

Covid-19: an analysis of an extended SEIR model and a comparison of different intervention strategies.

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

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

Modeling accurately the evolution and intervention strategies for the Covid-19 pandemic, that has now affected almost every country in the world, is a challenging problem. We present here an analysis of an extended Susceptible-Exposed-Infected-Recovered (SEIR) model, that takes into account the presence of asymptomatic carriers, and explore the effects 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 ($R_0 > 1$) 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 target value $R_0^{\text{target}} < 1$, 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 the ratio of tests/per day to the number of new cases/per day has to be scaled with the mean number of contacts that an infectious person has, which would be high in regions with high population density and low levels of social distancing. We point out that, apart from R_0 , an important quantity is the largest eigenvalue of the linearized dynamics which provides a more complete understanding of the disease progression both pre- and post- intervention and explains observed data. Weak extended intervention strategies (that reduce R_0 but not to a value less than 1) can reduce the peak values of infections and the asymptotic affected population. We provide simple expressions for these in terms of the disease parameters and apply them in the Indian context to obtain heuristic projections for the course of the pandemic. Looking at real data, we find that for many countries, several broad qualitative features are captured well by the model.

[*The authors in this work are listed alphabetically]

I. INTRODUCTION

The Covid-19 pandemic, that started in Wuhan (China) around December 2019, has now affected almost every country in the world. The total number of confirmed case on May 13 was close to 4.5 million with close to 300,000 deaths. 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 the ideal intervention strategy that a government should implement, while taking into account also economic and social factors. The role of mathematical models has been to provide guidance for policy makers [1–10].

One of the standard epidemiological model is the SEIR model which has four classes 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 in any well-mixed population). The SEIR model is parameterized by the rates β (infectivity), σ , specifying $E \rightarrow I$ transitions and γ specifying $I \rightarrow R$ transitions. 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 would be some fraction ($\approx 5\%$) of R while the number of hospital beds required at any time would be \approx (new cases \times typical days to recovery). An important

parameter characterizing the disease growth is the reproductive number R_0 — when this has a value > 1 , the disease grows exponentially. Typical values reported in the literature for Covid-19 are in the range $R_0 = 2 - 7$ [11]. For the basic SEIR model one has $R_0 = \beta/\gamma$.

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 intervention schemes have to be incorporated in the model in distinctive ways [3, 4] — SD effectively changes the infectivity parameter β while TQ changes the parameter γ . Intervention schemes attempt to reduce this 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 R_0 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 intervention strategies in an extended version of the SEIR model which incorporates the fact that asymptomatic or mildly symptomatic individuals [3–5, 12] are believed to play a significant role in the transmission of Covid-19. The extended model considers eight compartments of Susceptible (S), Exposed (E), asymptomatic Infected (I_a), presymptomatic Infected (I_a), and a further four compartments (U_a, D_a, U_p, D_p),

two each corresponding to 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 discuss the performance of two different intervention strategies (namely SD and TQ) in the disease dynamics and control. We specifically point out that strong interventions ($R_0^{\text{target}} < 1$, aimed at disease suppression) can be understood using the linearized dynamics, while for weak interventions ($R_0^{\text{target}} \gtrsim 1$, aimed at disease mitigation), one has to go beyond the linear theory.

Main conclusions: 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. For the case where only mitigation is achieved ($R_0^{\text{target}} \gtrsim 1$), we present analytic expressions for peak infection numbers, time to reach peak values, and asymptotic values of total affected populations, for the extended SEIR model. These provide useful guidance on disease progression and we apply it in the Indian context. We also show 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. We give conditions 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. We show that the above picture gives us a comprehensive understanding of data from several countries which have either achieved disease suppression or mitigation.

A note on the Indian situation: The number of daily new cases in India continues to rise and it is clear that only mitigation has been achieved, unlike in Europe and the US which have succeeded in suppression ($R_0 < 1$). The current disease doubling time is around 14 days. For two different choices for the the fraction of asymptomatics (and typical values of disease parameters) our estimates suggest a current value of $R_0^{\text{eff}} \approx 1.3$ and that the disease would peak between July to September. The predicted number of hospitalizations and deaths per day (assuming 1% deaths for symptomatic cases) have a large uncertainty but could be quite large (see Tables (I-III)), and there is an urgent need of preparing for this. However, the lockdown in India is now being eased.

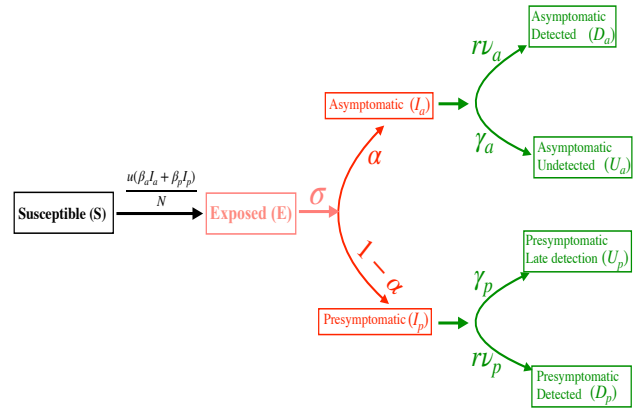


FIG. 1. A schematic description of the extended SEIR dynamics studied in this work. The parameters $\beta_a, \beta_p, \sigma, \gamma_a, \gamma_p, \alpha$ are intrinsic to the disease, u quantifies the degree of social distancing while ν_a, ν_p, r are related to intervention arising from testing-quarantining.

Given the huge economic and social costs of implementing SD, it is clear that a combination of weaker SD but intense TQ might be the only practical way of controlling the pandemic in India. A sustained and targeted testing and quarantining strategy (assuming community spreading is still limited), combined with some level of social-distancing has to be implemented at the earliest and to the fullest extent. Community transmission is unlikely to have taken place in all states and cities in India. Increasing the testing-to-detected ratio to a value around $\gtrsim 100$ from the current value of ≈ 25 could result in lowering of R_0 and the μ value or at least in not letting them increase further.

The paper is structured as follows. In Sec. (II) we state our main results. In Sec. (III) we make qualitative comparisons of the predictions of the SEIR model with real data on confirmed number of cases and make some heuristic predictions in the Indian context. In Sec. (IV) we define and then present the analysis of the extended SEIR model with and without interventions. We present some useful analytic results and closed form expressions as well as numerical results on different intervention protocols. We summarize our results in Sec. (V). Technical details and various analytical results are presented in two appendices (Sec. (A) and Sec. (B)).

II. RESULTS

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 ($\beta_a, \beta_p, \sigma, \gamma_a, \gamma_p, \alpha, \nu_a, \nu_p, r, u$). The details of the model are defined and explained in Sec. (IV). We note 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 $R = U_a + D_a + U_p + D_p$, the

reported total confirmed cases is $C = D_a + D_p + U_p$, and the reported new daily cases is $D = dC/dt = r\nu_a I_a + (\gamma_p + r\nu_p) I_p$.

The parameter u quantifies the degree of social distancing while r is related to the rate at which testing-quarantining is done. These are in general time-dependent, u changing from the free (without interventions) value $u = 1$ to a target value $u_l < 1$, while r is a rate that changes from 0 to a value $r_l > 0$. The time-scale for the change depends on how efficiently the control measures are implemented.

How is r related to testing rates? It is easy to see that with random testing of the population, intervention can be helpful only if a finite fraction of the population is tested, which is typically impossible to implement. Thus testing has to be based on contact tracing of the new detected cases. Suppose that the number of tests per day is T while the number of new cases is D and A the typical number of contacts made by a person over the period when the person is infectious and before detection.

Main result (I): We argue that TQ intervention can be successful only if one achieves

$$T(t) \approx \frac{r(t)AD(t)}{\gamma_p}, \quad (1)$$

where $r(t)$, our control rate function changes from the value 0 to a value r_l , which should be at least of the same order as γ_p , over the time scales of a week or so. This means that we need $T(t) \approx AD(t)$, that is *the number of tests/per day has to be to number of new detections/per day and in fact the ratio T/D 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. Hence, while $T(t)/D(t) \approx 25$ 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/D thus has to be at least ≈ 100 . This is the minimum value of testing-to-detected ratio that has to be targeted at the hot zones. *The argument presented here is largely independent of the details of the particular SEIR model that we study*

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

$$R_0^{\text{eff}}(t) = \alpha \frac{u(t)\beta_a}{\gamma_a + r(t)\mu_a} + (1 - \alpha) \frac{u(t)\beta_p}{\gamma_p + r(t)\nu_p}. \quad (2)$$

At long times this goes to the targeted reproduction number

$$R_0^{\text{target}} = \frac{\alpha u_l \beta_a}{\gamma_a + r_l \mu_a} + \frac{(1 - \alpha) u_l \beta_p}{\gamma_p + r_l \nu_p}. \quad (3)$$

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.

