 doi:10.20944/preprints202006.0148.v1

Are models useful? Reflections on simple epidemic projection models and the Covid-19 pandemic

Marc Artzrouni

marc.artzrouni@univ-pau.fr

University of Pau and Pays de l'Adour

Department of Mathematics (CNRS-UMR 5142)

64000 Pau, France.

Paper accepted for publication in *The Mathematical Intelligencer, 2020*

Abstract

In this paper we provide an "expository overview" of classic epidemic projection models. Starting with the simple case of an epidemic that grows exponentially we then investigate "compartmental" models. These assume that the growth of an infected population is limited endogenously by the size of the underlying pool of susceptibles. We then describe a new family of so-called "Exo-r" statistical models, which hinge on an exogenously driven growth rate of the infected population. This family, which can be used to model both infections and deaths, captures parsimoniously both the depletion of susceptibles and the effect of interventions such as lockdowns and "social distancing". The model is used to fit numbers of Covid-19 infections in China. It is also used to model and project deaths in the United States. Results are used to inform a discussion on i) the challenges at hand and ii) the extent to which epidemic projection models may be useful despite being wrong.

Keywords: epidemic model; covid

 doi:10.20944/preprints202006.0148.v1

1 Introduction

"Prediction is very difficult, especially if it's about the future" is a quote one might expect from Groucho Marx or Yogi Berra. Yet, it is attributed to the 20th century Nobel prize-winning Danish physicist Nils Bohr. The quote may be apocryphal but makes a valid point, particularly when it comes to epidemiological predictions.

Fast forward to April 2020 in the midst of the Covid-19 pandemic. Dr. Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, and a lead member of the White House Coronavirus Task Force, is quoted on April 2 as saying "I've looked at all the models. I've spent a lot of time on the models. They don't tell you anything. You can't really rely upon models" (Wan et al, 2020). This quote is too recent to be called apocryphal, but it is surprising and possibly out of context. Indeed, to end with a final aphorism, Dr Fauci surely knows that "all models are wrong, but some are useful" - a pronouncement attributed to the statistician George Box. This is another principle that is particularly true in the case of epidemiological models.

The first goal of this paper is to introduce mathematically inclined readers to a few simple epidemic projection models. We will start with the basic notion of an epidemic growing exponentially. We will then move on to a brief review of "compartmental" models which capture the demographic dynamics of an infected population whose growth is limited endogenously by the size of the underlying population. We next introduce a simple yet novel variant of these models which is driven by an exogenously determined growth rate of the infected population. We call it the "Exo-r" model and fit it to data on infections in China and deaths in the United States.

Informed by these projections we will close by reflecting on the questions raised above: why are epidemiological predictions so difficult, and how do we reconcile an understandable dose of skepticism with the fact that projection models may be useful despite being wrong?

2 Basic epidemic projection models

2.1 The exponential model

Epidemic models require an initial pool of infected individuals. A single person, known as "Patient 0", is enough to trigger an outbreak. We assume that every day an infected person is in contact

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

with c randomly chosen uninfected individuals (the "susceptibles"). A contact in itself does not guarantee transmission. There is a probability p, which depends on the virulence of the disease, that a contact leads to the susceptible becoming infected. If we define $r = cp$ then during a time interval dt each infected person generates on average rdt new infections[1]. The total (cumulative) number of infected individuals $I(t)$ up to day t then satisfies the differential equation

$$\frac{dI/dt}{I} = \frac{\dot{I}}{I} = r \tag{1}$$

where we use a physicist's dot notation over a variable to indicate its derivative with respect to time. Equation (1) is the world's simplest differential equation and has the exponential solution

$$I(t) = I(0)\exp(rt). \tag{2}$$

The derivative $\dot{I}(t)$ is the daily number of new infections on day t, also known as the incidence, the density of new infections, or the "epidemic curve" in epidemiological jargon. Because the numerical value of r does not have an intuitive interpretation we can reparameterize the model by defining the doubling time DT of the total number infected; DT is defined through the equation $I(0)\exp(rDT) = 2I(0)$ with root $DT = \ln(2)/r$.

Most epidemics, including the current Covid-19 pandemic, initially grow exponentially in the same way each dollar invested at a constant savings rate generates every year the same number of new dollars. Barring an economic downturn, nothing prevents your bank account from increasing without bound. The size of an epidemic however, has an upper bound K equal to the size of the whole population - or of a sub-population susceptible to the disease, e.g. with no natural immunity. In population biology K is called the"carrying capacity" of the system.

The next step is therefore a model that accounts for a simple biological reality: at some point those infected run out of susceptibles to infect and the total number infected cannot exceed the

[1] A perceptive non-specialist may wonder how one generates infinitesimal numbers of infected individuals - particularly if we start off with a single "Patient 0". We focus here on *deterministic* models which assume that large numbers of infected individuals generate expected, average numbers of new infections. For example if each one of 100 infected individuals generates a new case with probability 0.123 we say there will be 12.3 new infections - which combines the law of large numbers with a disregard for the fact that humans do not come in decimal numbers. Another way of dealing with a single "Patient 0" is with a *stochastic* model which hinges on a discrete probability distribution for the number of new infections generated by said "Patient 0".

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

carrying capacity K.

2.2 The logistic model: two compartments

A simple way of incorporating the upper limit K into the model is by transforming the growth rate r on the right-hand side of Eq. (1) into a quantity that decreases as the susceptible population is depleted while preventing the infected population from exceeding K. This can be achieved by transforming r into a cleverly defined decreasing function of I, for example

$$\frac{\dot{I}}{I} = r\left(1 - \frac{I}{K}\right). \tag{3}$$

The decrease of the growth rate $\dot{I}/I$ has been "endogenized", i.e. it is driven by the number $I(t)$ of infectives. The fact that $I(t)$ is a cumulative number of infected individuals means that its growth rate on the right-hand side of Eq. (3) must remain positive - which happens if and only if $I(t)$ is less than K for all t. This suggests that we are on the right track. (We assume that the intial number infected $I(0)$ is much smaller than K).

Before giving the solution to this differential equation we examine its "mechanistic" interpretation, i.e. the biological process and human behavior described by Eq. (3). New infections result from contacts between the $I(t)$ infected (a first compartment/variable) and the $S(t) = K - I(t)$ susceptibles (second compartment/variable). To determine how many new infections arise each day from these contacts we make two simple assumptions: 1) as with the exponential model every person (infected or not) each day comes into contact with the same number c of randomly chosen individuals, and 2) these contacts are "homogeneous", i.e. do not discriminate between susceptible and infected status. This means that if John is infected then only a fraction $S(t)/K$ of his c contacts are susceptible - a fraction that decreases as the susceptible population gets depleted. John, who is infected, is therefore in contact with $cS(t)/K$ susceptible individuals per day. If the probability of transmission is p then John generates $pcS(t)/K$ "secondary cases" (new infections) per day. Since there are $I(t)$ Johns at time t, they generate a total of

$$I(t+dt) - I(t) = I(t)pc\frac{S(t)}{K}dt = I(t)pc\left(1 - \frac{I(t)}{K}\right)dt \tag{4}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

new infections during the time interval $(t, t+dt)$ (think of dt as being equal to 1). When $dt \to 0$ this is Eq. (3) with $r = pc$.

