Conclusions

The paper investigates the SIR model-based current Covid-19 spread pattern and its growth in four prominent regions of India. The study encompasses various aspects of the disease spread pattern and a comparative analysis with few other disease models of the past. The study demonstrates the progression of real-time data by the SIR model with high accuracy. Though the model holds few assumptions, still it is good enough to validate a few facts and get reasonable predictions related to the rise and fall of the growth rate related parameters of the infection. This is evident from our experimental analysis over the real-time Covid-19 data. The gap between the theoretical explanations and experimental results enable us to find the reasons for deviations.



References

1. Coronavirus (COVID-19) outbreak, WHO Homepage: <https://www.who.int/westernpacific/emergencies/covid-19>. Accessed 16 Nov 2020
2. WorldOmeter Homepage: <https://www.worldometers.info/coronavirus/>. Accessed 17 Nov 2020
3. Bulletin of World Health Organization, WHO Homepage: <http://www.who.int/bulletin/volumes/89/7/11-088815/en/>. Accessed 18 Nov 2020
4. Quartz India Homepage: <https://qz.com/india/1827925/how-will-coronavirus-impact-indias-economy-as-per-moodys-fitch/> Accessed 19 Nov 2020
5. BBC news Homepage: <https://www.bbc.com/news/business-51706225>. Accessed 20 Nov 2020
6. Dworkin, J., Tan, S.Y.: Ronald ross delamater (1857–1932): discoverer of malaria's life cycle. *Singapore Med. J.* **52**(7), 466–467 (2011)
7. Anderson, R.M.: Discussion: the kermack-mckendrick epidemic threshold theorem. *Bull. Math. Biol.* **53**(1–2), 1 (1991)
8. Hethcote, H.W.: Three basic epidemiological models. In: *Applied Mathematical Ecology*, pp. 119–144. Springer (1989)
9. Andersson, H.; Britton, T.: *Stochastic Epidemic Models and Their Statistical Analysis*, vol. 151. Springer Science & Business Media (2012)
10. Berge, T., Lubuma, J.M.-S., Moremedi, G.M., Morris, N., Kondrasha, R.: A simple mathematical model for ebola in africa. *J. Biol. Dyn.* **11**(1), 42–74 (2017)
11. Rachah, A.; Torres, D.F.M.: Analysis, simulation and optimal control of a seir model for ebola virus with demographic effects. Preprint [arXiv:1705.01079](https://arxiv.org/abs/1705.01079) (2017)
12. Djimbanjankou, S.D., Nyabadza, F.: An optimal control model for ebola virus disease. *J. Biol. Syst.* **24**(01), 29–49 (2016)
13. Kato, F., Tainaka, K.-I., Sone, S., Morita, S., Iida, H., Yoshimura, J.: Combined effects of prevention and quarantine on a breakout in sir model. *Sci. Rep.* **1**, 10 (2011)
14. Mkhathshwa, T.; Mummert, A.: Modeling super-spreading events for infectious diseases: case study sars. Preprint [arXiv:1007.0908](https://arxiv.org/abs/1007.0908) (2010)
15. Mpeshe, S.C., Nyerere, N., Sanga, S.: Modeling approach to investigate the dynamics of zika virus fever: a neglected disease in africa. *Int. J. Adv. Appl. Math. Mech.* **4**(3), 14–21 (2017)
16. Riou, J., Poletto, C., Boëlle, P.-Y.: A comparative analysis of chikungunya and zika transmission. *Epidemics* **19**, 43–52 (2017)
17. Sultana, J., Podder, C.N., et al.: Mathematical analysis of nipah virus infections using optimal control theory. *J. Appl. Math. Phys.* **4**(06), 1099 (2016)
18. Biswas, M.H.A.; Haque, M.M.; Duvvuru, G.: A mathematical model for understanding the spread of nipah fever epidemic in bangladesh. In: *2015 International Conference on Industrial Engineering and Operations Management (IEOM)*, pp. 1–8. IEEE (2015)
19. Liu, X.-X.; Fong, S.J.; Dey, N.; Crespo, R.J.; Herrera-Viedma, E.: A new seaird pandemic prediction model with clinical and epidemiological data analysis on covid-19 outbreak. *Appl. Intell.* 1–37 (2021)
20. Shinde, G.R., Kalamkar, A.B., Mahalle, P.N., Dey, N., Chaki, J., Hassanien, A.E.: Forecasting models for coronavirus disease (covid-19): a survey of the state-of-the-art. *SN Comput. Sci.* **1**(4), 1–15 (2020)
