

# Studying the progress of COVID-19 outbreak in India using SIRD model

S Chatterjee, A Sarkar, S Chatterjee, M Karmakar and R Paul\* 

Indian Association for the Cultivation of Science, Kolkata 700032, India

Received: 15 May 2020 / Accepted: 03 June 2020 / Published online: 23 June 2020

**Abstract:** We explore a standard epidemiological model, known as the SIRD model, to study the COVID-19 infection in India, and a few other countries around the world. We use (a) the stable cumulative infection of various countries and (b) the number of infection versus the tests carried out to evaluate the model. The time-dependent infection rate is set in the model to obtain the best fit with the available data. The model is simulated aiming to project the probable features of the infection in India, various Indian states, and other countries. India imposed an early lockdown to contain the infection that can be treated by its healthcare system. We find that with the current infection rate and containment measures, the total active infection in India would be maximum at the end of June or beginning of July 2020. With proper containment measures in the infected zones and social distancing, the infection is expected to fall considerably from August. If the containment measures are relaxed before the arrival of the peak infection, more people from the susceptible population will fall sick as the infection is expected to see a threefold rise at the peak. If the relaxation is given a month after the peak infection, a second peak with a moderate infection will follow. However, a gradual relaxation of the lockdown started well ahead of the peak infection, leads to a nearly twofold increase of the peak infection with no second peak. The model is further extended to incorporate the infection arising from the population showing no symptoms. The preliminary finding suggests that random testing needs to be carried out within the asymptomatic population to contain the spread of the disease. Our model provides a semi-quantitative overview of the progression of COVID-19 in India, with model projections reasonably replicating the current progress. The projection of the model is highly sensitive to the choice of the parameters and the available data.

**Keywords:** COVID-19; SARS-CoV-2; SIRD; Epidemiology; Computer Model

## 1. Introduction

In the post-WW-2 era, the world probably has not witnessed such catastrophic morbidity and the looming threat of severe economic challenges caused by the worldwide outbreak of the disease COVID-19 caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The detection of the disease in the human host was first reported in Wuhan, China, on December 31, 2019 as a cluster of cases of pneumonia. As the highly contagious disease transmitted rapidly all over the globe, the outbreak was declared as a pandemic by the WHO on March 11, 2020. Tackling the outspread of the disease is found to be very challenging across the world for the following reasons:

(a) conventional flu-like symptoms in human carriers and (b) human-to-human transmission via asymptomatic human hosts and (c) the absence of a proper clinical doctrine (e.g., vaccines, drugs, concrete ideas about the immunological response, etc.). Extensive testing and the imposition of containment measures to maintain social distancing turn out to be the effective remedies to prevent disease transmission at the current stage of the epidemic at several places. To evaluate the impact of these preventive measures on infection spread, recovery, death tolls, and various other associated factors, mathematical models become useful in predicting realistic, quantitative estimates. A preliminary analysis suggests that the classic mean-field Susceptible-Infected-Recovery-Death (SIRD) model by Kermack and McKendrick [1, 2], can be used to obtain a quantitative picture of the epidemic [3–8]. In this article, implementing the SIRD model, we report the temporal progress of COVID-19 transmission in India, various Indian states and compare it with some other

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s12648-020-01766-8>) contains supplementary material, which is available to authorized users.

\*Corresponding author, E-mail: raja.paul@iacs.res.in

countries around the world. A similar model used by Fernandez-Villaverde et al [5] provides a detailed overview of the pandemic situation in the USA and many other countries.

India implemented a nation-wide lockdown from March 25, 2020. On the day of the announcement of nation-wide lockdown, India had about 657 corona positive cases, while the first COVID-19 positive was detected on January 30, 2020. The socioeconomic constraints in the Indian context alludes that: (a) ‘too-prolonged’ lockdown is difficult to sustain; (b) the sole imposition of containment measures without a manifold increase in testing capacity is a futile endeavor; (c) if the implementation of the lockdown measures is lenient, containment of the spread is highly improbable. Henceforth, the feasible solution for limiting the spread lies in carefully balancing various key epidemiological factors. That is where the importance of the current model predictions becomes relevant.

This study further highlights the effect of lockdown on the disease spread and predictions about the variability in the infection peak upon the severity of the containment measures (and/or the lack of it). The model predicts that, in India, the height of the peak infection decreases with stricter lockdown, but at the cost of ‘time’ (position of the peak shifts to a later month). Thus, with a large susceptible, the infection will stay for a long time if existing infections are not quarantined immediately or no proper medicine/vaccine is employed. The key is to quarantine the infection in small pockets while in lockdown and prevent inter-pocket transmission. The model further underlines that in the highly contagious zones (‘red’ zones where COVID-19 positive cases continue to grow), if the lockdown is extended and enforced with proper quarantine measures, the new infections will gradually plummet down flattening the COVID-19 curve at a much faster rate.

Our study also explores the plausibility of universality in the spread of the COVID-19 outbreak amongst different countries [6] and compares the situation in India with few other countries (e.g., Germany, South Korea, USA, Spain) in the relevant time window (February–April). Due to the simplicity of the SIRD model, we found it difficult to fit the observed patterns of the pandemic using the available data. The real data for analysis in India’s context is collected from the repository with an interactive interface hosted at <https://www.covid19india.org>. The data for other countries are taken from the repository with an interactive interface hosted at <https://www.worldometers.info/coronavirus>. The purpose of this article is not to make any quantitative prediction that should be used to design policies, but for the research purpose only.

## 2. Model

We employ the standard SIRD model where the population  $N$  is divided into sub-population of susceptible ( $S$ ), infected ( $I$ ), recovered ( $R$ ) and dead ( $D$ ) for all times  $t$ . Thus,  $N = S + I + R + D$ . The following set of mean-field differential equations governs the temporal dynamics of the population of susceptible ( $S$ ), infected ( $I$ ), recovered ( $R$ ), dead ( $D$ ) and describes a comprehensive picture of the SIRD epidemic evolution:

$$\frac{dS(t)}{dt} = -\beta SI/N \quad (1)$$

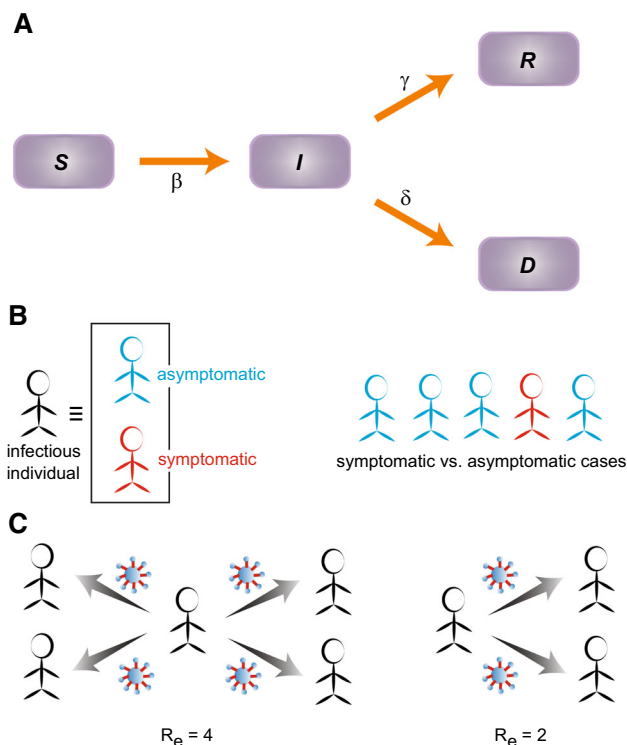
$$\frac{dI(t)}{dt} = \beta SI/N - \gamma I - \delta I \quad (2)$$

$$\frac{dR(t)}{dt} = \gamma I \quad (3)$$

$$\frac{dD(t)}{dt} = \delta I \quad (4)$$

Here,  $\beta$ ,  $\gamma$ ,  $\delta$  are the parameters determining the characteristics of infection, recovery and deaths respectively (Fig. 1a). Note that, in the current scenario  $I$  represents the population of symptomatic infection. When a susceptible person interacts with an infectious person, the susceptible become infected at a rate  $\beta SI/N$ . Large variability is observed in the rate  $\gamma$  that an infected individual is no longer infectious or equivalently has recovered in this simplified model. Literature [9–11] suggests that, on the average, infectiousness appears to start from 2 to 3 days before the symptoms are visible. The infectiousness increases to its peak before the arrival of the symptoms and remains for about 7–9 days after the peak infection. Thus, an infected individual remains infectious for about 12 days on the average and then recover. In our preliminary analysis, we set the recovery rate  $\gamma \sim 1/12$ , which however does not give the best fit for all the cases we studied. In essence, the numerical values of the model parameters are obtained from the best fit. Initial values (time  $t = 0$  days) of the number of infected, recovered and deaths ( $I_0, R_0, D_0$ ), are chosen from real data.

The choice for the initial number of susceptible ( $S_0$ ) is quite difficult. In the absence of antibody, the entire population can be susceptible to the COVID-19 pandemic. Nevertheless, the geographical, social, and economic characteristics of a region (and various other demographic factors) can substantially influence this number. We used two different approaches to get an estimate of  $S_0$ . First, we study the data for the large countries where the cumulative positive cases have reached closer to a plateau. Though, the infected population at the plateau can be determined only when the epidemic is over. Dividing this number by the total population of the country gives a fraction that appears



**Fig. 1** The transmission schematic representing the SIRD model. (a) The arrows indicate the flow between the populations of susceptible (S), infected (I), recovered (R), dead (D). (b) Infected individuals can be classified into two categories: symptomatic and asymptomatic. As per the WHO and the Indian Council of Medical Research (ICMR), India, the asymptomatic cases appears to be about 80% compared to the 20% that are symptomatic cases. (c) Effective reproduction number ( $R_e$ ): The average number of new infections transmitted by a single infectious person. For example,  $R_e = 4$  means one person is infecting 4 others on the average. Smaller the value of  $R_e$ , lesser the transmission capacity of the infection

to be of the order for  $10^{-3}$  for Germany, USA, Spain, Italy, and  $10^{-4}$  and  $10^{-5}$  for South Korea and China, respectively. Thus, an estimate of the susceptible may be obtained by multiplying the population of a country by this fraction. The number of susceptible obtained in this way, however, indicates a lower bound as many individuals with mild or no symptoms go unreported. Another possibility to estimate the fraction would be to test the number of positive cases by the number of tests carried out. This number would be an upper bound since there are many regions within a country that remains completely isolated and the populations in such pockets would not be susceptible. The ratio between the number of positive cases and the total number of tests for different countries are given in the following; the fraction is 0.159 for the USA, 0.016 for South Korea (as per data up to May 7), 0.1072 for Spain, 0.063 for Germany (as per data up to May 10). Conventionally, in epidemiological modeling  $S_0 \sim N$ . In our simulation, we have reasonably varied  $S_0$  within this range to

obtain the best fit with real data in a case by case manner (i.e., for India, few Indian states and other countries).