A daily number $\dot{I}(t)$ of new infections that is proportional to $I(t)(K - I(t))$ provides an epidemiological example of the "mass action principle" familiar to chemists and physicists: loosely speaking a product of components feeds back into the process that generates these components.

Equation (3) is a rare case of a differential equation with a closed form solution:

$$I(t) = \frac{K}{1 + A\exp(-rt)} \tag{5}$$

with

$$A = \frac{K - I(0)}{I(0)}. \tag{6}$$

Equation (5) is known as the logistic function. It has had a long and illustrious career - including with recent applications, extensions, and generalizations used to fit and project the Covid-19 epidemic in China and elsewhere (Wu et al., 2020). As one might expect this function is S-shaped and tends to K when $t \to \infty$. Differentiating (5) one finds that the daily number of new infections is

$$\dot{I}(t) = \frac{AKr\exp(-rt)}{(1 + A\exp(-rt))^2}. \tag{7}$$

This is a symmetric bell-shaped curve - the one politicians desperately want to "flatten" during the Covid-19 pandemic of 2020. To see how this can be done we differentiate Eq. (7) and find that $\dot{I}(t)$ reaches its maximum at the critical value $t_c = \ln(A)/r$. The corresponding maximum incidence, denoted $\dot{I}_{max}$, is $\dot{I}(t_c) = Kr/4$.

These results show that there are two way of flattening the $\dot{I}(t)$ curve. We can reduce $\dot{I}_{max}$ by decreasing r which delays the peak time t_c without changing the ultimate size of the epidemic. We can also lower the eventually infected population K (quarantine, vaccination, etc) which flattens the curve without delaying the peak t_c.

A first drawback of the logistic function is the symmetry of the curve, which does not allow for an epidemic with a rapid rise to its peak followed by a slower decline. (This was the case for the Covid-19 epidemics in China and South Korea in early 2020). Another more serious conceptual

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

drawback is that the model has only two "compartments" (variables) with individuals flowing from the susceptible compartment to the infected one. This means that once infected, a person can immediately and indefinitely spread the disease. In particular, no one ever recovers, becomes immune, or dies from the disease. We continue with a quick review of models with three and four compartments, which correct the aforementioned shortcomings.

2.3 SIR and SEIR models: three and fours compartments

An "SIR" model will add to the susceptible and infected compartments one that has $R(t)$ recovered individuals. An "SEIR" model will add another compartment of $E(t)$ "exposed" individuals who have been infected but cannot yet transmit the disease. These models result in systems of three or four differential equations, which are based on the mass action principle and have variants of Eq. (3) at their core.

These models are particularly useful when a large fraction of the population becomes infected and runs out of susceptibles to infect. This can lead to "herd immunity", which occurs when roughly 60 or 70 percent of the entire population is immune either because they have recovered and have developed an immunity, or have been vaccinated. This effect saturates the population with "uninfectables" and acts as a rate-limiting factor that can "snuff out" the epidemic.

3 The phenomenological Exo-r model

Think now of the Covid-19 epidemic in China, which started in January and had run its course by the end of March. Even if the official total count around 84,000 cases is under-reported by a factor of 10, this means less than a million people were infected out of a population of 1.4 billion. In other words well under one in a thousand Chinese became infected. There was no mass action principle to slow down the spread of the disease - and even less herd immunity. Rather, the spread was brought under control through exogenously imposed control measures which reduced the rate of infection (social distancing, lockdowns, quarantines, etc).

We will capture this effect with a growth rate $\dot{I}/I$ that is no longer an endogenously decreasing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

function of time (right-hand side of (3)) but simply a function $r(t)$ of time:

$$\frac{\dot{I}}{I} = r(t). \tag{8}$$

This type of model is called "phenomenological". This is a fancy word to describe mathematical formulations of real-life problems that are consistent with the data, *without attempting to describe the underlying mechanism.*[2] The only constraint is that $r(t)$ be positive since $I(t)$ is a cumulative number of infections and is therefore necessarily increasing. The solution $I(t)$ of (8) is

$$I(t) = I(0)e^{\int_0^t r(s)ds} \tag{9}$$

which has a closed form if $r(t)$ can be integrated. The derivative of $I(t)$ is

$$\dot{I}(t) = I(0)r(t)e^{\int_0^t r(s)ds}. \tag{10}$$

Because this model is driven by an exogenously determined growth rate $r(t)$ we will call it for brevity the "Exo-r" model.

3.1 Specification of growth rate $r(t)$

Before specifying a functional form for the growth rate we turn our attention to the doubling time $DT(t) = \ln(2)/r(t)$. This metric is popular because it is easy to interpret, even in the case of a time-varying rate $r(t)$: $DT(t)$ is the time it would take for the infected population to double if the growth rate after time t were frozen at the value $r(t)$.

Figure 1 depicts on a logarithmic scale the doubling times of total confirmed Covid-19 cases for three typical countries[3]. They all show a similar pattern. There is an early period of erratic logarithms of the doubling time $LDT(t) \stackrel{\text{def.}}{=} \ln(DT(t))$ due to small numbers and reporting problems/delays. We will consider that $LDT(t)$ is constant during this early stage of the epidemic.

[2]Such models can justifiably be criticized for ignoring demographic mechanisms which are fully exploited in SIR and SEIR models. Still, phenomenological models expand our methodological toolbox and can be useful. They are sometimes called "statistical" because they can be fitted to data with statistical methods.