In our numerical study we choose, as an illustrative example, the parameter set $\alpha = 0.67$ while 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 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.

A. Strong intervention: $R_0^{\text{target}} < 1$

For the *strong intervention* case the exponential growth stops around the time $t^{(\text{int})}$ when $R^{\text{eff}}(t)$ crosses the value 1, provided it is imposed early enough. 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.

We work with a population $N = 10^7$ and initial conditions $E(0) = 100, I_a(0) = I_p(0) = U_a(0) = D_a(0) = U_p(0) = D_p(0) = 0$ 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 interventions are attained over a time scale of 5 days.

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 correspond 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.

Parameter set II [$R_0^{\text{target}} = 0.904$]: We choose three SD and TQ strengths as (i) SD: $u_l = 0.24, r_l = 0$, (ii) TQ: $u_l = 1, r_l = 0.809$ and (iii) SD-TQ: $u_l = 0.6, r_l = 0.373$. This choice correspond to changing the free value

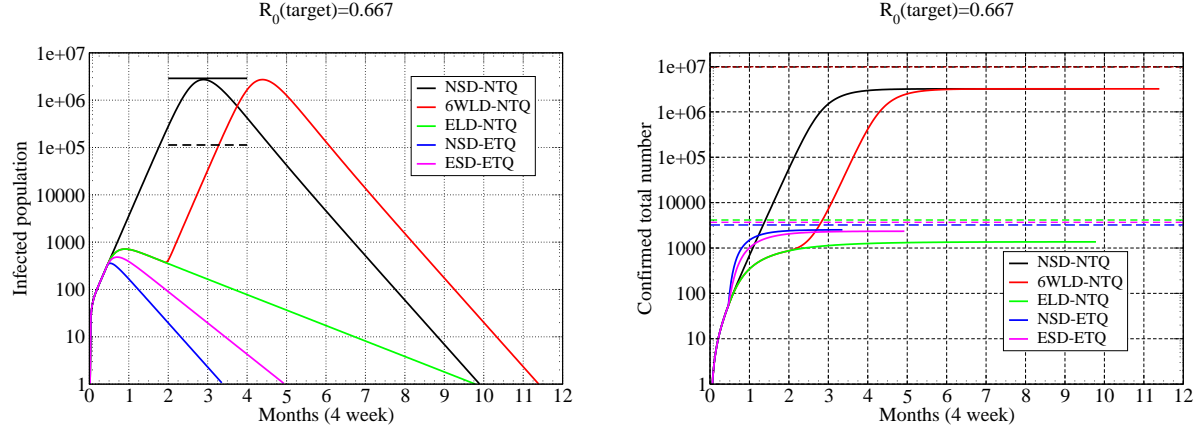


FIG. 2. **Parameter set I** [$R_0^{\text{target}} = 0.667$]: (left) Total number of infected cases $I = I_a + I_p$ for different intervention strategies. The solid and dashed black lines indicate the peak infected cases ($I^{(m)}$) as given by Eq. (5) and the corresponding value of $I_p^{(m)}$. (right) 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. (7). The total population was taken as $N = 10^7$.

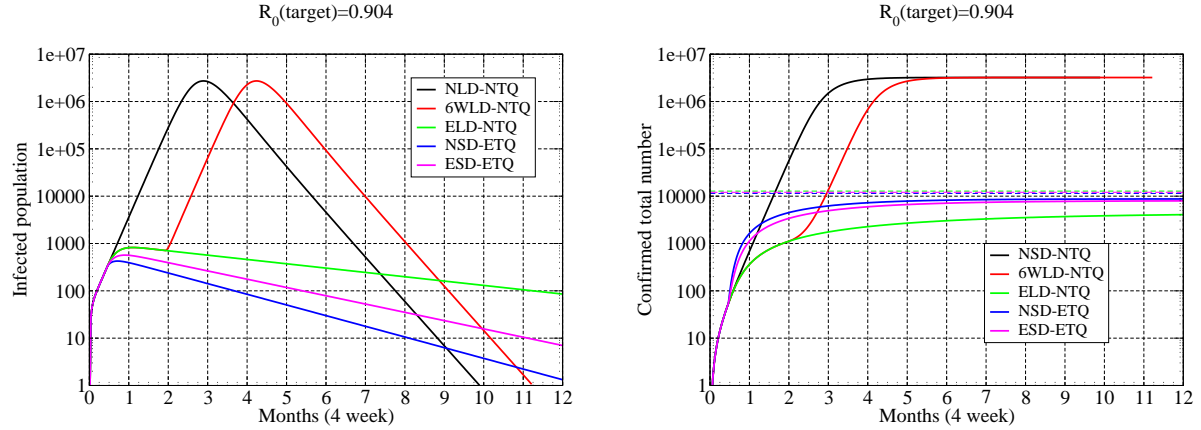


FIG. 3. **Parameter set II** [$R_0^{\text{target}} = 0.904$]: (left) Total number of infected cases $I = I_a + I_p$ for different intervention strategies. (right) Total number of confirmed cases $C = U_p + D_a + D_p$. The dashed lines indicate the total affected population $R = C + U_a$ at the end of one year, for the different strategies. Total population was taken as $N = 10^7$.

of $R_0 = 3.766$ to a fixed target value $R_0^{\text{target}} = 0.904$ for all the three different strategies. The largest eigenvalue μ changes from the free value $\mu = 0.158$ to the values (i) $\mu = -0.00744$, (ii) $\mu = -0.0184$ (iii) $\mu = -0.0142$ respectively.

The results of the numerical solution of the SEIR equations are presented Figs. (2,3).

Observations: A six weeks (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. (2), intervention is switched on after ≈ 2

weeks and the peak in infections shows 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).

Main result (II): 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.

1. For the case of the strongest intervention (I), 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

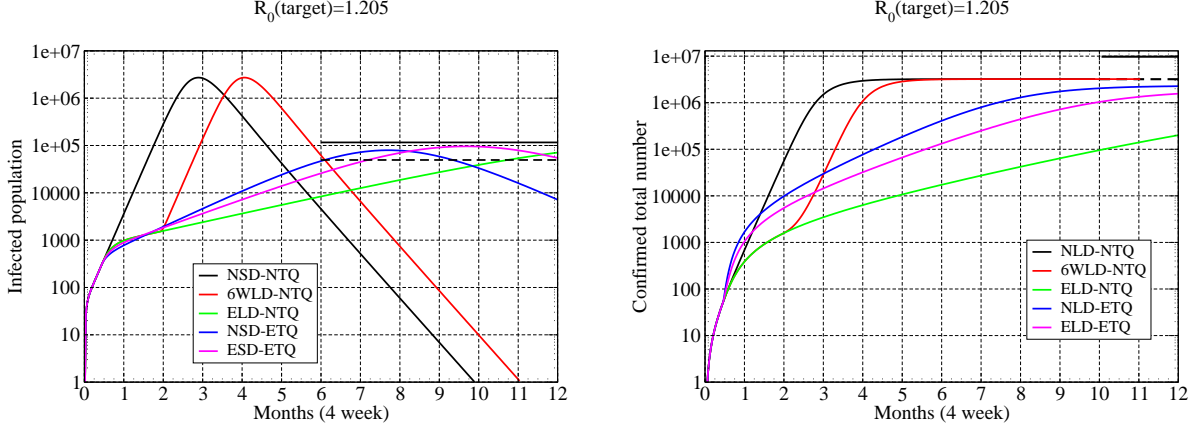


FIG. 4. **Parameter set III** [$R_0^{\text{target}} = 1.205$]: (left) Total number of infected cases $I = I_a + I_p$ for different intervention strategies. (right) Total number of confirmed cases $C = U_p + D_a + D_p$. The dashed lines indicate the total affected population $R = C + U_a$ at the end of one year, for the different strategies. Total population was taken as $N = 10^7$.

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.

2. For the case of the less stronger intervention (II), we see that NSD-ETQ ends the pandemic in about a years time while the other strategies would take much longer. The post-intervention μ values indicate that ELD-NTQ would take about 28 months while the ESD-ETQ combination would take about 15 months.
3. The expected time for the pandemic to die would be roughly given by

$$t_{\text{end}} \sim \frac{\ln(\text{Peak infection number})}{|\mu^{\text{post-intervention}}|}, \quad (4)$$

and so it is important that intervention schemes are implemented early and as strongly as possible.

B. Weak intervention: $R_0^{\text{target}} \gtrsim 1$

Parameter set III [$R_0^{\text{target}} = 1.205$]: 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.

