Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.
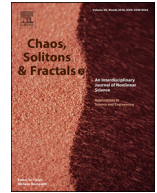
COVID-19: Analytic results for a modified SEIR model and comparison of different intervention strategies

Arghya Das, Abhishek Dhar*, Srashti Goyal, Anupam Kundu, Saurav Pandey

International Center for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore 560089, India

A B S T R A C T

The Susceptible-Exposed-Infected-Recovered (SEIR) epidemiological model is one of the standard models of disease spreading. Here we analyse an extended SEIR model that accounts for asymptomatic carriers, believed to play an important role in COVID-19 transmission. For this model we derive a number of analytic results for important quantities such as the peak number of infections, the time taken to reach the peak and the size of the final affected population. We also propose an accurate way of specifying initial conditions for the numerics (from insufficient data) using the fact that the early time exponential growth is well-described by the dominant eigenvector of the linearized equations. Secondly we explore the effect of different intervention strategies such as social distancing (SD) and testing-quarantining (TQ). The two intervention strategies (SD and TQ) try to reduce the disease reproductive number, R_0, to a target value $R_0^{\text{target}} < 1$, but in distinct ways, which we implement in our model equations. We find that for the same $R_0^{\text{target}} < 1$, TQ is more efficient in controlling the pandemic than SD. However, for TQ to be effective, it has to be based on contact tracing and our study quantifies the required ratio of tests-per-day to the number of new cases-per-day. Our analysis shows that the largest eigenvalue of the linearised dynamics provides a simple understanding of the disease progression, both pre- and post- intervention, and explains observed data for many countries. We apply our results to the COVID data for India to obtain heuristic projections for the course of the pandemic, and note that the predictions strongly depend on the assumed fraction of asymptomatic carriers.

© 2020 Elsevier Ltd. All rights reserved.

The COVID-19 pandemic, that started in Wuhan (China) around December 2019 [1,2], has now affected almost every country in the world. The total number of confirmed cases on July 30, 2020 were close to 17.5 million with close to 686,000 deaths [3]. One of the serious concerns presently is that there is as yet no clear picture or consensus on the future evolution of the pandemic. It is also not clear as to what is the ideal intervention strategy that a government should implement, while also taking into account the economic and social factors. The role of mathematical models has been to provide guidance for policy makers [4–18].

One of the standard epidemiological model is the SEIR model [19] which has four compartments of Susceptible (S), Exposed (E), Infected (I) and Recovered (R) individuals with $S + E + I + R = N$ being the total population of a region (the model can be applied at the level of a country or a state or a city and is expected to work better for well-mixed populations). The SEIR model is pa-

rameterized by the three parameters β, σ and γ that specify the rates of transitions from $S \rightarrow E$, $E \rightarrow I$ and $I \rightarrow R$ respectively. In terms of the data that is typically measured and reported, R corresponds to the total number of cases till the present date, while γI would be the number of new cases per day. The number of deaths across different countries is some fraction ($\approx 1 - 10\%$) of R [20] while the number of hospital beds required at any time would be $\approx$ new cases per day $\times$ typical days to recovery. An important parameter characterizing the disease growth is the reproductive number R_0 [21,22] — when this has a value greater than 1, the disease grows exponentially. Typical values reported in the literature for COVID-19 are in the range $R_0 = 2 - 7$ [23]. For the SEIR model one has $R_0 = \beta/\gamma$ [22].

The two main intervention schemes for controlling the pandemic are social distancing (SD) and testing-quarantining (TQ). Lockdowns (LD) impose social distancing and effectively reduce contacts between the susceptible and infected populations, while testing-quarantining means that there is an extra channel to remove people from the infectious population. These two interven-

* Corresponding author.
E-mail address: abhishek.dhar@icts.res.in (A. Dhar).

tion schemes have to be incorporated in the model in distinctive ways [7,8] — SD effectively changes the infectivity parameter β while TQ changes the parameter γ. Intervention schemes attempt to reduce R_0 to a value less than 1. In the context of the SEIR model with $R_0 = \beta/\gamma$, it is clear that we can reduce it by either decreasing β or by increasing γ. In this work we point out that for the same reduction in R_0 value, the effect on disease progression can be quite different for the two intervention strategies.

Here we analyze an extended version of the SEIR model which incorporates the fact that asymptomatic or mildly symptomatic individuals [7–9,24] are believed to play a significant role in the transmission of COVID-19. Our extended model considers eight compartments of Susceptible (S), Exposed (E), asymptomatic Infected (I_a), presymptomatic Infected (I_p), and a further four compartments (U_a, D_a, U_p, D_p), two corresponding to each of the two infectious compartments. These last four classes comprise of individuals who have either recovered (at home or in a hospital) or are still under treatment or have died — they do not contribute to spreading the infection. We do not include separate compartments for the number of hospitalized and dead since these extra details would not affect our main conclusions.

For this extended SEIR model we first provide, for the case with $R_0 > 1$, analytic expressions for peak infection numbers, time to reach peak values, and asymptotic values of total affected populations. These would provide useful guidance on disease progression. The linearized dynamics is accurate when the total affected population is small. We propose an accurate way of specifying initial conditions for the numerics (from insufficient data) using the fact that the early time exponential growth is well-described by the dominant eigenvector of the linearized equations. We next discuss the performance of two different intervention strategies (namely SD and TQ) in the disease dynamics and control. The aim of interventions is to reduce the reproductive number from it's free value R_0 to a target value $R_0^{\text{target}} < R_0$. We study both the cases, of strong interventions ($R_0^{\text{target}} < 1$ aimed at disease suppression,) and that of weak interventions ($R_0^{\text{target}} \gtrsim 1$, aimed at disease mitigation). Apart from the reproductive number, R_0, an important parameter is the largest eigenvalue of the linear dynamics, which we denote as μ. For $R_0^{\text{target}} > 1$, we have $\mu > 0$ and this gives us the exponential growth rate (doubling time $\approx 0.7/\mu$). On the other hand, for $R_0^{\text{target}} < 1$, the corresponding μ is less than 0 and this tells us that infections will decrease exponentially. We find that, for the same reduction of R_0^{target} to a value less than 1, the corresponding μ magnitude can be very different for different intervention schemes. A larger magnitude of μ, corresponding to a faster suppression of the pandemic, is obtained from TQ than that from SD. An important question for disease control is as to how much testing is required. In our work we relate the parameters of the model related to TQ to testing rates and point out that for TQ to be successful: (a) it has to be based on contact-tracing and (b) it is necessary that testing numbers are scaled up according to the number of new detected cases. Finally, we show that many of our results provide a qualitative understanding of COVID-19 data from several countries which have either achieved disease suppression or mitigation.

The rest of the paper is structured as follows. In Section 1 we define the extended SEIR model. A number of analytic results for the linearized model as well as the full nonlinear system are presented in Section 2. In Section 3, we discuss in detail how intervention strategies such as social distancing and testing-quarantining can be incorporated into the model, we comment on how real testing numbers enter the model parameters. We also present numerical studies comparing different intervention protocols and make qualitative comparisons of the predictions of the SEIR model with real data on the COVID-19 pandemic. In Section 4 we make some heuristic predictions in the Indian context. We summarize our results in Section 5.

1. Definition of the extended SEIR model

The extended SEIR model studied here is schematically described in Fig. 1. It has eight variables ($S, E, I_a, I_p, U_a, D_a, U_p, D_p$) and ten parameters ($\beta_a, \beta_p, \sigma, \gamma_a, \gamma_p, \alpha, \nu_a, \nu_p, r, u$), of which α represents the fraction of asymptomatic carriers while u, r, ν_a, ν_p are related to intervention strategies.

We consider a population of size N that is divided into eight compartments:

1. S = Susceptible individuals.
2. E = Exposed but not yet contagious individuals.
3. I_a = Asymptomatic, either develop no symptoms or mild symptoms.
4. I_p = Presymptomatic, those who would eventually develop strong symptoms.
5. U_a = Undetected asymptomatic individuals who have recovered.
6. D_a = Asymptomatic individuals who are detected because of directed testing-quarantining, may have mild symptoms, and would have been placed under home isolation (few in India).
7. U_p = Presymptomatic individuals who are detected at a late stage after they develop serious symptoms and report to hospitals. Here we assume that all individuals who develop significant symptoms are eventually detected.
8. D_p = Presymptomatic individuals who are detected because of directed testing-quarantining.

We have the constraint that $N = S + E + I_a + I_p + U_a + D_a + U_p + D_p$. A standard dynamics for the population classes is given by the following set of equations:

$$\frac{dS}{dt} = -\frac{u(\beta_a I_a + \beta_p I_p)}{N}S \tag{1}$$

$$\frac{dE}{dt} = \frac{u(\beta_a I_a + \beta_p I_p)}{N}S - \sigma E \tag{2}$$

$$\frac{dI_a}{dt} = \alpha\sigma E - \gamma_a I_a - r\nu_a I_a \tag{3}$$

$$\frac{dI_p}{dt} = (1-\alpha)\sigma E - \gamma_p I_p - r\nu_p I_p \tag{4}$$

$$\frac{dU_a}{dt} = \gamma_a I_a \tag{5}$$

$$\frac{dD_a}{dt} = r\nu_a I_a \tag{6}$$

$$\frac{dU_p}{dt} = \gamma_p I_p \tag{7}$$

$$\frac{dD_p}{dt} = r\nu_p I_p. \tag{8}$$

The parameters in the above equations correspond to

- α: fraction of asymptomatic carriers.
- β_a: infectivity of asymptomatic carriers.
- β_p: infectivity of presymptomatic carriers.
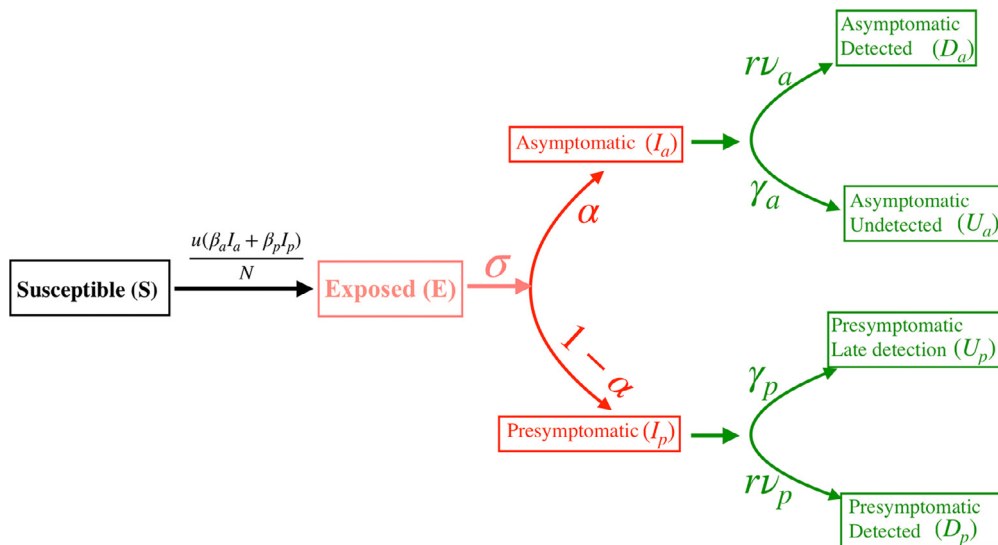- σ: transition rate from exposed to infectious.

