medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Covid-19: analysis of a modified SEIR model, a comparison of different intervention strategies and projections for India

Arghya Das, [c]Abhishek Dhar, Srashti Goyal, Anupam Kundu
International Center for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore-560089, India

To understand the spread of Covid-19, we analyse an extended Susceptible-Exposed-Infected-Recovered (SEIR) model that accounts for asymptomatic carriers, and 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 $R_0^{\rm target} < 1$, but in distinct ways, which we implement in our model equations. We find that for the same $R_0^{\rm target} < 1$, 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 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. Weak intervention strategies (that reduce R_0 but not sufficiently) reduce the peak values of infections and the asymptotic affected population and we provide analytic expressions for these in terms of the disease parameters. We apply them in the Indian context to obtain heuristic projections for the course of the pandemic, noting that the predictions strongly depend on the assumed fraction of asymptomatic carriers.

[c]Corresponding author: abhishek.dhar@icts.res.in

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 30 was close to 5.9 million with close to $360,000$ 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 [1–13].

One of the standard epidemiological model is the SEIR model 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 parameterized by the three parameters β, σ and γ that specify the rates of transitions from $S \to E$, $E \to I$ and $I \to R$ 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 [18] 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 — 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 $R_0 = 2 - 7$ [14]. For the 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 [4, 5] — 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 [4–6, 15] 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 (see Sec. (III A) for details). We do not include separate compartments for the number of hospitalized and dead since these extra details

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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 study both the cases, of strong interventions ($R_0^{\rm target} < 1$ aimed at disease suppression,) and that of weak interventions ($R_0^{\rm target} \gtrsim 1$, aimed at disease mitigation).

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 $R_0^{\rm target} > 1$, we have $\mu > 0$ and this gives us the exponential growth rate (doubling time $\approx 0.7/\mu$). On the other hand, for $R_0^{\rm target} < 1$, the corresponding μ is less than 0 and this tells us that infections will decrease exponentially. For the case where only mitigation is achieved ($R_0^{\rm target} \gtrsim 1$), we present analytic expressions for peak infection numbers, time to reach peak values, and asymptotic values of total affected populations. These provide useful guidance on disease progression and we apply it in the Indian context. We also show that, for the same reduction of $R_0^{\rm target}$ 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 disease doubling time is around 14 days (as on May 20). For two different choices of the fraction of asymptomatics (and typical values of disease parameters) our estimates suggest a value of $R_0^{\rm eff} \approx 1.3$ 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. Given the huge economic and social costs of implementing hard 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 value of ≈ 25 (as on May 15) could result in lowering of R_0 and the μ value or at least in not letting them increase further.*

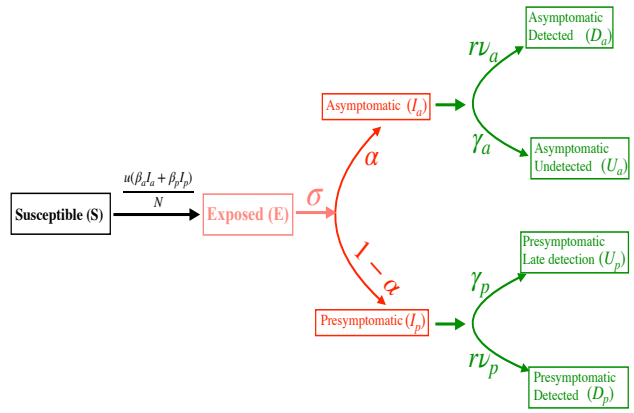

FIG. 1. A schematic description of the extended SEIR dynamics studied in this work. The parameters $\beta_a, \beta_p, \sigma, \gamma_a, \gamma_p, \alpha$ are intrinsic to the disease, u quantifies the degree of social distancing while ν_a, ν_p, r are related to intervention arising from testing-quarantining.

The rest of the paper is structured as follows. In Sec. I we state our main results. In Sec. I C we make 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. We summarize our results in Sec. II. All the technical details including the precise definition of the extended SEIR model with and without interventions and our analytic results are given in Sec. III.

I. 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 $(\beta_a, \beta_p, \sigma, \gamma_a, \gamma_p, \alpha, \nu_a, \nu_p, r, u)$, of which α represents the fraction of asymptomatic carriers while u, r, ν_a, ν_p are related to intervention strategies. The model details are given in Sec. III A. For the present, 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 interven-

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

tions) 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.

A. Comparison of different intervention strategies

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

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

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 Sec. III B). 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 $E(0) = 100$, $I_a(0) = I_p(0) = U_a(0) = D_a(0) = U_p(0) = D_p(0) = 0$ 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.

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 Figs. (2a) and (2b).

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 Figs. 2(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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