1. In this case, a finite fraction of the population is affected. The intervention succeeds in reducing the fraction that is eventually affected and the peak number of infections and in delaying considerably the date at which the peak occurs.

Main result (III): We point out that these modified values can be obtained from the following simple expressions in terms of basic disease parameters. One can use these formulas either using the pre-intervention or post-intervention values of R_0 and μ . Assuming that we start with a small seed infected or exposed population and almost the entire population susceptible, i.e $S(0) \approx N$, the peak value of infections, $I^{(m)}$, (which is proportional to the number of hospitalizations required) and the number of days, $t^{(m)}$, to reach this peak value are given by the simple general relations:

$$I^{(m)} \approx \frac{\sigma}{\gamma_e + \sigma} \left(1 - \frac{1 + \ln R_0}{R_0} \right) N. \quad (5)$$

$$t^{(m)} \approx \frac{\ln[I^{(m)}/I(0)]}{\mu} \approx \frac{\ln(N/c)}{\mu}, \quad (6)$$

where γ_e is an effective recovery rate [see Eq. (23) in Sec. (IV B)] and c is a constant that depends on initial infected population and other disease parameters. The fraction of population, $\bar{x} = R(t \rightarrow \infty)/N$, that is eventually affected is given by the solution of the equation

$$1 - \bar{x} - e^{-R_0 \bar{x}} = 0, \quad (7)$$

this result being valid for very general SEIR-type models with multiple compartments (see Appendix (A)). These relations are useful — for example, they give good estimates for the typical numbers for peak infections and when they happen (see

below). We also provide relations for estimating the number of asymptomatic infected and recovered individuals.

2. We find that the peak infection numbers are smallest for the case with ELD-NTQ and occur at a later stage. Again these results can be understood mathematically from the expressions in Eq. (5) and Eq. (6) 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. (7) with $R^{\text{target}} = 1.205$ predicts that only about 31% of the population is affected.

C. Linearized theory

Main result (IV): As already observed the linear theory is very useful to understand the growth and also decay time scales of the pandemic following strong interventions. Another observation that we make is that, independent of initial conditions, the vector describing all the system variables quickly points along the direction of the eigenvector corresponding to the largest eigenvalue. Hence (at such longish times) if we know one variable (or a linear combination), then the full vector is completely specified. This leads to an accurate way of specifying initial conditions for the numerics (from insufficient data). This implies that different initial conditions (such as different seed infections) will only cause a temporal shift of the observed evolution. This means that if we plot data for different countries, starting from the same initial value of say the confirmed number of cases (normalized by the population), we should see a collapse of the data. We test this idea and find that indeed an approximate collapse of data is obtained for a number of countries (see next section).

III. DISCUSSION

We do not attempt a detailed comparison of the model predictions with real data since there are too many poorly known parameters and possibly quite inaccurate knowledge of the initial conditions of the variables themselves. We make some overall qualitative observations relating real data to the predictions from SEIR-type models and find that in many cases, several broad qualitative features are remarkably well captured by the model.

Observation of strong and weak intervention in Covid-19 data: In Figs. (5) 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 Figs. (2,3,4). 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 Russia continue to show a positive μ and it is clear that intervention schemes need to be strengthened.

One issue is that different countries start with different initial conditions (for example the seed exposed population could be very different between countries). As discussed in Sec. (A 2), as long as the number of confirmed cases is much smaller than the population size, a description in terms of the linearized dynamics is accurate. This would predict an initial exponential growth and then as intervention schemes begin to operate, the reproductive number and the corresponding growth exponent would decrease till eventually one is able to achieve $R_0^{\text{target}} < 1$ and correspondingly $\mu < 0$. In Fig. (7) we show data for the reported number of new cases in 12 different countries and approximately see these features. Most countries have succeeded in disease suppression $R_0^{\text{target}} < 1$, but show a slow exponential decay of the disease. A few Asian countries (India, Pakistan, Indonesia) have not yet entered the decaying phase which means that intervention has been weak and only disease mitigation has been achieved. This means that with the same level of intervention strategy, a finite fraction of the population will eventually be affected in these countries. We discuss later below in some more detail the Indian situation.

Comparing data across different countries: The linearized SEIR dynamics also predicts that (see Sec. (A 2)), if one uses similar parameters and intervention parameters, then all countries should follow the same trajectory provided they start with the same value for the normalized fraction of confirmed new cases (D_0/N). We illustrate this idea, for the extended SEIR dynamics, in Fig. (6) where we show plots of $I(t) = I_a(t) + I_p(t)$ and $C(t) = D_a(t) + U_p(t) + D_p(t)$ for 5 different initial conditions. The right panel 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 the right panel of Fig. (7) we plot the data with this normalization and initial condition and see a rough collapse for several countries. We notice in particular that three of the Asian countries (India, Pakistan, Indonesia) follow a distinctly different trajectory — this could indicate either that the disease parameters are different or that the intervention strategies have been different, or the reporting of cases is inaccurate.

Predictions for India from extended SEIR model: In the following we make some heuristic predictions, based on the analytic results in Eqs. (5-7) and

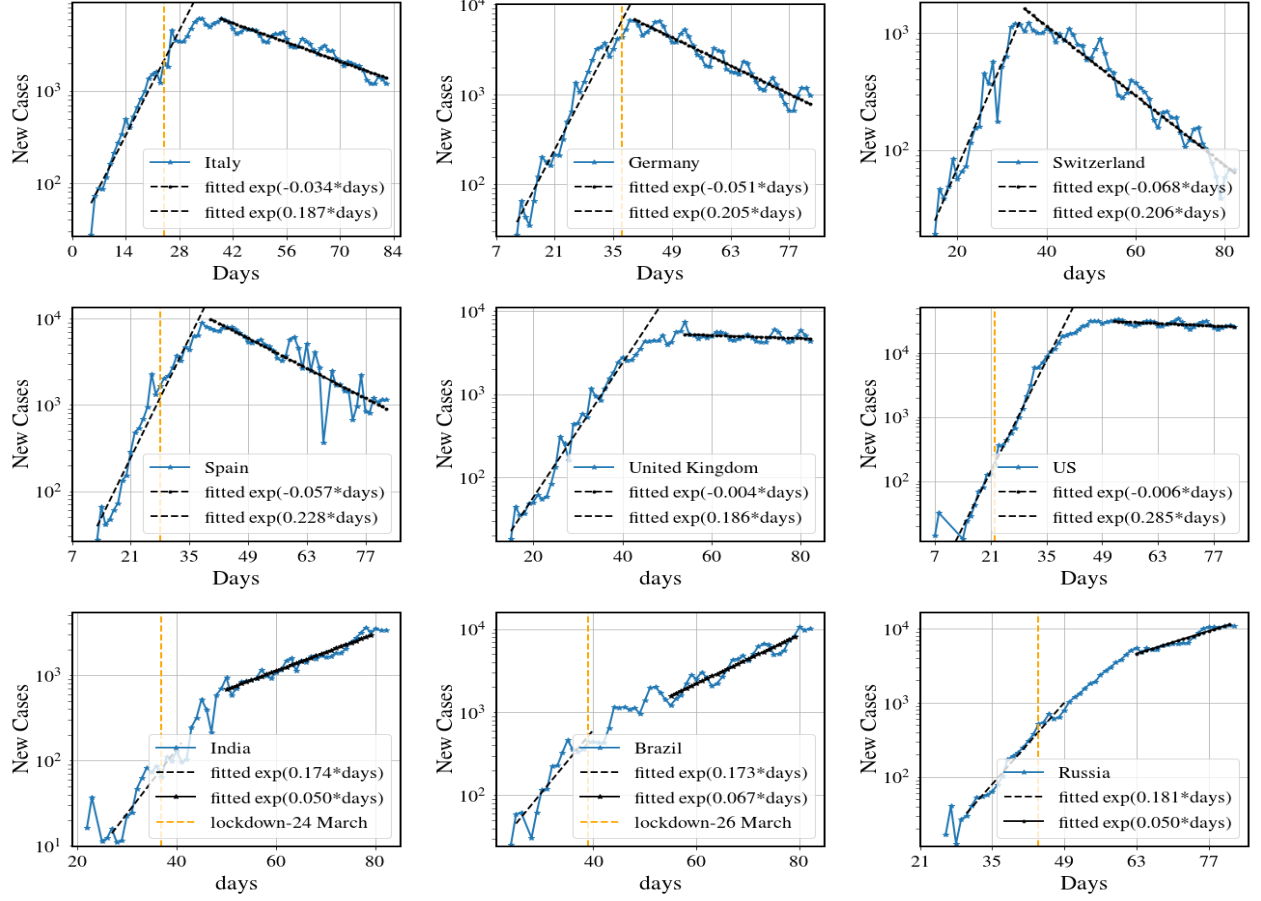


FIG. 5. Number of new cases per day for nine different countries. We note that the first six data sets exhibits the same broad features that we see for the model predictions in Fig. (2,3). In particular we see the fast exponential growth and slow exponential decrease in new cases (following strong interventions). 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 Russia show the behavior corresponding to model predictions in Fig. (4) and have only been able to achieve mitigation so far ($R_0^{\text{target}} > 1, \mu > 0$). Data from [15]