Fig. 1. A schematic description of the extended SEIR dynamics studied in this work. The parameters $\beta_a, \beta_p, \sigma, \gamma_a, \gamma_p, \alpha$ are intrinsic to the disease, u quantifies the degree of social distancing while ν_a, ν_p, r are related to intervention arising from testing-quarantining.

- γ_a: transition rate of asymptomatic carriers to recovery or hospitalization.
- γ_p: transition rate of presymptomatics to recovery or hospitalization.
- ν_a, ν_p: detection probabilities of asymptomatic carriers and symptomatic carriers.
- u: intervention factor due to social distancing (time dependence will be specified later).
- r: intervention factor due to testing-quarantining (time dependence to be specified later). This is a rate and depends on testing-quarantining rates.

With our definitions, the total number of confirmed cases, C, and the number of daily recorded new cases F would be

$$C = D_a + D_p + U_p, \quad F = \frac{dC}{dt} = r\nu_a I_a + (\gamma_p + r\nu_p)I_p. \quad (9)$$

Note that we include U_p because these are people who are not detected through directed tests but eventually get detected (after $\sim 1/\gamma_p$ days) when they get very sick and go to hospitals. On the other hand the class D_p get detected because of directed testing, even before they get very sick.

In the next section we will discuss the case where the intervention parameters u and r are kept fixed, and present a number of analytic results. In Section 3 we will discuss the case where u and r are time-dependent.

2. Analytic results for model with constant parameters

2.1. Linear analysis of the dynamical equations

Since at early times $S \approx N$ and all the other populations $E, I_a, I_p, D_a, D_p, U_a, U_p \ll N$, one can perform a linearization of the above equations. This tells us about the early time growth of the pandemic, in particular the exponential growth rate. Let us define new variables to characterize the linear regime: $x_1 = S - N, x_2 = E, x_3 = I_a, x_4 = I_p, x_5 = U_a, x_6 = D_a, x_7 = U_p, x_8 = D_p$. At early times when $x_i \ll N$, the dynamics is captured by linear equations

$$\frac{dX}{dt} = MX, \quad \text{with } X = (x_1, x_2, \ldots, x_8), \quad (10)$$

$$M = \begin{pmatrix} 0 & 0 & -\tilde{\beta}_a & -\tilde{\beta}_p & 0 & 0 & 0 & 0 \\ 0 & -\sigma & \tilde{\beta}_a & \tilde{\beta}_p & 0 & 0 & 0 & 0 \\ 0 & \alpha\sigma & -\tilde{\gamma}_a & 0 & 0 & 0 & 0 & 0 \\ 0 & (1-\alpha)\sigma & 0 & -\tilde{\gamma}_p & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_a & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & r\nu_a & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r\nu_p & 0 & 0 & 0 & 0 \end{pmatrix} \quad (11)$$

where $\tilde{\beta}_a = u\beta_a$, $\tilde{\beta}_p = u\beta_p, \tilde{\gamma}_a = \gamma_a + r\nu_a, \tilde{\gamma}_p = \gamma_p + r\nu_p$. For the present we ignore the time dependence of the SD factor u and the TQ factor r. The matrix has 5 zero eigenvalues while the 3 non-vanishing ones are given by the roots of the following cubic equation for λ:

$$\lambda^3 + (\tilde{\gamma}_a + \tilde{\gamma}_p + \sigma)\lambda^2$$
$$+ [\tilde{\gamma}_a\tilde{\gamma}_p + \tilde{\gamma}_a\sigma + \tilde{\gamma}_p\sigma - \alpha\tilde{\beta}_a\sigma - (1-\alpha)\tilde{\beta}_p\sigma]\lambda$$
$$+ \sigma[\tilde{\gamma}_a\tilde{\gamma}_p - (1-\alpha)\tilde{\beta}_p\tilde{\gamma}_a - \alpha\tilde{\beta}_a\tilde{\gamma}_p] = 0. \quad (12)$$

This can be re-written in the form

$$\lambda^3 + (\tilde{\gamma}_a + \tilde{\gamma}_p + \sigma)\lambda^2 + [\tilde{\gamma}_a\tilde{\gamma}_p + \sigma(\tilde{\gamma}_a + \tilde{\gamma}_p)(1 - Q)]\lambda$$
$$+ \sigma\tilde{\gamma}_a\tilde{\gamma}_p(1 - R_0) = 0, \quad (13)$$

where $\tilde{\beta}_a = u\beta_a, \tilde{\beta}_p = u\beta_p, \tilde{\gamma}_a = \gamma_a + r\nu_a, \ \tilde{\gamma}_p = \gamma_p + r\nu_p, \ Q = \alpha\tilde{\beta}_a/(\tilde{\gamma}_a + \tilde{\gamma}_p) + (1-\alpha)\tilde{\beta}_p/(\tilde{\gamma}_a + \tilde{\gamma}_p)$, and

$$R_0 = \alpha\frac{\tilde{\beta}_a}{\tilde{\gamma}_a} + (1 - \alpha)\frac{\tilde{\beta}_p}{\tilde{\gamma}_p}$$
$$= \alpha\frac{u\beta_a}{\gamma_a + r\nu_a} + (1 - \alpha)\frac{u\beta_p}{\gamma_p + r\nu_p} \quad (14)$$

is the expected form for the reproductive number for the disease. One can intuitively see this as follows. The reproductive number is the average number of secondary infection from one infected individual at the initial phase of the outbreak. In our model, an infected individual may either be asymptomatic or presymptomatic with probabilities α or $(1 - \alpha)$ respectively. On average, While an asymptomatic individual infects $\tilde{\beta}_a/\tilde{\gamma}_a$ number of people, a presymptomatic carrier infects $\tilde{\beta}_p/\tilde{\gamma}_p$ number of people. Consequently the expected number of secondary infection from an arbitrarily chosen infected individual will be R_0 with expression given

by Eq. (14). Noting the fact that $Q < R_0$, it follows that the condition for at least one positive eigenvalue is

$$R_0 > 1. \tag{15}$$

We denote the largest eigenvalue by μ and note that this is uniquely related to the reproductive number R_0 by Eq. (13). At early times the number of cases detected would grow as $\sim e^{\mu t}$. For $R_0 \approx 1$, we expect that the largest eigenvalue is close to zero and, after neglecting the λ^3 and λ^2 terms in Eq. (13), we can read off the value as

$$\mu \approx \frac{\sigma(R_0 - 1)}{1 + \sigma(\tilde{\gamma}_a^{-1} + \tilde{\gamma}_p^{-1})(1 - Q)}. \tag{16}$$

Initial conditions: We discuss here the fact that all initial conditions (which satisfy the condition $S(0) \approx N$) quickly move along the direction of the dominant eigenvector and how this provides us a way to choose the correct initial conditions from the knowledge of one variable (e.g confirmed cases) at an early time. We denote the right and left eigenvectors corresponding to an eigenvalue λ_q by $\phi_q(i)$ and $\chi_q(i)$ respectively. The largest eigenvalue is denoted by μ with corresponding right and left eigenvectors $\phi_m(i)$ and $\chi_m(i)$ respectively. The vector $X = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ can be written as [27]

$$x_i(t) = \sum_q c_q e^{\lambda_q t} \phi_q(i), \tag{17}$$

where the coefficients c_q are determined by putting $t = 0$ and then taking the inner product with the left eigenvector say $\chi_q(i)$. Doing so, we get $c_q = \sum_j \chi_q(j) x_j(0)$ which when substituted back in Eq. (17) gives

$$x_i(t) = \sum_j \sum_q \phi_q(i) \chi_q(j) e^{\lambda_q t} x_j(0)$$
$$\approx \sum_j \phi_m(i) \chi_m(j) e^{\mu t} x_j(0),$$
$$\approx c_m \phi_m(i) e^{\mu t}, \text{ where } c_m = \sum_j \chi_m(j) x_j(0) \tag{18}$$

where the second last line is true at sufficiently large times when only one eigenvalue μ dominates. This proves that the direction of the vector X is independent of initial conditions. In particular, using the explicit form of the dominant eigenvector we find the following relation in the growing phase of the pandemic:

$$\frac{I_a(t)}{I_p(t)} = \frac{\phi_m(3)}{\phi_m(4)} = \frac{\alpha(\mu + \tilde{\gamma}_p)}{(1-\alpha)(\mu + \tilde{\gamma}_a)}. \tag{19}$$

