Ramakrishna Mission
Vivekananda Centenary College, Rahara

Project Work On

Mathematical modeling and forecasting the COVID-19 pandemic in India

Name: Atanu Das

Guidance: Dr.Phonindra Nath Das

Examination roll no. : 2021141035

Registration no: A01-1112-223-002-2019

College Roll no. : 02

Class: M.Sc.4th semester

Paper name: MTM CC-303

Mathematical modeling and forecasting the COVID-19 pandemic in India

ATANU DAS

Ramakrishna Mission Vivekananda Centenary College,Rahara,Kolkata-700118,India

Department of Mathematics

Keywords

Abstract

COVID-19 epidemic is declared as the public health emergency of international concern by the World Health Organisation in the last week of March 2020. This disease originated from China in December 2019 has already caused havoc around the world, including India. The first case in India was reported on 30th January 2020, with the cases crossing 4 million on the day paper was written. This pandemic has caused more than 80,000 fatalities with 3 million recoveries. Strict lockdown of the nation for two months, immediate isolation of infected cases and app-based tracing of infected are some of the proactive steps taken by the authorities. For a better understanding of the evolution of COVID-19 in the world, study on evolution and growth of cases in India could not be avoided. To understand the same, one of the compartment model: Susceptible-Infectious-Quarantined-Recovered (SIQR) is used. Recovery rate and doubling rate of the total reported positive cases in the country had crossed 75% and 25 days, respectively. It is also estimated that there is a strong positive correlation between testing rate and detection of new cases up to 6 million tests per day. Using the SIQR modelling effective reproduction number, epidemic doubling rate and infected to quarantined ratio is determined to check the temporal evolution of the pandemic in the country. Effective reproduction number that was at its peak during first half of the April is gradually converging to 1. It is also estimated using this model that with each detected cases in India, there could be 10–50 undetected cases. Like every mathematical model, this model also has some assumptions. To make this model more robust, a technique with weighted parameter that can avoid a person with a strong immune system to be equally vulnerable to the infection, can be worked out. Machine learning algorithms can also be used to train our model with the data of other countries to make the analysis and prediction more precise and accurate.

E-mail address:dasa90552@gmail.com

ATANU DAS

1. Introduction

Coronavirus disease 2019 (COVID-19) is originated from the Wuhan city of the Hubei province (China) in December 2019. This outbreak has caused global pandemic with more than 28.5 million positive cases and 0.92 million deaths. India, with a population of more than 1.3 billion, has witnessed its first positive case on 30th January 2020. At present more than 4 million individuals are reported positive with 77,461 fatalities and 3 million recoveries [1]. Government of India has imposed strict nationwide lockdown in four phases, starting from 25th March, of more than two months ending on 31st May. Also, the country had initiated the gradual unlockdown under different phases that include the opening of industries, railways, metros, shopping malls, schools and colleges. In the tally of the number of positive cases, India is at the second number just behind the USA and is reporting more than 90,000 cases each day [1]. Despite the huge number of positive cases per day, deaths per million of the population is at the lowest (56 per million of the population) [1]. A detailed mathematical analysis for the evolution of the pandemic in the country has the potential to assist medical researchers, policy makers and other stake holders. In this work, analysis of the COVID-19 cases in India assisted with a mathematical modelling is performed with cases up to 22nd August 2020.
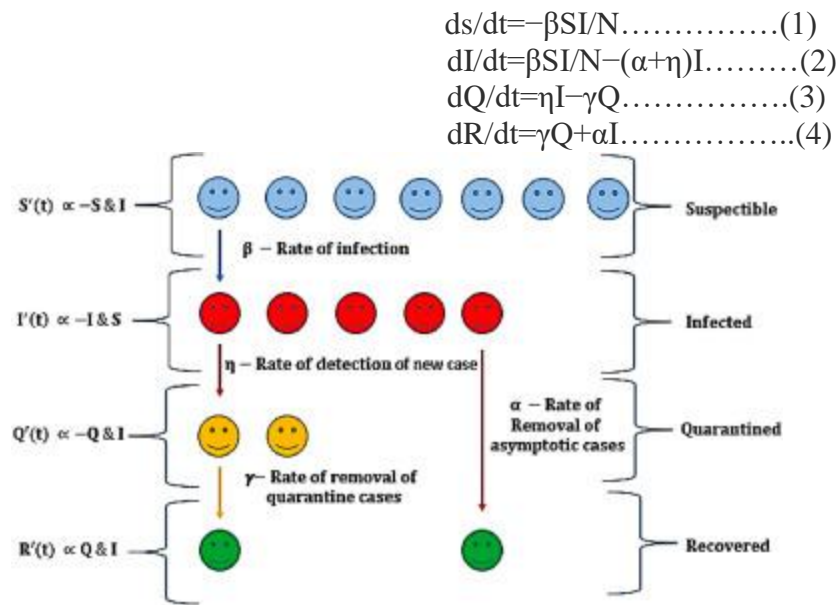
Mathematical equations are widely used to model the nature and impact of global pandemics in the society. The SIR model [2] is the classically adopted mathematical model to analyse and predict the evolution of a disease. Its one of the variant SIQR [3] is considered to be the best modelling technique for COVID-19, where isolation of infectious plays an important role. SIQR technique can be used to determine the parameters that quantify the growth and evolution of cases in a region using the past trend. This modelling approach to analyse the impact of this disease is performed for few affected nations, including Brazil [4] and Italy [5].

Particularly about India, studies regarding power-law growth [6] and the effect of lockdown on the spread of disease [7] have been reported. However, none of the papers has quantified the evolution of the disease in terms of epidemiological parameters. Present work has analysed the evolution of COVID-19 in India using SIQR model. Parameters and indicators that quantify the temporal evolution and growth of the disease are calculated.