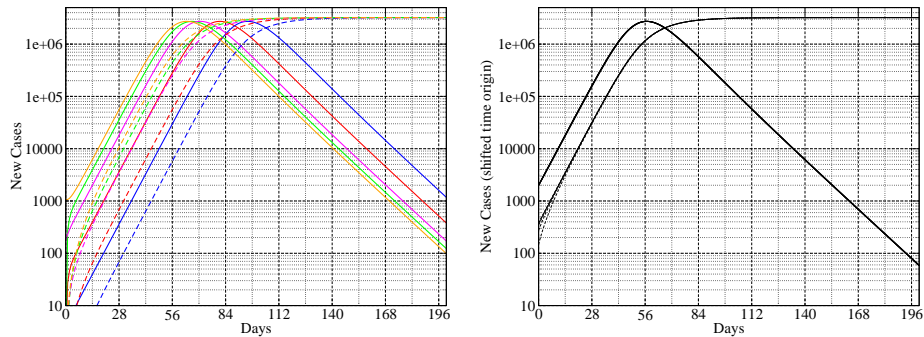


FIG. 6. **Role of initial conditions:** (left) Plots showing $I(t)$ and $C(t)$, for a fixed population of size $N = 10^7$, with 5 very different initial conditions : (1) $E(0) = 100, I_a(0) = 0, I_p(0) = 0$, (2) $E(0) = 10, I_a(0) = 0, I_p(0) = 0$, (3) $E(0) = 1000, I_a(0) = 0, I_p(0) = 0$, (4) $E(0) = 233, I_a(0) = 100, I_p(0) = 75$, (5) $E(0) = 233, I_a(0) = 1000, I_p(0) = 75$. (right) A collapse of all the curves obtained by translating all the trajectories so that they start with the same value of I .

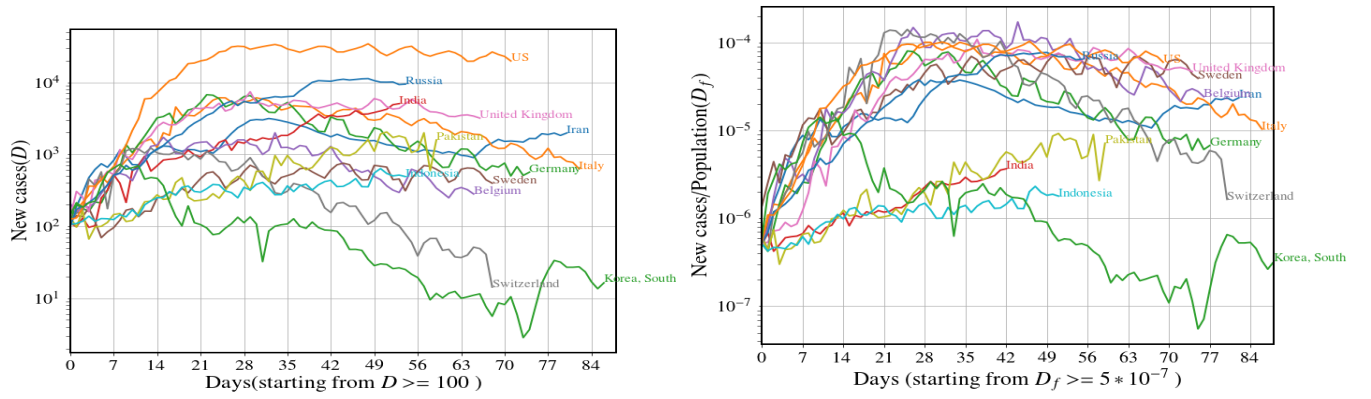


FIG. 7. (left) Number of new cases per day for different countries. (right) Number of new cases normalized by the total population, with the time axis shifted so that every country starts with the same normalized value. Data from [15]

TABLE I. Predictions for India with different choices of the asymptomatic fraction α .

α	R_0	Peak daily cases (PDC)	Time of peak	Total affected	Total deaths
0.67	1.33	2,456,630	2nd week September	45 %	1,936,000
0.9	1.3	686,767	3rd week August	42 %	550,700

TABLE II. Predictions for Delhi with different choices of the asymptomatic fraction α .

α	R_0	Peak daily cases (PDC)	Time of peak	Total affected	Total deaths
0.67	1.33	35,904	1st week August	45 %	28,296
0.9	1.3	10,037	2nd week July	42.3 %	8,048

TABLE III. Predictions for Mumbai with different choices of the asymptomatic fraction α .

α	R_0	Peak daily cases (PDC)	Time of peak	Total affected	Total deaths
0.67	1.33	24,566	3rd week July	45 %	19,360
0.9	1.3	6,867	4th week June	42.3 %	5,507

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 ($N \approx 1.3 \times 10^7$). We consider the following choice of parameter values which appears to be quite typical: $\sigma = 1/2$, $\beta_p = \beta$, $\beta_a = 2\beta/3$, $\gamma_p = \gamma$, $\gamma_a = 3\gamma/2$ (i.e, assume that asymptomatics are less infectious and recover faster). This gives us [using Eq. (23)] $\gamma_e = \gamma/(1 - \alpha/3)$ and the effective reproductive number as $R_0^{\text{eff}} = (1 - 5\alpha/9)\beta/\gamma$. From this last relation we can write $\beta = \gamma R_0^{\text{eff}}/(1 - 5\alpha/9)$. Plugging this into the equation for the eigenvalues, Eq. (17), and replacing λ by the observed mean exponential growth rate $\mu = 0.05$ (the value observed for India since around April 10 [14]), we see that we basically get an equation for R_0^{eff} in terms of σ , γ and μ .

For our analysis we need to know the total infections $I(0)$ on some day and we estimate it on the date April 10 in the following way. Suppose that the daily

observed cases on this day was $D_p(0)$ (assuming that only the symptomatics are detected). Then we have $I_p(0) = D_p(0)/\gamma_p$. From Eq. (24) 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. (5) to compute the peak number of infections $I^{(m)}$ and the peak daily cases (PDC) is then obtained as $\text{PDC} = D_p^{(m)} = \gamma_p I_p^{(m)} = (1 - \alpha) \times \gamma_e \times I^{(m)}$. The total affected population fraction \bar{x} , can be computed from Eq. (7), using only the knowledge of R_0 . If we assume the number of deaths is 1% of all asymptomatic cases this gives us an estimate for the total death statistics as $N\bar{x}(1 - \alpha)/100$.

The observed new cases in India, Delhi and Mumbai on April 10 were around $D_p(0) = 900$, $D_p(0) = 115$ and $D_p(0) = 195$ respectively. Choosing $\sigma = 1/2$ and $\gamma = 1/5$ and two different choices of $\alpha = 0.67$ and 0.9 , we compute the corresponding values obtained for R_0^{eff} . This

and the estimates for PDC= $D_p^{(m)}, t^{(m)}$ and \bar{x} are given in Tables (I), (II) and (III) 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 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 (I) are an over-estimate of the numbers for the true future trajectory. For the state of Delhi and the city of Mumbai, these should be more accurate, *however we see that the uncertainty in the parameter values, for example the true value of α , leads to a huge uncertainty in the predictions.*

IV. MATERIALS AND METHODS

Definition of the extended SEIR model: 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.
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:

$$\frac{dS}{dt} = -\frac{u(\beta_a I_a + \beta_p I_p)}{N} S \quad (8)$$

$$\frac{dE}{dt} = \frac{u(\beta_a I_a + \beta_p I_p)}{N} S - \sigma E \quad (9)$$

$$\frac{dI_a}{dt} = \alpha \sigma E - \gamma_a I_a - r \nu_a I_a \quad (10)$$

$$\frac{dI_p}{dt} = (1 - \alpha) \sigma E - \gamma_p I_p - r \nu_p I_p \quad (11)$$

$$\frac{dU_a}{dt} = \gamma_a I_a \quad (12)$$

$$\frac{dD_a}{dt} = r \nu_a I_a \quad (13)$$

$$\frac{dU_p}{dt} = \gamma_p I_p \quad (14)$$

$$\frac{dD_p}{dt} = r \nu_p I_p. \quad (15)$$

The parameters in the above equation 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.
- 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. Here we choose $\nu_a = 1/3, \nu_p = 1/2$,
- u : intervention factor due to social distancing (time dependence specified below).
- r : intervention factor due to testing-quarantining (time dependence specified below). 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 D would be

$$C = D_a + D_p + U_p, \quad D = \frac{dC}{dt} = r \nu_a I_a + (\gamma_p + r \nu_p) I_p. \quad (16)$$

Note that we include U_p because these are people who are not detected through directed tests but eventually get detected (after $1/\gamma_p$ 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.