Let us consider the initial condition $X = (-\epsilon, 0, 0, \epsilon, 0, 0, 0, 0)$ so that (noting that $\chi_m(1) = 0$)

$$x_i(t) \approx \epsilon \phi_m(i) \chi_m(4) e^{\mu t} = a_i \epsilon e^{\mu t}, \tag{20}$$

where $a_i = \phi_m(i) \chi_m(4)$. At a sufficiently large time t_l (but still in the very early phase of the pandemic) we equate the observed confirmed number C_0 on some day to $x_6(t_l) + x_7(t_l) + x_8(t_l)$ which therefore gives us the relation

$$\epsilon e^{\mu t_l} = \frac{C_0}{a_6 + a_7 + a_8}. \tag{21}$$

This then tells us that we should start with the following initial conditions, (now counting time $t = 0$ from the day of the observation C_0):

$$x_i(0) = \frac{\phi_m(i)}{\phi_m(6) + \phi_m(7) + \phi_m(8)} C_0. \tag{22}$$

The crucial point is that the leading eigenvector fixes the direction of the growth and then knowledge of linear combination fixes all the
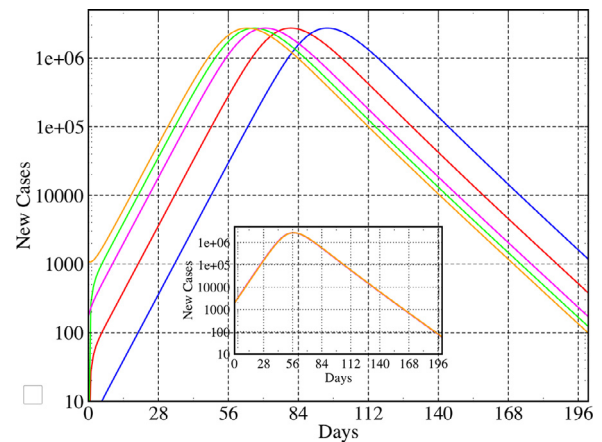

Fig. 2. Role of initial conditions: Plot showing $I(t)$, for a fixed population of size $N = 10^7$, with 5 very different initial conditions : (1)$E(0) = 100, I_a(0) = 0, I_p(0) = 0$, (2)$E(0) = 10, I_a(0) = 0, I_p(0) = 0$, (3)$E(0) = 1000, I_a(0) = 0, I_p(0) = 0$, (4) $E(0) = 233, I_a(0) = 100, I_p(0) = 75$, (5) $E(0) = 233, I_a(0) = 1000, I_p(0) = 75$. (Inset) A collapse of all the curves obtained by translating all the trajectories so that they start with the same value of I.

other coordinates. Thus, independent of initial conditions, the vector describing all the system variables quickly points along the direction of the eigenvector corresponding to the largest eigenvalue [15,28]. Hence if we know any one variable (or a linear combination of all the variables) at sufficiently large times in the growing phase, then the full vector is completely specified. *This leads to an accurate way of specifying initial conditions for the numerics (from insufficient data) and will help in reducing the number of fitting parameters in modeling studies, thereby increasing their accuracy in predicting.*

This fact also implies that different initial conditions (such as different seed infections) will only cause a temporal shift of the observed evolution. This also means that trajectories for different initial conditions are identical up to a time translation. If one uses identical parameters and intervention strategies, then all countries should follow the same trajectory provided they start with the same value for the normalized fraction of confirmed new cases F_0/N. We illustrate this idea, for the extended SEIR dynamics, in Fig. 2 where we show a plot of $I(t) = I_a(t) + I_p(t)$ for 5 different initial conditions. The inset shows a collapse of all the trajectories after an appropriate time translation of the different trajectories. Can we see a similar collapse of the real data for different countries (after normalizing by the respective populations and with appropriate time translation of the data)? In Fig. 3 we plot the data with this normalization and initial condition and see a rough collapse for several countries. The differences can be attributed to different parameter values and different control strategies in different countries. We notice in particular that three of the Asian countries (India, Pakistan, Indonesia) follow a distinctly different trajectory.

2.2. Final affected population

Let us define the asymptotic populations (i.e the populations at very long times) in the different compartments as $\bar{U}_a, \bar{D}_a, \bar{U}_p, \bar{D}_p$, and let $\bar{R}_a = \bar{U}_a + \bar{D}_a$, $\bar{R}_p = \bar{U}_p + \bar{D}_p$, $\bar{R} = \bar{R}_a + \bar{R}_p$. The total population that would eventually be affected by the disease (and either recover or die) is given by $\bar{R}$ and would have developed immunity. A fraction $\bar{U}_a$ (see below) would be undetected and uncounted.

It is possible to compute the final affected population $\bar{R}$ from the dynamical equations in Eqs. (1)–(8). For the moment let us assume that u and r do not have any time dependence. We also assume that $U_a(0) = 0$, $U_p(0) = 0$, $D_a(0) = 0$, $D_p(0) = 0$ and
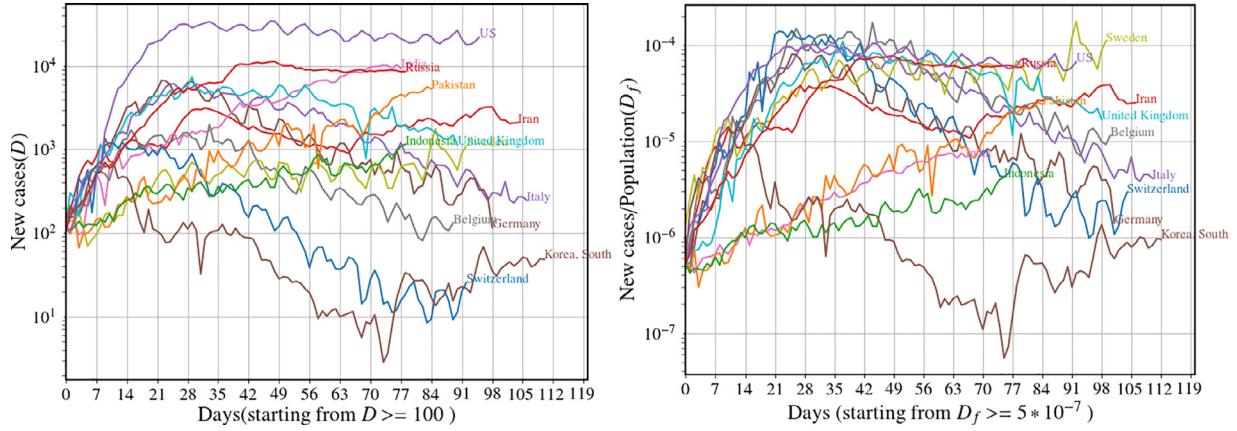
Fig. 3. (left) Number of new cases per day for different countries. (right) Number of new cases normalized by the total population, with the time axis shifted so that every country starts with the same normalized value. Data from [29]

$S(0) \approx N$. Then solving Eq. (1), we get

$$\bar{S} = Ne^{-\int_0^\infty dt \left[\tilde{\beta}_a I_a(t) + \tilde{\beta}_p I_p(t)\right]/N}, \tag{23}$$

where $\tilde{\beta}_a$ and $\tilde{\beta}_p$ are given after Eq. (13). Adding Eqs. (5) and (6) and then multiplying both sides by $\tilde{\beta}_a/\tilde{\gamma}_a$ we get $\frac{\tilde{\beta}_a}{\tilde{\gamma}_a} \frac{dR_a}{dt} = \tilde{\beta}_a I_a$ where $R_a = U_a + D_a$. Similarly, we also get $\frac{\tilde{\beta}_p}{\tilde{\gamma}_p} \frac{dR_p}{dt} = \tilde{\beta}_p I_p$ where $R_p = U_p + D_p$. Plugging these two equations into Eq. (23) then gives

$$\bar{S} = Ne^{-[(\tilde{\beta}_a/\tilde{\gamma}_a)\bar{R}_a/N + (\tilde{\beta}_p/\tilde{\gamma}_p)\bar{R}_p/N]}. \tag{24}$$

Next we note that $(d/dt)(I_a + R_a) = \alpha\sigma E$ and $(d/dt)(I_p + R_p) = (1-\alpha)\sigma E$. Hence, for the initial condition $I_a = I_p = R_a = R_p = 0$, we find that the ratio $[I_a(t) + R_a(t)]/[I_p(t) + R_p(t)] = \alpha/(1-\alpha)$ at all times. Since at large times $I_{a,p} \to 0$, this means that the asymptotic values of R_a and R_p are given by

$$\bar{R}_a = \alpha\bar{R}, \quad \text{and} \quad \bar{R}_p = (1-\alpha)\bar{R}. \tag{25}$$

Using this in Eq. (24), noting that $\bar{S} + \bar{R} = N$ and defining $\bar{x} = \bar{R}/N$, we then get the following simple equation that determines the asymptotic total affected population:

$$1 - \bar{x} = e^{-R_0 \bar{x}}, \tag{26}$$

where $R_0 = \alpha \frac{\tilde{\beta}_a}{\tilde{\gamma}_a} + (1-\alpha)\frac{\tilde{\beta}_p}{\tilde{\gamma}_p}$ is the reproductive number as stated earlier. We note that Eq. (26) has a non-zero solution only when $R_0 > 1$. In Fig. 4 we show the dependence of $\bar{x}$ on R_0, obtained from a numerical solution of Eq. (26). For the simple SIR model the result of Eq. (26) is well known [25], here we show that this is valid for an extended model as well generally. This computation of the asymptotic population can be straightforwardly extended to a more general model where one can have arbitrary number of compartments for the infected and recovered populations.

The asymptotic population of all four compartments within R are thus given by

$$\bar{R}_a = \alpha\bar{R}, \quad \bar{R}_p = (1-\alpha)\bar{R}$$
$$\bar{U}_a = \alpha \frac{\gamma_a}{\gamma_a + rv_a}\bar{R}, \quad \bar{D}_a = (1-\alpha)\frac{rv_a}{\gamma_a + rv_a}\bar{R},$$
$$\bar{U}_p = \alpha \frac{\gamma_p}{\gamma_p + rv_p}\bar{R}, \quad \bar{D}_p = (1-\alpha)\frac{rv_p}{\gamma_p + rv_p}\bar{R}. \tag{27}$$

2.3. Peak infections and the time to reach the peak

The peak infection numbers and the time at which the peak occurs are two important quantities that one would like to know during a pandemic. Here we provide heuristic analytic formulas for these, which are derived after making some simple physical assumptions.

