Are Models Useful? Reflections on Simple Epidemic Projection Models and the Covid-19 Pandemic

MARC ARTZROUNI

Prediction is very difficult, especially if it’s about the future” is a quotation one might expect from Groucho Marx or Yogi Berra. Yet it is attributed to the twentieth-century Nobel laureate Danish physicist Niels Bohr. The quotation may be apocryphal, but it 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 USA’s National Institute of Allergy and Infectious Diseases and a leading 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” [4]. This quotation 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 that 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?

Basic Epidemic Projection Models

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 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, \( r dt \) new infections.\(^1\) The total (cumulative) number of infected individuals \( I(t) \) up to day \( t \) then satisfies the differential equation

\[
\frac{dI}{dt} \frac{1}{I} = r - 1 = r,
\]

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

\[
I(t) = I(0) \exp(rt).
\]

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.” 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 that a dollar invested at a constant rate of interest generates every year an ever increasing number of new dollars as interest is compounded. As long as you keep your money in your account and the rate of interest remains positive, 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 entire population, or of a subpopulation susceptible to the disease, e.g., those 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 of infecteds cannot exceed the carrying capacity \( K \).

**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 (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

\[
\frac{\dot{I}}{I} = r \left( 1 - \frac{I}{K} \right),
\]

\( I(r) \) has been “endogenized,” i.e., it is driven by the number \( K \) of infecteds. The fact that \( K \) is a cumulative number of infecteds means that its growth rate on the right-hand side of (3) must remain positive, which happens if and only if \( K(t) \) is less than \( K \) for all \( t \). This suggests that we are on the right track. (We assume that the initial number of infecteds \( I(0) \) 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 (3). New infections result from contacts between the \( R(t) \) infecteds (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.
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 is 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 \( R(t) \) Johns at time \( t \), they generate a total of

\[
I(t + dt) - I(t) = I(t)pc \frac{S(t)}{K} dt = I(t)pc \left( 1 - \frac{I(t)}{K} \right) dt
\]

new infections during the time interval \( (t, t + dt) \) (think of \( dt \) as being equal to 1). As \( dt \to 0 \) this is (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 those components.

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

\[
I(t) = \frac{K}{1 + A \exp(-rt)}
\]

with

\[
A = \frac{K - I(0)}{I(0)}.
\]

\( ^1 \)A perceptive nonspecialist may wonder how one generates fractional 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 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 human beings do not come in fractions. 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.
Equation (5) is known as the logistic function. It has had a long and illustrious career, including recent applications, extensions, and generalizations used to fit and project the Covid-19 epidemic in China and elsewhere [6]. As one might expect, this function is S-shaped and tends to $K$ as $t \to \infty$. On differentiating (5), one finds that the daily number of new infections is

$$i(t) = \frac{AKr \exp(-rt)}{(1 + A \exp(-rt))^2}.$$  

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

These results show that there are two ways of flattening the $i(t)$ curve. We can reduce $I_{\text{max}}$ 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 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 that correct the aforementioned shortcomings.

**SIR and SEIR Models: Three and Four Compartments**

An SIR (susceptible, infected, recovered) 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 that are based on the mass action principle and have variants of (3) at their core.

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

**The Phenomenological Exo-r Model**

Think now of the Covid-19 epidemic in China. The first wave began in January 2020 and had run its course by the end of March. Even if the official total count of around 84,000 cases is underreported by a factor of 10, this means that fewer 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 that reduced the rate of infection (social distancing, lockdowns, quarantines, etc.).

We will capture this effect with a growth rate $I(t)$ that is no longer an endogenously decreasing function of time (right-hand side of (3)) but simply a function $r(t)$ of time:

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

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. The only constraint is that $r(t)$ must be positive, since $R(t)$ is a cumulative number of infections and is therefore necessarily increasing. The solution $R(t)$ of (8) is

$$I(t) = I(0) \exp\left(\int_0^t r(s) \, ds\right).$$

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

$$I(t) = I(0) r(t) \exp\left(\int_0^t r(s) ds\right).$$

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

**Specification of the 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. They all show a similar pattern. There is an early period of erratic logarithms of the doubling time, log($DT$), due to small numbers and reporting problems/delays. We will consider that log($DT$) is constant during this early stage

---

2Such models can justifiably be criticized for ignoring demographic mechanisms that 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.

3These Creative Commons data and visualizations are freely available and usable at https://ourworldindata.org/grapher/doubling-time-of-covid-cases?yScale=log; accessed May 9, 2020.
of the epidemic. This means an exponential increase, which is common when a disease initially spreads unhindered. Figure 1 shows that at some point, the logarithm of the doubling time increases in a roughly linear manner, which means that the growth rate decreases exponentially, reflecting a diminishing pool of susceptibles and/or the onset of interventions.