A. 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. For the present let us ignore the time dependence of the SD factor u and the TQ factor r . As shown in App. (A2), the system has three non-zero eigenvalues given by the roots of the cubic equation:

$$\lambda^3 + (\tilde{\gamma}_a + \tilde{\gamma}_p + \sigma)\lambda^2 + [\tilde{\gamma}_a\tilde{\gamma}_p + \sigma(\tilde{\gamma}_a + \tilde{\gamma}_p)(1 - Q)]\lambda + \sigma\tilde{\gamma}_a\tilde{\gamma}_p(1 - R_0) = 0, \quad (17)$$

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

$$R_0 = \alpha\frac{\tilde{\beta}_a}{\tilde{\gamma}_a} + (1 - \alpha)\frac{\tilde{\beta}_p}{\tilde{\gamma}_p} = \alpha\frac{u\beta_a}{\gamma_a + r\nu_a} + (1 - \alpha)\frac{u\beta_p}{\gamma_p + r\nu_p} \quad (18)$$

is the expected form for the reproductive number for the disease. Noting the fact that $Q < R_0$, it follows that the condition for at least one positive eigenvalue is

$$R_0 > 1. \quad (19)$$

We denote the largest eigenvalue by μ . At early times the number of cases detected would grow as $\sim e^{\mu t}$.

Initial conditions: In Appendix. (A2) we explain the fact that all initial conditions (which satisfy the condition $S(0) \approx N$) will quickly move along the direction of the dominant eigenvector and so all the trajectories for different initial conditions are *identical up to a time translation*. We also discuss in Appendix (A2) how one can choose the correct initial conditions for all the dynamical variables given knowledge of just one of the observed variables (e.g the number of daily new reported infections).

B. Final affected population and peak infections

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 recovered or died) is given by \bar{R} and would have developed immunity. A fraction \bar{U}_a (see below) would be undetected and uncounted.

Here for the moment let us assume that u and r do not have any time dependence. As shown in App. (A) the asymptotic fraction $\bar{x} = \bar{R}/N$ is simply given by the solution of the equation

$$1 - \bar{x} = e^{-R_0\bar{x}}, \quad (20)$$

with R_0 being the reproductive number given by Eq. (18). We note that Eq. (20) has a non-zero solution only when $R_0 > 1$. For the simple SIR model this result is well known [13], here we show that this is valid quite generally.

The asymptotic population of the individual populations are then given by

$$\begin{aligned} R_a &= \alpha R, \quad R_p = (1 - \alpha)R \\ U_a &= \frac{\gamma_a}{\gamma_a + r\nu_a} R_a, \quad D_a = \frac{r\nu_a}{\gamma_a + r\nu_a} R_a, \\ U_p &= \frac{\gamma_p}{\gamma_p + r\nu_p} R_p, \quad D_p = \frac{r\nu_p}{\gamma_p + r\nu_p} R_p. \end{aligned} \quad (21)$$

As shown in App. (B1) for the SEIR model, the peak value of the infection number ($I = I_a + I_p$) can be found from a heuristic argument and is very accurately given by the formula

$$I^{(m)} = \frac{\sigma}{\gamma + \sigma} \left(1 - \frac{1 + \ln R_0}{R_0} \right) N. \quad (22)$$

We find that this also describes accurately the peak value for the extended SEIR dynamics with γ now replaced by

$$\gamma_e = [\alpha\gamma_a^{-1} + (1 - \alpha)\gamma_p^{-1}]^{-1}. \quad (23)$$

In Fig. (8) we show the dependence of \bar{x} on R_0 (as obtained from a numerical solution of Eq.(20) and provide a numerical verification of the result in Eq. 22. The peak values of the asymptomatic and presymptomatic populations are respectively given by

$$\begin{aligned} I_a^{(m)} &= \alpha\gamma_e I^{(m)}/\gamma_a, \\ I_p^{(m)} &= (1 - \alpha)\gamma_e I^{(m)}/\gamma_p. \end{aligned} \quad (24)$$

An estimate of the time to reach this peak value can be obtained by noting that we can use the linearized dynamics (see previous section) till the time $I(t)$ reaches its peak I_{\max} . Hence we write $I^{(m)} = I(t^{(m)}) = I(0) e^{\mu t^{(m)}}$ which provides $t^{(m)} = \frac{\ln[I^{(m)}/I(0)]}{\mu}$. We naturally expect that $I^{(m)}$ is of the order $O(N)$ which implies

$$t^{(m)} \sim \frac{\ln[I^{(m)}/I(0)]}{\mu} \sim \frac{\ln(N/c)}{\mu}, \quad (25)$$

where c is a constant that depends on the initial infected population and the parameters. A verification of this result, obtained by solving the basic SEIR equations numerically, is provided in Fig. (8).

C. Interventions: Social distancing and Testing-Quarantining

We discuss here the choices of the intervention functions u and r introduced in the dynamical equations in Sec. (IV). 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.

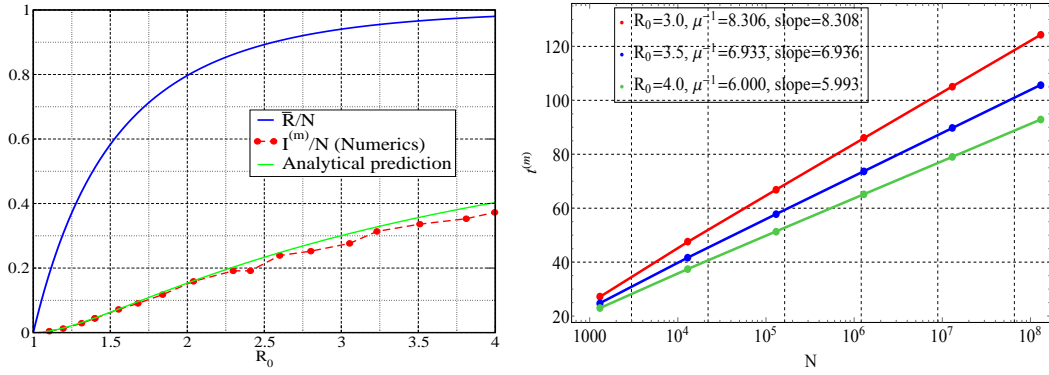


FIG. 8. (left) Plot of the asymptotic total affected population fraction, \bar{R}/N , as a function of the reproductive number R_0 . We also plot the quantity $(I^{(m)}/N)(\sigma + \gamma_e)/\sigma$, obtained numerically from many different parameter sets, and compare it with the theoretical predicted curve $1 - (1 + \ln R_0)/R_0$ (green line). (right) Verification of the $\ln(N)$ dependence of $t^{(m)}$ in Eq. (25) for different choices of R_0 . The slopes of the straight lines compares well with μ^{-1} as stated in Eq. (25).

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 a 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

$$u(t) = 1 \quad t < t_{on}, \\ = u_l + (1 - u_l)e^{-(t-t_{on})/t_w}, \quad t > t_{on}. \quad (26)$$

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 $r\nu_a I_a$ and $r\nu_p I_p$ in the dynamical equations. A reasonable choice for the TQ function is perhaps to take

$$r(t) = 0 \quad t < t'_{on}, \\ = r_l - r_l e^{-(t-t'_{on})/t'_w}, \quad t > t'_{on}. \quad (27)$$

where we 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

$$R_0^{\text{eff}}(t) = \alpha \frac{u(t)\beta_a}{\gamma_a + r(t)\mu_a} + (1 - \alpha) \frac{u(t)\beta_p}{\gamma_p + r(t)\nu_p}. \quad (28)$$

At long times this goes to the targeted reproduction number

$$R_0^{\text{target}} = R_0^{\text{eff}}(t \rightarrow \infty) = \alpha \frac{u_l \beta_a}{\gamma_a + r_l \nu_a} + (1 - \alpha) \frac{u_l \beta_p}{\gamma_p + r_l \nu_p}. \quad (29)$$