We start with the evolution equations for the infected populations, given by:

$$\frac{dI_a}{dt} = \alpha\sigma E - \tilde{\gamma}_a I_a,$$
$$\frac{dI_p}{dt} = (1-\alpha)\sigma E - \tilde{\gamma}_p I_p. \tag{28}$$

Let us assume that in comparison to the time scale of the pandemic, E, I_a and I_p all peak at roughly the same time say t_m, and let us indicate by $E^{(m)}, I_a^{(m)}, I_p^{(m)}$ the respective peak values. We then obtain

$$\tilde{\gamma}_a I_a^{(m)} = \alpha\sigma E^{(m)}$$
$$\tilde{\gamma}_p I_p^{(m)} = (1-\alpha)\sigma E^{(m)}. \tag{29}$$

Defining $I^{(m)} = I_a^{(m)} + I_p^{(m)}$, Eq. (29) gives

$$E^{(m)} = \frac{\gamma_e}{\sigma}I^{(m)}, \tag{30}$$

where $\gamma_e = [\alpha\tilde{\gamma}_a^{-1} + (1-\alpha)\tilde{\gamma}_p^{-1}]^{-1} \tag{31}$

is an effective recovery rate. Substituting this in Eq. (29) yields

$$I_a^{(m)} = \alpha \frac{\gamma_e}{\tilde{\gamma}_a}I^{(m)},$$
$$I_p^{(m)} = (1-\alpha)\frac{\gamma_e}{\tilde{\gamma}_p}I^{(m)}. \tag{32}$$

Let us further assume that Eq. (32) is valid in general (for all time) and not just at the peak. Then, after defining $I = I_a + I_p$, we have

$$I_a = \alpha \frac{\gamma_e}{\tilde{\gamma}_a}I, \quad I_p = (1-\alpha)\frac{\gamma_e}{\tilde{\gamma}_p}I. \tag{33}$$

Substituting Eq. (33) in the model Eqs. (1)–(8), and defining $R = U_a + D_a + U_p + D_p$, we obtain the standard SEIR equations,

$$\frac{dS}{dt} = -\frac{\beta_e I}{N}S, \tag{34}$$

$$\frac{dE}{dt} = \frac{\beta_e I}{N}S - \sigma E, \tag{35}$$

$$\frac{dI}{dt} = \sigma E - \gamma_e I, \tag{36}$$

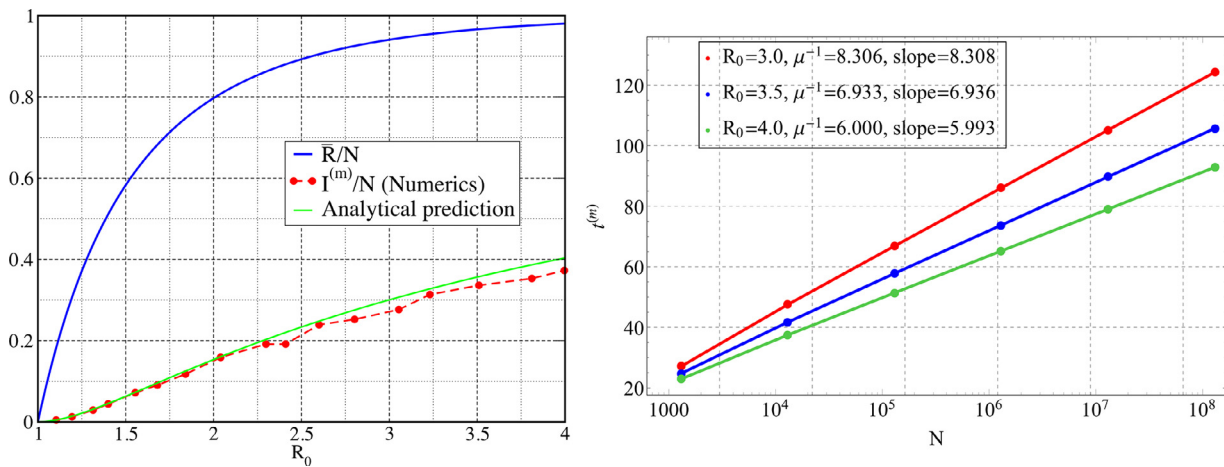$$\frac{dR}{dt} = \gamma_e I, \tag{37}$$

Fig. 4. (left) Plot of the asymptotic total affected population fraction, $\bar{R}/N$, as a function of the reproductive number R_0. The parameters used are $\beta_a = 0.33$, $\beta_p = 0.5$, $\sigma = 0.33$, $\gamma_a = 0.125$, $\gamma_p = 0.083$, $r = 0.0$ and $u = 1.0$. We also plot the quantity $(I^{(m)}/N)(\sigma + \gamma_e)/\sigma$, obtained numerically from many different parameter sets, and compare it with the theoretical predicted curve $1 - (1 + \ln R_0)/R_0$ (green line). (right) Verification of the $\ln(N)$ dependence of $t^{(m)}$ in Eq. (40) for different choices of R_0. The slopes of the straight lines compares well with μ^{-1} as stated in Eq. (40). The parameters used which correspond to Eqs. (34)–(37) are $\sigma = 1/3$, $\gamma_e = 0.1$ for all values of R_0 and $\beta = R_0\gamma_e$. The dominant eigenvalue μ are computed using the linearised version of Eqs. (34)–(37) and slopes are obtained from the data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Here, γ_e is given by Eq. (31) and $\beta_e = R_0\gamma_e$ where the reproductive number R_0 is given by Eq. (14). Assuming the initial conditions: $S(0) \approx N$, $R(0) = 0$, we can solve Eqs. (34) and (37) to get:

$$S = Ne^{-R_0 R/N}. \tag{38}$$

At the peak, our assumption of $dE/dt = 0$ gives from Eq. (35), $S^{(m)}/N = \sigma E^{(m)}/(\beta_e I^{(m)}) = \gamma_e/\beta_e = R_0^{-1}$ [using Eq. (30)]. The peak value $R^{(m)}$ is then given from Eq. (38) by $R^{(m)}/N = R_0^{-1}\ln(R_0)$. Finally, using Eq. (30) and the fact that $S/N + E/N + I/N + R/N = 1$ gives

$$I^{(m)} = \frac{\sigma}{\gamma_e + \sigma}\left(1 - \frac{1 + \log R_0}{R_0}\right)N. \tag{39}$$

For somewhat simpler models it is possible to obtain an analytical estimate of the peak size [16] and for the time, $t^{(m)}$, required to reach the infection peak [26]. In our case we estimate $t^{(m)}$ by noting that the linearized dynamics is approximately valid (see previous section) up to the time $I(t)$ reaches its peak $I^{(m)}$. Hence we write $I^{(m)} = I_0 e^{\mu(t^{(m)} - t_0)}$, where $I_0 = I(t = t_0)$ is the infection number at some early time (but already in the exponentially growing regime). Hence we get

$$t^{(m)} - t_0 = \frac{\ln[I^{(m)}/I_0]}{\mu} \sim \frac{\ln(N/c)}{\mu}, \tag{40}$$

with $I^{(m)}$ given by Eq. (39) and c is a constant that depends on initial conditions and model parameters.

In Fig. 4, we provide a numerical verification of the results in Eqs. (39) and (40) by solving the extended SEIR equations numerically. One of the main assumptions required for the proof is the validity of Eq. (33). In Fig. (5) we check this assumption in a numerical example with a particular parameter set.

3. Interventions: social distancing and testing-quarantining

We next consider the effect of different intervention strategies which are incorporated into the extended SEIR dynamical Eqs. (1)–(8), through the parameters u and r which we will now make time-dependent. We discuss here the choices of the intervention functions u and r. Note that u is a dimensionless number quantifying the level of social contacts, while r is a rate which, as we will see, is closely related to the testing rate.

Social distancing (SD): We multiply the constant factors $\beta_{a,p}$ by the time dependent function, $u(t)$, the "lockdown" function that incorporates the effect of social distancing, i.e reducing contacts between people. A reasonable form is one where $u(t)$ has the constant value ($= 1$) before the beginning of any interventions, and then from time t_{on} it changes to a value $0 < u_l < 1$, over a characteristic time scale $\sim t_w$. Thus we take a form

$$u(t) = 1 \quad t < t_{on},$$
$$= u_l + (1 - u_l)e^{-(t - t_{on})/t_w}, \quad t > t_{on}. \tag{41}$$

The number u_l indicates the lowering of social contacts.

Testing-quarantining (TQ): We expect that testing and quarantining will take out individuals from the infectious population and this is captured by the terms rv_aI_a and rv_pI_p in the dynamical equations. A reasonable choice for the TQ function is perhaps to take

$$r(t) = 0 \quad t < t'_{on},$$
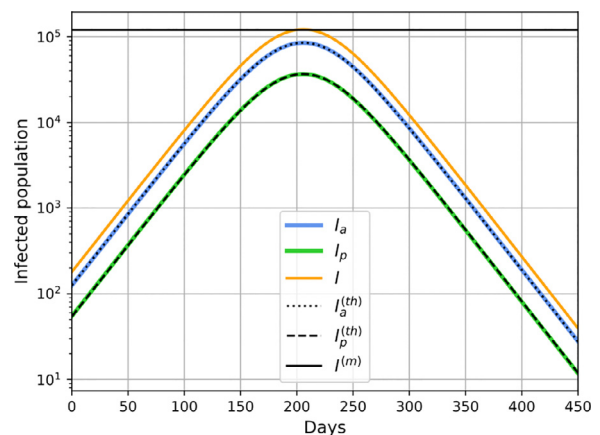$$= r_l - r_le^{-(t - t'_{on})/t'_w}, \quad t > t'_{on}. \tag{42}$$