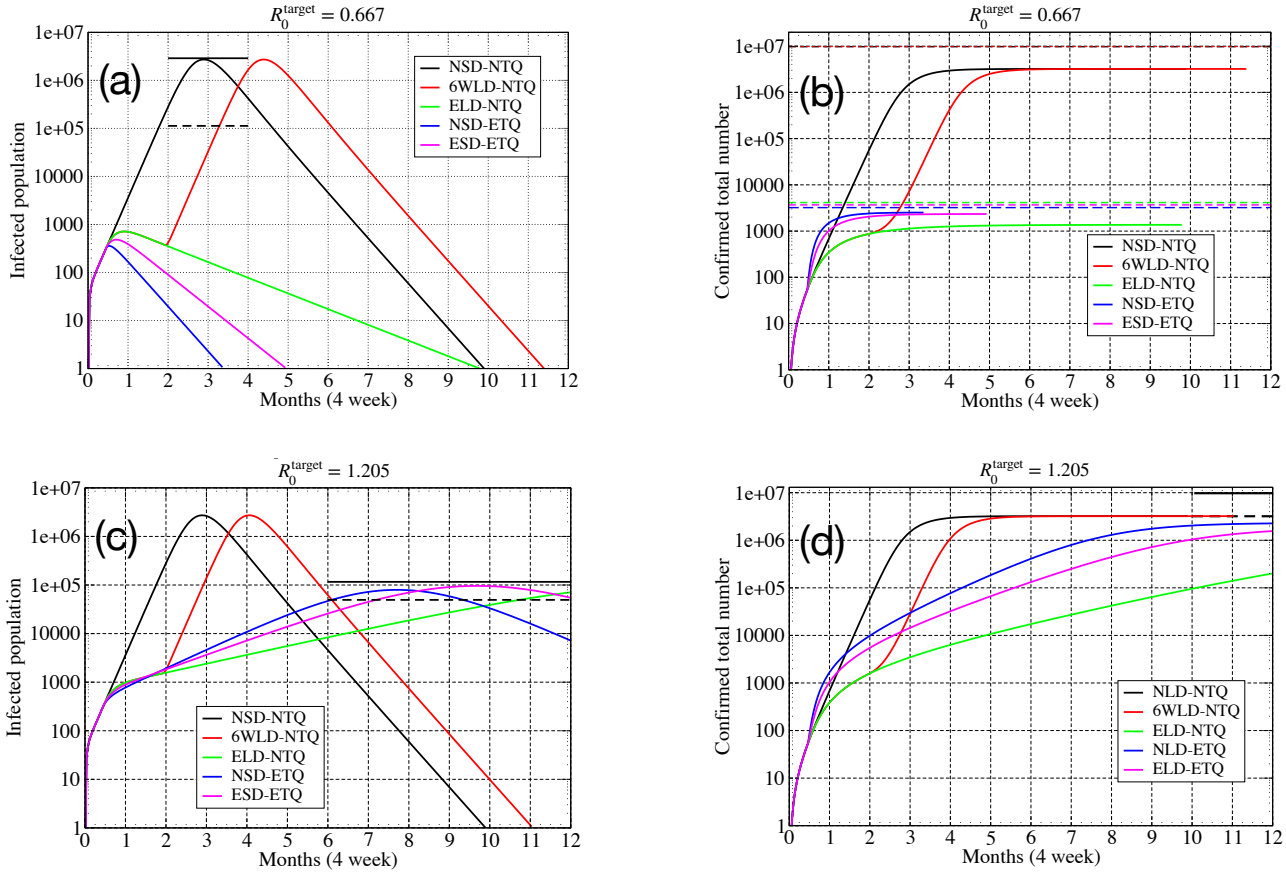

FIG. 2. **Parameter set I** [$R_0^{\text{target}} = 0.667$]: (a) Total number of infected cases $I = I_a + I_p$ for different intervention strategies. The solid and dashed black lines indicate the peak infected cases ($I^{(m)}$) as given by (3) and the corresponding value of $I_p^{(m)}$. (b) Total number 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 (5). 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 $N = 10^7$.

be roughly given by

$$t_{\text{end}} \sim \frac{\ln(\text{Peak infection number})}{|\mu^{\text{post-intervention}}|}, \qquad (2)$$

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

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 Figs. (2c) and (2d).

Main observations:

1. *We find that in this case the peak infections, peak infection time and the final affected population can be obtaied from analytic expressions in terms of μ, R_0 and a few other 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 with 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)

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

and the number of days, $t^{(m)}$, to reach this peak value are given by the simple general relations

$$I^{(m)} \approx \frac{\sigma}{\gamma_e + \sigma}\left(1 - \frac{1 + \ln R_0}{R_0}\right) N. \tag{3}$$

$$t^{(m)} \approx \frac{\ln[I^{(m)}/I(0)]}{\mu} \approx \frac{\ln(N/c)}{\mu}, \tag{4}$$

where γ_e is an effective recovery rate (see Eq. 39 in Sec. III C] and c is a constant that depends on initial infected population and other disease parameters. The fraction of population, $\bar{x} = R(t \to \infty)/N$, that is eventually affected is given by the solution of the equation

$$1 - \bar{x} - e^{-R_0\,\bar{x}} = 0, \tag{5}$$

this result being valid for very general SEIR-type models with multiple compartments. We also provide relations for estimating the number of asymptomatic infected and recovered individuals. These analytic results are very useful to obtain quick heuristic estimates for the typical numbers for peak infections and for the time when the peak occurs. In Sec. I C we use these to make predictions for India, which is an example where only mitigation ($R_0^{\text{target}} \gtrsim 1$) has been achieved.

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 (3) and (4) 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 (5) with $R_0^{\text{target}} = 1.205$ predicts that only about 31% of the population is affected.

B. Results from the linearized theory

As already observed, the linear theory is very useful in understanding the growth and also the 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 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. Plotting the data for number of confirmed cases (normalized by the population), starting with the same initial value, should therefore lead to a collapse of the data for different countries. We test this idea and find that indeed an approximate collapse of data is obtained for a number of countries (see next section).

C. Comparisons with observed data for Covid-19

In this section we discuss a comparison of our results with real data on evolution of the Covid pandemic in different countries. We do not attempt a detailed comparison of the model predictions with the data since there are too many poorly known parameters and possibly quite inaccurate knowledge of the initial conditions of the variables themselves. Rather we make some overall qualitative observations relating 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. For the case of India we discuss predictions, based on our analytic formulas, on the disease evolution for a range of choice of parameter values.