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. (9) 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. In our extended SEIR model the number of detected cases per day is given by $D(t) = r\nu_a I_a + (\gamma_p + r\nu_p)I_p$. Then, the number of contacts of these individuals would be $AD(t)$ where A is the number of contacts one infected person made. A good assumption is to say that the infected people are from this pool. Hence, if we conduct a total of $T = NT_r$ tests per day *on only this set of people*, then the number of detected cases per day (through contact tracing) would be

$$D_{a,p} \approx \frac{NT_r}{AD(t)} \gamma_{a,p} I_{a,p} = \frac{T}{AD(t)} \gamma_{a,p} I_{a,p}. \quad (30)$$

Hence we see that TQ intervention can be successfully implemented if we can achieve

$$T \approx \frac{r(t)AD(t)}{\gamma_p}, \quad (31)$$

where 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 need

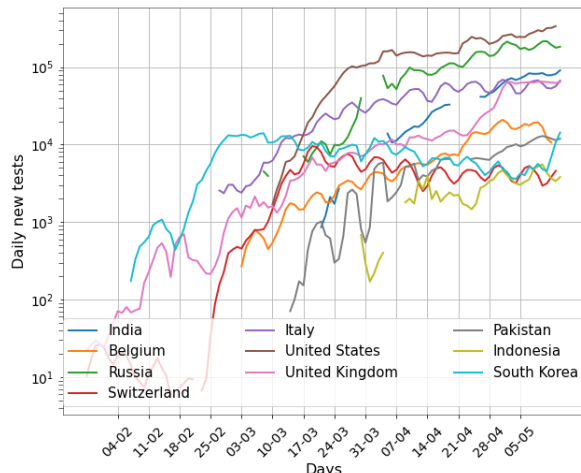


FIG. 9. Data of number of tests per day per thousand in several countries on a log-scale. Notice in particular the large testing numbers in the early period in Korea which has been very successful at controlling the disease. Data from [16].

$T(t) \sim AD(t)$. The implications of this is discussed after Eq. (1). In Fig. (9) we show data for daily new tests for a set of countries. A noteworthy case is the data for South Korea where we see the large testing rate at early days of the pandemic. Perhaps this explains the quick control of the pandemic in that country. The table in Fig. (10) shows data for the ratio $T(t)/D(t)$ for a set of countries and also how this ratio has evolved over time. As discussed in Sec. (II) the number A is expected to depend on the population density and also how well SD is being implemented and hence, for a country like India $T(t)/D(t) \approx 25$ may not be sufficient.

V. CONCLUSIONS

A modified version of the SEIR model, incorporating asymptomatic individuals, was used for analyzing the effectiveness of different intervention protocols in control-

ling the growth of the Covid-19 pandemic. Non-clinical interventions can be either through social distancing or through testing-quarantining. Our results indicate that a combination of both, implemented over an extended period may be the most effective and practical strategy. We point out that short-term lock-downs cannot stop a recurrence of the pandemic if interventions are completely relaxed and developing herd immunity is not a practical solution either since this would affect a very large fraction of the population.

We have provided numerical examples to illustrate the basic ideas and in addition have stated a number of analytical results which can be useful in making empirical estimates of various important quantities that provide information on the disease progression. Looking at real data for new Covid-19 cases in several countries, we find that the SEIR model captures some important qualitative features and hence could provide guidance in policy-making. We use 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, but point out that these predictions could be incorrect for India (due to big inhomogeneity in disease progression across the country) and perhaps more reliable for the cases of Delhi and Mumbai. In general we believe that our formulas are easy to use and give quick heuristic estimates on disease progression, which would be reliable when applied to local populations (in towns, cities and perhaps smaller countries). Lack of precise knowledge of the disease parameters (e.g the fraction of asymptomatic carriers) of course leads to rather large uncertainties in the predictions.

ACKNOWLEDGMENTS

We thank Jitendra Kethepalli and Kanaya Malakar for very helpful discussions and Ranjini Bandyopadhyay, Siddhartha Chatterjee, Joel Lebowitz and Sriram Shastri for a careful reading of the draft and making useful suggestions. We acknowledge support of the Department of Atomic Energy, Government of India, under project no.12-R&D-TFR-5.10-1100.

-
- [1] Kucharski, A. J. et al. *Early dynamics of transmission and control of COVID-19: a mathematical modelling study*. *Lancet Global Health*, [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4) (2020).
 - [2] N. Ferguson et al., *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand*, <https://doi.org/10.25561/77482> (2020).
 - [3] B. Tang, N. Luigi Bragazzi, Qian Li, S. Tang, Y. Xiao, Jianhong Wu, *An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov)*, *Infectious Disease Modelling* **5**, 248 (2020).
 - [4] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri, *Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy*, *Nature Medicine* (2020) <https://doi.org/10.1038/s41591-020-0883-7>.
 - [5] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, A. Rinaldo, *Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures*, *Proceedings of the National Academy of Sciences*, 202004978. DOI:10.1073/pnas.2004978117.
 - [6] D. Ray et al., *Predictions, role of interventions and ef-*

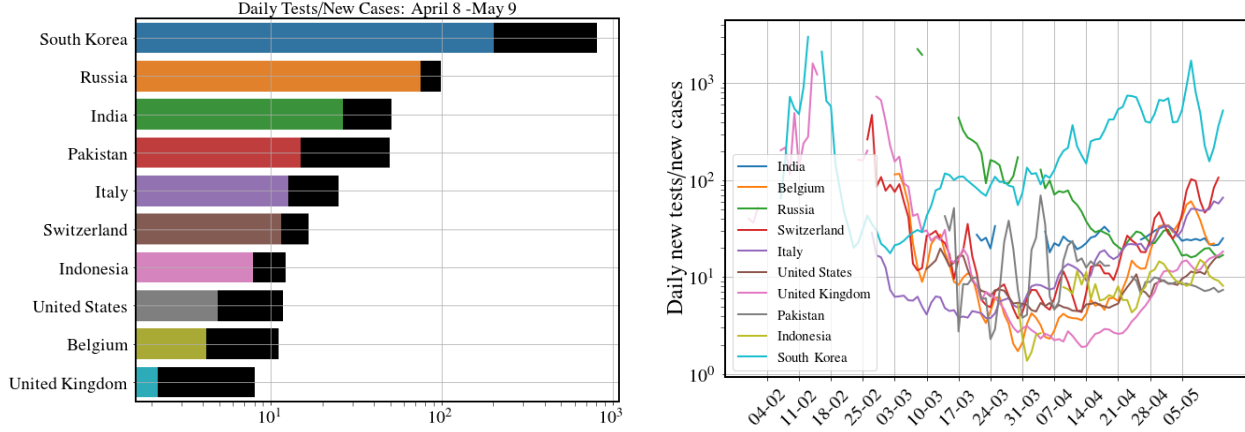


FIG. 10. Data from different countries on the number of tests per detected case $T(t)/D(t)$ for 10 countries (left) on the dates April 8 (coloured bar), May 9 (black bar) and (right) the change over time of this ratio. Data from [16]

fects of a historic national lockdown in India's response to the COVID-19 pandemic: data science call to arms, <https://doi.org/10.1101/2020.04.15.20067256>.

- [7] K. Sarkar and S. Khajanchi, *Modeling and forecasting of the COVID-19 pandemic in India*, arXiv:2005.07071 (2020).
- [8] S. Shekatkar et al, *INDSCI-SIM: A state-level epidemiological model for India*, <https://indscicov.in/indscisim>.
- [9] Agrawal et al., *COVID-19 Epidemic: Unlocking the lockdown in India (working paper)*, IISc-TIFR Technical Report, <https://covid19.iisc.ac.in/wp-content/uploads/2020/04/Report-1-20200419-UnlockingTheLockdownInIndia.pdf>.
- [10] D. Meidan et al., *Alternating quarantine for sustainable mitigation of COVID-19*, <https://arxiv.org/pdf/2004.01453.pdf> (2020).
- [11] Y. Liu, A. A. Gayle, A. Wilder-Smith, J. Rocklöv, *The reproductive number of COVID-19 is higher compared to SARS coronavirus*, Journal of Travel Medicine, Volume 27, Issue 2, March 2020, <https://doi.org/10.1093/jtm/taaa021>.
- [12] K. Y. Leung, P. Trapman, T. Britton, *Who is the infector? Epidemic models with symptomatic and asymptomatic cases*, Mathematical Biosciences, **301**, 190 (2018).
- [13] F Brauer, P. V. Driessche, and J. Wu, *Lecture notes in Mathematical Epidemiology*, Springer (2008).
- [14] Covid-19 data analysis: <https://www.icts.res.in/COVID-data-commentary>
- [15] Johns Hopkins Center for Systems Science and Engineering Coronavirus COVID-19 Global Cases. <https://systems.jhu.edu/research/public-health/ncov/>
- [16] Coronavirus (COVID-19) Testing, <https://ourworldindata.org/coronavirus-testing>.