[3]These Creative Commons data and visualizations are freely available and usable at `https://ourworldindata.org/grapher/doubling-time-of-covid-cases?yScale=log`.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

This means an exponential increase, which is common when a disease initially spreads unhindered. Figure 1 shows that at some point the logarithm $LDT(t)$ increases roughly in a linear fashion, i.e. $LDT(t) \approx qt+u$, with $q > 0$. Therefore the growth rate $r(t) = \ln(2)/DT(t) \approx \ln(2)e^{-qt-u}$ decreases exponentially, which reflects a diminishing pool of susceptibles and/or the onset of intervention

Given these observations we propose a stylization of the growth rate $r(t)$ as a constant r_0 between time 0 and some $t^* \geq 0$ followed by an exponential decrease with a negative decay rate s $(= -q)$:

$$r(t) = \begin{cases} r_0 & \text{if } t \leq t^* \quad (11a) \\ r_0 \exp(s(t - t^*)) & \text{if } t > t^*. \quad (11b) \end{cases}$$

As before we define the doubling time $DT_0 = \ln(2)/r_0$ during the early exponential phase. We can reparameterize s through the half-life HL of the growth rate $r(t)$ after time t^*; HL is the time it takes for $r(t)$ to drop by a half after t^*: $HL = -\ln(2)/s$. Equation (11) can be paraphrased by noting that the doubling time $DT(t) = \ln(2)/r(t)$ is then

$$DT(t) = \begin{cases} DT_0 & \text{if } t \leq t^* \quad (12a) \\ DT_0 \exp\left(\frac{\ln(2)}{HL}(t - t^*)\right) & \text{if } t > t^*. \quad (12b) \end{cases}$$

This equation shows that the half-life HL of the growth rate is also the doubling time of the doubling time $DT(t)$. The logarithm of the doubling time is then

$$LDT(t) = \begin{cases} \ln(DT_0) & \text{if } t \leq t^* \quad (13a) \\ \ln(DT_0) + \frac{\ln(2)}{HL}(t - t^*) & \text{if } t > t^*. \quad (13b) \end{cases}$$

The functions $DT(t)$ on a logarithmic scale and $LDT(t)$ on a linear scale are both linearly increasing for $t > t^*$ (with slopes $\log(e)\ln(2)/HL$ and $\ln(2)/HL$ respectively).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

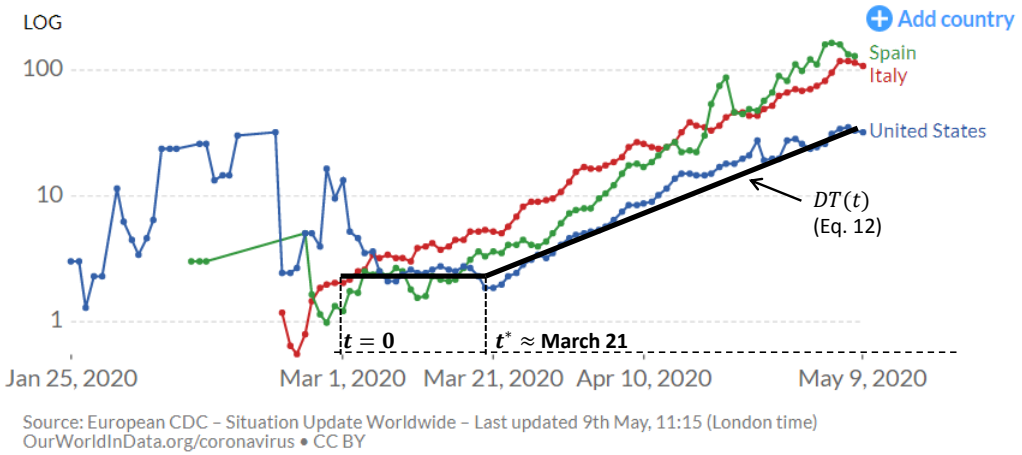

Figure 1: Doubling time of total confirmed Covid-19 cases for three countries - up to May 9, 2020 (logarithmic scale). The superimposed modeled $DT(t)$ (black line) captures for the U.S. the pattern of a constant value DT_0 roughly from March 1 ($t = 0$) to March 21 ($t^* = 21$). For $t > t^*$ the logarithm of $DT(t)$ increases roughly linearly with slope $\ln(2)/HL$. (In Spain and Italy, where there appears to be little or no initial exponential stage, we might take a duration $t^* = 0$ for that stage).

3.2 Results

Routine integrations left as an exercise show that with $r(t)$ of Eq. (11) the infections $I(t)$ of (9) are

$$I(t) = \begin{cases} I(0)\exp(r_0 t) & \text{if } t < t^* \quad (14a) \\ I(0)\exp\left(r_0 \dfrac{\exp(s(t-t^*)) - 1 + st^*}{s}\right) & \text{if } t \geq t^*. \quad (14b) \end{cases}$$

Because $s < 0$ the ultimate number infected is equal to

$$I_\infty \stackrel{\text{def.}}{=} \lim_{t\to\infty} I(t) = I(0)\exp\left(r_0 \frac{st^* - 1}{s}\right). \quad (15)$$

Bearing in mind Eq. (11b) we note that for $t > t^*$ we have $I(t)/I_\infty = \exp\left(\dfrac{r(t)}{s}\right)$: This fractional number of total infections at time $t > t^*$ depends only on the current growth rate $r(t)$ and the rate

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

$s < 0$ at which $r(t)$ decays. The derivative of $I(t)$ in (14) is

$$\dot{I}(t) = \begin{cases} I(0)r_0 \exp(r_0 t) & \text{if } t < t^* \quad (16a) \\ I(t)r_0 \exp(s(t-t^*)) & \text{if } t \geq t^* \ . \quad (16b) \end{cases}$$

This expression tells us that $\dot{I}(t) \sim \exp(st)$ for large t. Therefore the number of daily cases $\dot{I}(t)$ decays asymptotically at the rate s, i.e. with the same half-life $HL = -\ln(2)/s$ as $r(t)$. We can therefore have an asymmetric epidemic curve $\dot{I}(t)$ which grows at the positive exponential rate r_0 early in epidemic and at the negative exponential rate s for $t \to \infty$.

With an exponential growth until t^* the density of new cases $\dot{I}(t)$ can only peak after t^*. If we differentiate (16b) we find that there are two scenarios depending on the values of the growth and decay rates r_0 and s.

1. Scenario A: $s < -r_0$ ($\Leftrightarrow HL = -\ln(2)/s < DT_0 = \ln(2)/r_0$), i.e. $r(t)$ decays faster than the initial infected population increases. Then $\dot{I}(t)$ reaches its maximum $\dot{I}_{max}$ at time t^*, the end of the exponential stage:

$$\dot{I}_{max} \overset{\text{def.}}{=} \dot{I}(t^*) = I(0)r_0 \exp(r_0 t^*). \quad (17)$$

   At t^* the second derivative $\ddot{I}(t)$ goes discontinuously from 0 to a negative value causing an unconventional Alpine peak of a maximum. Still, this is of interest: if the growth rate starts decreasing fast enough then the number of new cases $\dot{I}(t)$ immediately starts decreasing.