With the formulation of the model, comes the quantitative estimate of the speed at which the disease spreads across a population. In other words, from the deterministic SIRD model, the objective is to assess how fast a human carrier would infect people belonging to the population of susceptible. The quantity that determines the transmission speed of the pandemic is the effective reproduction number or replacement number ( $R_e$ ) [12]. Often the basic reproduction number  $R_0$ , defined as the average number of secondary infections that occur when an infectious person (primary or source of infection) is placed into a susceptible population, is used in the epidemiological models.  $R_0$  can be estimated from the very early stage of the infection when the infectious person mixes freely with the susceptible population. Estimating  $R_0$  is often challenging due to lack of unbiased data as all secondary infections cannot be determined exactly; especially for COVID-19, where asymptomatic cases are hardly identified (Fig. 1b). The effective reproduction number ( $R_e$ ), which we used in this study, evaluates the mean number of new infections (infected from the susceptible pool) directly transmitted/induced by a typical infected person and can vary over the entire duration of the infection (Fig. 1c). In the SIRD model,  $R_e$  can be represented as  $\beta/(\gamma + \delta)$ . From the best fit of the data, we find that  $\gamma > \delta$ , yielding  $R_e \sim \beta/\gamma$ . If  $R_e > 1$ , the disease starts spreading in a population infecting more and more people, but spreading does not occur if  $R_e$  falls below 1. It is easy to notice that longer a person remains infectious (i.e.  $1/\gamma$  days), can give rise to very large  $R_e$  even if the number of infectious interactions per day (i.e.,  $\beta$ ) is small.

### 2.1. Incorporating the effect of containment-measures

Containment measures in terms of social distancing and lockdown have been implemented world-wide to mitigate the transmission speed of the outbreak. We implemented the effect of lockdown in the model by modifying the infection rate and obtained the best-fit. We chose the following functional form of time-dependent infection rate  $\tilde{\beta}(t)$  where it gradually decreases after the containment measures are enforced [5, 6]. Before lockdown, the infection rate is  $\beta$  which is constant. When the lockdown is imposed on day  $\tau$  (counted from the initial time point  $t = 0$  or day 0 as chosen in the simulation), the time-dependent infection rate  $\tilde{\beta}(t)$  diminishes with every progressing day which is assumed to vary exponentially in the following manner [6]:

$$\tilde{\beta}(t) = \beta(1 - \zeta)e^{-[(t-\tau)/T]} + \zeta\beta \quad \text{for } t > \tau \quad (5)$$

$$\tilde{\beta}(t) = \beta \quad \text{for } t \leq \tau \quad (6)$$

Here,  $\zeta \in [0, 1]$  is the infection parameter (or interaction parameter) and  $T$  is the delay in the number of days before the effect of lockdown is visible in the propagation of infection. Without lockdown  $\zeta = 1$ , referring to rapid infection while  $\zeta = 0$  means that infection is contained (e.g., no interaction between infected and susceptible population, hence no transmission).  $\zeta \in [0, 1]$  reflects the asymptotic mitigation of the infection rate  $\tilde{\beta}(t)$ , when containment measures are imposed. Lower the value of  $\zeta$ , stricter is the containment measures (or the manifestation of the same) [5, 6]. Here,  $\tilde{\beta}(t)$  has an initial value  $\beta$  which is constant.  $\tilde{\beta}(t)$  diminishes over time and reaches a value  $\zeta\beta$  as containment measures continue. Essentially, the initial value of  $\beta$  determines the characteristic properties of the disease which depends on the effective interaction of people in a region, social behavior, density of population, etc. The terminal value  $\zeta\beta$  reflects the effect of the containment and how the social distancing is being maintained. In the current model setup,  $\tilde{\beta}(t)$  is meant to account for the changes in the behavior of infection spread due to social distancing and containment measures. This is an external parameter that is expected to decay with time to a smaller value  $\zeta\beta$  when physical contact is avoided. Taking a cue from the previous studies [5, 6], the functional form of  $\tilde{\beta}(t)$  is constructed to account for the apparent changes on the infection growth due to containment measures and social distancing.  $T$  controls the effective speed at which the slow down in the disease transmission occurs due to the enforced containment measures.

The model simulation, data analysis, and plotting are carried out in python. The analysis of the COVID-19 data, using the deterministic compartmental SIRD model, sheds light on the primary characteristics of the temporal evolution of the pandemic. Relevant parameter values chosen for the India and few Indian states are listed in Table S1–S2. The best-fit parameters chosen for foreign countries are listed in Table S3.

### 3. Results

We carried out the SIRD model analysis on COVID-19 progression in India's context (and few other countries) with realistic variations in following parameters: rates of infection ( $\beta$ ), recovery ( $\gamma$ ) and deaths ( $\delta$ ), the initial number of susceptible ( $S_0$ ) and the effective reproduction number ( $\mathcal{R}_e$ ). Detailed results are described in the following and illustrated in Figs. 2, 3, 4, 5, 6, 7, 8, 9 and 10, Fig. S1–S4.

In a nutshell, we start with the initial susceptible population ( $S_0$ ) varied within the range  $\sim 1$ –3 million, keeping the effective reproduction number  $\mathcal{R}_e$  fixed at  $\sim 4.0$ , and show how the model prediction fits with the Indian data without a lockdown, the location of the infection peak and the relative deviation from the real data (Fig. 2a). The best fit is obtained by tuning the rates of infection ( $\beta$ ), recovery ( $\gamma$ ), and deaths ( $\delta$ ) keeping  $S_0$  constrained in the mentioned range. Then, we incorporate the effect of containment-measures/lockdown in the functional form of time-dependent  $\tilde{\beta}(t)$  and show how the effect of the containment-measures has altered the location and the height of the infection peak (Fig. 2b). Next, we explore how the variability in the effective reproduction number  $\mathcal{R}_e$  influences the infection peak (Figs. 3, 4). Furthermore, we analyze the COVID-19 progression in few Indian states e.g., Kerala, Maharashtra, Delhi, and West Bengal (Figs. 5, 6) and foreign countries e.g., South Korea, Germany, USA, Spain (Figs. 7, 8). Lastly, we explore, in brief, what happens to the outspread, if the lockdown is lifted (in other words, containment measures are relaxed) in the Indian context (Figs. 9, 10).

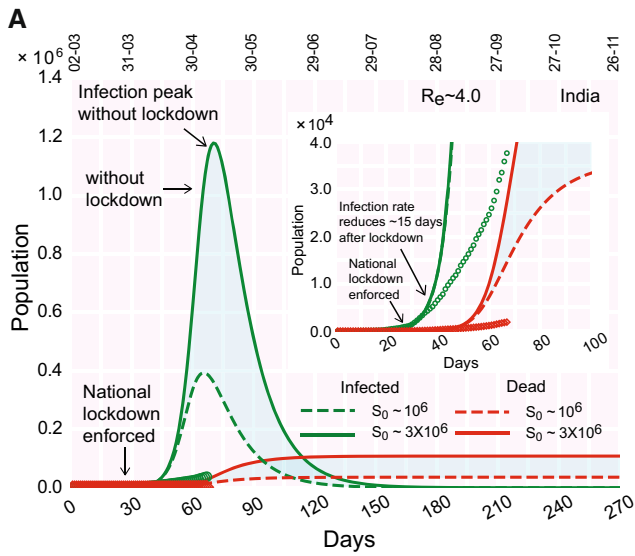
#### 3.1. India without lockdown: what could have happened?

The first COVID-19 positive human host was reported in India on January 30, 2020. The exponential growth of the number of infections, from 30th January onward, reached a number 657 on March 25, 2020, the day on which India imposed a nation-wide lockdown (Fig. S1A).

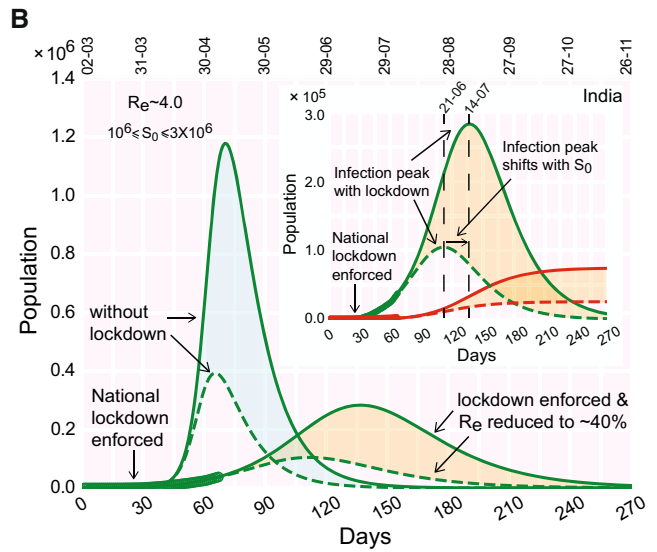
Using the SIRD model, we first explored what could have happened, if the containment measures had not been undertaken. As mentioned earlier, we chose the factor  $\sim 10^{-3}$  (obtained in case of Germany and few other countries by dividing the cumulative population at infection peak by the actual population of the country) and multiplied it with the Indian population of  $\sim 10^9$  to estimate the lower bound of the susceptible population ( $S_0$ ). It turns out that it would be a 'good' estimation to have a 'working'  $S_0$  in the range  $\sim 10^6$ . With susceptible population  $S_0$  varied in the range  $\sim 1$ –3 million (for fixed  $\mathcal{R}_e \sim 4.0$ ), the peak of the infection occurs in the first half of May (Fig. 2a). As expected, the peak height (infected population at the peak) increases with increasing  $S_0$ . For an initial susceptible pool of  $S_0 \sim 10^6$ , the peak reaches a height of 0.4 million, whereas the peak jumps to  $\sim 1.2$  million for  $S_0 \sim 3$  million (Fig. 2a). The total death toll is estimated to reach about 30,000–100,000 for  $S_0$  in range  $\sim 1$ –3 million, during July–August, 2020 (Fig. 2a).