Appendix A: Extended SEIR model

1. Asymptotic total affected population

Let us consider a more general form of the SEIR equations with n compartments for the infectious population with I_1, I_2, \dots, I_n , n compartments for the recovered population with R_1, R_2, \dots, R_n and the other 2 compartments of S, E, R with the following dynamics

$$\frac{dS}{dt} = - \sum_{i=1}^n \beta_i \frac{I_i S}{N}, \quad (\text{A1})$$

$$\frac{dE}{dt} = \sum_{i=1}^n \beta_i \frac{I_i S}{N} - \sigma E, \quad (\text{A2})$$

$$\frac{dI_i}{dt} = \sigma_i E - \gamma_i I_i, \quad i = 1, 2, \dots, n, \quad (\text{A3})$$

$$\frac{dR_i}{dt} = \gamma_i I_i, \quad i = 1, 2, \dots, n, \quad (\text{A4})$$

where $\sigma = \sum_{i=1}^n \sigma_i$.

Let us assume that $R_i(0) = 0$ for all i , and $S(0) \approx N$. Then solving Eq. (A1)), we get

$$\bar{S} = N e^{-\sum_{i=1}^n \beta_i \int_0^\infty dt I_i(t)/N}. \quad (\text{A5})$$

Multiplying Eqs. (A4) by β_i/γ_i , summing over i and integrating time from 0 to ∞ , we get $\sum_{i=1}^n \beta_i \int_0^\infty dt I_i(t) = (\beta_i/\gamma_i) \bar{R}_i$. Plugging this into the previous equation then gives

$$\bar{S} = N e^{-\sum_{i=1}^n \frac{\beta_i}{\gamma_i} \bar{R}_i/N}. \quad (\text{A6})$$

Next we note that $(d/dt)(I_i + R_i) = \sigma_i E$. Hence for the initial condition $I_i = R_i = 0$ we find that the ratio $[I_i(t) + R_i(t)]/[I_j(t) + R_j(t)] = \sigma_i/\sigma_j$ at all times. Since at large times $I_i \rightarrow 0$, this means that the asymptotic values

of R_i s are given by

$$\bar{R}_i = \frac{\sigma_i}{\sigma} \bar{R}. \quad (\text{A7})$$

Using this in Eq. (A6), noting that $\bar{S} + \bar{R} = N$ and defining $x = \bar{R}/N$, we then get the following simple equation that determines the asymptotic total affected population:

$$1 - x = e^{-R_0 x}, \quad (\text{A8})$$

$$\text{where } R_0 = \sum_{i=1}^n \frac{\beta_i \sigma_i}{\gamma_i \sigma} \quad (\text{A9})$$

is the reproductive number.

2. Linear analysis of extended SEIR model

We now again focus on the special case with the $n = 8$ variable dynamics described by Eqs. (8-15). Let us denote the variables by $x_1 = S - N, x_2 = E, x_3 = I_a, x_4 = I_p, x_5 = U_a, x_6 = D_a, x_7 = U_p, x_8 = D_p$. At early times when $x_i \ll N$, the dynamics is captured by linear equations

$$\frac{dx}{dt} = Mx, \quad (\text{A10})$$

$$\text{with } M = \begin{pmatrix} 0 & 0 & -\tilde{\beta}_a & -\tilde{\beta}_p & 0 & 0 & 0 & 0 \\ 0 & -\sigma & \tilde{\beta}_a & \tilde{\beta}_p & 0 & 0 & 0 & 0 \\ 0 & \alpha\sigma & -\tilde{\gamma}_a & 0 & 0 & 0 & 0 & 0 \\ 0 & (1-\alpha)\sigma & 0 & -\tilde{\gamma}_p & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_a & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & r\nu_a & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r\nu_p & 0 & 0 & 0 & 0 \end{pmatrix}, \quad (\text{A11})$$

where $\tilde{\beta}_a = u\beta_a, \tilde{\beta}_p = u\beta_p, \tilde{\gamma}_a = \gamma_a + r\nu_a, \tilde{\gamma}_p = \gamma_p + r\nu_p$. This has 5 zero eigenvalues while the remaining 3 ones are given by the roots of the cubic equation for λ :

$$\begin{aligned} & \lambda^3 + (\tilde{\gamma}_a + \tilde{\gamma}_p + \sigma)\lambda^2 \\ & + [\tilde{\gamma}_a\tilde{\gamma}_p + \tilde{\gamma}_a\sigma + \tilde{\gamma}_p\sigma - \alpha\tilde{\beta}_a\sigma - (1-\alpha)\tilde{\beta}_p\sigma]\lambda \\ & + \sigma[\tilde{\gamma}_a\tilde{\gamma}_p - (1-\alpha)\tilde{\beta}_p\tilde{\gamma}_a - \alpha\tilde{\beta}_a\tilde{\gamma}_p] = 0. \end{aligned} \quad (\text{A12})$$

This can be written in the form

$$\begin{aligned} & \lambda^3 + (\tilde{\gamma}_a + \tilde{\gamma}_p + \sigma)\lambda^2 \\ & + [\tilde{\gamma}_a\tilde{\gamma}_p + \sigma(\tilde{\gamma}_a + \tilde{\gamma}_p)(1-Q)]\lambda + \sigma\tilde{\gamma}_a\tilde{\gamma}_p(1-R_0) = 0, \end{aligned} \quad (\text{A13})$$

$$\text{where } Q = \alpha \frac{\tilde{\beta}_a}{\tilde{\gamma}_a + \tilde{\gamma}_p} + (1-\alpha) \frac{\tilde{\beta}_p}{\tilde{\gamma}_a + \tilde{\gamma}_p} \quad (\text{A14})$$

$$\text{and } R_0 = \alpha \frac{\tilde{\beta}_a}{\tilde{\gamma}_a} + (1-\alpha) \frac{\tilde{\beta}_p}{\tilde{\gamma}_p}. \quad (\text{A15})$$

We identify R_0 with the reproductive number of the disease. Noting the fact that $Q < R_0$, it is easy to prove that the necessary condition for at least one positive eigenvalue is

$$R_0 > 1. \quad (\text{A16})$$

Let us denote the largest eigenvalue by μ . For $R_0 \approx 1$, we expect that the largest eigenvalue is close to zero and from Eq. (A15) we can read off the value as

$$\mu \approx \frac{\sigma(R_0 - 1)}{1 + \sigma(\tilde{\gamma}_a^{-1} + \tilde{\gamma}_p^{-1})(1-Q)}. \quad (\text{A17})$$

A way to choose correct initial conditions from knowledge of one variable (e.g confirmed cases) at an early time: We denote the right and left eigenvectors corresponding to the eigenvalue μ by $\phi_m(i)$ and $\chi_m(i)$ respectively. The time evolution of the vector $X = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ is given by

$$\begin{aligned} x_i(t) &= \sum_j \sum_q \phi_q(i) \chi_q(j) e^{\lambda_q t} x_j(0) \\ &\approx \sum_j \phi_m(i) \chi_m(j) e^{\mu t} x_j(0), \end{aligned} \quad (\text{A18})$$

where the last line is true at sufficiently large times when only one eigenvalue λ dominates. Let us consider the initial condition $X = (-\epsilon, 0, 0, \epsilon, 0, 0, 0, 0)$ so that (noting that $\chi_m(1) = 0$)

$$x_i(t) \approx \epsilon \phi_m(i) \chi_m(4) e^{\mu t} = a_i \epsilon e^{\mu t}, \quad (\text{A19})$$

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

$$\epsilon e^{\mu t_l} = \frac{C_0}{a_6 + a_7 + a_8}. \quad (\text{A20})$$

This then tells us that we should start with the following initial conditions, counting now time from $t = 0$:

$$x_i(0) = \frac{\phi_m(i)}{\phi_m(6) + \phi_m(7) + \phi_m(8)} C_0. \quad (\text{A21})$$

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

Appendix B: Analysis of the basic SEIR model

In the standard SEIR model one divides a population of size N into four compartments of

1. S = Number of Susceptible individuals.

2. E = Number of Exposed but not yet contagious individuals.
3. I = Number of Infected contagious individuals
4. R = Number of Recovered, hospitalized or dead individuals.

The dynamics of this model can be described as follows:

- The infected individuals, I , come in contact with the susceptible population, S , and cause transitions $S \rightarrow I$.
- People who are Exposed carry the virus, do not yet show symptoms and cannot infect others.
- After a latency period T_L the Exposed people become Infected and can now infect others, so $E \rightarrow I$ happens at a rate $\sigma = 1/T_L$. These people could either be symptomatic or asymptomatic and their diseases are yet un-detected.
- We assume that infected people typically either recover or are detected after T_R days, so $I \rightarrow R$ happens at a rate $\sigma = 1/T_R$.