2. Scenario B: $-r_0 \leq s < 0$ ($\Leftrightarrow DT_0 \leq HL$), i.e. $r(t)$ decays slower than the initial infected population increases. Then $\dot{I}(t)$ reaches its maximum at the critical value

$$t_c \overset{\text{def.}}{=} \frac{\ln\left(\frac{-s}{r_0}\right)}{s} + t^* > t^* \quad (18)$$

   at which time the second derivative $\ddot{I}$ vanishes. The corresponding maximum daily incidence is

$$\dot{I}_{max} \overset{\text{def.}}{=} \dot{I}(t_c) = -I(0)\frac{s}{e} \exp\left(r_0 \frac{st^* - 1}{s}\right) = \frac{-s}{e} I_\infty. \quad (19)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

3.3 Insights

3.3.1 Ultimate infected population

If we recall that $-s = \ln(2)/HL$ where HL is the half-life of the growth rate $r(t)$ then Eq. (19) shows that the ultimate number infected is

$$I_\infty = \frac{e}{\ln 2/HL}\dot{I}_{max} = \overbrace{3.922HL}^{\text{Penetration } P} \times \dot{I}_{max}. \tag{20}$$

This result is surprisingly simple and highlights the role of HL: each extra day by which the half life HL of $r(t)$ can be shortened reduces the overall size of the epidemic by almost four times the maximum incidence. Equation (20) shows that the initial population $I(0)$, the growth rate r_0, and the duration t^* effect I_∞ through the maximum incidence $\dot{I}_{max}$. Multiplying the half-life by 3.922 yields the dimensionless disease penetration (or "protraction") factor $P = 3.922HL$ equal to the ratio of ultimate to maximum infected populations: the larger HL (or P) is the more protracted the outbreak.

We now illustrate the use of Eq. (20) with an attempted back-of-the-envelope estimate of the ultimate number of cases I_∞ in the United States. Indeed, if we can estimate HL and are sufficiently advanced in an epidemic to observe $\dot{I}_{max}$, then Eq. (20) is a window into the future. During March and April the slope $\ln(2)/HL$ for the logarithm of the doubling time (Figure 1) yields a half-life HL around 12 days, i.e. a penetration factor $P = 3.922HL \approx 47$.

In May the daily incidence $\dot{I}(t)$ had plateaued in the 25,000 to 30,000 range with a slight downward trend[4]. Simple epidemic curves usually peak and go back down fairly quickly. A long plateau means something more complicated is going on. In the case of Covid-19 in the U.S. this could mean reporting or testing problems. This plateau could also reflect a complex and spatially heterogeneous epidemic that is rippling through the country over several months. It would therefore be a miracle if its dynamics could be described accurately with our simple three-parameter model. Still, going on the high side with 30,000 for $\dot{I}_{max}$ we find that Eq. (20) yields $\dot{I}_{max} = P \times 30,000 = 1.4M$. By May 10 the number of confirmed cases had reached 1.3 M. Our 1.4 M is the right order

[4]In an abuse that would horrify statisticians we use throughout the same notations $\dot{I}(t)$, $DT(t)$ etc. for observed and modeled values. We do this to avoid cumbersome notations. Observed daily incidences should really be noted something like $\dot{I}^o(t)$ to distinguish them from modeled values $\dot{I}(t)$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

of magnitude but obviously falls short.

The reason for the low estimate of 1.4 M can be found in the U.S. data of Figure 1. As a first approximation the trend for $LDT(t)$, the logarithm of $DT(t)$, appears to be roughly linear beyond time $t^* \approx$ March 21. However, a closer look at Figure 1 suggests that beyond April 15 $LDT(t)$ continues as a straight line but with a smaller slope. This means a larger HL and implies a growth rate $r(t)$ that decreases more slowly. The resulting more protracted epidemic could be caused by a virus that spreads unhindered into new areas.

This effect can be captured by adding a third stage to the model starting April 15, say; $LDT(t)$ continues linearly after that date with a smaller slope $\ln(2)/HL'$ and $HL' > HL$. This amounts to "restarting" the model with time $t = 0$ on April 15. The doubling time on April 15 is the new DT_0. We take a new exponential duration $t^{*'}$ of 0 (no exponential phase) and a new half-life HL' for the growth rate after April 15. In the present case HL' would be larger than the early HL. It could also be smaller than HL if intervention measures are increased or a vaccine removes large numbers of potential susceptibles. Those are just a few ways in which exogenous effects are captured through a time-varying growth rate.

3.3.2 Timing and intensity of interventions

With closed form expressions for epidemiological parameters one can explore analytically the effect of the timing t^* and intensity HL of interventions on the disease dynamics. In particular one can compare ways of flattening the curve $\dot{I}(t)$, while not losing track of the timing t_c of this peak and the total number infected I_∞[5]. With a fixed $\dot{I}_{max}$ in mind we solve Eq. (17) for t^*:

$$t^* = \frac{DT_0 \left(1 + \ln\left(\frac{HL \times \dot{I}_{max}}{I(0)\ln(2)}\right)\right) - HL}{\ln(2)}. \tag{21}$$

This equation can be viewed as the locus of values (HL, t^*) corresponding to specified values of $I(0), DT_0$ and $\dot{I}_{max}$. A detailed examination of this equation is beyond the scope of this paper. We will just illustrate its use with a numerical example inspired by the U.S. data considered above. We

[5]When journalists talk of "flattening the curve" they seem to think this can be achieved without changing its timing t_c or the total area I_∞ under the curve. We will see that these parameters do change.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

take $DT_0 = 2.3$, a typical doubling time for the initial stage and $I(0) = 70$. In Figure 2 we plotted the two (HL, t^*) loci for a baseline $\dot{I}_{max} = 30,000$ and a significantly flatter $\dot{I}_{max} = 10,000$.

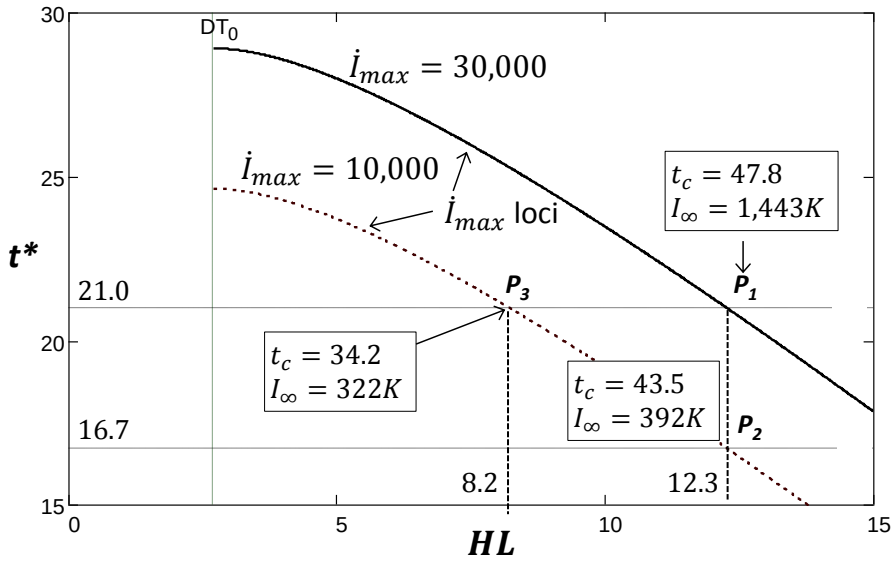

Figure 2: Loci of (HL, t^*) values for which the maximum incidence is $\dot{I}_{max} = 10,000$ and $\dot{I}_{max} = 30,000$, with $DT_0 = 2.3$, $I(0) = 70$. A reduction from 30,000 to 10,000 is possible with (HL, t^*) moving from P_1 to P_2 with a 20% reduction of t^*. A move from P_1 to P_3 requires a 33% drop in HL but with more favorable outcomes concerning the timing t_c of the peak and the total number infected I_∞.