Fig. 5. Comparison of values of I_a, I_p, I obtained from direct numerics with those from the approximate assumed forms, I_a^{th}, I_p^{th} from Eq. (33) (I taken from the numerics). We also plot the predicted peak value of $I^{(m)}$. The parameters in the numerics were taken as $N = 10^7$, $\alpha = 0.67$, $\beta_a = 0.333$, $\beta_p = 0.5$, $\gamma_a = 1/8$, $\gamma_p = 1/12$, $\sigma = 1/3$, $v_a = 1/3$, $v_p = 1/2$, $u = 1$, $r = 0.5$, which gives $R_0 = 1.26$, $\mu = 0.039$. We see in this case the assumption is valid to a high degree of accuracy.
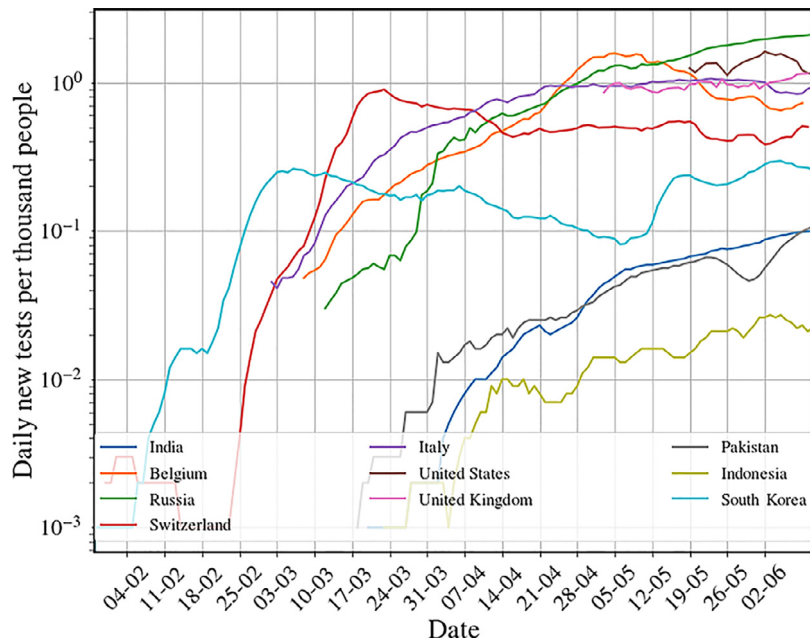
Fig. 6. Data of number of tests per day per thousand in several countries on a log-scale. Data from [20].

where one needs a final rate $r_l > 0$. In general the time at which the TQ begins to be implemented t'_{on} and the time required for it to be effective t'_w could be different from those used for SD.

A useful quantity to characterize the system with interventions is the time-dependent effective reproductive number given by

$$R_0^{\text{eff}}(t) = \alpha \frac{u(t)\beta_a}{\gamma_a + r(t)\nu_a} + (1-\alpha)\frac{u(t)\beta_p}{\gamma_p + r(t)\nu_p}. \qquad (43)$$

At long times this goes to the targeted reproduction number

$$R_0^{\text{target}} = R_0^{\text{eff}}(t \to \infty) = \alpha\frac{u_l\beta_a}{\gamma_a + r_l\nu_a} + (1-\alpha)\frac{u_l\beta_p}{\gamma_p + r_l\nu_p}. \qquad (44)$$

The time scale for the intervention target to be achieved is given by t_w and t'_w.

Relation of the TQ function $r(t)$ to the number of tests done per day: Let us suppose that the number of tests per person per day is given by T_r. We show in Fig. 6 the data for the number of tests per 1000 people per day across a set of countries and see that this is around 0.05 for India which means that $T_r = 0.00005$. If tests are done completely randomly, then the number of detected people (assuming that the tests are perfect) would be $T_r \times I$ and so it is clear that we can identify $r(t) = T_r(t)$. It is then clear that this would have no effect on the pandemic control. To have any effect we would need $r \gtrsim \gamma_p \approx 0.1$ which means around 100 tests per 1000 people per day which is clearly not practical.

However, a better strategy is to do focused tests on the contacts of all those who have been detected on a given day. We now give an estimate of the rate r if we followed this strategy. For simplicity of presentation of our argument we here assume $\nu_a = \nu_p = 1$ and $\gamma_a = \gamma_p$. From our extended SEIR model the number of detected cases per day is given by $F(t) = r\nu_a I_a + (r\nu_p + \gamma_p)I_p = rI + \gamma_p I_p$. In the growing phase we have, from Eq. (19), that $I_a = \alpha I$ and $I_p = (1-\alpha)I$. Hence we get $F(t) = \hat{\gamma}I$ with $\hat{\gamma} = r + (1-\alpha)\gamma_p$. The total number of contacts of the $I = F(t)/\hat{\gamma}$ individuals would be $AF(t)/\hat{\gamma}$, where A is the mean number of contacts of a single infected person. If we perform T tests per day *on this pool*, then the rate of detections will be given by

$$r = \frac{T\hat{\gamma}}{AF(t)} \qquad (45)$$

Denoting $c = T/(AF)$ and noting that $\hat{\gamma} = r + (1-\alpha)\gamma_p$, we self-consistently solve the above equation to find

$$r = \frac{c(1-\alpha)\gamma_p}{1-c}. \qquad (46)$$

Now it is clear that unless r and $\gamma_a = \gamma_p$ are of the same order, TQ will not have much effect on the dynamics and the change in R_0 will be small. Setting $r \gtrsim \gamma_p$ then gives us the condition

$$T \gtrsim \frac{AF(t)}{2-\alpha}. \qquad (47)$$

Note that in our model we identify $r(t)$ as our control rate function that changes from the value 0 to a value $r_l \approx \gamma_p$ over the time scales of a week or so. This means that we would need to change the testing rate in a controlled way such that the condition $T(t) \sim AF(t)$ is maintained. Thus *the number of tests/per day has to be proportional to number of new detections/per day and in fact the ratio T/F has to be larger than the average number of contacts, A, that each infectious person makes*. The number A is expected to depend on the population density and also how well SD is being implemented. The table in Fig. 7 shows data for the ratio $T(t)/F(t)$ for a set of countries and also how this ratio has evolved over time. While the value of $T(t)/F(t) \approx 25$ (around May 15) for India appears to be large, it may not be sufficient given that the population densities are much larger than in many other countries and implementation of SD may be less effective. If we assume 20 contacts a day and the number of days before isolation of the individual to be 5 we get the rough estimate of $A \approx 100$ and then the ratio T/F thus has to be at least ≈ 100. This is the minimum value of testing-to-detected ratio that has to be targeted at localities with high infection rates. The details of the arguments presented here are largely *independent of the specifics of the particular SEIR model that we study*.

3.1. Comparision of intervention strategies

The model details are given in Section 1. We recall that at any given time the total infectious population size is $I = I_a + I_p$, the cumulative affected population (recovered, in hospital or dead) is
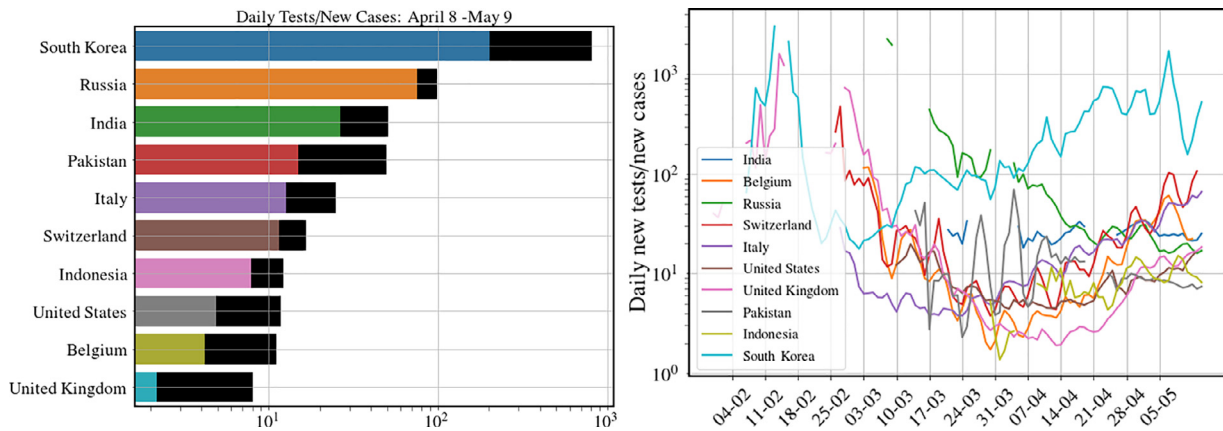
Fig. 7. Data from different countries on the number of tests per detected case $T(t)/F(t)$ for 10 countries (left) on the dates April 8 (colored bar), May 9 (black bar) and (right) the change over time of this ratio. Data from [20]

$R = U_a + D_a + U_p + D_p$, the reported total confirmed cases is $C = D_a + D_p + U_p$, and the reported new daily cases is $F = dC/dt = rv_a I_a + (\gamma_p + rv_p)I_p$.

A useful quantity to characterize the system with interventions is the targeted reproduction number [see Eqs. (43) and (44)]

$$R_0^{\text{target}} = \frac{\alpha u_l \beta_a}{\gamma_a + r_l v_a} + \frac{(1-\alpha)u_l \beta_p}{\gamma_p + r_l v_p}. \tag{48}$$

We classify intervention strategies by the targeted R_0^{target} value. A **strong intervention** is one where $R_0^{\text{target}} < 1$ and will achieve suppression of the disease while a **weak intervention** is one with $R_0^{\text{target}} \gtrsim 1$ and will only mitigate the effects of the disease.

Other than R_0, an important quantity to characterize the disease growth is the largest eigenvalue μ of the linearized dynamics (see Section 2.1). In the early phase of the pandemic, all populations other than S grow exponentially with time as $\sim e^{\mu t}$. As we will see, for the case of strong intervention, μ becomes negative and gives the exponential decay rate of the disease.

In our numerical study we choose, for the purpose of illustration, the following parameter set: $\alpha = 0.67$ and the rates $\beta_a = 0.333, \beta_p = 0.5, \sigma = 1/3, \gamma_a = 1/8, \gamma_p = 1/12$ all in units of day^{-1}. For the specified choice of parameter values (free case with $u = 1.0, r = 0.0$) we get $\mu = 0.158$ which is close to the value observed for the early time data for confirmed cases in India. The corresponding free value of R_0 is 3.7665. Note that μ is not uniquely fixed by R_0 (and vice versa) and different choices of parameters can give the same observed μ but different values of R_0

Choosing these typical parameter values for COVID-19, we now compare the efficacy of strong and weak interventions implemented in four different ways: (1) 6WLD-NTQ: Six weeks lockdown (strong value of SD parameter) and no testing-quarantining, (2) ELD-NTQ: Extended lockdown and no testing-quarantining,(3)NSD-ETQ: No social distancing and extended testing-quarantining, (4) ESD-ETQ: Extended social distancing and extended testing-quarantining. The case with no social distancing and no testing-quarantining is indicated as NSD-NTQ.