1. *Observation of strong and weak intervention in Covid-19 data*

In Figs. (3) 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. 2(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.

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. III B, 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 $R_0^{target} < 1$ and correspondingly $\mu < 0$. In Fig. 5 we show data for the reported number of new cases in 12 different countries and approximately see these features. Most coun-

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

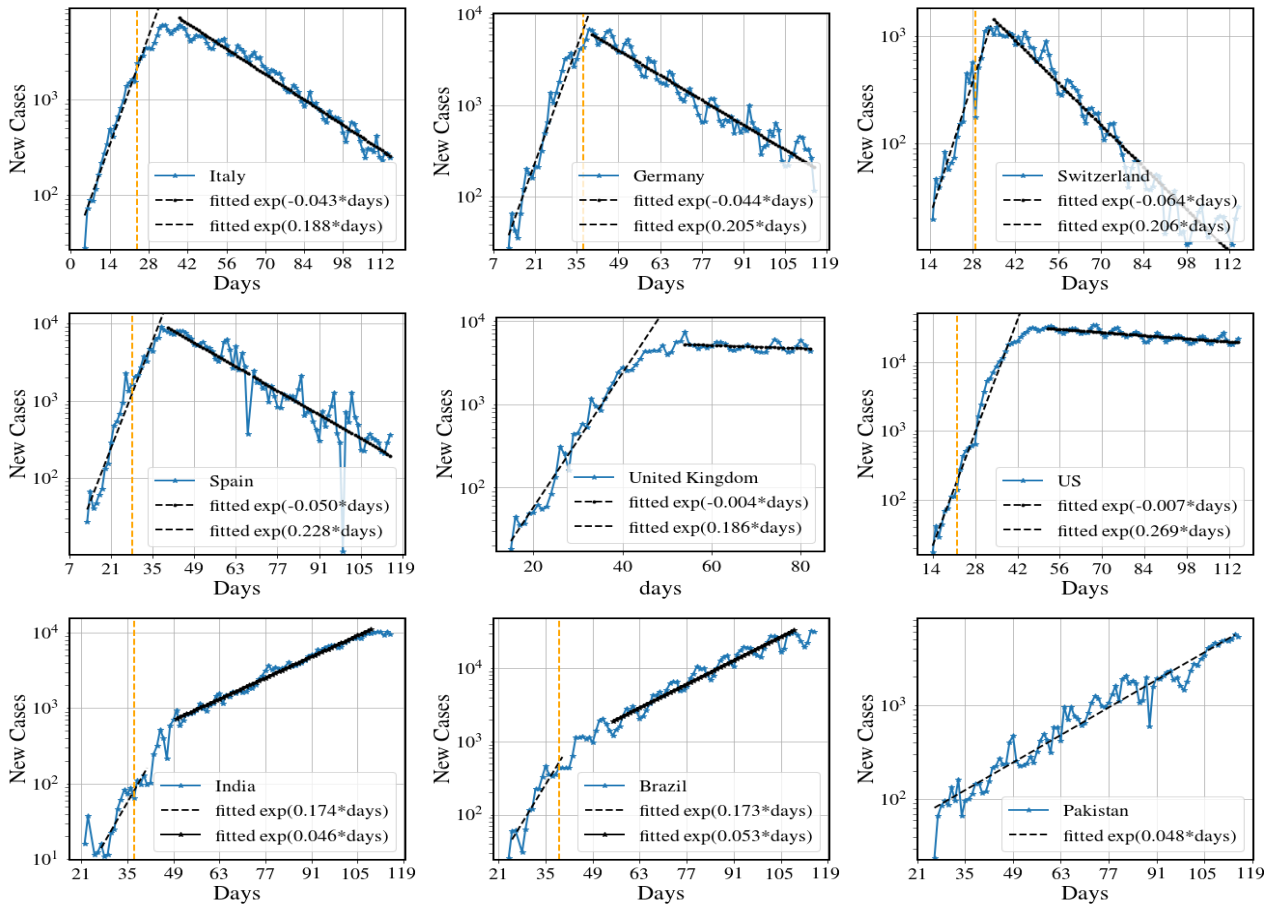

FIG. 3. Number of new cases per day for nine different countries. We note that the first six data sets exhibits the same broad features that we see for the model predictions in Fig. 2(a,b). In particular we see the fast exponential growth and slow exponential decrease in new cases (following strong interventions). 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. 2(c,d) and have only been able to achieve mitigation so far ($R_0^{\text{target}} > 1, \mu > 0$). Data from [19] and the end date is June 10.

tries have succeeded in disease suppression $R_0^{(target)} < 1$, 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 in some more detail the Indian situation.

2. *Comparing data across different countries*

The linearized SEIR dynamics also predicts that (see Sec. III B), 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. 4 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 the right panel of Fig. 5 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.

D. How much testing is required?

How is r related to testing rates ? It is easy to see that with random testing of the population, intervention

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

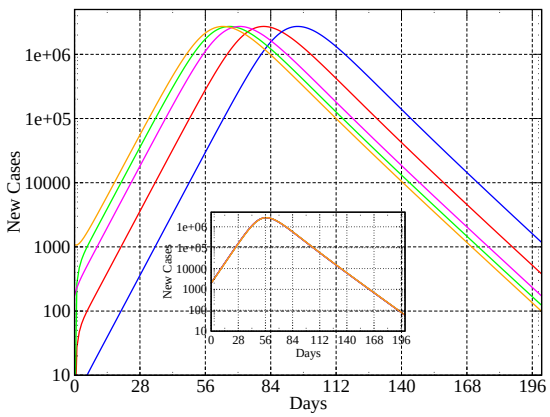

FIG. 4. **Role of initial conditions**: Plot showing $I(t)$, for a fixed population of size $N = 10^7$, with 5 very different initial conditions : (1)$E(0) = 100, I_a(0) = 0, I_p(0) = 0$, (2)$E(0) = 10, I_a(0) = 0, I_p(0) = 0$, (3)$E(0) = 1000, I_a(0) = 0, I_p(0) = 0$, (4) $E(0) = 233, I_a(0) = 100, I_p(0) = 75$, (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.