In our hypothetical example authorities want to flatten the curve because they fear a maximum daily incidence of 30,000 would overwhelm the health system. They want to know how to bring the number down to 10,000. Although there are infinitely many points to choose from on the $\dot{I}_{max} = 10,000$ locus we will just consider the two straight-line paths. We can take (HL, t^*) vertically down from P_1 to P_2 by bringing t^* down from 21.0 to 16.7 days (a 20% reduction in t^*). We can also move horizontally from P_1 to P_3 by shifting HL leftward from 12.3 to 8.2 days (a 33% reduction in HL).

It could be tempting to conclude that the vertical path down to P_2 is the better solution because it requires a 20% reduction in t^* as opposed to a 33% reduction of HL along the horizontal path. However this is simplistic. First, this line of reasoning implicitly assumes that the cost of a one percent reduction is the same for both parameters - which is unlikely to be the case. Secondly, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

horizontal path to P_3 may require a larger percentage drop in HL but the metrics other than an $\dot{I}_{max} = 10,000$ are more favorable than with the vertical path down to P_2. Indeed, the peak incidence time t_c occurs 43.5-34.2=9.3 years earlier and the total number infected I_∞ is 392-322=70K smaller than at P_2.

This example shows that a mathematical model simple enough to provide closed form expressions for the outputs of interest is only a first step. Results need to be fed into carefully crafted economic models in order to reach truly optimal decisions in the management of an infectious disease.

We next apply the model to (mainland) China, where the epidemic originated but was rapidly brought under control through strictly enforced lockdowns and social distancing measures.

4 Applications

4.1 China: a completed epidemic with stringent "Non-pharmaceutical interventions" (NPIs)

We plotted in Panel A of Fig. 3 the doubling times of numbers infected from January 22 to May 11, for mainland China (black dots). This panel is analogous to the one in Fig. 1. Black and red dots in Panel B represent reported total and daily numbers of infections.

Our primary goal is to fit as well as possible the modeled total numbers $I(t)$ to observed counts. This is because total numbers carry with them the entire history of the epidemic. Our secondary goal was to keep modeled doubling times $DT(t)$ and daily numbers $\dot{I}(t)$ as close as possible to observed values. Parameter values $t^*, DT_0, I(0)$ and $HL = \ln(2)/r_0$ satisfying these conditions were found without much difficulty by trial and error and are given in Fig. 3. In particular the coefficient of penetration $P = 3.922HL$ is $3.922 \times 4.3 \approx 17$.

A (semi) logarithmic scale is often used to describe the spread of Covid-19 because it captures well a range of infected numbers from very small to very large. It also highlights the exponential stages of the epidemic. Indeed, the straight upward (red) line for $\dot{I}(t)$ between 0 and $t^* = 10$ reflects the initially exponential growth. The downward straight line in the latter part of the projection reflects the asymptotically exponential decay at rate $s < 0$.

Panel B of Fig. 3 shows a good fit for $I(t)$ except for a divergence between reported and modeled

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

total cases for a couple of weeks following the end of the first stage at time $t^* = 10$. The slowdown in reported cases was followed by a large increase on the 23rd day (Feb 13). This discontinuity in the data reflected a broadening of reporting criteria. Still, the trials and errors of a human (as opposed to a software's number crunching) produced a fitted model that smoothly corrected for the shortfall of reported cases before the 23rd day. Leaving aside this spurious discontinuity on Feb 13, the fit is better with cumulative numbers than with daily ones. This is because the former are strictly increasing and smooth out daily numbers which can be erratic (Panel B).

Panel A shows that a good fit for total numbers $I(t)$ comes at the cost of a modeled $DT(t)$ that is in agreement with observed doubling times only between the 10th and approximately 55th day. Panel B shows that this was a month and half during which much of the spread took place. After the 55th day observed doubling times drift downwards, remain around 1000 for a few weeks, and then climb into the thousands. This reflects a late "endemic" stage with a very low growth rate that makes little difference to an epidemic that had essentially ended.

We note in Panel B the corresponding late daily numbers under 100 after the 55th day (red dots). They diverge from the expected exponential path to 0 and drop to low double digits by the 100th day. This divergence is magnified with a logarithmic scale but makes only a small dent to a total number of cases approaching 82,000. A linear scale would be more forgiving because late stage discrepancies of a few dozen cases would be practically invisible in view of a maximum number of daily cases $\dot{I}_{max}$ close to 5,000.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