2. Modelling approach

Susceptible Infected Quarantine Recovered (SIQR), a variant of classical SIR model is proved to be particularly convenient for the modelling of COVID-19 [5]. This model considers two categories for infected individuals, one who gets quarantine and others who don't (asymptotic or negligence). Susceptible in the models are the individuals who are at the risk of getting infected. Infected are those susceptible who gets affected by the virus; they may be asymptotic or have symptoms. Infected individuals who develop signs and gets isolated is considered as quarantine. Recovered are those infectious or quarantine individuals who recovered or died from the disease.

Each component of this model (S, I, Q, R) [Fig. 1] can be framed as the set of ODE using four proportionality constants (α, β, γ and η) [3]:

$$ds/dt = -\beta SI/N \ldots\ldots\ldots\ldots(1)$$
$$dI/dt = \beta SI/N - (\alpha+\eta)I \ldots\ldots(2)$$
$$dQ/dt = \eta I - \gamma Q \ldots\ldots\ldots\ldots(3)$$
$$dR/dt = \gamma Q + \alpha I \ldots\ldots\ldots\ldots..(4)$$

1. Fig. 1. Pictorial representation of the dependence of each component of the SIQR model, where S stands for Susceptible, I for infected, Q for quarantined and R for recovered. Left side of the picture represents the dependence of the rate of change on S, I, Q, R. Four proportionality constants relating each parameter of the model is mentioned.

In the equations mentioned above: β denotes the rate of infection, η determines the rate at which new cases are detected from the infected population. γ is the rate at which quarantines are getting removed (recovered or died). α is the rate of removal of infectious individuals who are asymptotic and didn't get quarantined (for any reasons). N is the total size of the population. 'Flattening of the curve' a term which is widely used during this pandemic is nonetheless about optimization of these four rate constants. By imposing lockdowns, authorities are trying to minimise the rate of infection (β) [Fig. 2]. In India during the lockdown, authorities have tried significantly to minimise the rate of infections and while the gradual opening focus is on to increase the number of tests per million in population to maximise η.
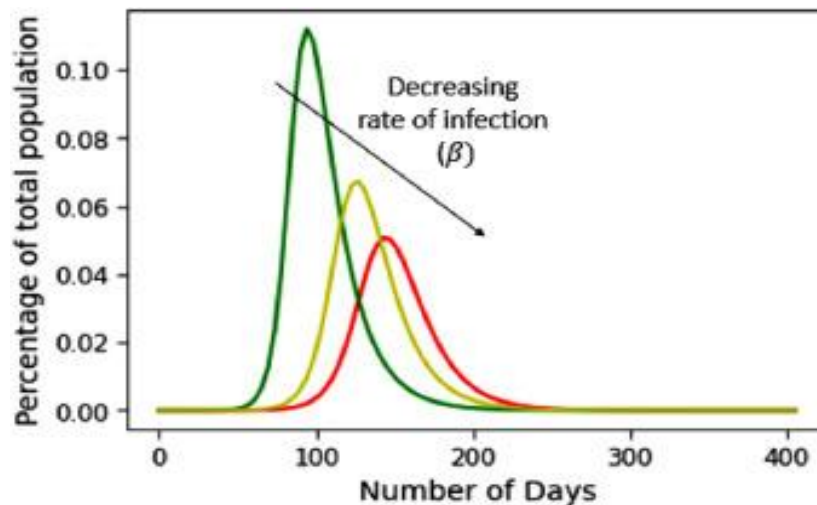
Fig. 2. Reduction of rate of infection (β) lead to the flattening of the curve of the active cases (Q) in a population. This graph is obtained using the numerical integration of Eqs. (1)–(4). Such models, no doubt, is used frequently to study the impact of large scale pandemics and proven to be beneficial for medical researchers and policymakers. However, like every mathematical modelling tool, this also works on a few assumptions. SIQR model considered that everyone in the population has an equal probability of getting infected.

   Government future steps to control the pandemic is always uncertain like weekly lockdown, or partial opening in the future. Due to uncertainty in the future actions by authorities, it is challenging to make the future prediction of this pandemic using such mathematical tool which depends heavily on the previous data. For the same reason this work has concentrated on the analysis of the evolution of the infection.

3. **Results and analysis**

This section of the paper is broadly divided into two parts: analysis of the available data and modelling of the pandemic to obtain pandemic parameters. Active cases (Q) in India has crossed 0.6 million and active cases per day have touched 20,000 during the start of last week of July [Fig. 3]. It is interesting to note that after the last week of July, active cases per day is mostly less than 10,000 cases (and seems to be going down). If this trend continues, that may correspond to peak or some interim plateau. Doubling rate of the total positive cases on n^{th} day is defined as the number of days required for the case on n^{th} day to get double. From 4 days of doubling rate during the first week of March to 26 days in the last week of August, doubling rate is improved significantly [Fig. 4].

   Recovery rate, i.e. defined as the ratio of the number of cases recovered on n^{th} day to the total active cases on the same day. Recovery rate is increased from around 10% in the first week of March to more than 75% in the last week of August (Fig. 4). India, during the initial days of the pandemic, was blamed for testing less among its population. Correlation coefficient between reported positive cases (Q+R) and the test per million [9] is determined [Fig. 3(c)]. High positive correlation coefficient between positive cases on a day with the test on the same day is as

expected. However, it is interesting to note, with more than 6 lakhs test per million this coefficient comes down to 0.72. Based on this, it can be concluded that the test of 6 lakhs per day would be sufficient to detect most of the infections.