We then have the following equations for the dynamics for the system

$$\frac{dS}{dt} = -\frac{\beta I}{N} S, \quad (B1)$$

$$\frac{dE}{dt} = \frac{\beta I}{N} S - \sigma E, \quad (B2)$$

$$\frac{dI}{dt} = \sigma E - \gamma I, \quad (B3)$$

$$\frac{dR}{dt} = \gamma I. \quad (B4)$$

In this case the reproductive number is simply given by $R_0 = \beta/\gamma$.

1. Final fraction of infected population in the absence of intervention

To determine the fraction of the population that would be affected finally if there was no intervention, we first note from Eqs. (B1,B4) that

$$\frac{dS}{dR} = -\frac{\beta}{N\gamma} S = -\frac{R_0}{N} S, \quad (B5)$$

hence

$$S(t) = S(0)e^{-R_0 R(t)/N}, \quad (B6)$$

where we have assumed $R(0) = 0$. In the steady state we should have $E = I = 0$ while $\bar{S} = S(\infty)$, $\bar{R} = R(\infty)$ are determined from the condition $N = S + E + I + R$, which gives

$$N = \bar{S} + \bar{R} = S(0)e^{-R_0 \bar{R}/N} + \bar{R}. \quad (B7)$$

Denoting the fraction of initially infected by $\epsilon = I(0)/N$ and the fraction of total eventually affected fraction by $\bar{x} = \bar{R}/N$, we see that x can be determined from solution the following equation

$$1 = (1 - \epsilon)e^{-R_0 \bar{x}} + \bar{x}. \quad (B8)$$

Typically $\epsilon \ll 1$ and so see that the final fraction of affected population is given by the solution of the equation

$$1 - \bar{x} - e^{-R_0 \bar{x}} = 0. \quad (B9)$$

2. Size of infection population peak $I^{(m)}$ and the number of days to reach the peak

We now evaluate the peak value I_{\max} of the infected population in the course of the outbreak. We first note that the equation (B6) allows us to express the susceptible population at any time t as a function of $R(t)$. In fact, one can express all the other populations in terms of $R(t)$ or its time derivatives, such as

$$I = \frac{1}{\gamma} \frac{dR}{dt}$$

$$E = \frac{1}{\sigma} \frac{dI}{dt} + \frac{\gamma}{\sigma} I = \frac{1}{\gamma\sigma} \frac{d^2 R}{dt^2} + \frac{1}{\sigma} \frac{dR}{dt}. \quad (B10)$$

Hence, after expressing all variables in terms of R and inserting them in the constraint equation $S + E + I + R = N$ we get

$$N e^{-R_0 \frac{R}{N}} + \frac{1}{\gamma\sigma} \frac{d^2 R}{dt^2} + \frac{1}{\sigma} \frac{dR}{dt} + \frac{1}{\gamma} \frac{dR}{dt} + R = N$$

$$\Rightarrow \frac{d^2 x}{dt^2} + (\gamma + \sigma) \frac{dx}{dt} + \gamma\sigma (x + e^{-R_0 x} - 1) = 0, \quad (B11)$$

where, $x = R/N$. Defining $v = \frac{dx}{dt} = \gamma I/N$, we see that the four dimensional SEIR-dynamics is equivalent to a two-dimensional dynamical system specified by the equations

$$\frac{dx}{dt} = v \quad (B12)$$

$$\frac{dv}{dt} = -(\gamma + \sigma)v - \gamma\sigma (x + e^{-R_0 x} - 1). \quad (B13)$$

The above equation resembles a damped oscillator constrained to move in the positive half line and in a potential $U(x) = \gamma\sigma(x^2/2 - x) + (\gamma^2\sigma/\beta)e^{-R_0 x}$ so that $F(x) = -U'(x) = -\gamma\sigma(x + e^{-R_0 x} - 1)$. The nontrivial fixed point, which is the steady state, is given by the zero of $F(x)$, as already obtained in earlier section.

On the other hand, the peak of the infected population is given by setting $dI/dt = 0$ or $dv/dt = 0$, which implies $v^{(m)} = -\frac{\gamma\sigma}{\gamma+\sigma} (x_m + e^{-R_0 x_m} - 1)$, where $x^{(m)}$, $v^{(m)}$ denote the values of x, v at the time when I

peaks. To determine $(x^{(m)}, v^{(m)})$ we need another equation which could be obtained for example from a solution of the equation for dv/dx . This is difficult to calculate exactly. However we can obtain a second equation if we make the reasonable assumption that I and E peak at around the same time, which simply gives $S^{(m)}/N = 1/R_0$. On the other hand From Eq. (B6) we get $R^{(m)} = -R_0^{-1} \ln(S/N) = R_0^{-1} \ln R_0$. Then using the overall constraint $N = S + E + I + R$ we finally obtain

$$I^{(m)} = \frac{\sigma}{\gamma + \sigma} \left(1 - \frac{1 + \ln R_0}{R_0} \right) N. \quad (\text{B14})$$

An estimate of the time to reach this peak value can be obtained by noting that we can use the linearized dynamics (see previous section) till the time $I(t)$ reaches its peak $I^{(m)}$. Hence we write $I^{(m)} = I(t_{\max}) = I(0) e^{\mu t_{\max}}$ which provides $t^{(m)} = \frac{\ln[I_{\max}/I(0)]}{\mu}$. Hence

$$t^{(m)} \sim \frac{\ln[I^{(m)}/I(0)]}{\mu} \sim \frac{\ln(N/c)}{\mu}, \quad (\text{B15})$$

where c is a constant that depends on initial infection numbers and parameter values.

3. Linear analysis of the SEIR model and fixing of initial conditions

To get the growth at early time regime let us define the variables $S = N + s, E = e, I = i, R = r$. Inserting these in Eqs. (B1, B2, B3) and (B4), and then expanding the right hand sides of each equations to linear order, we get

$$\frac{dX}{dt} = M X, \quad \text{where, } X = \begin{pmatrix} s \\ e \\ i \\ r \end{pmatrix}, \quad (\text{B16})$$

$$\text{and } M = \begin{pmatrix} 0 & 0 & -\beta & 0 \\ 0 & -\sigma & \beta & 0 \\ 0 & \sigma & -\gamma & 0 \\ 0 & 0 & \gamma & 0 \end{pmatrix}. \quad (\text{B17})$$

This set of linear equations can be solved by diagonalizing the matrix M . It has eigenvalues

$$\begin{aligned} \lambda_1 &= 0, \\ \lambda_2 &= 0, \\ \lambda_3 &= \frac{-(\sigma + \gamma) - \sqrt{(\sigma - \gamma)^2 + 4\beta\sigma}}{2} \\ \lambda_4 &= \frac{-(\sigma + \gamma) + \sqrt{(\sigma - \gamma)^2 + 4\beta\sigma}}{2}. \end{aligned} \quad (\text{B18})$$

Let us denote the right and left eigenvectors corresponding to the eigenvalue λ_q by $\phi_q(i)$ and $\chi_q(i)$ respectively.

The right eigenvectors are given by

$$\begin{aligned} \phi_1 &= (0, 0, 0, 1), \\ \phi_2 &= (1, 0, 0, 0), \\ \phi_3 &= (-\beta/\gamma, \lambda_3(\lambda_3 + \gamma)/(\sigma\gamma), \lambda_3/\gamma, 1), \\ \phi_4 &= (-\beta/\gamma, \lambda_4(\lambda_4 + \gamma)/(\sigma\gamma), \lambda_4/\gamma, 1). \end{aligned} \quad (\text{B19})$$

We denote the largest eigenvalue $\lambda_4 \equiv \mu = [-(\sigma + \gamma) + \sqrt{(\sigma - \gamma)^2 + 4\beta\sigma}]/2$ and it is easy to see that this is positive for $\beta/\gamma = R_0 > 1$.

It is instructive to examine the structure of μ near $R_0 = 1$. For this we rewrite this in the form

$$\mu = \frac{-(\sigma + \gamma) + \sqrt{(\sigma + \gamma)^2 + 4\gamma\sigma(R_0 - 1)}}{2} \approx \frac{\gamma\sigma}{\gamma + \sigma}(R_0 - 1). \quad (\text{B20})$$

One qualitative aspect that this equation tells us is the following. Suppose we start with free parameters β, γ such that $R_0 = 1.8$ and want to change (through interventions) the reproductive number to a target value $R_0(\text{target}) = R_0/2 = 0.9$. We can do this either (a) by decreasing β to $\beta' = \beta/2$ or (b) by increasing γ to a value $\gamma' = 2\gamma$. It is clear from the above expression that (b) would lead to a negative eigenvalue of larger magnitude and so a faster decay of the disease.