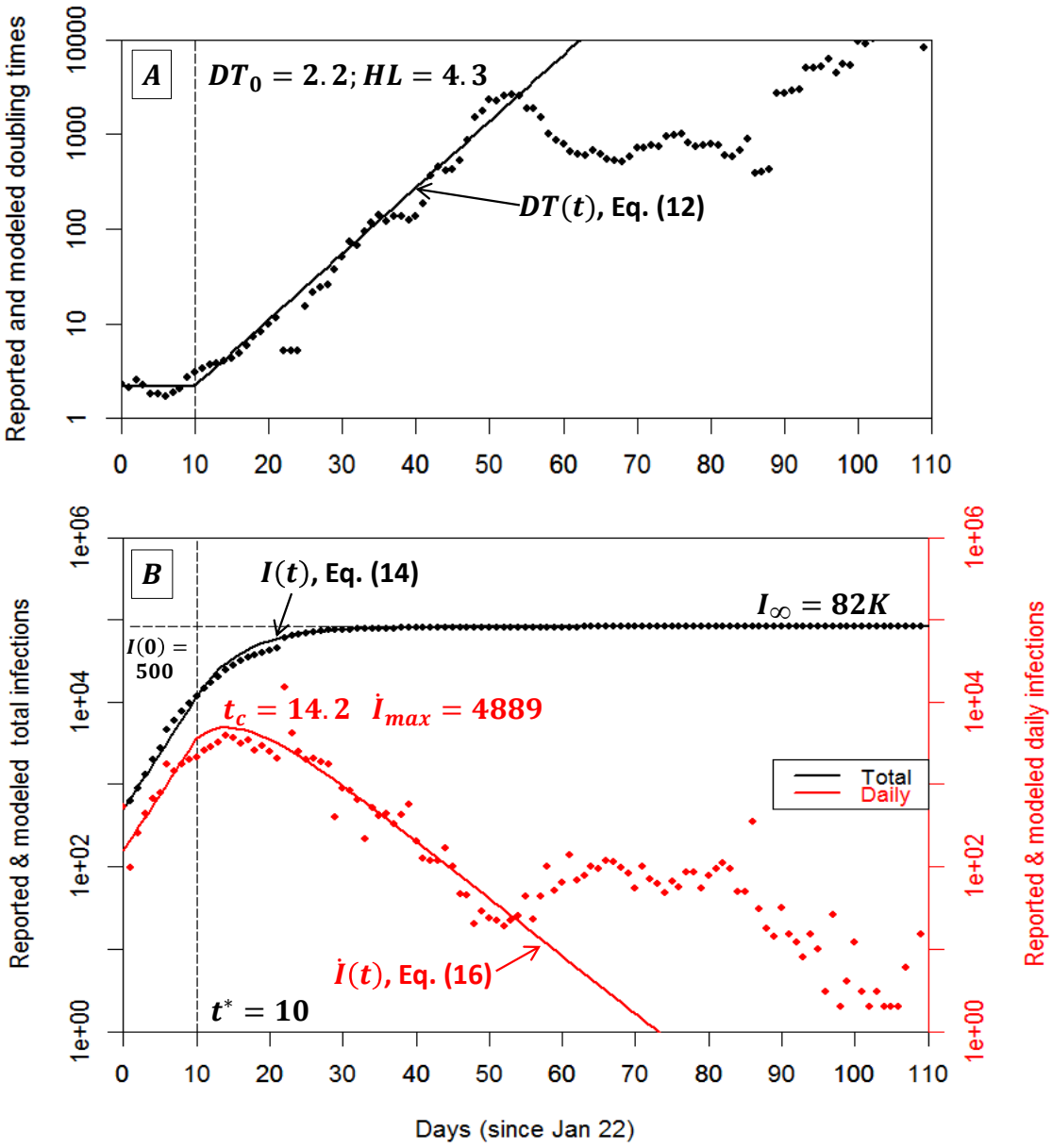

Figure 3: Panel A: Reported and modeled doubling times $DT(t)$ in mainland China, from January 22 to May 11, 2020 (110 days). Panel B: Total reported and modeled numbers $I(t)$ in black; daily reported and modeled numbers $\dot{I}(t)$ in red (semi-logarithmic scales).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

4.2 Exo-r modeling of deaths

4.2.1 From infections to deaths

In the case of Covid-19 true numbers of infections are several times the reported ones. Given these uncertainties and the keen interest health authorities have in the number of deaths, we ask whether our phenomenological Exo-r model could be used to model and project Covid-19 deaths as well as infections. This would require some sort of lagged relationship between infections and deaths. To explore this possibility we let $D(t)$ and $\dot{D}(t)$ be the total and daily numbers of deaths on day t. We assume i) the *true* numbers of infections is a_1 times the *reported* one modeled as $I(t)$ and $\dot{I}(t)$ in Section 4); ii) the probability of death due to Covid-19 is a_2; and iii) the time between infection and death (when it occurs) is a constant a_3. Then the number of deaths $\dot{D}(t)$ on day t is the "real" number $a_1\dot{I}(t-a_3)$ of infections a_3 days earlier multiplied by the probability of death a_2:

$$\dot{D}(t) = a_1 a_2 \dot{I}(t-a_3). \tag{22}$$

This is a highly simplified relationship which captures the essence of a more complex reality[6]. Still, Eq. (22) shows that if infections and deaths can be considered at least roughly as expansions/translations of one another, and if reported infections follow an Exo-r model (as in China), then it may be possible to also fit the Exo-r model to deaths.

We begin by replacing all I's of Section 3.2 with D's. The expressions are recalled here for convenience. Total deaths are modeled as

$$D(t) = \begin{cases} D(0)\exp(r_0 t) & \text{if } t < t^* \quad (23a) \\ D(0)\exp\left(r_0 \dfrac{\exp(s(t-t^*)) - 1 + st^*}{s}\right) & \text{if } t \geq t^* \quad (23b) \end{cases}$$

where t^* and r_0 are now the duration of the exponential stage and the corresponding (initial) growth rate of the number of deaths; s is the rate of decay of $r(t)$ after t^*. The derivative is

$$\dot{D}(t) = \begin{cases} D(0) r_0 \exp(r_0 t) & \text{if } t < t^* \quad (24a) \\ D(0) r_0 \exp\left(r_0 \dfrac{\exp(s(t-t^*)) - 1 + st^*}{s} + s(t-t^*)\right) & \text{if } t \geq t^* \ . \quad (24b) \end{cases}$$

[6]The main problem is that the lag a_3 differs for each person, i.e. is random. This means a_3 has a probability distribution which makes expected deaths $\dot{D}(t)$ really a moving average of past infections.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

We will apply this approach to Covid-19 deaths in the United States - a country where the epidemic consists of a myriad superimposed outbreaks with different sizes, timings, intervention strategies and death rates (which depend on age, race, and occupation among other things). Will such a patchwork be amenable to a simple model that assumes a reasonable degree of homogeneity?

4.2.2 United States: an ongoing epidemic with some NPIs

We use the doubling times of the total number of deaths, and total and daily numbers of deaths between March 4 and May 11 to model and project U.S. deaths. Panel A of Fig. 4 shows that doubling times were initially erratic but started to climb around day $t^* = 28$. To assess precisely the exponential growth of the numbers of deaths during the first 28 days (i.e. r_0 or equivalently $DT_0 = \ln(2)/r_0$) we used an ordinary least square approach to obtain the linear trend for $\ln(D(t))$. The resulting intercept $D(0)$ and DT_0 are given in the plot. We chose by trial and error values of HL equal to 11 and 14, which provide high and low trajectories of deaths beyond the 28th day. The eventual infected numbers D_∞, the peak times t_c and corresponding maximum incidences for the two variants are given in Fig. 4.

These projections are for illustrative purposes and are not based on carefully estimated parameters. Still, projected half-lives HL of 11 and 14 (penetration $P = 43$ and 55) that are significantly larger than $DT_0 = 3.08$ yield asymmetrical functions $\dot{D}(t)$ (Panel B) consistent with daily numbers of deaths that have decreased only slowly since mid-April. The logarithmic scale distorts the incidence curve in a way that makes the epidemic appear more protracted than it really is. Indeed, despite considerable uncertainties, these projections suggest that this first wave of Covid-19 in the U.S. could taper off in June or July.