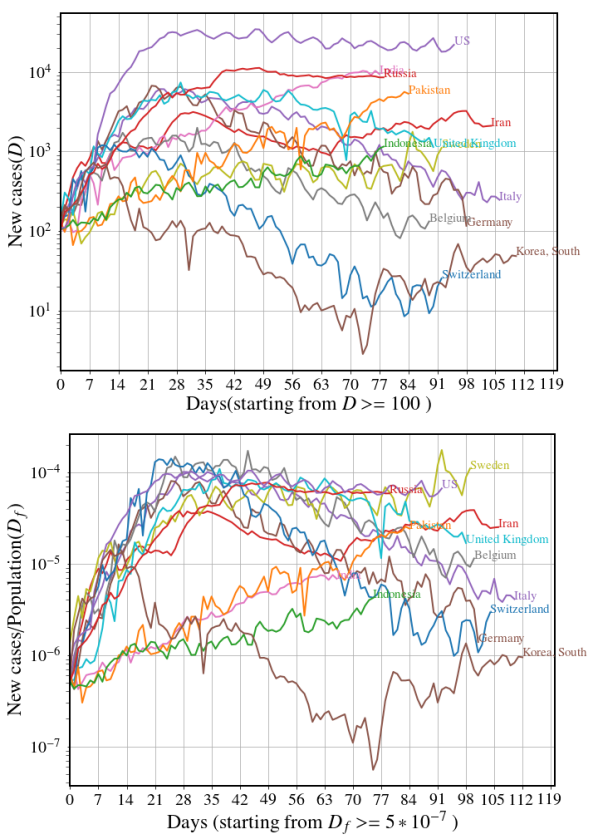

FIG. 5. (left) Number of new cases per day for different countries. (right) Number of new cases normalized by the total population, with the time axis shifted so that every country starts with the same normalized value. Data from [19]

can be helpful only if a finite fraction of the population is tested, which is practically very difficult 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 daily new cases is ΔD and A the typical number of contacts made by a person over the period when the person is infectious and before detection. Following arguments described in Sec. (III E) and Eqs. (45-46), we find that TQ intervention can be successful only if one achieves

$$T(t) \approx \frac{r(t) A \Delta D(t)}{\gamma_p}, \tag{6}$$

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, the recovery rate of symptomatic carriers. This means that we need $T(t) \approx A\Delta D(t)$, that is *the number of tests/per day has to be proportional to number of new detections/per day and in fact the ratio $T/\Delta D$ 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 the value of $T(t)/\Delta D(t) \approx 25$ (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 $T/\Delta D$ 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 our arguments, described later (see Sec. (III E)), are largely *independent of the specifics of the particular SEIR model that we study.*

E. Predictions for India from extended SEIR model

In the following we make some heuristic predictions, based on the analytic results in Eqs. (3-5) 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 ($N \approx 1.3 \times 10^7$). The analysis here is based on the assumption of a best case scenario where the value of R_0^{target} (achieved after a nationwide lockdown of 6 weeks) will be maintained.

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 (39)] $\gamma_e = \gamma/(1-\alpha/3)$ and the effective reproductive number as $R_0^{\text{target}} = (1 - 5\alpha/9)\beta/\gamma$. From this last relation we

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

TABLE I. 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	2nd week September	45 %	1,936,000
0.5	0.2	0.9	2.16	1.3	686,770	3rd week August	42 %	550,700
0.4	0.143	0.67	2.82	1.45	2,956,600	3rd week September	55 %	2,363,700
0.4	0.143	0.9	2.65	1.41	832,890.	4th week August	52 %	676,140

TABLE II. 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	1st week August	45 %	28,296
0.5	0.2	0.9	2.16	1.3	10,037	2nd week July	42.3 %	8,048
0.4	0.143	0.67	2.82	1.45	43,212	2nd week August	55 %	34,546
0.4	0.143	0.9	2.65	1.41	12,173	2nd week July	52.3 %	9,882

TABLE III. 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	3rd week July	45 %	19,360
0.5	0.2	0.9	2.16	1.3	6,867	4th week June	42.3 %	5,507
0.4	0.143	0.67	2.82	1.45	29,566	3rd week July	55 %	23,637
0.4	0.143	0.9	2.65	1.41	8,328	4th week June	52 %	6,761

can write $\beta = \gamma R_0^{\text{target}}/(1-5\alpha/9)$. Plugging this into the equation for the eigenvalues, (19), and replacing λ by the observed mean exponential growth rate $\mu \approx 0.05$ (the value observed for India since around April 10, see Fig. 3), 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 $D_p(0)$ (assuming that only the symptomatics are detected). Then we have $I_p(0) = D_p(0)/\gamma_p$. From (40) 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. (3) to compute the peak number of infections $I^{(m)}$ and the peak daily cases (PDC) is then obtained as 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. (5), 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 $D_p(0) \approx 900$, $D_p(0) \approx 115$ and $D_p(0) \approx 195$ respectively [20]. For a range of choice of the parameters with $\sigma = 0.5,\ 0.4$, $\gamma = 0.2,\ 0.143$ and of $\alpha = 0.67,\ 0.9$, we compute the corresponding values obtained for R_0 and R_0^{target}. These 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 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 [see Fig. (9) in Sec. (III)] 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 I 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 leads to a huge uncertainty in the predictions.*

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

II. DISCUSSION

In summary a modified version of the SEIR model, incorporating asymptomatic individuals, was used 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 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 extended 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. 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). While the lack of precise knowledge of the disease parameters (e.g the fraction of asymptomatic carriers) leads to rather large uncertainties in the predictions, they could perhaps be used to obtain reasonable bounds.