Next, we introduced the effect of containment-measures in the infection rate  $\tilde{\beta}(t)$  (Eqs. 5–6). Numerical analysis is





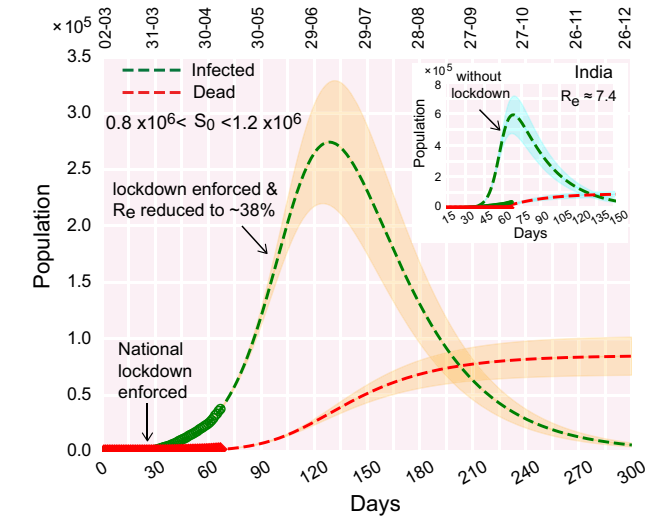
**Fig. 2** Evolution of the COVID-19 and effect of lockdown in India's context. (a) Time progression of the infection growth curve, if no lockdown is imposed. (inset) The infection growth (as observed from the real data, points) reduces 20 days after the national lockdown is implemented deviating from the theoretical curve with no lockdown (dashed line). (b) Time evolution of the infected population (and



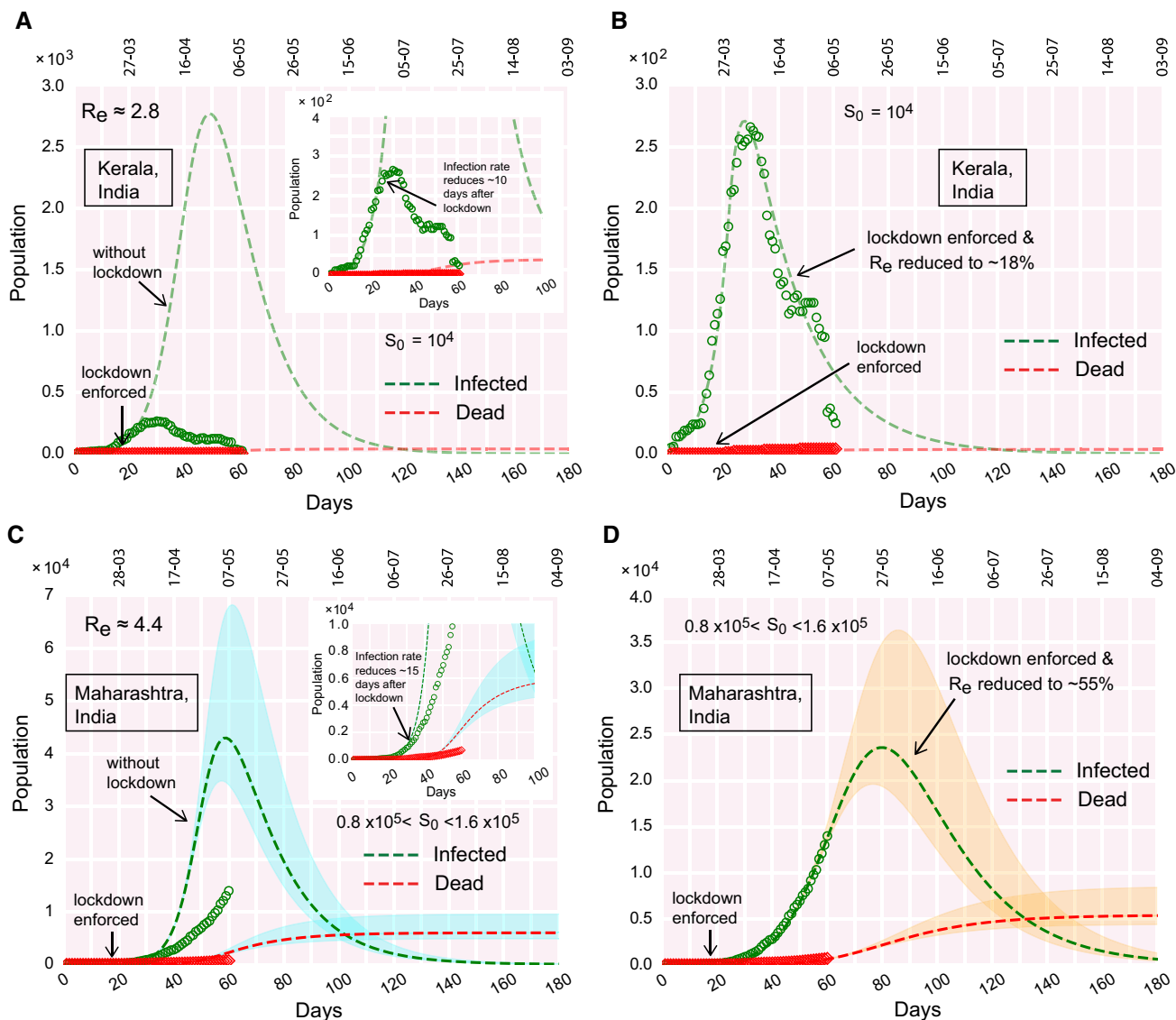
**Fig. 3** Effect of variations in  $R_e$  on the infection curve (and death curve). (a) Time progression of the infection growth curve when lockdown is enforced; (inset) if no lockdown was enforced. The color shades enveloping the curves denote the variation in  $R_e$ . The real data considered for fitting are from March 2, 2020.

carried out to investigate whether the progression of the outbreak is mitigated after the lockdown is imposed.

deaths, inset), when containment measures (lockdown) are enforced. The color shades enveloping the curves denote the variation in susceptible population  $S_0$  ('cyan' shades: without lockdown, 'orange' shades: with lockdown). The real data considered for fitting are from March 2, 2020 (color figure online)



**Fig. 4** Best fit of the infection and death curves with the real data freely varying the effective  $R_e$  which appears to be  $\sim 7.4$  in the early stage. The color shades enveloping the curves denote variation in susceptible population within a range of  $\sim 0.8-1.2 \times 10^6$ . The real data considered for fitting are from March 2, 2020. Relevant parameters for the analysis are  $\beta = 0.257 \text{ day}^{-1}$ ,  $\gamma = 0.0315 \text{ day}^{-1}$ ,  $\delta = 0.0032 \text{ day}^{-1}$ ,  $\tau = 29 \text{ days}$ ,  $T = 7 \text{ days}$ ,  $\zeta = 0.375$



**Fig. 5** COVID-19 progression in few Indian states. (a)–(b) Time evolution of the population of infected and dead for Kerala, (a) if lockdown was not imposed, (b) due to effect of lockdown. (c)–(d)

Time evolution of the population of infected and dead for Maharashtra, (c) if lockdown was not imposed, (d) due to effect of lockdown

### 3.2. Effect of containment measures: how well is India doing?

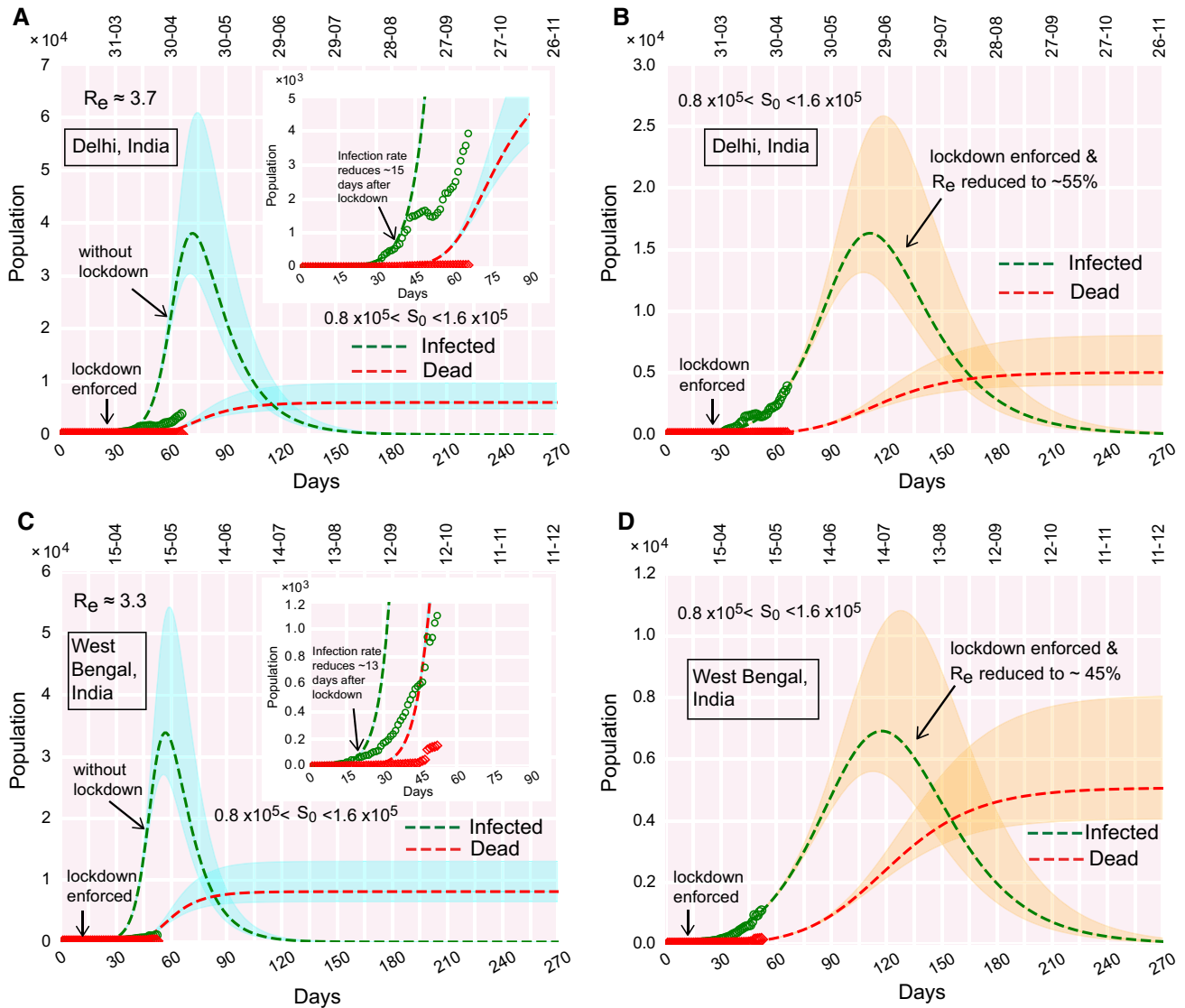
#### 3.2.1. Straightening of the growth curve