Fig. 3. (a) Reported active cases (Q) and active cases per day in India from 15th March to 22nd August 2020. (b) Doubling rate of the
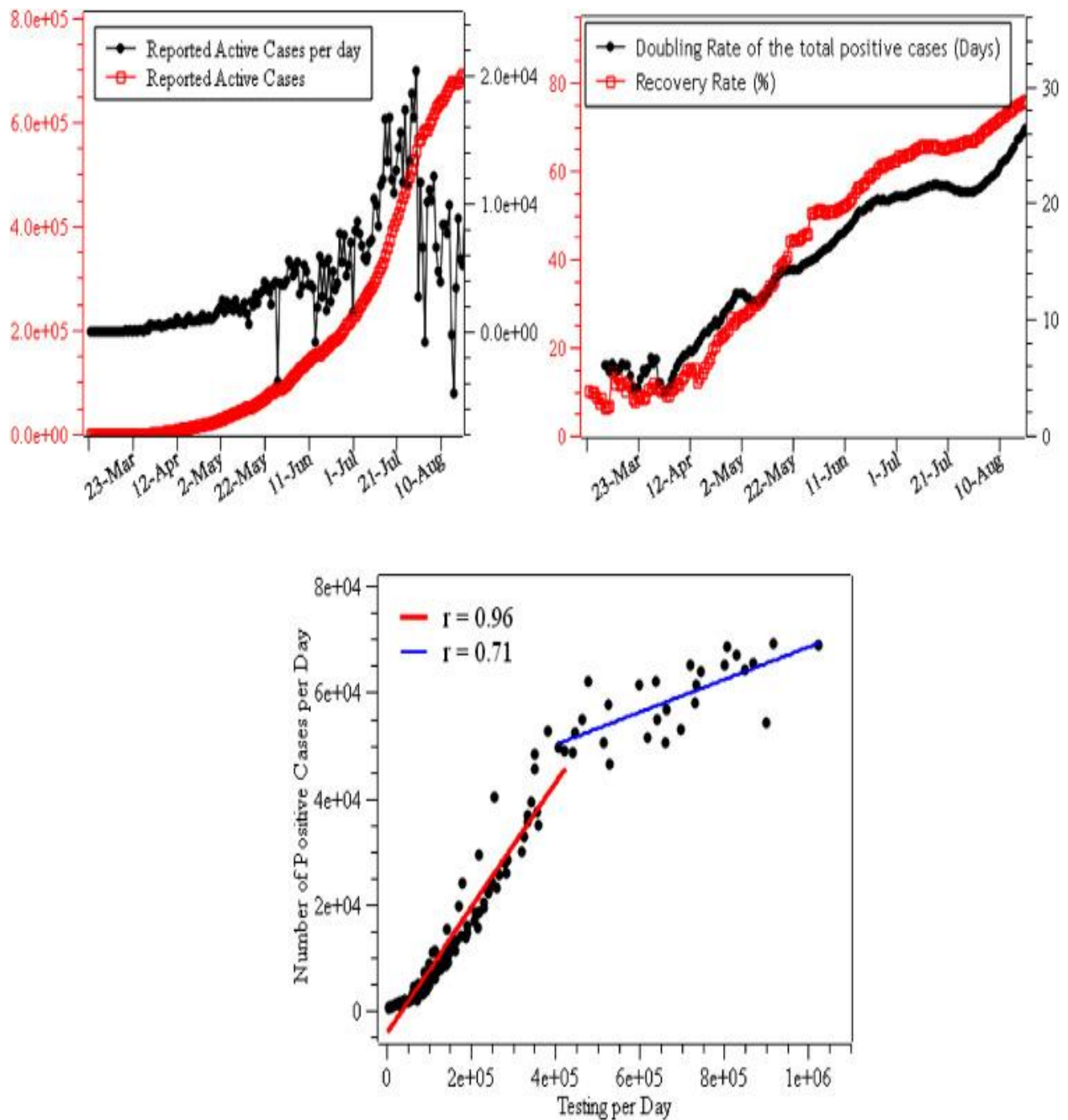
total reported positive cases (days) and recovery rate (%) from the reported data of India from 15th March to 22nd August 2020. **(c)Determination of correlation coefficient between number of positive cases per day and the tests per day.**
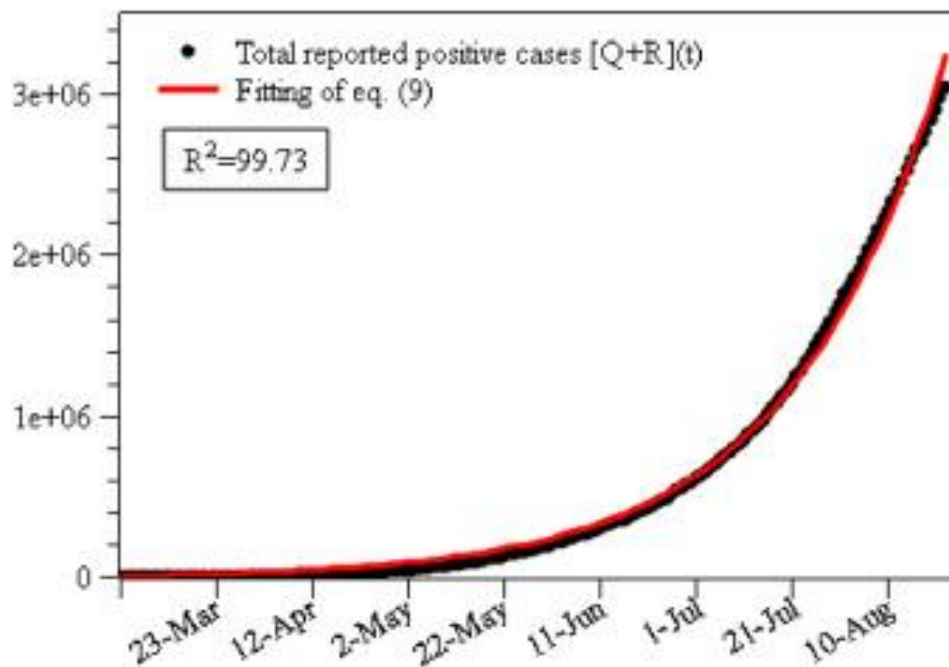

Fig. 4. Fitting of Eq. (9) to the total reported positive cases. Fitted parameters a and b is obtained with errors.