[1] Kucharski, A. J. et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis* **20**, 533-38 (2020) https://doi.org/10.1016/S1473-3099(20)30144-4.

[2] 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.

[3] Ferguson, N. et al. Report 9:Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. https://doi.org/10.25561/77482 (2020).

[4] Tang, B. et al. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov)., *Infectious Disease Modelling* **5**, 248 (2020).

[5] Giordano, G. et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine* https://doi.org/10.1038/s41591-020-0883-7 (2020).

[6] 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. DOI:10.1073/pnas.2004978117.

[7] 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. *medRxiv* https://doi.org/10.1101/2020.04.15.20067256 (2020).

[8] Sarkar, K. and Khajanchi, S. Modeling and forecasting of the COVID-19 pandemic in India. *arXiv*:2005.07071 (2020).

[9] Shekatkar, S. et al. INDSCI-SIM: A state-level epidemiological model for India. https://indscicov.in/indscisim.

[10] Agrawal, S. 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.

[11] Venkateswaran, J. Damani, O. Effectiveness of Testing, Tracing, Social Distancing and Hygiene in Tackling Covid-19 in India: A System Dynamics Model, *arXiv*:2004.08859 (2020).

[12] Meidan, D. et al., Alternating quarantine for sustainable mitigation of COVID-19, *arxiv*.org:2004.01453 (2020).

[13] George N. Wong et al, *arxiv*:2006.02036 (2020)

[14] Liu, Y., Gayle, A. A., Wilder-Smith, A., Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus, *Journal of Travel Medicine* **27**, Issue 2, March 2020, https://doi.org/10.1093/jtm/taaa021.

[15] Leung, K. Y., Trapman, P., Britton, T., Who is the infector? Epidemic models with symptomatic and asymptomatic cases, *Mathematical Biosciences*, **301**, 190 (2018).

[16] Brauer, F., Driessche, P. V., and Wu, J. *Lecture notes in Mathematical Epidemiology* , Springer (2008).

[17] Li J. and Lou Y., Characteristics of an epidemic outbreak with a large initial infection size, Journal of biological dynamics **10**, 366 (2016). https://doi.org/10.1080/17513758.2016.1205223

[18] Coronavirus (COVID-19) Testing, https://ourworldindata.org/coronavirus-testing.

[19] Johns Hopkins Center for Systems Science and Engineering Coronavirus COVID-19 Global Cases. https://systems.jhu.edu/research/public-health/ncov/

[20] 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

III. METHODS

A. 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:

$$\frac{dS}{dt} = -\frac{u(\beta_a I_a + \beta_p I_p)}{N} S \tag{7}$$

$$\frac{dE}{dt} = \frac{u(\beta_a I_a + \beta_p I_p)}{N} S - \sigma E \tag{8}$$

$$\frac{dI_a}{dt} = \alpha\sigma E - \gamma_a I_a - r\nu_a I_a \tag{9}$$

$$\frac{dI_p}{dt} = (1-\alpha)\sigma E - \gamma_p I_p - r\nu_p I_p \tag{10}$$

$$\frac{dU_a}{dt} = \gamma_a I_a \tag{11}$$

$$\frac{dD_a}{dt} = r\nu_a I_a \tag{12}$$

$$\frac{dU_p}{dt} = \gamma_p I_p \tag{13}$$

$$\frac{dD_p}{dt} = r\nu_p I_p. \tag{14}$$

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

$$C = D_a + D_p + U_p, \;\; D = \frac{dC}{dt} = r\nu_a I_a + (\gamma_p + r\nu_p) I_p \; . \tag{15}$$

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.

B. 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 $x_i << N$, the dynamics is captured by linear equations

$$\frac{dX}{dt} = MX, \quad \text{with } X = (x_1, x_2, \ldots, x_8), \tag{16}$$

$$M = \begin{pmatrix} 0 & 0 & -\tilde{\beta}_a & -\tilde{\beta}_p & 0 & 0 & 0 & 0 \\ 0 & -\sigma & \tilde{\beta}_a & \tilde{\beta}_p & 0 & 0 & 0 & 0 \\ 0 & \alpha\sigma & -\tilde{\gamma}_a & 0 & 0 & 0 & 0 & 0 \\ 0 & (1-\alpha)\sigma & 0 & -\tilde{\gamma}_p & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_a & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & r\nu_a & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_p & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r\nu_p & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{17}$$

where $\tilde{\beta}_a = u\beta_a, \; \tilde{\beta}_p = u\beta_p, \tilde{\gamma}_a = \gamma_a + r\nu_a, \tilde{\gamma}_p = \gamma_p + r\nu_p$. 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:

$$\begin{aligned} &\lambda^3 + (\tilde{\gamma}_a + \tilde{\gamma}_p + \sigma)\lambda^2 \\ &+ [\tilde{\gamma}_a\tilde{\gamma}_p + \tilde{\gamma}_a\sigma + \tilde{\gamma}_p\sigma - \alpha\tilde{\beta}_a\sigma - (1-\alpha)\tilde{\beta}_p\sigma]\lambda \\ &\quad + \sigma\left[\tilde{\gamma}_a\tilde{\gamma}_p - (1-\alpha)\tilde{\beta}_p\tilde{\gamma}_a - \alpha\tilde{\beta}_a\tilde{\gamma}_p\right] = 0. \end{aligned} \tag{18}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