21. Teles, P.: Predicting the evolution of sars-covid-2 in portugal using an adapted sir model previously used in south korea for the mers outbreak. Preprint [arXiv:2003.10047](https://arxiv.org/abs/2003.10047) (2020)
22. Russo, L.; Anastassopoulou, C.; Tsakris, A.; Bifulco, G.N.; Campana, E.F.; Toraldo, G.; Siettos, C.; Tracing DAY-ZERO. forecast- ing the fade out of the covid-19 outbreak in lombardy. Italy: a compartmental modelling and numerical optimization approach. *medRxiv* (2020)
23. Li, L., Yang, Z., Dang, Z., Meng, C., Huang, J., Meng, H., Wang, D., Chen, G., Zhang, J., Peng, H., et al.: Propagation analysis and prediction of the covid-19. *Infect. Dis. Model.* **5**, 282–292 (2020)
24. Lai, S.; Bogoch, I.I.; Ruktanonchai, N.W.; Watts, A.; Lu, X.; Yang, W.; Yu, H.; Khan, K.; Tatem, A.J.: Assessing spread risk of wuhan novel coronavirus within and beyond china, january-april 2020: a travel network-based modelling study. *MedRxiv* (2020)
25. Magal, P.; Webb, G.: Predicting the number of reported and unreported cases for the covid-19 epidemic in south korea, italy, france and germany. Italy, France and Germany (March 19, 2020) (2020)
26. Victor, A.: Mathematical predictions for covid-19 as a global pandemic. Available at SSRN 3555879 (2020)
27. Wang, H.; Zhang, Y.; Lu, S.; Wang, S.: Tracking and forecasting milestone moments of the epidemic in the early-outbreak: framework and applications to the covid-19. *F1000Research*, 9 (2020)
28. Batista, M.: Estimation of the final size of the second phase of coronavirus epidemic by the logistic model. *medrxiv* (2020)
29. Hu, Z.; Ge, Q.; Li, S.; Jin, L.; Xiong, M.: Evaluating the effect of public health intervention on the global-wide spread trajectory of covid-19. *medrxiv* (2020)
30. Jia, L.; Li, K.; Jiang, Y.; Guo, X.; et al.: Prediction and analysis of coronavirus disease 2019. Preprint [arXiv:2003.05447](https://arxiv.org/abs/2003.05447) (2020)
31. Cooper, I., Mondal, A., Antonopoulos, C.G.: A sir model assumption for the spread of covid-19 in different communities. *Chaos, Solitons & Fractals* **139**, 110057 (2020)
32. Nesteruk, I.: Estimations of the coronavirus epidemic dynamics in south korea with the use of sir model. Preprint.] *ResearchGate* (2020)
33. COVID-19 Corona Virus India Dataset, Homepage: <https://www.kaggle.com/imdevskp/covid19-corona-virus-india-dataset?select=complete.csvbyDevakumarKp>. Accessed 25 Nov 2020
34. Healthline, Homepage : <https://www.healthline.com/health/r-nought-reproduction-number#meaning>. Accessed 11 April 2021
35. Sanche, S., Lin, Y.T., Xu, C., Romero-Severson, E., Hengartner, N., Ke, R.: High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **26**(7), 1470–1477 (2020)
36. Delamater, P.L., Street, E.J., Leslie, T.F., Yang, Y.T., Jacobsen, K.H.: Complexity of the basic reproduction number (R_0). *Emerg. Infect. Dis.* **25**(1), 1 (2019)
37. Gupta, D.; Biswas, D.; Kabiraj, P.: Covid-19 outbreak and urban dynamics: regional variations in india. *GeoJournal* 1–19 (2021)
38. Rajasthan Population, Homepage: <http://www.populationu.com/in/rajasthan-population>. Accessed 20 Nov 2020
39. Maharashtra Population, Homepage: <https://www.census2011.co.in/census/state/maharashtra.html>. Accessed 13 June 2020
40. Gujarat Population, Homepage: <http://www.populationu.com/in/gujarat-population>. Accessed 22 Nov 2020
41. Delhi Population, Homepage: <http://www.populationu.com/in/delhi-population>. Accessed 22 Nov 2020
42. Basic Reproduction Number, Homepage: https://en.wikipedia.org/wiki/Basic_reproduction_number. Accessed 24 Nov 2020
43. Global Virus Network, Homepage: <https://gvn.org/update-on-the-nipah-virus-outbreak-in-kerala-india/>. Accessed 24 Nov 2020
44. Visualizing the History of Pandemics, Homepage: <https://www.visualcapitalist.com/history-of-pandemics-deadliest/>, by Nicholas LePan, 14/03/2020, Accessed 24 Nov 2020
45. Science Daily, Homepage: <https://www.sciencedaily.com/releases/2019/04/190415081956.htm>. Accessed 24 Nov 2020
46. World Health Organization, Homepage: <https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease>. Accessed 24 Nov 2020