From the real data, it appears that the infection rate begins to reduce, 15–20 days after the national lockdown is implemented (Fig. 2a, inset). We further observed that the growth curve for the infected population displays a straightening feature during the lockdown time frame. This is expected to be observed if containment measures are initiated; the unhindered exponential growth before the lockdown slows down due to the effect of containment measures during the lockdown. While slowing down and

deviating from the exponential trajectory, the infection growth curve (time progression of the infected population size) acquires a distinctive straightening feature until the very recent surge (Fig. 2a, inset).

#### 3.2.2. Dwarfing the infection peak

Next, adding the lockdown effects into the picture, we fit the theoretically obtained infection growth curve with the real data. The best fit with the current set of parameters demonstrates that, due to the effect of the present lockdown, the infection peak dwarfs down to about 0.10 million from about 0.40 million in ‘without lockdown’ scenario (dashed curve, Fig. 2b and inset). The infection



**Fig. 6** COVID-19 progression in few Indian states. (a)–(b) Time evolution of the population of infected and dead for Delhi, (a) if lockdown was not imposed, (b) due to effect of lockdown. (c)–(d)

Time evolution of the population of infected and dead for West Bengal, (c) if lockdown was not imposed, (d) due to effect of lockdown

peak is projected to reach a peak at the end of June, tentatively (Fig. 2b, 2b, inset). The estimated death toll also reduces substantially compared to the earlier scenario without containment measures.

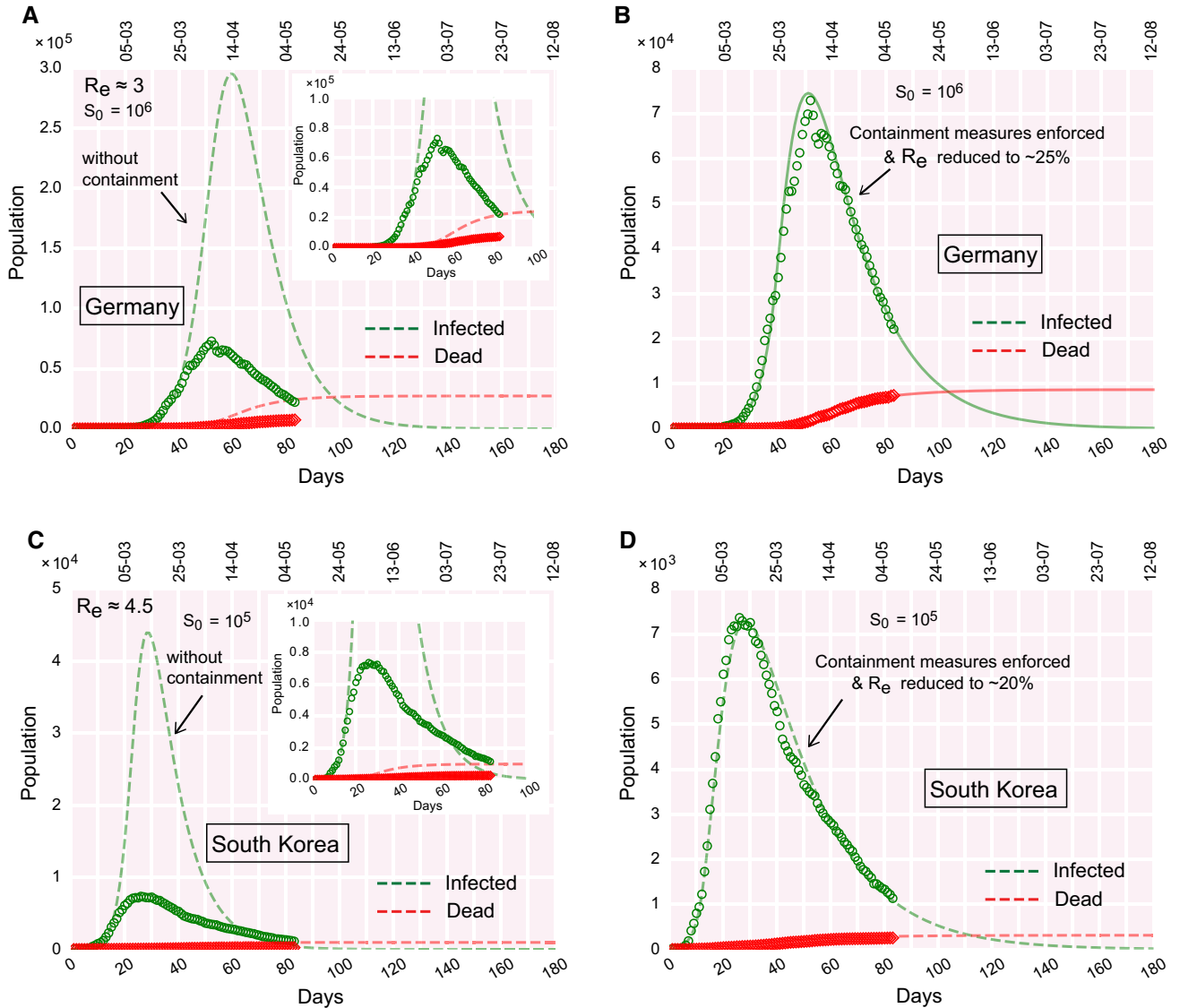
However, the model also shows that the situation can be improved further. The infection growth curve can be dwarfed down further if the lockdown is extended and reinforced stringently in COVID-19 prone zones. In that case, the infection growth curve noticeably flattens with the infection peak reduced further.

As mentioned earlier,  $S_0$  is a very crucial parameter in governing the position and the height of the infection peak. In the following, we summarize how the variations in the size of the susceptible population  $S_0$  influence the infection growth curve.

### 3.2.3. Variability in the susceptible population

Keeping  $\mathcal{R}_e$  fixed at  $\sim 4.0$ , we varied the size of the total susceptible population within a range of 1–3 million (Fig. 2a, b). The model analysis shows that the larger the size of the susceptible population, the higher the infection peak (Fig. 2). Moreover, for the larger size of the susceptible population, attainment of the infection peak is delayed with the infection peak shifted to a later time zone (Fig. 2B, 2B, inset). These characteristic features are consistent in both without and with lockdown scenarios.

Thus, it is evident that the key to containing the outbreak lies in keeping  $S_0$  small. This is feasible only when interactions between a demographic region with the recent occurrence of infections and a region with no ‘latest’



**Fig. 7** COVID-19 progression in Germany and South Korea. (a)–(b) Time evolution of the population of infected and dead for Germany, (a) if containment measures were not undertaken, (b) due to the effect of containment measures implemented. (c)–(d) Time evolution of the

population of infected and dead for South Korea, (c) if containment measures were not undertaken, (d) due to the effect of containment measures implemented. The real data considered for fitting are from February 15, 2020