We work with a population $N = 10^7$ and initial conditions $E(0) = 100, \quad I_a(0) = I_p(0) = U_a(0) = D_a(0) = U_p(0) = D_p(0) = 0$ and $S(0) = N - E - I_a - I_p - U_a - D_a - U_p - D_p$. In all cases, we will assume that intervention strategies are switched on when the confirmed number of cases reaches 50 and after that the full intervention values are attained over a time scale of 5 days.

3.1.1. Strong intervention ($R_0^{\text{target}} < 1$)

In this case, the exponential growth stops around the time when $R_0^{\text{eff}}(t)$ crosses the value 1. After this time, the infection numbers will start decaying exponentially. Since the infection numbers

are still small compared to the total population, one can work with the linearized theory and the magnitude of the largest eigenvalue μ (now negative) determines the exponential decay rate. For illustrating this case, we take:

Parameter set I [$R_0^{\text{target}} = 0.667$] — We choose three SD and TQ strengths as (i) SD: $u_l = 0.177, r_l = 0$, (ii) TQ: $u_l = 1, r_l = 1.2$ and (iii) SD-TQ: $u_l = 0.461, r_l = 0.4$. This choice corresponds to changing the free value of $R_0 = 3.766$ to a target value $R_0^{\text{target}} = 0.667$, for all the three different strategies. The largest eigenvalue μ changes from the free value $\mu = 0.158$ to the values (i) $\mu = -0.027$, (ii) $\mu = -0.077$ (iii) $\mu = -0.0546$ respectively. The results of the numerical solution of the extended SEIR equations are presented Fig. 8(a) and (b).

Main observations:

1. A six week (or eight week) lockdown is insufficient to end the pandemic and will lead to a second wave. If the interventions are carried on indefinitely, the pandemic is suppressed and only affects a very small fraction of the population (less than 0.1%). We can understand all features of the dynamics from the linear theory. In Fig. 8(a,b), intervention is switched on after ≈ 2 weeks and the peak in infections appears roughly after a period of 5 days. Thereafter however, the decay in the number of infections occurs slowly, the decay rate being given by the largest eigenvalue μ (now negative and smaller in magnitude than μ in the growth phase).

2. *We find that for the same target $R_0^{\text{target}} < 1$, different intervention schemes (ELD-NTQ, NSD-ETQ, or ESD-ETQ) can give very different values of the decay rate μ and, in general we find that TQ is more effective than SD. We see that ELD-NTQ ends the pandemic in about 10 months while NSD-ETQ would take around 3.5 months. This can be understood from the fact that the corresponding μ values (post-intervention) are given by $\mu = -0.027$ and $\mu = -0.077$ respectively, i.e, they differ by a factor of about 3. With a mixed strategy where one allows almost three times more social contacts ($u_l = 0.431$) than for LD case and that requires three times less testing ($r_l = 0.4$) than for TQ case, we see that the disease is controlled in about 5 months. Hence this appears to be the most practical and effective strategy.*

3. The expected time for the pandemic to die would be roughly given by

$$t_{\text{end}} \sim \frac{\ln(\text{Peak infection number})}{|\mu^{\text{post-intervention}}|}, \tag{49}$$

and so it is important that intervention schemes are implemented early and as strongly as possible.
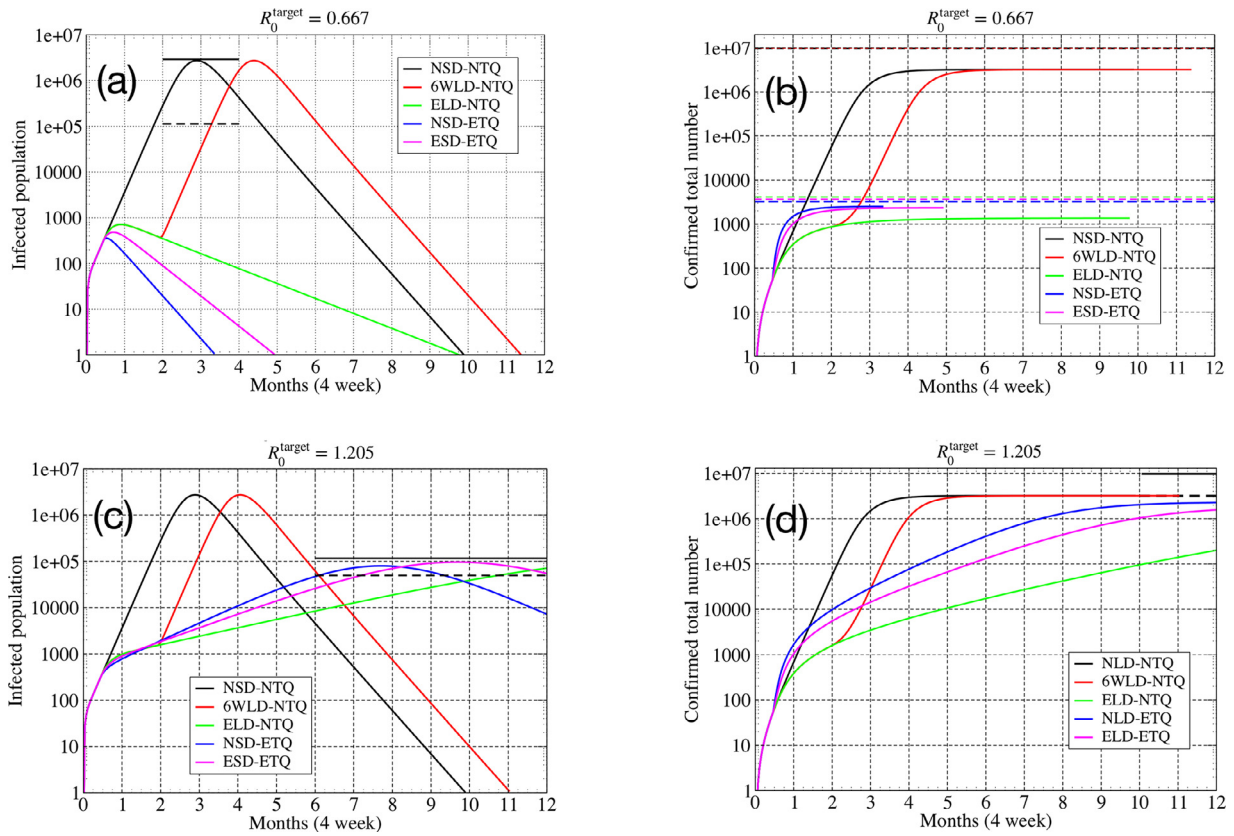
Fig. 8. Parameter set I [$R_0^{\text{target}} = 0.667$]: (a) Total number of infected cases $I = I_a + I_p$ for different intervention strategies. The solid and dashed black lines indicate the peak infected cases $I^{(m)}$ as given by Eq. (39) and the corresponding value of $I_p^{(m)}$. (b) Total number of confirmed cases $C = U_p + D_a + D_p$. The dashed lines indicate the total affected population $R = C + U_a$ at the end of one year, for the different strategies. In the absence of interventions this is close to 96% and is given by Eq. (26). The total population was taken as $N = 10^7$. **Parameter set II [$R_0^{\text{target}} = 1.205$]:** (c) Total number of infected cases $I = I_a + I_p$ for different intervention strategies. (d) Total number of confirmed cases $C = U_p + D_a + D_p$. The dashed lines indicate the total affected population $R = C + U_a$ at the end of one year, for the different strategies. Total population was taken as $N = 10^7$.

3.1.2. Weak intervention ($R_0^{\text{target}} \gtrsim 1$)

In this case, a finite fraction of the population is eventually affected, but the intervention succeeds in reducing this from its original free value and in delaying considerably the date at which the infections peak. We take the following parameter set for this study:

Parameter set II [$R_0^{\text{target}} = 1.205$] — we choose three SD and TQ strengths as (i) SD: $u_l = 0.32, r_l = 0$, (ii) TQ: $u_l = 1, r_l = 0.536$ and (iii) SD-TQ: $u_l = 0.634, r_l = 0.24$. This choice corresponds to changing the free value of $R_0 = 3.766$ to a fixed target value $R_0^{\text{target}} = 1.205$ for all the three different strategies. The largest eigenvalue μ remains positive and changes from the free value $\mu = 0.158$ to the values (i) $\mu = 0.0152$, (ii) $\mu = 0.032$ (iii) $\mu = 0.0248$ respectively. The results of the numerical solution of the extended SEIR equations are presented Fig. 8(c) and (d).

Main observations:

1. We find that in this case the peak infections, peak infection time and the final affected population can be obtained from the analytic expressions given by Eqs. (26), (39) and (40) in terms of the basic disease parameters, using their values after interventions are introduced. We assume that the intervention parameters change from the values $u = 0, r = 0$ to their full strength $u = u_l, r = r_l$ over a short time scale and thereafter remain constant.
2. We find that the peak infection numbers are smallest for the case with ELD-NTQ and they occur at a later stage. These results can also be understood mathematically from the expressions in Eqs. (39) and (40) using the post-intervention values of γ and μ (from the linear theory).

3. We note that while weak interventions can slow down and reduce the impact of the pandemic, they do not lead to development of herd immunity of the population (assuming that all the recovered people develop immunity). It is well known that herd immunity is attained when a fraction $1 - 1/R_0$ of the population has developed immunity. Thus herd immunity in the above example would require that $1 - R_0^{-1} \approx 0.74$, i.e 74% of the population be affected, while Eq. (26) with $R_0^{\text{target}} = 1.205$ predicts that only about 31% of the population is affected.

3.2. Observation of strong and weak intervention in COVID-19 data

In Fig. 9 we give some examples of data for number of new cases for nine countries where we see that some of the qualitative features seen in the model results in Fig. 8(a,b). In particular we see the fast exponential growth phase and then a much slower decay phase for the first six countries which have succeeded in controlling the disease with various levels of success. On the other hand we see that India, Brazil and Pakistan continue to show a positive μ and it is clear that intervention schemes need to be strengthened.