Modelling of the available data with the SIQR technique is done to get the parameters that quantify the growth of a pandemic. Eqs. (1–4) is simplified and fitted with the reported data of the country to calculate parameters. First positive case in the country was reported on 30th January 2020 with almost no growth till 4th March. For the analysis purpose, cases from 4th March is used. Considering the whole population to be susceptible, i.e. S≈N and approximating Eq. (2) to(5)

$$dI/dt=[β−(α+η)]I\ldots\ldots\ldots\ldots(5)$$

Integrating Eq. (5) to obtain

$$I(t)=I_0exp[β−(α+η)]t\ldots\ldots.(6)$$

I0 is the number of infected individuals at the start of the disease in the country. For modelling purpose, this study has considered the number of initial cases as I0=29, since cases were almost stagnant before that.

Adding Eqs. (3) and (4) will give us the rate of change of the total confirmed positive cases [Q+R] in the country with the day (t).

$$d[Q+R]/ d(t)=(α+η)I(t)\ldots\ldots\ldots\ldots (7)$$

Putting I(t) from Eqs. (6) to (7) will give us

$$d[Q+R]/d(t)=(\alpha+\eta)I_0\exp[\beta-(\alpha+\eta)]t\ldots\ldots\ldots-(8)$$

Integrating Eq. (8) over t

$$[Q+R](t)=\{(\alpha+\eta)/\beta-(\alpha+\eta)\}\times I_0[\exp\{\beta-(\alpha+\eta)\}t-1]\ldots\ldots\ldots-(9)$$

Above equation (Eq. 9) is in the form of $y(t)=AB[\exp(Bt)-1]$ where $y(t)$ is the number of confirmed positive cases on a day (t).

This form of equation is fitted to the total positive reported cases at every eighth day (starting from 15-March) using least square fitting (Fig. 4) to give data of fitting parameter $(\alpha+\eta)I_0$ and $\beta-(\alpha+\eta)$ with errors. Twenty-one fitted parameters at regular interval of 8 days starting from 15th March till 22nd August is calculated to get β and $\alpha+\eta$ for the calculation of parameters defined in the later sections.

Determination of the rate of detection of new cases from infectious individuals (η) is not straight forward due to insufficient data of asymptotes in the infectious Indian population. However, a range of value of (η) from the studies based on the incubation period with the assumption on a fraction of infectious population is widely used [4,5].

Rate of detection of new cases from infectious individuals depends not only to the incubation period but also to the fraction of infected individuals getting positive and quarantined just after the incubation period [5]. On average, it takes 6.93 days for an infectious individual to get symptoms in India [10]. If that individual gets quarantined on that day only, this rate can be written as $\eta=16.93\times\mu$, where μ is the fraction of the infected individuals that get quarantined. It has been reported that in Japan 50% of the population is asymptotic [11]. However, no similar studies can be found for India. In this case three different value of μ is considered to get the worst and best-case scenario. It is assumed that either 10%, 30% or 50% of the total infected population will get quarantined, i.e. μ can be considered as 0.1, 0.3 and 0.5 that corresponds to η=0.144×0.1(0.3and0.5)=0.014 (0.043 and 0.072).

Rate of removal of quarantine cases (γ) from the reported positive cases is calculated based on the data of the number of reported positive cases and number of removed (death + recovered) from the reported cases. γ is calculated based on the following equation.

$$\gamma=(Q_i-Q_{i-1})/R_{i-1}\ldots\ldots\ldots\ldots(10)$$

Here, Q_i and Q_{i-1} is the total number of reported quarantine cases on a day and its previous day, respectively. R_{i-1} is the total number of removed cases on the previous day. Average of last 8 days have been considered for the value of γ (with error) on a particular day [Fig. 5].
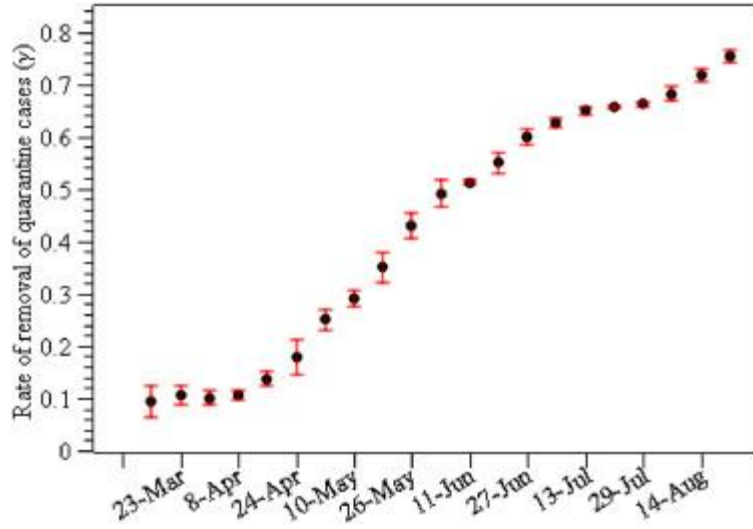
Fig. 5. Plot of rate of removal of quarantine cases (γ) (with error bound) at every eighth day starting from 15th March using Eq. (10).

All these model parameters are used to determine indicators that quantify the transmissibility of disease in the country [4]. These indicators are defined as follows, and its derivations are discussed in the appendix:

Effective reproduction number (Reff): This number is used to quantify the transmission ability of a disease. It is defined as the average number of individuals that can get infected from a single individual. It is formulated as follows:(11)

$$R_{eff}=\beta\alpha+\eta$$

Temporal evolution of effective reproduction number in India plotted in Fig. 6. It can be noted that up to the last week of April, Reff is on the higher side (higher than 2). From May, it started gradually decreasing but it is still greater than 1. This means a single person on an average is affecting more than 1 person. As it is known, for a pandemic to die down, this number should be less than 1.
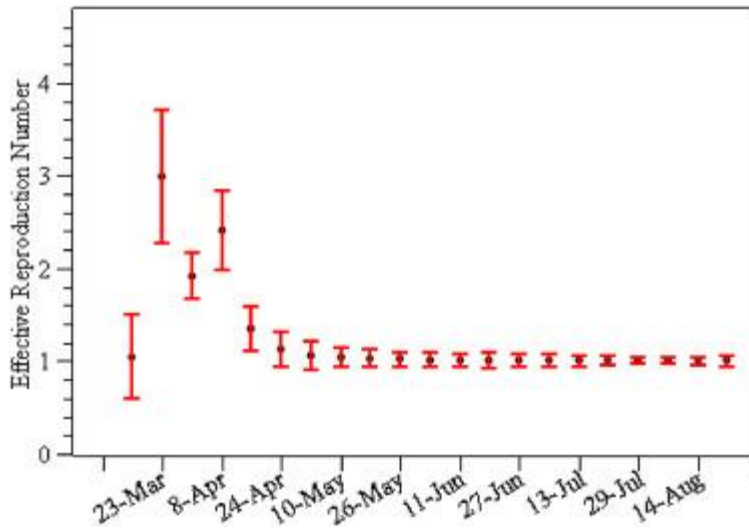
Fig. 6. Temporal evolution of effective reproduction number at every eight day from 15-March to 22-March with error bounds.