Instead of having a low and a high variant fit the entire epidemic we could have a single optimal set of parameter values fit the data up to May 11, e.g. with $HL = 11$. As discussed earlier we could then "restart" the model on May 11, by extrapolating the straight line $LDT(t)$ of Eq. (13) with two different slopes $-s = \ln(2)/HL$. For example $HL_{low} = 10$ for a quicker conclusion and $HL_{hi} = 12$ for a more protracted affair.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

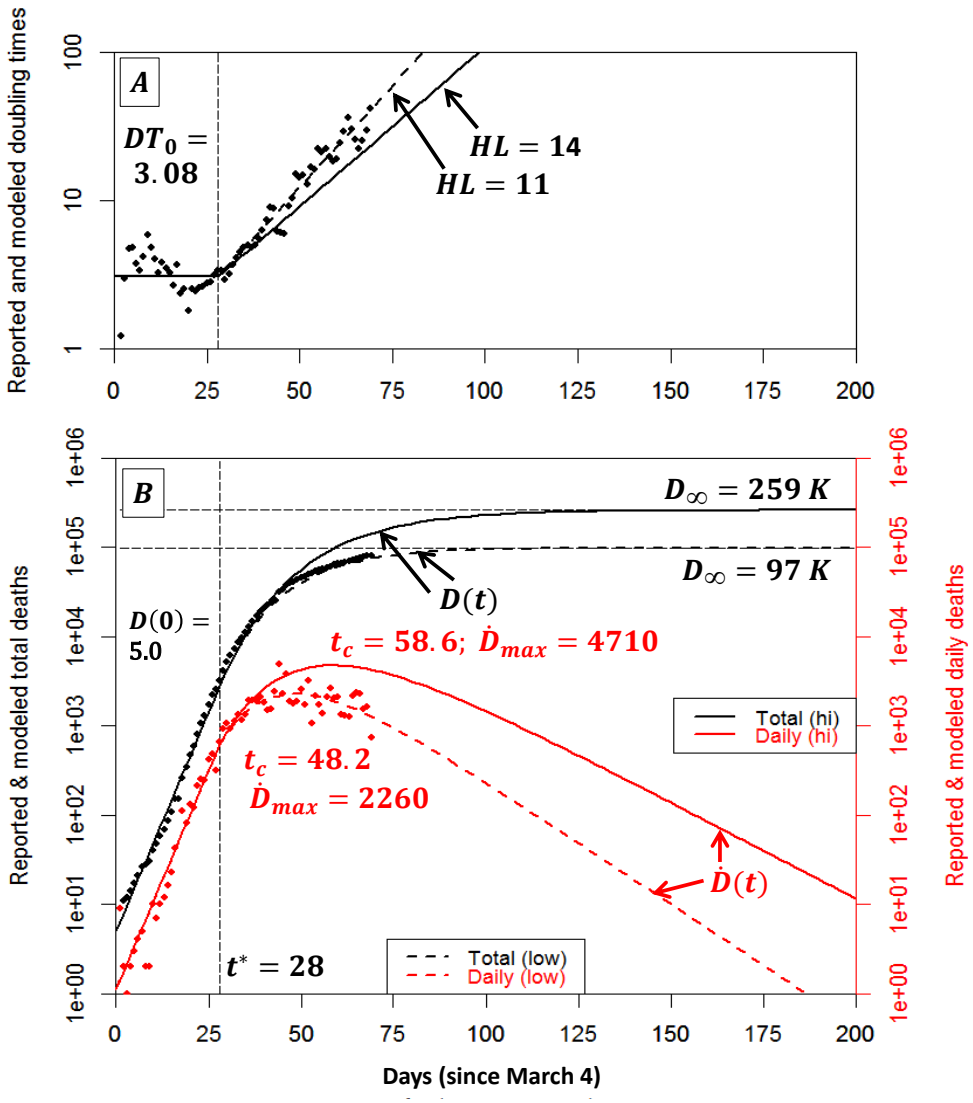

Figure 4: Panel A: Reported and low/high projected doubling times $DT(t)$ for U.S. deaths from March 4 to May 11, 2020 (68 days). Panel B: Total reported and low/high projected total deaths $D(t)$ in black; daily reported and low/high projected daily deaths $\dot{D}(t)$ in red (semi-logarithmic scales).

 doi:10.20944/preprints202006.0148.v1

5 Discussion

We had three goals in this paper: i) to review a few simple epidemic projection models; ii) to contribute an easily understood "toy model" based on first principles, namely that the growth rate of an infectious disease is initially constant and then decreases; and iii) to reflect on the challenges outlined in the introduction: why is it so difficult to predict the future course of an epidemic and can flawed models still be useful? Ours and other projections for Covid-19 deaths in the United States will illustrate the discussion.

5.1 The usefulness of epidemiological projections

The Exo-R model was fitted to a completed Chinese epidemic with infections conforming well to an early exponential stage followed by an exponentially decreasing growth rate. This approach needs to be tested with more countries or states - and rigorous statistical techniques will need to be applied in order to estimate the model's parameters.

Applying the model to the U.S. left us with projected ultimate numbers of deaths between 97,000 and 259,000. This broad range is dismaying but reflects the many unknowns concerning the growth rate $r(t)$ beyond May 11. We do not know how long it will take for potential contacts to either i) make themselves scarce through depletion/social distancing, or ii) become immunized following infection and recovery. Not knowing if/when intervention measures will kick in, how long they will last, and what effect they will have, further complicates the modeler's task.

The only consolation to our broad ranges is that with a simple model that folds both the depletion of susceptibles and the inception of intervention measures into a plausible pattern of decrease for the growth rate, we obtain a range of outcomes with peak times in the latter part of April that is in general agreement with other projections made at the time. The Institute for Health Metrics and Evaluation (IHME, University of Washington) projected in March a range from 38,000 to 162,000 U.S. deaths[7]. Our figures are even consistent with the White House's massaged range of 100,000 to 240,000 deaths - perhaps based on the IHME modelling (Wu et al., 2020; Wan et al, 2020).

[7]Just a few weeks later in the second half of April the number of deaths had already surpassed 40,000. IHME's low estimate may have assumed unrealistically stringent social distancing/quarantine measures similar to those applied in China. Regardless, the speed at which their low estimate was overtaken by reality reflects the speed at which the epidemic is unfolding in the United States. HIV/AIDS modelers in the 80s and 90s had the opposite problem: they had to wait years, if not decades, to see if their projections panned out.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

These models assume various mitigation efforts such as quarantines, social distancing, and lockdowns, which keep entire nations at home. Earlier in March Imperial College London projected up to 2.2 million U.S. deaths in the absence of mitigation. (see Ferguson et al (2020) for Imperial's early modeling efforts; Wu et al. (2020) and Wan et al (2020) provide news coverage of these events, which were unfolding too rapidly and recently to be covered in real time by the academic literature).

These projections of Covid-19 deaths in the United States have weaknesses and strengths. Their weakness is the wide range of values for the numbers of deaths, which result from the many uncertainties concerning the transmission dynamics. One must then honestly ask "At what point is a projected range so broad as to be useless?[8]" This question alone deserves an entire article.