4. Difficulties in making predictions from the extended SEIR model: case study for India

In the following we make some heuristic predictions, based on the analytic results in Eqs. (26), (39) and (40) and the present observed data, for daily new cases in India ($N \approx 1.3 \times 10^9$), in the state of Delhi ($N \approx 1.9 \times 10^7$) and in the city of Mumbai
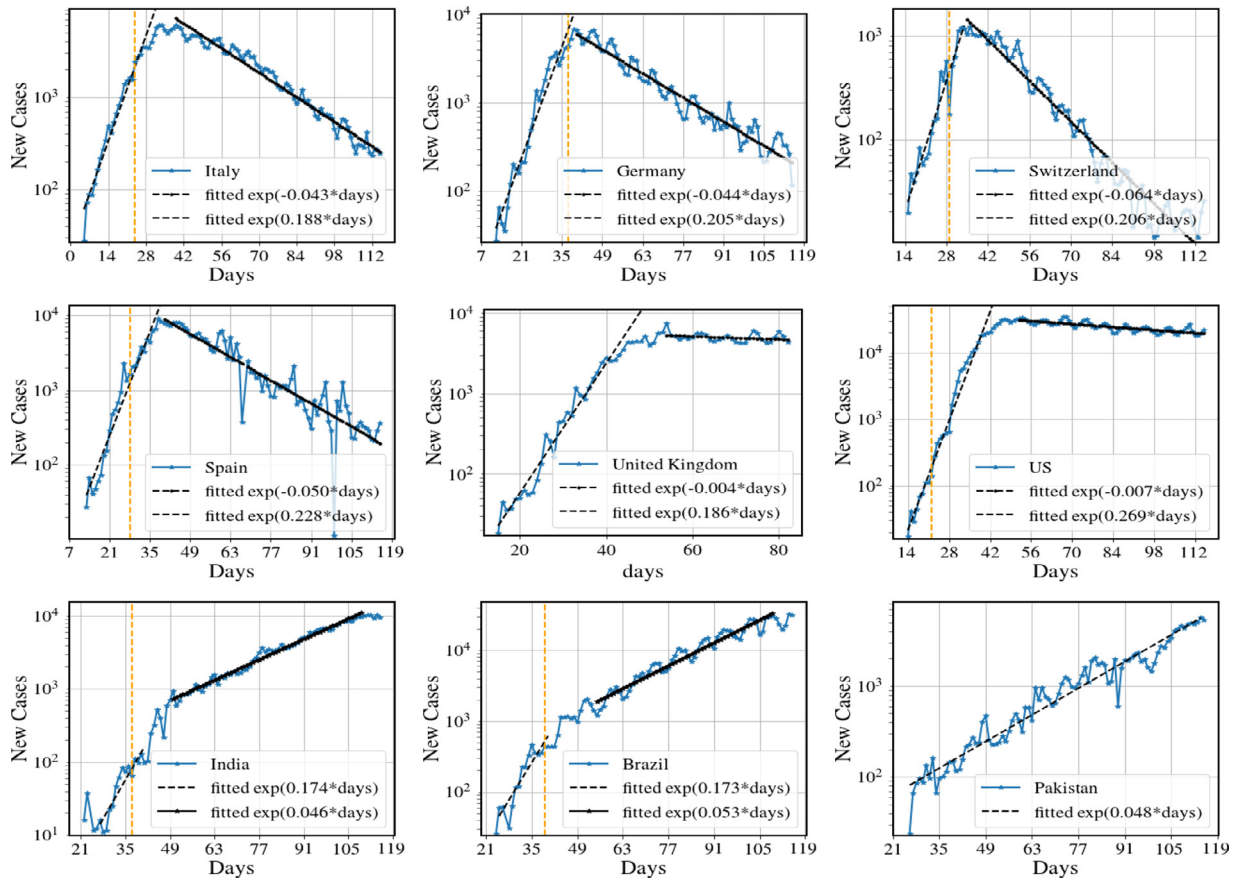
Fig. 9. Number of new cases per day for nine different countries. The dashed orange vertical lines in some of the figures denote the day of the implementation of lockdown. We note that the first six data sets exhibits the same broad features that we see for the model predictions in Fig. 8(a,b). In particular we see the fast exponential growth and slow exponential decrease in new cases (following strong interventions). The two countries UK and US show a very slow decay rate, indicating that disease suppression has barely been achieved. The data for India, Brazil and Pakistan show the behavior corresponding to model predictions in Fig. 8(c,d) and have only been able to achieve mitigation so far ($R_0^{\text{target}} > 1, \mu > 0$). Data from [29] and the end date is June 10. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

($N \approx 1.3 \times 10^7$). The analysis here is based on the assumption of a best case scenario where the value of R_0^{target}, achieved after various intervention schemes is maintained at a constant value.

Here we assume that intervention has effectively been through SD, with $r << \gamma$ being neglected. We consider the following choice of parameter values which appears to be reasonable for getting a conservative estimate: $\sigma = 1/2, \tilde{\beta}_p = \beta, \tilde{\beta}_a = 2\beta/3, \tilde{\gamma}_p = \gamma_p = \gamma, \tilde{\gamma}_a = \gamma_a = 3\gamma/2$, i.e, we assume that asymptomatics are 2/3rd less infectious and recover 3/2 times faster. This gives us [using Eq. (31)] $\gamma_e = \gamma/(1 - \alpha/3)$ and the effective reproductive number as $R_0^{\text{target}} = (1 - 5\alpha/9)\beta/\gamma$. From this last relation we can write $\beta = \gamma R_0^{\text{target}}/(1 - 5\alpha/9)$. Plugging this into the equation for the eigenvalues, Eq. (13), and replacing λ by the observed mean exponential growth rate (we choose the values of $\mu = 0.05$ and $\mu = 0.035$ which are representative of the values observed in India since around April 10), we see that we basically get an equation for R_0^{target} in terms of α, σ, γ and μ. For specific choices of α, σ and γ, the observed values of μ before and after intervention will then give us the corresponding values of R_0.

For our analysis we need to know the total infections $I(0)$ on some day (we take this as April 11) and we estimate it in the following way. Suppose that the daily observed cases on this day was $F_p(0)$ (assuming that only the symptomatics are detected). Then we have $I_p(0) = F_p(0)/\gamma_p$. From Eq. (32) we have $I_p^{(m)} = (1 - \alpha)\gamma_e I^{(m)}/\gamma_p$ and so the time to the peak can be estimated as $t^{(m)} = \mu^{-1} \ln[I_p^{(m)}/I_p(0)]$. We use Eq. (39) to compute the peak

number of infections $I^{(m)}$ and the peak daily cases (PDC) is then obtained as PDC$= F_p^{(m)} = \gamma_p I_p^{(m)} = (1 - \alpha) \times \gamma_e \times I^{(m)}$. The total affected population fraction $\bar{x}$, can be computed from Eq. (26), using only the knowledge of R_0^{target}. If we assume the number of deaths is 1% of all symptomatic cases this gives us an estimate for the total number of deaths as $N\bar{x}(1 - \alpha)/100$.

The observed daily new cases in India, Delhi and Mumbai on April 10 were around $F_p(0) \approx 900$, $F_p(0) \approx 115$ and $F_p(0) \approx 195$ respectively [30]. For a range of choice of the parameters with $\sigma = 0.5, 0.4, \gamma = 0.2, 0.143$ and of $\alpha = 0.67, 0.9$, and two representative values of the post-intervention growth rates, $\mu = 0.05, 0.035$, we compute the corresponding values obtained for R_0 and R_0^{target}. These and the estimates for PDC$= F_p^{(m)}, t^{(m)}$ and $\bar{x}$ are given in Tables 1–3 for $\mu = 0.05$ and in Tables 4–6 for $\mu = 0.035$, for India, Delhi and Mumbai. Note that while the peak numbers and total affected population and deaths simply scale with population size, the time to peak depends on the daily detected numbers on April 10, and this leads to the observed differences in the time to the peak for the three cases. We also note here that changing the initial conditions by about 10% causes a change of few days in the peak time while the other quantities remain unchanged. The full numerical solution in Fig. 10 also shows that the complete suppression of the disease takes more than 6 months after the peak.

In Fig. 10 we show results of a numerical solution of the dynamical equations in presence of intervention (SD) for one of the parameter sets in Table 3 and find excellent agreement with our analytic formula in Eqs. (26), (31) and (39). We see that the pre-

Table 1
Growth rate $\mu = 0.05$: Predictions for India with different choices of parameter values.

σ	γ	α	R_0 (free)	R_0^{target}	Peak daily cases (PDC)	Time of peak	Total affected	Total deaths
0.5	0.2	0.67	2.28	1.33	2,456,630	158 (2nd week September)	45%	1,936,000
0.5	0.2	0.9	2.16	1.3	686,770	132 (3rd week August)	42%	550,700
0.4	0.143	0.67	2.82	1.45	2,956,600	161 (3rd week September)	55%	2,363,700
0.4	0.143	0.9	2.65	1.41	832,890.	136 (4th week August)	52%	676,140

Table 2
Growth rate $\mu = 0.05$: Predictions for Delhi with different choices of the asymptomatic fraction α.

σ	γ	α	R_0 (free)	R_0^{target}	Peak daily cases (PDC)	Time of peak	Total affected	Total deaths
0.5	0.2	0.67	2.28	1.33	35,904	114 (1st week August)	45%	28,296
0.5	0.2	0.9	2.16	1.3	10,037	89 (2nd week July)	42.3%	8048
0.4	0.143	0.67	2.82	1.45	43,212	118 (2nd week August)	55%	34,546
0.4	0.143	0.9	2.65	1.41	12,173	93 (2nd week July)	52%	9882

Table 3
Growth rate $\mu = 0.05$: Predictions for Mumbai with different choices of the asymptomatic fraction α.

σ	γ	α	R_0 (free)	R_0^{target}	Peak daily cases (PDC)	Time of peak	Total affected	Total deaths
0.5	0.2	0.67	2.28	1.33	24,566	97 (3rd week July)	45%	19,360
0.5	0.2	0.9	2.16	1.3	6867	71 (4th week June)	42.5%	5507
0.4	0.143	0.67	2.82	1.45	29,566	100 (3rd week July)	55%	23,637
0.4	0.143	0.9	2.65	1.41	8328	75 (4th week June)	52%	6761

Table 4
Growth rate $\mu = 0.035$: Growth rate $\mu = 0.035$: Predictions for India with different choices of parameter values.