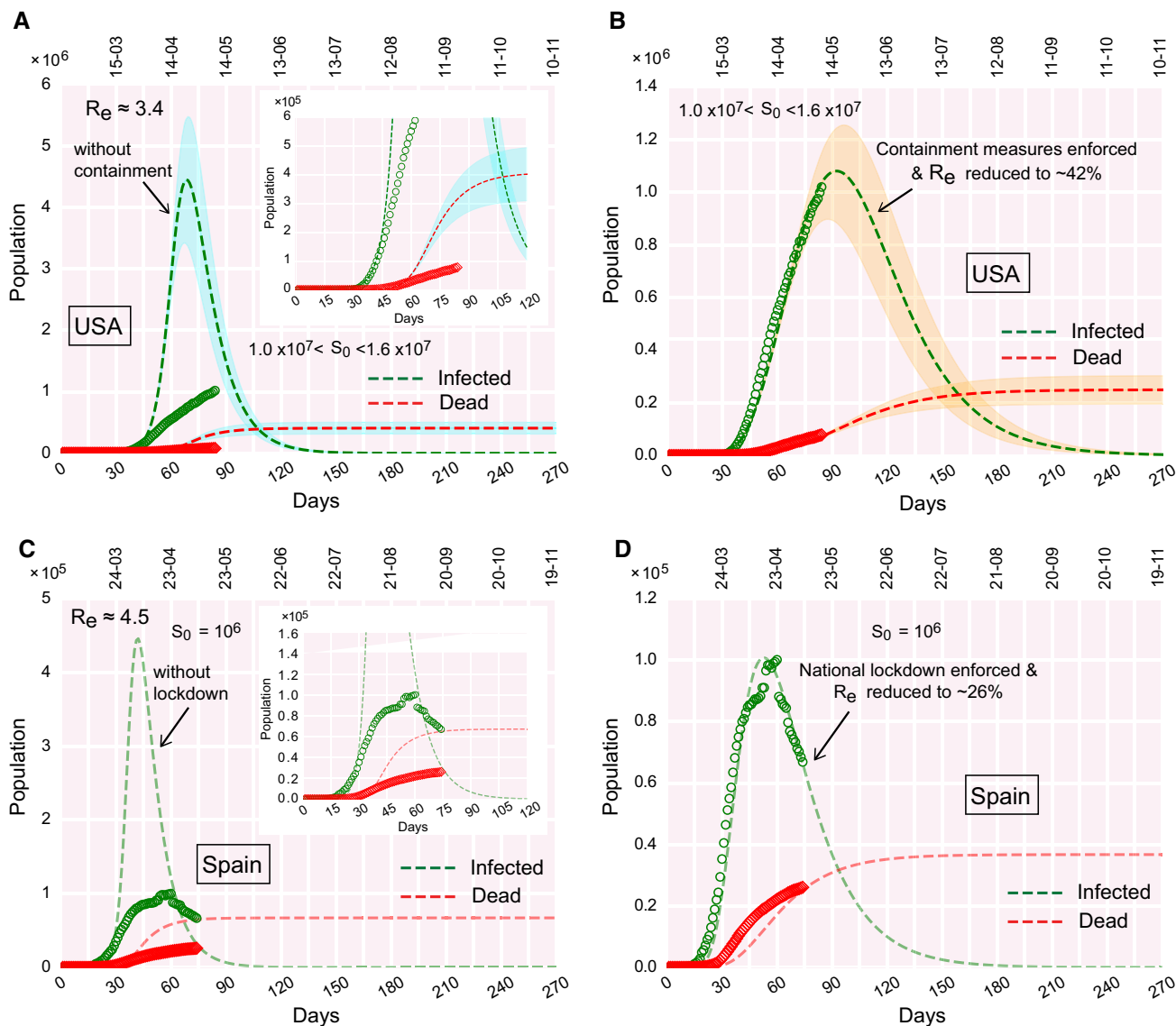
instance of infection are strictly barred. Besides the ‘global’ lockdown (in a nation-wide sense), locally keeping an infected region isolated from other proximal unaffected regions may help to keep  $S_0$  in check.

The next question that crops up is what happens to the magnitude of effective reproduction number ( $\mathcal{R}_e$ ) when containment measures are put in place. We discuss in the following, how the effective  $\mathcal{R}_e$  changes with time during the lockdown (Fig. 3).

### 3.2.4. Effective reproduction number ( $\mathcal{R}_e$ ) and lockdown

We start with  $\mathcal{R}_e$  in range  $3.5 \leq \mathcal{R}_e \leq 4.4$  in the beginning. As the lockdown is implemented, less number of people interact. Therefore, the effective infection rate  $\tilde{\beta}(t)$  starts decreasing over time. How much the reduction would be for  $\tilde{\beta}(t)$  in longer time regime, is determined by the factor  $\zeta$  in Eq. (5). The reduced  $\tilde{\beta}(t)$  settles at a value  $\zeta\beta$  due to the containment effects. Thus, if the recovery rate  $\gamma$  is fixed, the  $\mathcal{R}_e$  will diminish and reach a value  $\frac{\tilde{\beta}(t)}{\gamma} \sim \frac{\zeta\beta}{\gamma}$ . The decrease in  $\mathcal{R}_e$  due to the effect of lockdown is evident in Fig. 3 where the effective  $\mathcal{R}_e$  reduces to  $\sim 40\%$  of its initial value (before lockdown). The smeared color shades





**Fig. 8** COVID-19 progression in the USA and Spain. (a)–(b) Time evolution of the population of infected and dead for the USA, (a) if containment measures were not undertaken, (b) due to the effect of containment measures implemented. Since for the USA, the infection peak is yet to occur, the estimation of the peak value is predictive, and thereby, sensitive to relevant parameter choices. Hence, the susceptible pool ( $S_0$ ) for the USA is varied, to obtain an average estimation

of the peak value. The smeared color shades enveloping the dashed curves denote the variation in  $S_0$  in the context of the USA. (c)–(d) Time evolution of the population of infected and dead for Spain, (c) if containment measures were not undertaken, (d) due to the effect of containment measures implemented. The real data considered for fitting are from February 15, 2020, for the USA. For Spain, real data are considered from February 24

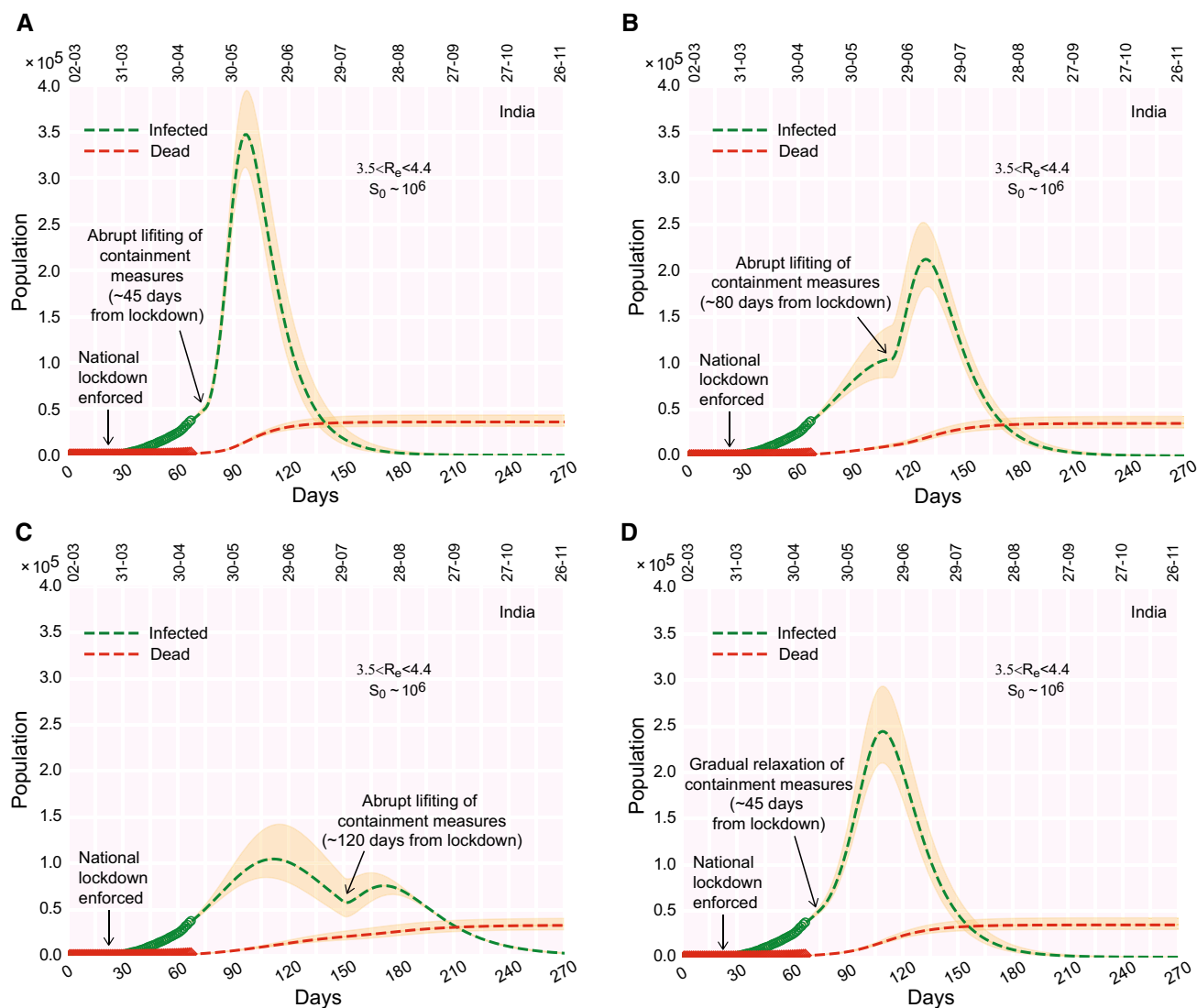
enveloping the dashed lines label the variations in  $R_e$  within the mentioned range. As expected, the higher the value of  $R_e$ , the taller the infection peak. This feature is consistent both in presence and absence of lockdown (Fig. 3, inset).

Next, we investigate, whether the value of  $R_e$ , extracted from the best fit with real data, is unique (of course with marginal variation) or the variation is non-marginal.

### 3.3. What would be the scenario if $R_e$ is large?

Instead of fixing  $R_e$  in the beginning, we varied the rates of infection ( $\beta$ ), recovery ( $\gamma$ ), and deaths ( $\delta$ ) without any restriction on the resulting value of  $R_e$ . We aimed to verify whether the best fit of real data with the theoretical curves (infection, recovery, and death) can be obtained for a set of ( $\beta, \gamma, \delta$ ), other than the already chosen values in Figs. 2 and 3, with no apparent constraint on the values of  $R_e$ .