The strength and usefulness of these models lie in the big picture that emerges from projections with and without mitigation: a maximum "do-nothing" 2.2 million deaths in the U.S. can be reduced by an order of magnitude (to low hundreds of thousands) with aggressive mitigation efforts. Crucially, we must hope that this a fairly "robust" conclusion - i.e. one that is not too sensitive to departures from the underlying assumptions and data. In fairness, the better efforts address this question, e.g. Imperial College provides reassurances on this issue. Still, it does not bear thinking of the disastrous implications of policies that rely on mathematical models but may overestimate the public health benefits of social distancing measures and lockdowns - interventions that are wreaking havoc on the world economy and the mental health of millions.

5.2 Complex vs simple models

Complex models (notably by IHME, Imperial College London and others) are ambitious and go beyond mere demographic projections by forecasting numbers of required hospital beds, ventilators etc. They are admirably detailed, but have many moving parts and can be structurally flawed.

In contrast, our Exo-r model has limited ambitions but captures the dynamics of an infectious disease fairly accurately and in closed form using elementary mathematics and a spreadsheet. We can now answer the question raised before fitting U.S. deaths to our Exo-r model: yes, a very simple model can plausibly capture *in the aggregate* a patchwork of complex local epidemics unfolding heterogeneously in time and space. Obviously it can also be used at the local and state levels.

[8]The U.S. Centers for Disease Control and Prevention tracks several short-term academic projections of U.S deaths - with a range of ranges that is truly dispiriting.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

The Exo-r approach has an advantage in its extreme simplicity: it has few moving parts that can go wrong. Indeed, it is incontrovertible that the growth rate of infections $\dot{I}/I$ is a (positive) quantity $r(t)$ that changes with time. The only moving part is the nature of this change. But much can be done with it. One can assess the impact of "stop-and-go" non-pharmaceutical interventions which are by definition exogeneous. The growth rate $r(t)$ might decreases in stages or even increases if social distancing measures are relaxed too soon - leading to the dreaded "second wave"[9].

5.3 Toward a taxonomy of Covid-19 epidemics

Covid-19 is the first pandemic in a globalized world. It has spread to many countries but at different rates and with different death tolls. By April China had completed its outbreak with fewer than 100,000 reported cases out of a total population of 1.4 billion. Several European countries have fared much worse, reaching 200,000 cases in May despite much smaller populations. The United States, with 1.3 M cases and no end in sight, is in a league of its own.

For years to come epidemiologists will be sifting through terabytes of data in order to understand and describe the variety of Covid-19 epidemics across the world. This is where a classification system (known as a taxonomy) could help sort out the different types of epidemics at the national or sub-national levels.

The three parameters of our proposed Exo-r can be used to describe and classify Covid-19 outbreaks in terms of both the numbers infected and deaths. The classification hinges on the three-dimensional vector $V = (t^*, DT_0, HL \text{ (or } P))$, combined with a common, fixed initial infected population $I(0) = 10$ (or 100)[10]. For example China, with $t^* = 10$ and $DT_0 = 2.2$, could be in an "average duration / high initial growth rate" category - and probably a low penetration (P) category. With $t^* = 0$, Spain and Italy (Figure 1) might be in a "zero duration first stage" category, etc.

If three parameters are not enough to describe all Covid-19 epidemics, one could, as discussed earlier, "restart" the model by adding two parameters: a time $t^{*'} > t^*$ at which the growth rate's half-life changes, and HL', the new half-life thereafter.

[9] These effects can be modeled through the *reproductive number* $R(t)$, the average number of new cases generated by each infected person; $R(t)$ is a proxy for social distancing and is in a simple relationship with $r(t)$.

[10] This normalization is not a restriction: regardless of their ultimate size all epidemics reach 10 (or 100) infected individual at some early point taken as time $t = 0$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 June 2020 doi:10.20944/preprints202006.0148.v1

6 Conclusion

We close on a puzzling but hopeful note. We first recall Dr Fauci's April 2 downbeat assessment about models not telling you anything. This skepticism (not to say frustration) was understandable in view of an initial onslaught of extremely diverse projections. Six days later Dr. Fauci is quoted as saying that "Models are good, they help us to make projections. But as you get data in, you modify your model" (Wan and Johnson, 2020). These words of wisdom were in sharp contrast with his earlier skeptical remarks. The apparent contradiction a few days apart could result from out-of-context quotes, inaccurate reporting, or perhaps a renewed faith in the virtues of modeling.

It is rewarding to hear from top authorities that models are "good" and help make projections. Today, the big challenge for modelers is to quantify the impact of unpredictable and at times politically motivated decisions concerning stay-at-home policies and the reopening of the economy. This dynamic creates an unstable feedback loop that causes models to "gyrate fairly significantly from week to week"(Aizenman and McMinn, 2020). This has led to recent attempts at combining the results of different models - with the modest goal of projecting deaths only a few weeks ahead. This is a Covid-era variant of the "Delphi Method" invented in the 1950s which aggregates "expert opinions". In an ideal world public health authorities and modelers would work together to produce useful projections based on reliable data on testing, prevalence rates, etc. These efforts would inspire governments to adopt wise and sensible policies that strike the right balance between saving lives and preventing an economic meltdown.

A Data source and Supplementary Material

Data on infections and deaths come from `https://ourworldindata.org`. The Supplementary Material (`https://bit.ly/3bAYFZr`) has a link to a google spreadsheet of the model that can be downloaded and modified. An Excel spreadsheet is also available from the author.

B Acknowledgments

The author wishes to thank an anonymous referee whose suggestions significantly improved the paper.

References

Aizenman, N., and MacMinn, S. (2020) How To Make Sense of All The COVID-19 Projections? A New Model Combines Them, 13 May, 2020, *National Public Radio*. Retrieved May 14, 2020.

Ferguson, NM, Laydon, D., Nedjati-Gilani, G. et al (2020) Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, Imperial College COVID-19 Reponse Team, https://doi.org/10.25561/77482

Rucker, P and Wan, W (2020) Trump projects up to 240,000 coronavirus deaths in U.S., even with mitigation efforts, 31 March, *The Washington Post.*

Wan, W, Dawsey, J, Parker, A, and Achenbach, J (2020) Experts and Trump's advisers doubt White House's 240,000 coronavirus deaths estimate, 2 April, *The Washington Post.*

Wan, W, and Johnson, CY (2020) America's most influential coronavirus model just revised its estimates downward. But not every model agrees. 8 April, *The Washington Post.*

Wu, K., Darcet, D. , Wang, Q., and Sornette, D. (2020) Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and the rest of the world. MedRχiv, doi: https://doi.org/10.1101/2020.03.11.2003436