This can be re-written in the form

$$\lambda^3 + (\tilde{\gamma}_a + \tilde{\gamma}_p + \sigma)\lambda^2 + [\tilde{\gamma}_a\tilde{\gamma}_p + \sigma(\tilde{\gamma}_a + \tilde{\gamma}_p)(1-Q)]\lambda + \sigma\tilde{\gamma}_a\tilde{\gamma}_p(1-R_0) = 0, \tag{19}$$

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

$$\begin{aligned} R_0 &= \alpha\frac{\tilde{\beta}_a}{\tilde{\gamma}_a} + (1-\alpha)\frac{\tilde{\beta}_p}{\tilde{\gamma}_p} \\ &= \alpha\frac{u\beta_a}{\gamma_a + r\nu_a} + (1-\alpha)\frac{u\beta_p}{\gamma_p + r\nu_p} \end{aligned} \tag{20}$$

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

$$R_0 > 1. \tag{21}$$

We denote the largest eigenvalue by μ. 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 from (19) we can read off the value as

$$\mu \approx \frac{\sigma(R_0 - 1)}{1 + \sigma(\tilde{\gamma}_a^{-1} + \tilde{\gamma}_p^{-1})(1-Q)}. \tag{22}$$

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 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

$$\begin{aligned} x_i(t) &= \sum_j \sum_q \phi_q(i)\chi_q(j)e^{\lambda_q t}x_j(0) \\ &\approx \sum_j \phi_m(i)\chi_m(j)e^{\mu t}x_j(0), \\ &\approx c_m\phi_m(i)e^{\mu t}, \text{ where } c_m = \sum_j \chi_m(j)x_j(0) \end{aligned} \tag{23}$$

and 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:

$$\frac{I_a(t)}{I_p(t)} = \frac{\phi_m(3)}{\phi_m(4)} = \frac{\alpha(\mu + \tilde{\gamma}_p)}{(1-\alpha)(\mu + \tilde{\gamma}_a)}. \tag{24}$$

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

$$x_i(t) \approx \epsilon\phi_m(i)\chi_m(4)e^{\mu t} = a_i\epsilon e^{\mu t}, \tag{25}$$

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

$$\epsilon e^{\mu t_l} = \frac{C_0}{a_6 + a_7 + a_8}. \tag{26}$$

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

$$x_i(0) = \frac{\phi_m(i)}{\phi_m(6) + \phi_m(7) + \phi_m(8)}C_0. \tag{27}$$

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. This also means that trajectories for different initial conditions are identical up to a time translation (See Fig. 4 and related discussion).

C. Final affected population

Let us define the asymptotic populations (i.e the populations at very long times) in the different compartments as $\bar{U}_a, \bar{D}_a, \bar{U}_p, \bar{U}_p$, 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.

It is possible to compute the final affected population $\bar{R}$ from the dynamical equations in 7 - 14. 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 $S(0) \approx N$. Then solving (7), we get

$$\bar{S} = Ne^{-\int_0^\infty dt[\tilde{\beta}_a I_a(t) + \tilde{\beta}_p I_p(t)]/N}. \tag{28}$$

where $\tilde{\beta}_a$ and $\tilde{\beta}_p$ are given after (19). Adding Eqs. (11) and (12) 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 (28) then gives

$$\bar{S} = Ne^{-[(\tilde{\beta}_a/\tilde{\gamma}_a)\bar{R}_a/N + (\tilde{\beta}_p/\tilde{\gamma}_p)\bar{R}_p/N]}. \tag{29}$$

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

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

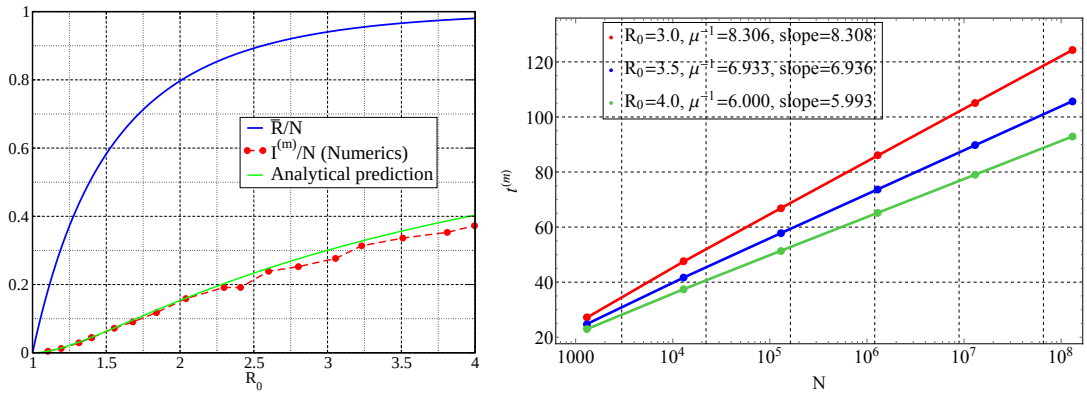

FIG. 6. (left) Plot of the asymptotic total affected population fraction, $\bar{R}/N$, as a function of the reproductive number R_0. We also plot the quantity $(I^{(m)}/N)(\sigma+\gamma_e)/\sigma$, obtained numerically from many different parameter sets, and compare it with the theoretical predicted curve $1-(1+\ln R_0)/R_0$ (green line). (right) Verification of the $\ln(N)$ dependence of $t^{(m)}$ in (38) for different choices of R_0. The slopes of the straight lines compares well with μ^{-1} as stated in (38).