We find that, for a fixed size of the susceptible population of about 0.8–1.2 million, the real data can still be



**Fig. 9** COVID-19 progression upon lifting the containment measures in India. (a)–(c) Time evolution of the population of infected, if the containment measures are ‘rapidly’ lifted  $\sim 45$  days (a),  $\sim 80$  days (b),  $\sim 120$  days (c) from lockdown (e.g., from the day lockdown was

fitted with the theoretical curves, even if the  $\mathcal{R}_e$  is large ( $\mathcal{R}_e \sim 7.4$ , Fig. 4). Similar to Figs. 2 and 3, the active infection cases deviate from the theoretical infection curve without lockdown, as the enforced lockdown effectively slows down the progression of infection (Fig. 4, inset). Note here, that in comparison with Figs. 2 and 3, the location and height of the infection peaks change (both in the cases of without lockdown and with lockdown), as effective  $\mathcal{R}_e$  is increased  $\sim$  twofold (Fig. 4). Consistent with the definition of  $\mathcal{R}_e$ , we observe that greater the value of  $\mathcal{R}_e$ , larger the size of the infected population (compare Fig. 2b,  $\mathcal{R}_e \sim 4.0$ ; and Fig. 4,  $\mathcal{R}_e \sim 7.4$ ). It is also noteworthy to mention that here the recovery rate  $\gamma = 0.0315 \text{ day}^{-1}$  corresponds to about  $\sim 31$  days compared to 12

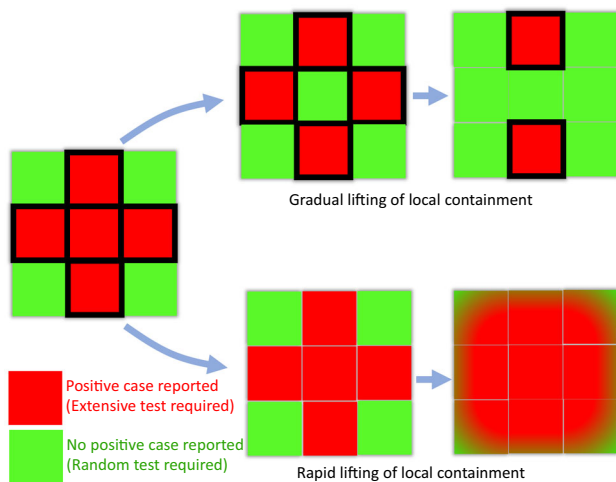
enforced). (d) Time evolution of the population of infected, if,  $\sim 45$  days from lockdown, the containment measures are gradually relaxed/phased out within a time window of  $\sim 30$  days. The real data considered for fitting are from March 2, 2020

days as discussed earlier. Prolonged infectiousness leads to the rise in the  $\mathcal{R}_e$  and consequently the total number of infected people.

From the above observations, we connote that the exactness of  $\mathcal{R}_e$  can be ascertained, when we have more data points in the time evolution of the infected, recovered, and dead population. The current model setup may not be able to precisely pinpoint the exact ‘real’  $\mathcal{R}_e$ .

### 3.4. How well the individual Indian states are doing?

In India, the first COVID-19 positive case was reported in Indian state Kerala on January 30, 2020, and now almost half of the active COVID-19 positive cases are from



**Fig. 10** Representative schematic illustrating the aftermath of slowly relaxing (phasing out) the local containment measures vs rapid lifting of the same. The boxes bordered in black depict the containment zones. Rapid lifting of the local containment paves the way for inter-mingling between regions with no positive cases (green) and regions with positive cases found lately (red). This, in turn, results in rapid transmission of the disease across zones (all zones becoming red), rendering the purpose of preceding lockdown futile. On the contrary, initially, the partial lifting of local containment only in the green zones bars the import of transmissions from red zones. When a red zone becomes green, the local containment can be lifted from that region. Due to the effect of this gradual lifting, the red zones diminish over time, with ‘greens’ taking over the ‘reds’ (color figure online)

another Indian state Maharashtra. In this note, we explore the COVID-19 progression in these Indian states along with Delhi and West Bengal and compare the features of pandemic progression with each other (Figs. 5, 6, S1B).

After the first case being detected in Kerala on January 30, the second and third cases were reported on February 2–3. After February 3, there was no new case detected in Kerala till March 7. The previous three cases were all recovered within February 20. The ‘second-wave’ of infections started from March 8. From March 8 onwards, there was a rapid upsurge of infections. However, about 2 weeks after the national lockdown is imposed, Kerala reached its infection peak. It is evident from Fig. 5a, b that the downfall of the infection is rapid, as the infection curve moved past its peak. If the lockdown was not enforced, the infection was projected to occur around mid-May. But, Fig. 5b alludes that Kerala implemented the containment measures so well that the infection peak occurred early at a much lower height (Fig. 5a, b). The model analysis further projects that due to the effect of lockdown,  $\mathcal{R}_e$  reduces to  $\sim 18\%$  of its initial value during the upsurge of infections before lockdown. The reduced value of  $\mathcal{R}_e$  is  $\ll 1$ , which means Kerala is on the way to become a COVID-19 free state soon if the trend continues.

In Maharashtra, the first case was detected on March 9. The total infected population is yet to attain its peak. The

projected infection peak would occur around the end of May or early June if the present trend continues and containment measures remain enforced in places (Fig. 5c, d).

Similarly, in Delhi and West Bengal, the infection growth curves are yet to attain their respective peaks (Fig. 6a–d). The first cases in these states were reported on March 2 and March 17, respectively. The peaks are projected to be reached at the end of June and mid-July for Delhi and West Bengal respectively, if the enforced lockdown remains deployed and the current trend continues (Fig. 6b, d).

It is important to note that, in Indian states, Maharashtra, Delhi and West Bengal, the estimated  $\mathcal{R}_e$  plummets down to value  $> 1.0$ , even after staying months under lockdown. Among the Indian states we analyzed, Kerala turns out to be the only exception where the effective  $\mathcal{R}_e$  reduces to a value  $< 1$ , meaning that further ‘out-of-bound’ spreading is unlikely to occur there if the current trend is followed.

### 3.5. Progression of COVID-19: where does India stand compared to other countries?

It is evident from Fig. 7a–d and S1A that both Germany and South Korea have moved past the infection peak. The infected population is decreasing day by day in those countries. The best fit with real data is obtained for the initial  $\mathcal{R}_e \sim 3.0$  and  $\sim 4.5$  for Germany and South Korea, respectively. However, as the containment measures were undertaken in those countries, the effective transmission (or  $\mathcal{R}_e$ ) reduced to  $\sim 20\text{--}25\%$  from the initial values for the respective countries (Fig. 7b, d). This observation, suggests that the counter-measures to fight the pandemic (e.g., containment measures, social distancing, quarantining, testing, etc.), undertaken in these countries, were reasonably successful in repressing the outspread. Moreover, the reduction of  $\mathcal{R}_e$  to 20–25 % of its initial values, rescales the  $\mathcal{R}_e$  for the respective countries to a value  $< 1$  which alludes that new infections are declining and any more ‘out-of-bound’ infection growth is unlikely to occur if the current trend is followed.

We analyzed COVID-19 progression data for two more countries: USA and Spain (Fig. 8a–d). The USA is approaching the infection peak and will reach its peak shortly if the current trend continues (Fig. 8b). However, contrary to the USA, Spain has already passed the infection peak and the infected population is decreasing gradually (Fig. 8d). Spain imposed a nation-wide lockdown on March 14. Model analysis (fitting parameter optimization) suggests that, due to the effect of lockdown,  $\mathcal{R}_e$  for Spain reduced to 25 % of its initial value. But in the USA, the reduction in  $\mathcal{R}_e$  is only  $\sim 42\%$  implying that the implementation of local containment was not that stringent.

Contrary to the infection curves for Germany, South Korea, Spain (Figs. 7, 8), India is yet to reach the infection peak (the USA is about to reach the peak.). The following remarks briefly summarize where India stands compared to these countries.

1. The effective containment during the present lockdown in India indicates that the infected population might reach its peak at the end of June (Fig. 2) whereas Germany, South Korea, and Spain have already moved past the peak and daily new infections are decreasing (Figs. 7, 8).
2. Since India has a large population, the infection is expected to stay for a longer duration. Germany, South Korea, and Spain might have the advantage of a smaller population of the susceptible. We allude that the higher the actual population of a country, the higher would be the effective size of the susceptible pool for that country while making the previous statement. The key is to contain the infections in small zones and prevent transmission between infectious and non-infectious zones.
3. The growth of the infected population in Germany, South Korea, and Spain were greater than that experienced in India which gave India an additional advantage of 'buying precious time'. Slow growth rate alludes to a smaller peak value at the zenith of the infection. However, as mentioned earlier, the height of the peak is subjected to the effective size of the initial susceptible pool ( $S_0$ ).

## 4. Discussion

For better clarity and wider accessibility to general readers, we discuss and summarize the important observations from our study in Q&A format in the following:

### 4.1. How universal the model predictions are for different countries?

To investigate the universality in the COVID-19 outbreak across different countries, we looked into iterative time lag maps for the cumulative confirmed infected ( $C = I+R+D$ ), recovered (R), and dead (D) population [6]. Using the iterative maps, we try to extract the correlation between a population on the day  $n$  and day  $n + 1$ . From the recurrence plots (population count on  $n^{\text{th}}$  day vs population count on  $(n + 1)^{\text{th}}$  day) in Fig. S2, we observe that the real data for all the cases follow the same power law of the following kind:  $f(x) = ax^b$ . The factor and exponent  $a$  and  $b$  are similar for all the countries considered in the plots. This

finding indicates that there exists an underlying universality in the outbreak of the pandemic across various countries.

#### 4.1.1. How accurate are the model predictions?

To check the predictiveness of the model, we compared the projected outcome with the real data accumulated over the last couple of weeks since our initial submission of the manuscript. We plotted the data and the model prediction in Fig. S3A-S3B (similar to Figs. 2b and 9d respectively) without altering the parameters. One can notice that, so far, the real data is bounded by the model predicted curves. Nevertheless, the SIRD model is a drastically simplified approach to thoroughly understand the dynamics of COVID-19 progression. From the available information, it is now becoming apparent that a susceptible person goes through a latent period of 2–3 days after coming in contact with an infected individual. Subsequently, the person remains infectious for several days ( $\sim 5$  days). The infectious individual may or may not develop symptoms. The current model does not incorporate any of these details and hence fitting is imperfect. Moreover, data used to fit with the model also vary between different locations leading to uncertain predictions. A compartmental model with multiple species may be useful to study the dynamics of the sub-population [13].

#### 4.1.2. Is the model implemented here specifically designed for the COVID-19 outbreak? What are the model limitations?