b

Epidemic doubling rate: Number of days required for a disease to double its infected population is termed as epidemic doubling time. It is different from the plot reported in Fig. 3(b) that is doubling of the reported positive cases (not the infected population). Epidemic doubling time can be given by following equation:

$$(12)\tau=\ln2[(\alpha+\eta)(R_{eff}-1)]$$

Doubling time from less than 5 days in the second week of March is gradually increased to 23 days during the last week of August [Fig. 7].

a

Infected to Quarantined ratio: This ratio gives us the estimate of the population which are infected but not quarantined. Infected population may be asymptotic or suffering from mild symptoms that get unnoticed. As the calculation of this parameter depends inversely on the calculation of η, values corresponding to three different value of η has been considered.(13)

$$I/Q=\{\gamma+\beta-(\alpha+\eta)\}/\eta$$
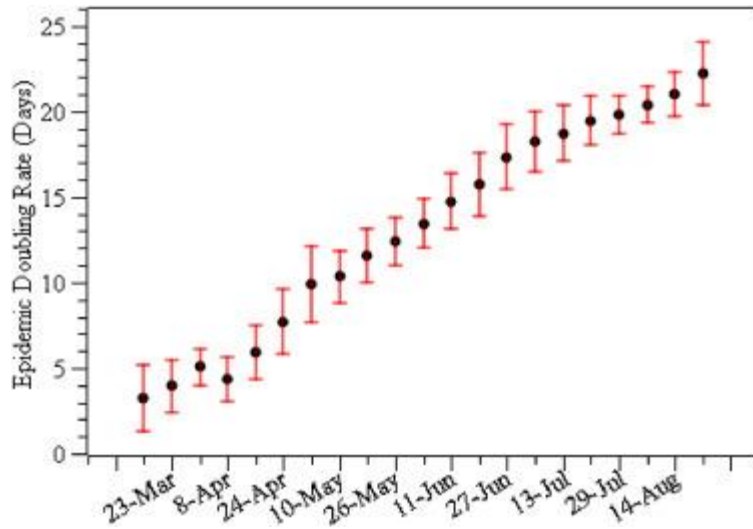
Fig. 7. Epidemic doubling rate determined from Eq. (12) is plotted at every eighth day starting from 15th March.

This ratio depends inversely on the rate of detection of new cases from the infected population. Higher the testing rate lower would be this ratio. In recent findings from national serosurvey May–June 2020 [12], it is reported that around 6.4 million are infected in total by early May 2020. However, in the same period, total reported active cases (Q) is about 50,000 (mid-May). It can be said that with every reported cases there were around 100–150 cases that went undetected in the population. Mathematically I/Q ratio determines a similar ratio. It can be deduced from Fig. 8 that at the end of August 2020, in the worst-case scenario there would be 50 times more cases then the reported active case (Q) and on the best side, this ratio could be 10.
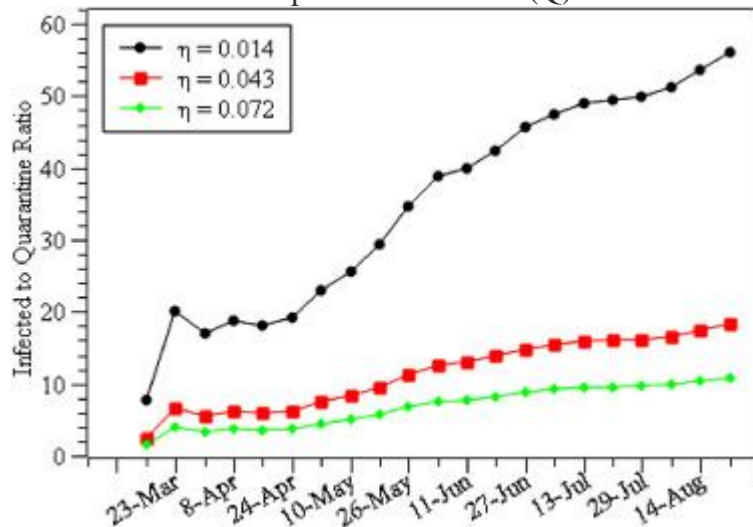
Fig. 8. Infected to Quarantined ratio determined from Eq. (13) at three different rate of detection of new cases (η) that are 0.014, 0.043 and 0.072 which corresponds to the assumption of 10%, 30% and 50% (respectively) of the infected population will get quarantined and others will get unnoticed (asymptotic).

Effective reproduction number, Epidemic doubling rate and Infected to quarantined ratio discussed and calculated in the previous section can be useful to retrace the growth of pandemic

in the country. Although the country's doubling rate is increasing, high effective reproduction number and infected to quarantined ratio confirms the spreading nature of the virus in the country. Prediction of the peak can be made using the numerical integration of Eqs. (1)–(4), but the focus of this paper is to discuss the method to determine the parameters that can be useful for medical researchers. Also, the prediction of peak done by many researchers in the past [7] has gone wrong. Accurate prediction of peak is challenging because of the difficulty in modelling the unpredictable government policies with the limited parameters in mathematical models.