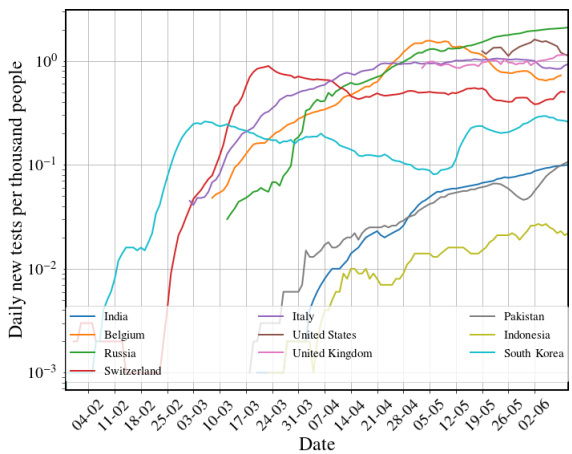

FIG. 7. Data of number of tests per day per thousand in several countries on a log-scale. Notice in particular the large testing numbers in the early period in Korea which has been very successful at controlling the disease. Data from [18].

$I_a = I_p = R_a = R_p = 0$, 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

$$\bar{R}_a = \alpha\bar{R}, \quad \text{and} \quad \bar{R}_p = (1-\alpha)\bar{R}. \tag{30}$$

Using this in (29), 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:

$$1-\bar{x} = e^{-R_0\bar{x}}, \tag{31}$$

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 (31) has a non-zero solution only when $R_0>1$. For the simple SIR model this result is well known [16], 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 the individual populations are then given by

$$\begin{aligned} R_a &= \alpha R,\ R_p = (1-\alpha)R \\ U_a &= \frac{\gamma_a}{\gamma_a+r\nu_a}R_a, D_a = \frac{r\nu_a}{\gamma_a+r\nu_a}R_a, \\ U_a &= \frac{\gamma_a}{\gamma_p+r\nu_p}R_p, D_p = \frac{r\nu_p}{\gamma_a+r\nu_p}R_p. \end{aligned} \tag{32}$$

D. Peak infections and the time at the peak

To compute the maximum of the infected population $I^{(m)}$ and the time $t^{(m)}$ at which this peak appears, let us first look at the basic SEIR model consisting of four compartments with populations S, E, I and R, and the dynamics of these variables described by

$$\frac{dS}{dt} = -\frac{\beta I}{N}S, \tag{33}$$

$$\frac{dE}{dt} = \frac{\beta I}{N}S - \sigma E, \tag{34}$$

$$\frac{dI}{dt} = \sigma E - \gamma I, \tag{35}$$

$$\frac{dR}{dt} = \gamma I. \tag{36}$$

where now β is the infectivity, $E \to I$ transitions happens at a rate σ and recovery $I \to R$ happens at a rate γ. In this case the reproductive number is simply given by $R_0 = \beta/\gamma$. As shown in the previous section it is easy to see that $S(t)$ and $R(t)$ are related by $\frac{dS}{dR} = -\frac{\beta}{N\gamma}S = -\frac{R_0}{N}S$ which gives $S(t) = S(0)e^{-R_0R(t)/N}$. The peak of

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

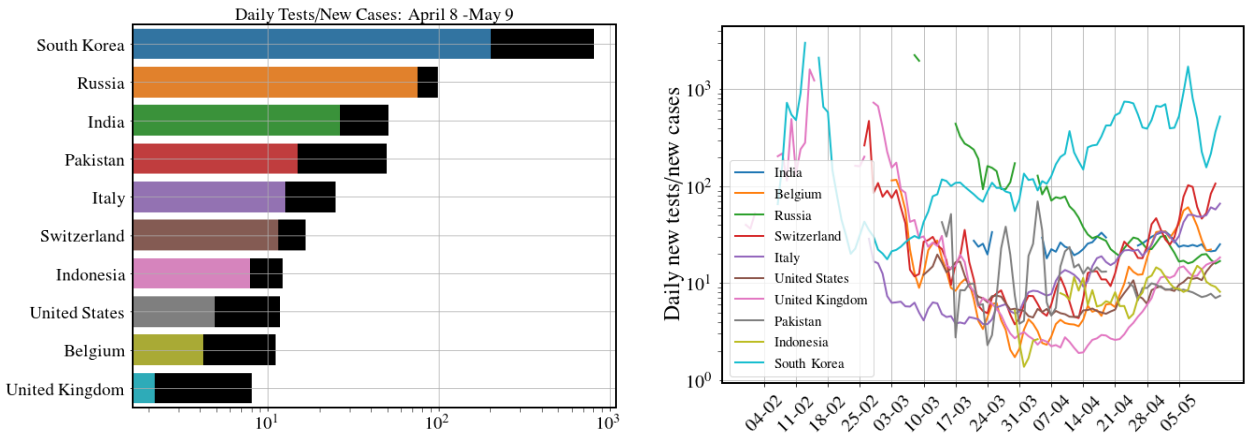

FIG. 8. Data from different countries on the number of tests per detected case $T(t)/D(t)$ for 10 countries (left) on the dates April 8 (coloured bar), May 9 (black bar) and (right) the change over time of this ratio. Data from [18]