The model used here to analyze the COVID-19 progression is the well-known SIRD model. This is a standard epidemiological model with three characteristic parameters for infection, recovery, and death. The model estimates the number of infections within a closed (conserved) population of susceptible bearing the risk of contagion. Note that this is 'NOT' a COVID-19 specific model by construction. For years, people have used this model to study several outbreaks all across the globe [14]. In our study, the word 'COVID-19' enters into the picture only through the best fitting of the 'real data', that determines the instantaneous rates. The biological and clinical nitty-gritty of COVID-19 is beyond the scope of the model. The limitations of this current model prescription are given in the following: (a) The mean-field SIRD equations do not include the spatial variation in the population density. By construction, there is no spatial degree of freedom in the standard SIRD model. The term  $\beta SI$  in Eq. (1) alludes that all the infected individuals are equally likely to interact with susceptibles and transmit the disease [14]. However, in a realistic scenario, this is unlikely to occur. In the current Indian

context, an individual living in a remote Himalayan village is far less likely to encounter an infected person compared to an individual living near Marine Drive, Mumbai. (b) The model assumes homogeneous exposure and response to the disease [14]. There are various key factors like different demographic features, ethnicity, lifestyle, socioeconomic strata of a sub-population that can induce variability in the exposure to the infection. (c) The model assumes that the contagiousness of an infected person remains constant throughout before his/her recovery and a person is capable of transmitting the disease right after s/he became infected which may not be realistic in the present context. (d) Naturally, an increase in the number of tests would also increase the number of infections as more and more undetected infections will be detected. The availability of hospital beds, critical care facilities would also influence the recovery rate in reality. The simplistic model cannot account for all these factors.

There are also a few “what if” paradigms that cannot be assessed via this model unless further compartments and rate parameters are added upon. Few “what if” scenarios are: (a) What if the SARS-CoV-2 immunity is not permanent? [15]; (b) what if there is seasonal variation in transmission rates of the disease? [15]; (c) what if the vaccine becomes available within a certain period? and many more. In essence, to address each of these aspects, we need to formulate elaborate models brick by brick in a context-dependent manner. The current working handle of the SIRD model marks the preliminary footstep along that direction.

#### 4.1.3. How sensitive are the model predictions to parameter variations?

We investigated the sensitivity of the model to parameter variations, focusing in particular on the parameters that change the rates of infection, recovery ( $\beta, \gamma$ ) and most importantly the effective reproduction number or replacement number ( $\mathcal{R}_e$ ) within a feasible range to see the effect the model prediction. The lockdown stringency, characterized by the time-dependent  $\tilde{\beta}(t)$ ,  $\zeta$ , was varied to get an estimate of the infected population. Corresponding data are shown in each figure by the shaded envelope around the mean curves. The model appears to be sensitive to the variation in the value of  $\zeta$  and  $S_0$  when compared with the real data. Increasing (decreasing) the value  $S_0$ ,  $\zeta$ , and  $\mathcal{R}_e$  rapidly increases (decreases) the population of total infection and death mostly around the peak and alters the position of the peak infection. These parameters can be decreased by enforcing local containment and social-distancing measures.

#### 4.1.4. Topological dependence of model predictions

The standard SIRD model does not contain any spatial degree of freedom. The mean-field nature of the current approach excludes the topological dependence of the model predictions. Hence, the spatially explicit modeling incorporating infection hotspots and disease transmission from the hotspots to the rest of the places would yield a more realistic reconstruction of the scenario. A compartmentalized SIRD model simulated on a virtual network of all Indian states where the network connectivity manifests the transmission spread from one state to another would be worthwhile. The spatial topology of the underlying network should also include details like the geographical topology of a region, human mobility, connectivity by transport system, health care facilities, etc. The detailed phenomenological reconstruction on a lattice-based model would shed light on the topological dependence of the predictions, robustly. A similar spatially explicit study using the SEIR model shows how the disease is transmitted from initial foci/local pockets of infections to entire Italy [16]. We plan to adopt a similar approach in the Indian context as a worthy future endeavor. We reiterate that the predictions based on the SIRD model are simple mean-field predictions. Given the public interest on this pandemic, we put it as a disclaimer that, including topological effects may quantitatively alter the picture of the COVID-19 progression in Indian context up to a reasonable degree.

#### 4.1.5. How effective are the quarantine measures? Can quarantining single-handedly contain the transmission?

A common perception of flu and other infectious diseases is that an infected individual spreads infection when symptoms appear. In the case of seasonal flu, infection mostly occurs when a person has symptoms [17]. However, as we understand from the literature survey, an individual with COVID-19 would be contagious before developing symptoms. The incubation period for COVID-19 is  $\sim 5$  days, and maximum infectiousness appears to be 2–3 days before the symptoms appear. Thus infection spread by an individual is maximum before he/she becomes sick [10, 18]. Due to limitations of the testing procedure, diagnosis takes about 5 days after symptoms are visible, i.e., 10 days from the day of infection. Clearly, on the average, an infected individual is beyond the peak of maximum infectiousness after this time. Thus, a reduced rate of infection demands early diagnosis and isolation of positive patients. This means that a COVID-19 patient needs to be identified in the pre-symptomatic stage as evidence suggests the infectiousness of the patient before developing symptoms which is extremely challenging



(effectively, RT-PCR needs to be carried out for every individual who might have come in contact with the patient). The epidemic becomes even more complex due to a majority of the infected individual who develops mild or no symptoms [18, 19]. Therefore, even with isolating/quarantining, all the infected COVID-19 would not be eliminated for two reasons: a) normally an individual would be tested after symptoms appear which is when he/she has passed the peak of the contagiousness, b) asymptotically infected person, in general, are not tested but he/she is also contagious like the symptomatic individual.

#### 4.1.6. Gradual relaxation of the containment versus extended lockdown?

We have investigated the possible effect of relaxing the containment measures at three different time points for India (Fig. 9a–d). We find that if the containment would have been relaxed in the middle of May, i.e. the before the projected peak infection is reached at the end of June, the infected population would rise rapidly to a great extent (Fig. 9a). The peak height reduces, if the containment measures are relaxed, when the infection is close to the peak, a time point around the 3rd week of June (Fig. 9b). However, if the relaxation occurs a month after the peak infection, a second peak arrives which is lower than the first infection-peak (Fig. 9c). A third possibility is to gradually relax the containment measures after May 17. The model shows that in this case the original peak does not shift its position but becomes twofold higher than before (Fig. 9d). A gradual relaxation could be carried out in steps: (a) First, identify all the sensitive (red) and safe (green) zones having positive and no cases respectively. Smaller the size of such zones, easier they can be managed by the administration, and necessary supplies can be arranged. It is important to seal the boundary of the red zones. (b) Test for new cases carrying symptoms and randomly test a few having no symptoms. (c) Dissolve the boundary between red and neighboring green zones once the red zone does not report a case for 2 weeks. This process will increase the size of the green zone where more and more people can communicate and business can restart. Successively extending the relaxation from the local neighborhood to the cities, districts, states, the containment measures can be relaxed across the country. Nevertheless, social distancing is mandatory even after the containment is officially lifted as there might be many undetected cases that can trigger the spread of the disease again. A schematic diagram in Fig. 10 summarizes the above-mentioned steps of relaxation and the consequential aftermaths, pictorially. In a nutshell, ‘too-early’ lifting of containment measures, long before the infection reached its peak, makes the purpose of lockdown ‘null-and-void’. The reduced

infection rate  $\tilde{\beta}(t)$  again starts increasing yielding a larger  $\mathcal{R}_e$  promoting the outspread.

It is imperative to note that there are certain differences in the concepts and implications of lockdown, containment measures, and social distancing. In the Indian context, ideally, lockdown implies that containment measures are enforced in every nook and corner of the country. However, the intervention of containment measures can be applied locally. For example, in principle, the nation-wide lockdown can be lifted, but containment measures can remain in action in places that are identified as infection hotspots. Hence, the word ‘lockdown’ refers to the restrictions (social, economical) applied over a very large region (e.g., a state or the whole country) whereas the ‘containment measures’ refers to implementing the same disciplines locally as well as universally. Social distancing, on the other hand, is rather a personalized affair. People can maintain social distancing even when a lockdown is not in place. In an ideal scenario, if all individuals within a closed community maintain social distancing without ‘lockdown’, new infections are unlikely to occur. In that note, our model projects that, with containment measures and social distancing in effect, the number of active infection (noncumulative) would be about 80,000–140,000 at the peak which is expected to appear sometime toward the end of June (Fig. 3). Nevertheless, to reduce the economic impacts, it is essential to relax the strict containment measures applied across the country. Through our model, we checked several scenarios for applying the relaxation and estimated the evolution of the infection. If the containment measures were suddenly relaxed after May 17, the peak infection would have increased sharply to 0.3–0.4 million (Fig. 9a). It is imperative to mention that, this number is estimated without altering the susceptible population which is about  $10^{-3}$  times the actual population observed for many large countries. The remaining population is considered to be shielded from the infection due to containment and demographic segregation. Note that, due to containment (and social distancing), the size of the susceptible population is kept at a value of only several million while the total population of India is about 1.35 billion. Upon lifting the containment measures, the effective susceptible population should also increase. In this simplistic model, since the attributes are based on a closed population, no additional increment in the number of susceptibles upon the lifting of containment measures is considered. In the event of unrestricted mixing of the population of the whole country, the peak infection might see a  $10^3$  fold rise which would be challenging for any health care system to deal with. Thus, social distancing measures must remain in place unless the infectious population is contained or drastically reduced.

In a nutshell, the four scenarios depicted in Fig. 9a–d are “what if” propositions. Previously, the model analysis had been carried out before May 17. There, we explored what would have happened if containment measures (and social distancing) are abruptly lifted, May 17 onwards. The model prediction was a peak infection of 0.3–0.4 million (Fig. 9a). But, in reality, in the Indian context, we observed that the lockdown was still in place after May 17. Thus, Fig. 9a remains as a ‘what could have been’ scenario. Note that when we say ‘lockdown’, we assume that both the containment measures and social distancing norms are enforced/ followed. Now the question is, do any of these plots (Fig. 9a–d) corroborate with the present situation (real data as of May 29)? In India, 4th phase of lockdown was enforced (rather extended) from May 18 onward. However, during this phase, various relaxation measures were also given. In that note, the closest theoretical consideration would be the scenario depicted in Fig. 9d. To analyze the situation, we compare two scenarios side by side from a more recent perspective (evolved when the manuscript was under revision): time evolution of infection curves (a) with lockdown (containment measures + social distancing in place) as depicted in Fig. 2b; (b) containment measures gradually relaxed as depicted in Fig. 9d. In Figs. 2b and 9d, the real data was up to May 7. In Fig. S3A–S3B, we re-plotted the real data up to May 29, with the theoretical curves of Figs. 2b and 9d. The real data up to May 29, in both Fig. S3A–S3B, reasonably falls within the range enveloped by the theoretical curves (with marginal deviation). The range of the active infection peak in Fig. S3A, is about 0.1–0.3 million occurring between the end of June to mid-July whereas in Fig. S3B, the infection peak is predicted to be in the range of 0.2–0.3 million also occurring at the end of June. Thus, from Fig. S3A–S3B combined, we gather that the active infection peak may be above 0.1 million and likely to be in a range of 0.2–0.3 million. According to this analysis, the peak is likely to occur during the end of June and mid-July. However, this is merely a model prediction; the dynamics, in reality, depends on a myriad of factors which are beyond the scope of this simple model.