4.Model prediction and some preventive measures

In this part, we shall explore the model prediction in Indian aspect and seek the preventive measure to control COVID-19 in India. The model parameter d denotes the proportion of population who maintain the social distance from other (or stay at home in safe). According to our estimation, 67.38% of the population maintain the social distance from other people in the lockdown period.

In Fig. 9, we presented the model predicted total out break size and the peak prevalence during the pandemic varying the value of d for the entire time period of prevalence. It is clear from the figures that if 80% people follow the lockdown Model effect, the disease is easy to control. Thus, to control the disease the administration should be strict to impose the lockdown properly. Since, as per our model prediction only 20% or less people does not follow the lockdown then disease may be controlled. So only the administrative, food and medicine supply persons may go outside breaking the lockdown.
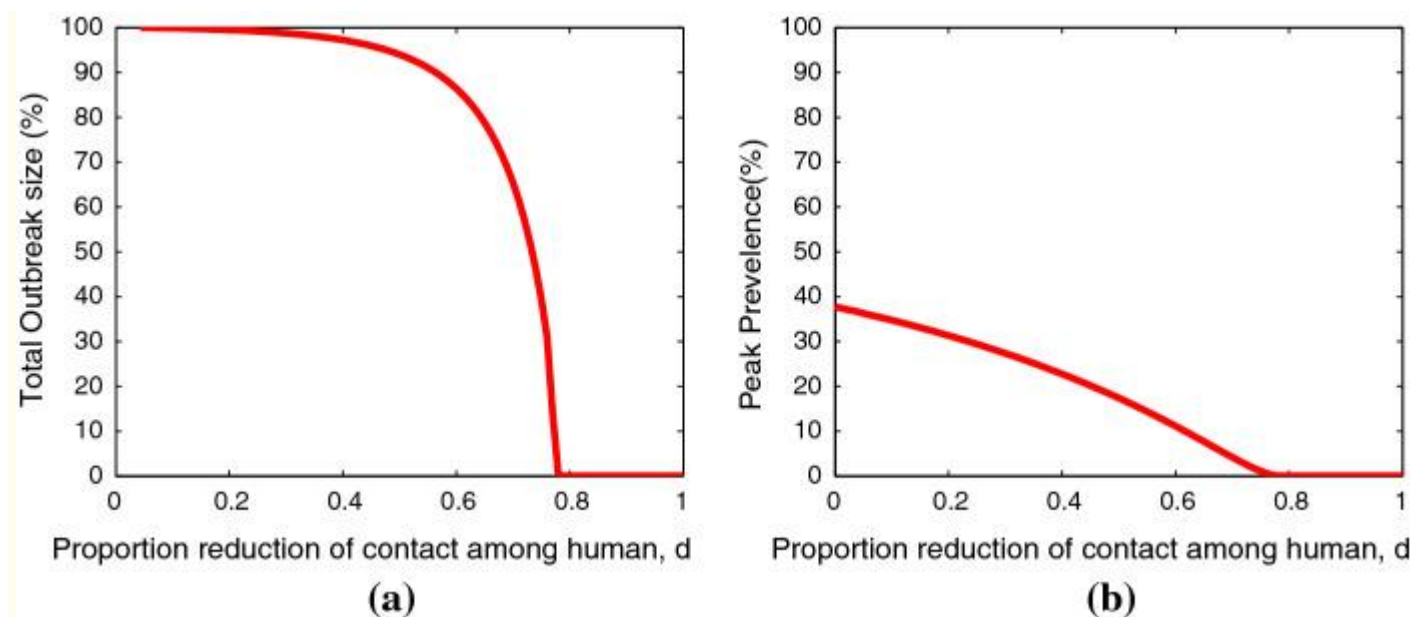

Fig. 9

a Total outbreak size and b Peak prevalence during the pandemic predicted by the model for the prevention programs on reducing proportion of social distance among human

5.Prevention measure on asymptomatic class

Due to high population density, in general human to human contact rate is high in India. On the other hand, an asymptomatic infected human has no symptom, but he/ she is able to transmit the virus to other susceptible human. So susceptible person will not take or take in late (after diagnostic test) prevention measure to protect from the virus transmission from an asymptomatic infected person. In this context, our finding shows that approximately 45% of the infected population are asymptomatic. The time series of different infected populations show that the number of daily asymptomatic infected population is larger than other infected population. The contribution on basic reproduction number of asymptomatic infected class is larger than other compartment. So protection of the virus transmission from asymptomatic infected class is a challenging problem for India like highly populated country.

In order to solve this problem, we propose a prevention policy that reduces contact rate between asymptomatic infected human and susceptible human. In this context, first of all we have to identify the asymptomatic infected human among the population and then keep them in quarantine. To identify asymptomatic infected human, we have to focus on the diagnostic test among major portion of population, specially in the affected area and quarantine them.