σ	γ	α	R_0 (free)	R_0^{target}	Peak daily cases (PDC)	Time of peak	Total affected	Total deaths
0.5	0.2	0.67	2.28	1.23	1,331,400	208 (1st week November)	34.5%	1,482,000
0.5	0.2	0.9	2.16	1.207	368,809	172 (1st week October)	32%	418,500
0.4	0.143	0.67	2.82	1.31	1,651,200	214 (1st week November)	43%	1,856,300
0.4	0.143	0.9	2.65	1.28	459,776	178 (1st week October)	40.5%	526,222

Table 5
Growth rate $\mu = 0.035$: Predictions for Delhi with different choices of the asymptomatic fraction α.

σ	γ	α	R_0 (free)	R_0^{target}	Peak daily cases (PDC)	Time of peak	Total affected	Total deaths
0.5	0.2	0.67	2.28	1.227	19,458	146 (1st week September)	34.5%	21,661
0.5	0.2	0.9	2.16	1.207	5390	109 (4th week July)	32%	6117
0.4	0.143	0.67	2.82	1.31	24,132	152 (1st week September)	43%	27,131
0.4	0.143	0.9	2.65	1.28	6719	116 (1st week August)	40.5%	7691

Table 6
Growth rate $\mu = 0.035$: Predictions for Mumbai with different choices of the asymptomatic fraction α.

σ	γ	α	R_0 (free)	R_0^{target}	Peak daily cases (PDC)	Time of peak	Total affected	Total deaths
0.5	0.2	0.67	2.28	1.23	13,316	120 (1st week August)	34.5%	14,821
0.5	0.2	0.9	2.16	1.21	3688	84 (1st week July)	32%	4185
0.4	0.143	0.67	2.82	1.31	16,512	126 (2nd week August)	43%	18,563
0.4	0.143	0.9	2.65	1.28	4597	90 (2nd week July)	40.5%	5262

dicted peak time is off by about 10%. The numerics also shows that the complete suppression of the disease takes more than 6 months after the peak.

We point out that the mixed-population assumption of the SEIR model is expected to be more accurate for a smaller population and so the estimates for Delhi and Mumbai would be more reliable than the one for India. For a big and highly in-homogeneous country like India, smaller regions (states or cities) would have different values of μ and R_0 and also different initial conditions, hence the global values would not capture the local dynamics correctly. It is likely that the numbers in Table 1 are an over-estimate of the true future trajectory. For the state of Delhi and the city of Mumbai

these would be more accurate, *however we see that the uncertainty in the true value of α and other parameters leads to a huge uncertainty in the predictions.*

5. Discussion

Several earlier work have discussed, using determinsitic compartmentalized models, the effect of asymptomatic affected population and the effect of intervention measures on the COVID pandemic [7–18]. Here we present a somewhat different choice of compartments and perform a careful quantitative comparison of different intervention strategies. A distinction from earlier studies
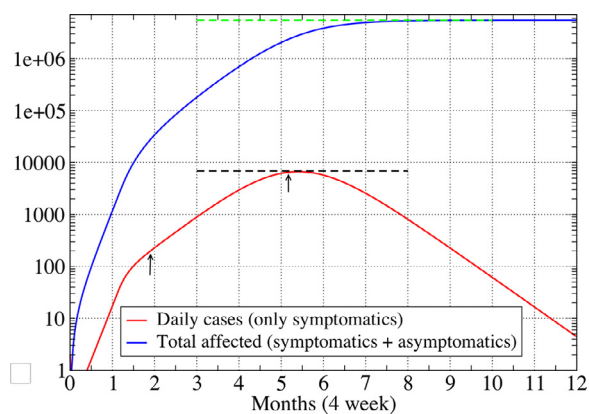

Fig. 10. Plot of the new daily cases ($\gamma_p I_p$) and total affected population fraction (R), as a function number of months for one of the parameter sets in Table 3 for the city of Mumbai. Parameter values were $\sigma = 0.5$, $\tilde{\gamma}_p = 0.2$, $\alpha = 0.9$, $R_0 = 2.16$ before intervention and $R_0^{\text{target}} = 1.3$. The dashed lines give the analytic predictions for the peak daily cases (black line) and the final affected population (green line), and show the good agreement with the numerics. The arrows indicate the date when initial condition was specified $\gamma_p I_p(0) = 195$ and the peak infection date, which occurs about 15 days after the date predicted from Eq. (40). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is that we present a number of analytic results which we believe will be useful beyond their immediate application to epidemiological modeling, in more general studies of population dynamics.

To summarize our findings, a modified version of the SEIR model, incorporating asymptomatic individuals, was analyzed in detail. We have obtained a number of analytical results for the full nonlinear model which can be useful in making empirical estimates of various important quantities that provide information on disease progression. We believe that the derivation of these results can be extended to more sophisticated SIR type models including more compartments and more complex interactions. From the linearized dynamics we point out a simple but important property, namely that at early times the motion of the system quickly settles along the direction of the dominant eigenvector. This allows one to determine accurately initial conditions from sparse data. We provided numerical examples to illustrate these ideas and in addition have provided comparisons with real COVID-19 data. Looking at COVID-19 data in several countries, we find that the extended SEIR model captures some important qualitative features and hence could provide guidance in policy-making

We used the extended SEIR model for analyzing the effectiveness of different intervention protocols in controlling the growth of the COVID-19 pandemic. Non-clinical interventions can be either through social distancing or testing-quarantining. Our results indicate that a combination of both, implemented over an extended period may be the most effective and practical strategy. We have attempted to relate real testing rates to the parameters of the model and comment on what the minimum testing rates should be in order for testing-quarantining to be an effective control strategy.

Finally we have used our analytic formulas to make predictions for disease peak numbers and expected time to peak for India, the state of Delhi and the city of Mumbai, pointing out that these predictions would be highly unreliable for India (due to big inhomogeneity in disease progression across the country) and perhaps more reliable for the cases of Delhi and Mumbai. Our main conclusion here is that the lack of precise knowledge of the disease parameters (e.g the fraction of asymptomatic carriers) and changing control strategies lead to rather large uncertainties in the predictions. Nevertheless, we believe that they could perhaps be used to obtain reasonable bounds.

References

[1] Huang C, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497–506.
[2] Coronavirus N.. (2019-nCoV) situation report dated 21 January 2020, world health organization. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf.
[3] https://www.worldometers.info/coronavirus/.
[4] Kucharski AJ, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020;20:533–8. https://doi.org/10.1016/S1473-3099(20)30144-4
[5] Prem K., Liu Y., Russell T.W., et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. lancet public health. 2020. (published online March 25.) https://doi.org/10.1016/S2468-2667(20)30073-6.
[6] Ferguson N., et al. Report 9:impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. 2020. https://doi.org/10.25561/77482.
[7] Tang B, et al. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCoV). Infect Dis Model 2020;5:248.
[8] Giordano G, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in italy. Nat Med 2020. https://doi.org/10.1038/s41591-020-0883-7
[9] Gatto M., et al. Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. Proceedings of the National Academy of Sciences 202004978. 10.1073/pnas.2004978117.
[10] Ray D., et al. Predictions, role of interventions and effects of a historic national lockdown in India's response to the COVID-19 pandemic: data science call to arms. 2020;medRxiv, https://doi.org/10.1101/2020.04.15.20067256.
[11] Sarkar K., Khajanchi S.. Modeling and forecasting of the COVID-19 pandemic in India. 2020. arXiv:2005.07071
[12] Shekatkar S., et al. INDSCI-SIM: a state-level epidemiological model for India. https://indscicov.in/indscisim.
[13] Agrawal S., et al. COVID-19 epidemic: Unlocking the lockdown in India (working paper), IISc-TIFR technical report. 2020. https://covid19.iisc.ac.in/wp-content/uploads/2020/04/Report-1-20200419-UnlockingTheLockdownInIndia.pdf.
[14] Venkateswaran J., Damani O.. Effectiveness of testing, tracing, social distancing and hygiene in tackling COVID-19 in India: a system dynamics model. 2020. arXiv:2004.08859
[15] Frank TD. COVID-19 order parameters and order parameter time constants of Italy and China: a modeling approach based on synergetics. J Biol Syst 2020;28:589–608.
[16] Piovella N. Analytical solution of SEIR model describing the free spread of the COVID-19 pandemic. Chaos Solitons Fractals 2020;140:110243.
[17] Meidan D., et al. Alternating quarantine for sustainable mitigation of COVID-19. 2020. arXiv.org:2004.01453
[18] Wong G.N. et al. Modeling COVID-19 dynamics in illinois under non-pharmaceutical interventions. 2020. arXiv:2006.02036
[19] Li MY. An introduction to mathematical modeling of infectious diseases. Springer; 2018.
[20] Coronavirus (COVID-19) testing. https://ourworldindata.org/coronavirus-testing.
[21] Giesecke J. Modern infectious disease epidemiology. CRC Press; 2017.
[22] Andersson H, Britton T. Stochastic epidemic models and their statistical analysis. New York: Springer; 2000.
[23] Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med 2020;27(2). https://doi.org/10.1093/jtm/taaa021

[24] Leung KY, Trapman P, Britton T. Who is the infector? Epidemic models with symptomatic and asymptomatic cases. Math Biosci 2018;301:190.

[25] Brauer F, Driessche PV, Wu J. Lecture notes in mathematical epidemiology. Springer; 2008.

[26] J Li J, Lou Y. Characteristics of an epidemic outbreak with a large initial infection size. J Biol Dyn 2016;10:366. https://doi.org/10.1080/17513758.2016.1205223

[27] Lax PD. Linear algebra and its applications. Wiley; 2013.

[28] Frank TD. COVID-19 interventions in some european countries induced bifurcations stabilizing low death states against high death states: an eigenvalue analysis based on the order parameter concept of synergetics. Chaos Solitons Fractals 2020;140:110194.

[29] Johns hopkins center for systems science and engineering coronavirus COVID-19 global cases. https://systems.jhu.edu/research/public-health/ncov/.

[30] Data for India and Delhi are from the website. https://www.kaggle.com/sudalairajkumar/covid19-in-india/data#covid_19_india.csv. The numbers for Mumbai are estimated from newspaper reports.