#### 4.1.7. *What is causing the local resurgence of positive cases even after the extended lockdown?*

Although the nation is under lockdown, it is observed that the number of positive cases is still growing at large. A distinct feature of this growth is the local resurgence of infections. As gleaned from various news reports, even after several days with a few new cases, suddenly, there had been jumps in the COVID-19 positive cases in quite a few places. In other words, ‘lull’ ‘green’ zones are, all of a sudden, turning into ‘red’ zones. We discuss a few

plausible factors behind the resurgence: (a) Cross-country reverse migration: Due to the lockdown, a large population of migrant workers reeling at the bottom of the economic barrel got stranded in different places without much subsistence. These people started returning to their homes taking desperate measures. During this migration, human-to-human transmission of COVID-19 might have occurred to a great extent due to a lack of social distancing adding fuel to the ‘resurgence’ of infection. (b) Lack of ‘test, trace and contain’: Interestingly, an important aspect of COVID-19 is the number of patients who do not develop any symptoms (Fig. 1b). In India, primarily the testing capacity was devoted to the persons showing typical symptoms of COVID-19. The asymptomatic pool largely remained unnoticed at the initial stages of infection outgrowth which probably contributed to the resurgence of infections. Moreover, it is not sufficient to only isolate the positive cases but to trace all those people who came in contact with the individual tested positive and find the source of infection. This is known as ‘contact tracing’. If the source of the infection is not traceable, this could indicate an insufficient testing or asymptotically positive source. Extensive use of the app-based modern technology may become useful to trace contacts, however, often at the cost of privacy. In India, where a large portion of the population has no ‘digital footprint’, contact tracing becomes even harder. South Korea flattened the infection curve with extensive testing and other mentioned measures. In India, a similar endeavor of a magnitude proportional to its humongous population seems extremely challenging. With the limited capacity and huge population, randomized testing, at least in the infectious neighborhood, is an immediate solution to detect and isolate the asymptomatic individuals.

In both the above-mentioned scenarios, the majority of the infection spreading is likely to be spearheaded by the asymptomatic human hosts who remain undetected due to a lack of randomized testing and come in social contact with others. This means that they would be infecting healthy people unknowingly. According to the WHO and the Indian Council of Medical Research (ICMR), as much as 80% of the infected individual can be asymptomatic. Thus, all the symptomatic cases reported so far contribute to only about 20% of the total infection. Going by the reported number of cumulative infections 61,356 as on May 9, almost all of which are symptomatic, this would correspond to about 245,424 people who also had the virus but did not show any symptoms. Together, about 306,780 people have actually been infected so far in India carrying symptoms or no symptoms. Therefore, the number of people in the country who are still susceptible to the infection is still in the order of billion. One can realize that, with so many active infections, extensive mixing of the countrywide population soon after the lockdown is over

(after 17 May) would cause a huge surge in the total number of infections which is nearly impossible to manage by any health care system. To estimate the asymptomatic population from the model, we rewrite the equations as follows:

$$\begin{aligned}\frac{dS(t)}{dt} &= -S(\beta I_s + \beta' I_a)/N; \\ \frac{dI_a(t)}{dt} &= S(\beta I_s + \beta' I_a)/N - (\alpha + \nu') I_a; \\ \frac{dI_s(t)}{dt} &= \alpha I_a - (\nu + \delta) I_s; \\ \frac{dR(t)}{dt} &= \nu' I_a + \nu I_s; \quad \frac{dD(t)}{dt} = \delta I_s\end{aligned}$$

Here, the total infectious population is segregated into two compartments: (a) symptomatic  $I_s$  and (b) asymptomatic or mildly symptomatic  $I_a$  population. A susceptible person can be infected upon contact with a symptomatic or asymptomatic individual with rates  $\beta, \beta'$  respectively. The infected individual can remain asymptomatic or mildly symptomatic and transit into a symptomatic state with rate  $\alpha$ . The asymptomatic and symptomatic persons can recover at rates  $\nu'$  and  $\nu$  respectively. For a symptomatic individual, death occurs with rate  $\delta$ . For simplicity, we assumed no death for asymptomatic population. We find that with an  $S_0$  in range  $\sim 4$ –6 million and  $R_e \sim 4.0$ , new model data (Fig. S4) matches with the previously plotted population of symptomatic infection (see Fig. 2). The asymptomatic infection peak appears to be about 4 fold larger than the symptomatic infection peak. Thus the current lockdown can only be relaxed in the presence of extensive testing of symptomatic and asymptomatic population and contact tracing.

It is noteworthy to mention that the total number of cases reported in all over India as well as in various Indian states are negligible compared to the total population of the country and states respectively. Besides, the severity of the infection with symptoms is relatively less in India than in the USA and other large European countries. Whether it is due to the effect of hot and humid weather of India or other meteorological parameters such as high UV index, future research would be able to evaluate.

## 5. Summary

The model predicts the infection peak for India at the end of June or the first half of July (Fig. 2) assuming that the social distancing measures will remain in place. The model shows that as the size of the susceptible population increases, the infection peak shifts to a later date (Fig. 2b, inset). Thus, if the human mobility between regions increases, that would lead to an effective expansion in the

number of susceptibles. This increase in susceptible population would not only lead to a surge in infections but also delay the occurrence of the infection peak by a few weeks. Lately, we have been observing an uprise in the new COVID-19 positive cases daily. This may be attributed to the transmission of the pathogen via asymptomatic carriers and reverse migration of migrant workers from one province to another within India. Whether asymptomatic transmission in tandem with elevated human mobility plays a crucial role in the recent infection surge is a topic of our ongoing work [20]. Our results also exemplify that, the best fit of the real data can be obtained for different  $R_e$  values with the difference being non-marginal (Figs. 3, 4). This may be a limitation of the current model setup to zero in on the exactness of characteristic  $R_e$  corresponding to the outspread.

As mentioned earlier, a spatially explicit model considering a network of Indian provinces connected by human mobility, domestic travel from one place to another, and corresponding disease transmission graphs may lead to deeper understandings of the dynamics of the ongoing pandemic in India.

**Acknowledgements** We thank Santanu Bhattacharya, Heiko Rieger, Soumitra Sengupta, Jayanta Kumar Bhattacharjee, Deb Shankar Ray, Sankar Prasad Bhattacharyya, Parongama Sen for exciting discussions and valuable suggestions. Sa.C. and A.S. were supported by a fellowship from the University Grants Commission (UGC), India. Sw.C. thanks Indian Association for the Cultivation of Science, Kolkata for financial support. M.K. was supported by a fellowship from CSIR, India. R.P. thanks IACS for support and Grant No. EMR/2017/001346 of SERB, DST, India for the computational facility.

## References

- [1] W O Kermack and A G McKendrick, Proceedings of the Royal Society of London. Series A **115**, 700 (1927)
- [2] W O Kermack and A G McKendrick, Proceedings of the Royal Society of London. Series A **138**, 55 (1932)
- [3] B Fred, C-C Carlos and F Zhilan, *Mathematical Models in Epidemiology* (Springer, Berlin 2019)
- [4] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction* (Cambridge University Press, Cambridge 2009)
- [5] J Fernández-Villaverde and C I Jones (2020)
- [6] D Fanelli and F Piazza, Chaos, Solitons and Fractals **134** (2020)
- [7] S Mukherjee, S Mondal and B Bagchi, [arXiv:2004.14787v](https://arxiv.org/abs/2004.14787v) (2020)
- [8] K Biswas, A Khaleque and P Sen, [arXiv:2003.07063](https://arxiv.org/abs/2003.07063) (2020)
- [9] Y M Bar-On, A Flamholz, R Phillips and R Milo, eLife (2020)
- [10] X He, E H Y Lau, P Wu, X Deng, J Wang, X Hao, Y C Lau, J Y Wong, Y Guan, X Tan, X Mo, Y Chen, B Liao, W Chen, F Hu, Q Zhang, M Zhong, Y Wu, L Zhao, F Zhang, B J Cowling, F Li and G M Leung, Nature Medicine **26**, 672 (2020)
- [11] R Li, S Pei, B Chen, Y Song, T Zhang, W Yang and J Shaman, Science **368-6490**, 489 (2020)
- [12] H W Hethcote, SIAM Review **42-4**, 599 (2000)

- [13] G Giordano, F Blanchini, R Bruno, P Colaneri, A D Filippo, A D Matteo and M Colaneri, *Nat Med* <https://doi.org/10.1038/s41591-020-0883-7> (2020)
- [14] A Huppert and G Katriel, *Clin Microbiol Infect* **19** (11), 999 (2013)
- [15] S M Kissler, C Tedijanto, E Goldstein, Y H Grad and M Lipsitch, *Science* **368(6493)**, 860 (2020).
- [16] M Gatto, E Bertuzzo, L Mari, S Miccoli, L Carraro, R Casagrandi and A Rinaldo, *PNAS* (2020)
- [17] D K M Ip, L L H Lau, N H L Leung, V J Fang, K-H Chan, D K W Chu, G M Leung, J S M Peiris, T M Uyeki and B J Cowling, *Clinical Infectious Diseases* **64-6**, 736 (2017)
- [18] L Ferretti, C Wymant, M Kendall, L Zhao, A Nurtay, L Abeler-Dörner, M Parker, D Bonsall, and C Fraser, *Science* **368-6491** (2020)
- [19] N G Davies, P Klepac, Y Liu, K Prem, M Jit, C C. working group and R M Eggo, *medRxiv* (2020)
- [20] S Chatterjee, A Sarkar, M Karmakar, S Chatterjee, and R Paul, [arXiv:2006.03034](https://arxiv.org/abs/2006.03034) (2020)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.