Let p with $0 \leq p \leq 1$ be an effectiveness of this prevention policy (that is the reduction in contact rate among asymptomatic infected human and susceptible human), applying such policy causes the following transformation in the model: $\rho \rightarrow (1-p)\rho$. The impact of such prevention (p) on total outbreak size presents in Fig. Fig.10a,10a, and impact on peak prevalence presents in Fig. Fig.10b.10b. Our finding suggests that enhancing p by at least 72% the total outbreak size will be reduced from 94.65% to nearly 2.772%, whereas the peak prevalence reduces by 18%.
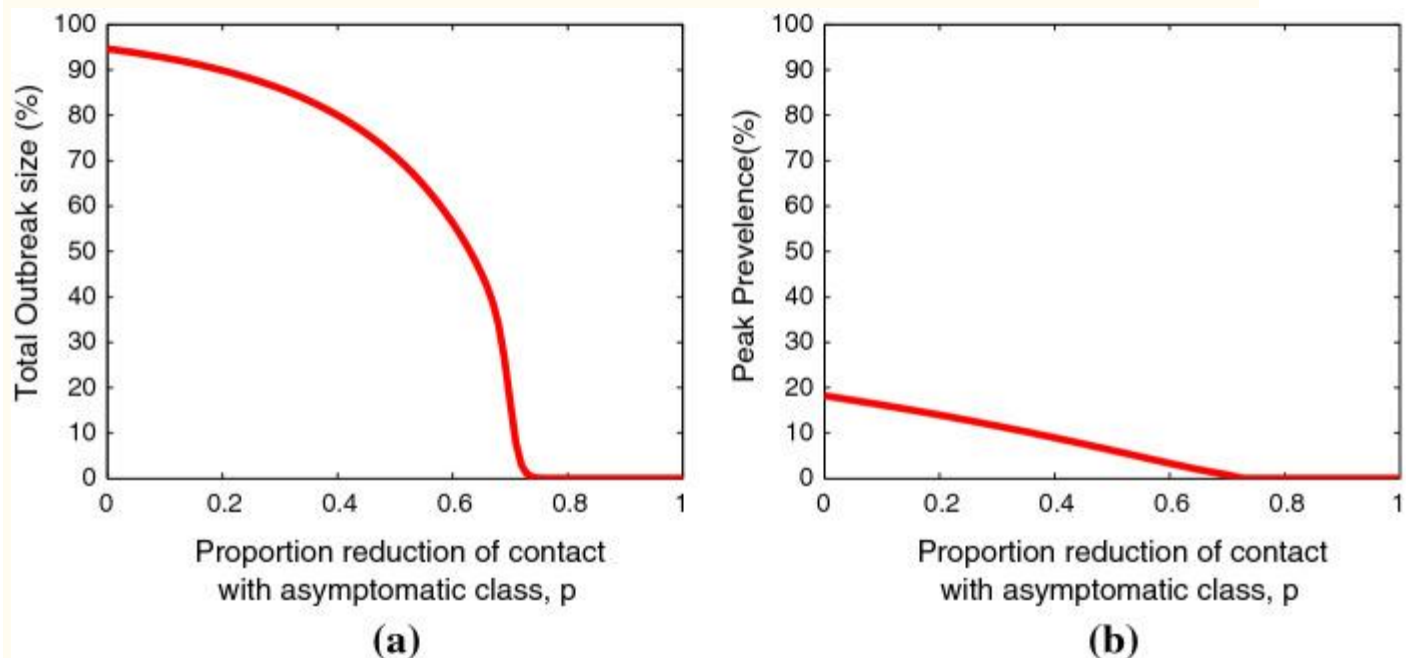

(a) (b)

Fig. 10

a Total outbreak size and **b** Peak prevalence during the pandemic predicted by the model for the prevention programs on reducing proportion of social distance among human

6.Prevention programme quarantined, isolation and rapid hospitalisation

In the real field of Indian scenario, we have observed that many staffs of quarantine and isolation centre (hospital) are infected by the quarantined and isolated infected person. In the model, q1 and q2 represent the effectiveness of quarantine and isolation, respectively. For perfect quarantine and isolation, q1=1 and q2=1 but real information suggests that q1<1,q2<1. So we have to focus on perfect quarantine and isolation and our model suggests that by increasing q1 from 0.94 to 1.0 and q2 from 0.9 to 1.0 simultaneously the total outbreak size reduces by 2.19% and this fact is presented graphically in Fig. Fig.11 a.
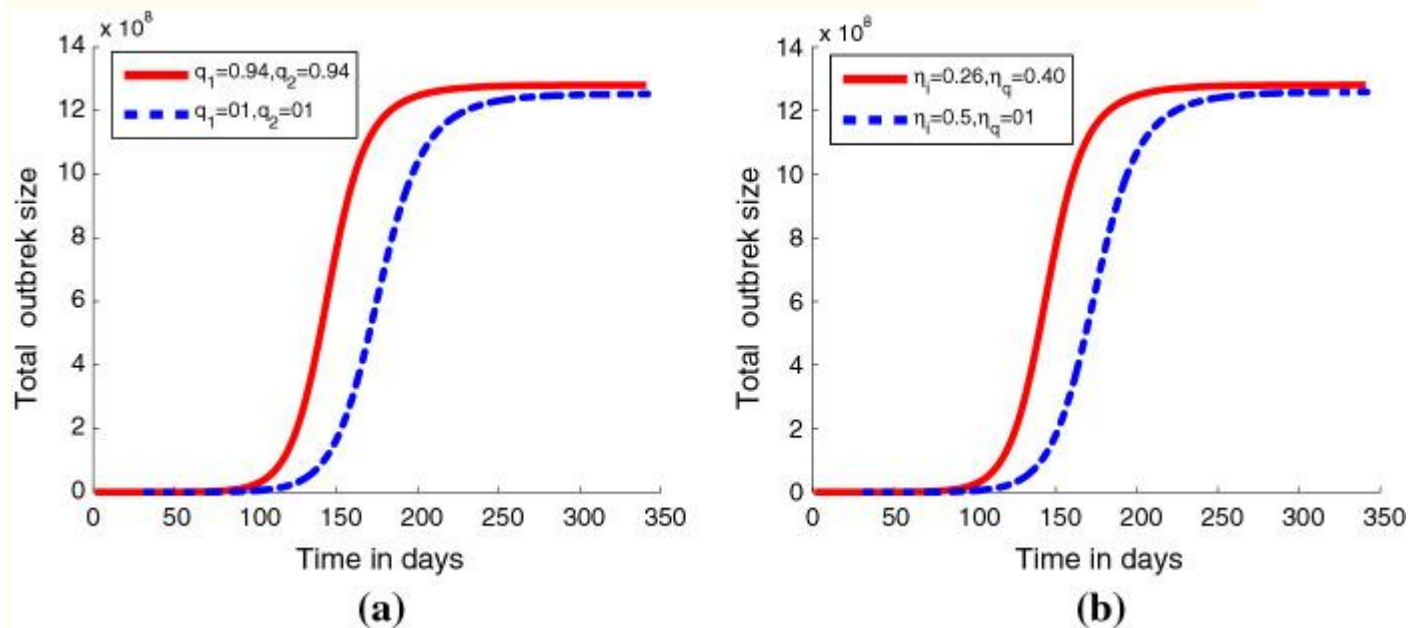

Fig.11

Time series of total outbreak size for **a** original parameter (solid line) and increase values (dash line) of q1 and q2 and **b** original parameter (solid line) and increase values (dash line) of η1 and η2

Our another observation is that according to our estimation the rate of hospitalisation from non-quarantine infected (η_i) and quarantine infected classes η_q are 0.26 and 0.40, respectively. We have to increase these rate to control the COVID-19 outbreak. If we increase these rates up to 0.5 and 1.0, respectively, then the total number of infected population will be decreased by 2% whose graphical presentation is given in Fig. Fig.10b.10b. By both the above said preventions, the outbreak can also be delayed.