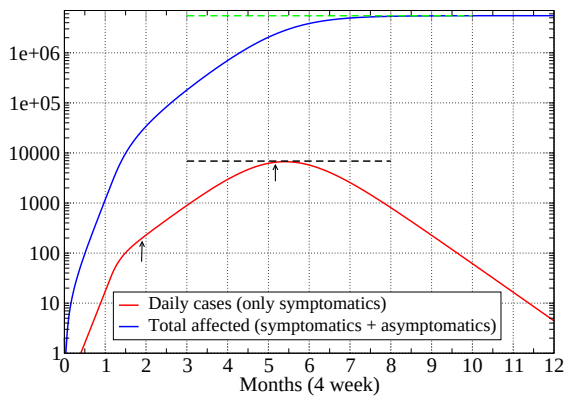

FIG. 9. Plot of the new daily cases ($\gamma_p I_p$) and total affected population fraction (R), as a function number of months for one of the parameter sets in Table. III for the city of Mumbai. Parameter values were $\sigma = 0.5, \tilde{\gamma}_p = 0.2, \alpha = 0.9$, $R_0 = 2.16$ before intervention 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 (38).

the infected population is given by setting $dI/dt = 0$ at $t = t^{(m)}$. By solving the equations (33)-(36) for several set of parameters we observe that E also achieves its peak around the same time. Hence setting also $dE/dt = 0$ at $t = t^{(m)}$, we find $S(t^{(m)}) = S^{(m)}/N = 1/R_0$. On the other hand from $S(t) = S(0)e^{-R_0 R(t)/N}$ 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 that the peak value of the infection number is given by

$$I^{(m)} = \frac{\sigma}{\gamma+\sigma}\left(1 - \frac{1+\ln R_0}{R_0}\right)N. \tag{37}$$

An estimate of the time to reach this peak value can be obtained [17] 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 $I^{(m)} = I(t^{(m)}) = I(0)\ e^{\mu t^{(m)}}$ which finally gives

$$t^{(m)} \sim \frac{\ln[I^{(m)}/I(0)]}{\mu} \sim \frac{\ln(N/c)}{\mu}, \tag{38}$$

where c is a constant that depends on initial infection numbers and parameter values. A verification of this result, obtained by solving the basic SEIR equations numerically, is provided in Fig. 6.

Interestingly, we find that the expression for $I^{(m)}$ also describes quite accurately the peak value for the extended SEIR dynamics with γ now replaced by

$$\gamma_e = [\alpha\gamma_a^{-1} + (1-\alpha)\gamma_p^{-1}]^{-1}. \tag{39}$$

In Fig. 6 we show the dependence of $\bar{x}$ on R_0 (as obtained from a numerical solution of (31)) and provide a numerical verification of the result in Eqs. (37,39) for the extended SEIR model. The peak values of the asymptomatic and presymptomatic populations are respectively given by

$$\begin{aligned} I_a^{(m)} &= \alpha\gamma_e I^{(m)}/\gamma_a, \\ I_p^{(m)} &= (1-\alpha)\gamma_e I^{(m)}/\gamma_p. \end{aligned} \tag{40}$$

In Fig. 9 we show results of a numerical solution of the dynamical equations in presence of intervention (SD) for one of the parameter sets in Table. III and find excellent agreement with our analytic formula in Eqs. (31,37-39). We see that the predicted 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

E. Interventions: Social distancing and Testing-Quarantining

We discuss here the choices of the intervention functions u and r introduced in the dynamical equations 7-14. 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 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

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

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

$$\begin{aligned} r(t) &= 0 \quad t < t'_{\rm on}, \\ &= r_l - r_l e^{-(t-t'_{\rm on})/t'_w}, \quad t > t'_{\rm on}. \end{aligned} \tag{42}$$

where we one needs a final rate $r_l > 0$. In general the time at which the TQ begins to be implemented $t'_{\rm on}$ 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

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

At long times this goes to the targeted reproduction number

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

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. (7) 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 $\Delta D(t) = r\nu_a I_a + (r\nu_p + \gamma_p)I_p = rI + \gamma_p I_p$. In the growing phase we have, from Eq. (24), that $I_a = \alpha I$ and $I_p = (1-\alpha)I$. Hence we get $\Delta D(t) = \hat{\gamma} I$ with $\hat{\gamma} = r + (1-\alpha)\gamma_p$. The total number of contacts of the $I = \Delta D/\hat{\gamma}$ individuals would be $A\Delta D(t)/\hat{\gamma}$, 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

$$r = \frac{T\hat{\gamma}}{A\Delta D(t)} \tag{45}$$

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

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

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

$$T \gtrsim \frac{A\Delta D(t)}{2-\alpha}. \tag{47}$$

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 $T(t) \sim A\Delta D(t)$ is maintained. The implications of this is discussed after (6). In Fig. 7 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. 8 shows data for the ratio $T(t)/\Delta D(t)$ for a set of countries and also how this ratio has evolved over time. As explained earlier (see Sec. I C), 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 $T(t)/\Delta D(t) \approx 25$ may not be sufficient.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.04.20122580; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ACKNOWLEDGMENTS

We thank Jitendra Kethepalli and Kanaya Malakar for very helpful discussions and Ranjini Bandyopadhyay, Siddhartha Chatterjee, Joel Lebowitz, Srujana Merugu, Alpan Raval, Sankar Das Sarma and Sriram Shastry 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.

AUTHOR CONTRIBUTIONS

All authors performed research; S. Goyal analyzed data; A. Das, A. Dhar, A. Kundu wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.