7. Conclusion

Susceptible-Infectious-Quarantined-Recovered (SIQR) model is used in this paper to estimate the parameters that can be used to quantify the temporal evolution of COVID-19 in the country. Effective reproduction number, epidemic doubling rate and infected to quarantined ratio is studied in this paper. India has successfully reduced the reproduction number from 2–3 to 1 by the last week of August, but it is still more than 1. In the worst-case scenario, there could be 50 times more infected cases with each reported active case. Serosurveys published from some part of the country has reported that there can be 150 undetected infections with evert detection. Epidemic doubling rate is improved significantly to more than 20 days. This mathematical modelling technique no doubt has been proved to be essential for the researchers, but it has limited applicability with some assumptions. This model can be used at state or district level to check the temporal evolution of pandemic. Weighted parameters can be added to the existing SIQR model to overcome the assumption of equal weightage of infection to each one in the population. Models can be trained with the data of different nations [13] using Machine Learning algorithms to make predictions accurate.

Our findings suggest that the dramatical reduction in total outbreak size and peak prevalence is possible by increasing d (the proportion of number of population who maintain safe distance from each other) in a certain level and p (the proportion of number of asymptomatic population who are detected and isolated from the other susceptible population) in a certain level. We also have to increase the detection (on the basis of diagnostic test)and hospitalize rate to a certain level.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Appendix

A1. Derivation of effective reproduction number (Eq. 11) -

A disease is said to be spreading if with time number of infections are increasing i.e. dI/dt>0. From Eq. (5):

$$dI/dt = [\beta - (\alpha + \eta)]I >$$

$$\Rightarrow [\beta-(\alpha+\eta)]>0$$
$$\Rightarrow \beta\alpha+\eta>1$$

Consider;

$$\beta\alpha+\eta=R_{eff}$$

So, for a disease to spread $R_{eff}>1$

A2. Derivation of Infected to Quarantined ratio (Eq. 13) –

Eq. (3) can be re-written as: $dQ/dt+\gamma Q=\eta I$
Here, γ and ηI are function of time.
Above equation is linear ODE of the form

$$dy/dx+P(x)Y=Q(x)$$
$$IF=e^{\gamma t}$$
$$\int d/dt(e^{\gamma t}Q)=\int e^{\gamma t}\eta I$$

Putting I from Eq. (6) in the above equation:

$$\int d/dt(e^{\gamma t}Q)=\int \eta I_o e(\beta-\alpha-\eta+\gamma)t$$
$$\Rightarrow e\gamma tQ=\eta I_0 \times e^{(\beta-\alpha-\eta+\gamma)t}/\beta-\alpha-\eta+\gamma$$
$$\Rightarrow Q=\eta \times I_o \times e^{(\beta-\alpha-\eta)t}/\beta-\alpha-\eta+\gamma$$
$$\Rightarrow Q(t)=\eta \times I(t)/\beta-\alpha-\eta+\gamma$$
$$\Rightarrow \gamma+\beta-(\alpha+\eta)/\eta=I(t)/Q(t)$$

References:

[1]
Worldometer, https://www.worldometers.info/coronavirus/ [Last accessed on 12.09.2020].
Google Scholar

[2]   R.M. Anderson, R.M. May

Infectious Diseases of Humans
Oxford University Press, Oxford (1991)
Google Scholar

[3]
H. Hethcote, M. Zhien, L. Shengbing

Effects of quarantine in six endemic models for infectious diseases
Math. Biosci., 180 (1–2) (2002), pp. 141-160, 10.1016/S0025-5564(02)00111-6
ArticleDownload PDFView Record in ScopusGoogle Scholar

[4]
N. Crokidakis

Modeling the early evolution of the COVID-19 in Brazil: results from a Susceptible – Infected – Quarantined- Recovered (SIQR) model
Int. J. Modern Phys. C (2020), 10.1142/S0129183120501351

Google Scholar

[5]

Pedersen, M.G. and Meneghini, M., Quantifying undetected COVID-19 cases and effects of containment measures in Italy, preprint 2020, available on-line at https://www.researchgate.net/publication/339915690.
Google Scholar

[6]

M.K. Verma, A. Asad, S. Chatterjee

COVID-19 epidemic: Power law spread and flattening of the curve
Indian Natl. Acad. Eng., 5 (2020), pp. 103-108, 10.1007/s41403-020-00104-y
CrossRefView Record in ScopusGoogle Scholar

[7]

Singh, R. and Adhikari, R., Age-structured impact of social distancing on the COVID-19 epidemic in India, arXiv: 2003.12055 (2020).
Google Scholar

[8]

To Wang, KK. et al., COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. Clin. Infect. Dis.. 10.1093/cid/ciaa1275
Google Scholar

[9]

COVID-19 India, https://www.covid19india.org/ [Last accessed on 12.09.2020]
Google Scholar

[10]

Patrikar S.R. et al., Incubation period and Reproduction number for novel coronavirus (COVID 19) infections in India, preprint 2020. 10.1101/2020.06.27.20141424
Google Scholar

[11]

M.V Murhekar, *et al.*
Prevalence of SARS-CoV-2 infection in India: findings from the national serosurvey
Indian J. Med. Res. (2020), 10.4103/ijmr.IJMR_3290_20
Epub ahead of print
Google Scholar