Given that the common and natural logarithms are the same up to a constant multiple, 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 $s$:

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

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 $t^*; HL = -\ln 2/s$. Equation (11) can be paraphrased by noting that the doubling time $DT(t) = \ln 2/r(t)$ is then

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

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

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

which increases linearly with the positive slope $\ln 2/HL = -s$.

**Results**

Routine integrations left as an exercise show that with $r(t)$ of (11), the number of infections $I(t)$ of (9) is

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

Because $s < 0$, the ultimate number of infecteds is equal to

$$ I_f = \frac{I(0) r_0}{s}.$$

Bear in mind (11b), we note that for $t > t^*$ we have

$$ I(t)/I_f = \exp(\frac{r_0}{s}(t - t^*)) = \exp(-\frac{HL}{DT(t)}). \quad (16) $$

This equation means that the fractional number of total infections at time $t > t^*$ depends only on the current growth rate $r(t)$ and the rate $s < 0$ at which $r(t)$ decays.

The derivative of $I(t)$ in (14) is

$$ \dot{I}(t) = \begin{cases} I(0)r_0 \exp(r_0 t) & \text{if } t < t^*, \\ I(t)r_0 \exp(r_0(t - t^*)) & \text{if } t \geq t^*. \end{cases} \quad (17a) $$

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 $\dot{I}(t)$ that grows at the positive exponential rate $r_0$ early in the epidemic and at the negative exponential rate $s$ for $t \to \infty$.

With exponential growth until $t^*$, the density of new cases $\dot{I}(t)$ can peak only after $t^*$. If we differentiate (17b), we find that there are two scenarios, depending on the values of the growth and decay rates $r_0$ and $s$.

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

$$ \dot{I}_{\text{max}} = \frac{I(0)r_0}{s} \exp(r_0 t^*). \quad (18) $$

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.

Scenario B: $-s_0 \leq s < 0$, i.e., $r(t)$ decays more slowly than the initial infected population increases. Then new infections $\dot{I}(t)$ keep growing after the intervention and reach a maximum at the critical value

$$t_c \equiv \frac{\ln(-s/s_0)}{s} + \tau^* > \tau^*.$$  \hfill (19)

In this case, a growth rate that decreases more slowly creates an inertia in the infections, which continue to increase until a time $t_c$ before turning around. Not surprisingly, the closer $s$ is to 0, the more protracted the epidemic, and the larger $t_c$ is. The corresponding maximum daily incidence is

$$I_{\text{max}} = \dot{I}(t_c) = -I(0) \frac{s}{e} \exp \left(\frac{s_0}{s} - 1\right) = \frac{-s}{e} I_\infty.$$  \hfill (20)

Insights

Ultimate Infected Population

If we recall that $-s = \ln 2/HL$, where $HL$ is the half-life of the growth rate $r(t)$, then we can rewrite (20) as

$$I_\infty = \frac{e}{\ln 2/HL} I_{\text{max}} = \frac{3.922HL}{s} \times I_{\text{max}}.$$  \hfill (21)

This result highlights the role of $HL$ and provides a simple relationship between $I_\infty$ and $I_{\text{max}}$. Indeed, 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.

We illustrate the use of (21) with an attempted back-of-the-envelope estimate of the ultimate number of cases $I_\infty$ in the United States. Indeed, if we can estimate $HL$ and are sufficiently advanced in an epidemic to observe $I_{\text{max}}$, then (21) 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 $HL$ of 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 25000 to 30000 range, with a slight downward trend.\(^4\) Simple epidemic curves usually peak and go back down fairly quickly. A long plateau means that something more complicated is going on. In the case of Covid-19 in the United States, 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 30000 for $I_\text{max}$, we find that (21) yields $I_{\text{max}} = P \times 30000 = 1.4 \times 10^6$. By May 10, the number of confirmed cases had reached 1.3 million. Our 1.4 million is the right order of magnitude but obviously falls short.

The reason for the low estimate of 1.4 million can be found in the United States data of Figure 1. As a first approximation, the trend for $\log(DT)$ appears to be roughly linear beyond time $t^* \approx$ March 21. However, a closer look at Figure 1 suggests that beyond April 15, $\log(DT)$ 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. This change in regime can be captured simply by “restarting”/updating the model on April 15.

Exo-r Model Update

In the present case, we can reset time $t = 0$ to April 15. For the initial number of infecteds $R(0)$ and corresponding growth rate $r(t)$, we take the values reported on April 15. In order to reflect an epidemic that becomes more protracted, we can assume no exponential stage ($\tau^* = 0$) and a growth rate that decreases from $r(t)$ with a half-life $HL$ larger than the early value of $HL$. ($HL$ could be smaller than $HL$ if intervention measures were increased or a vaccine removed large numbers of potential susceptibles). These are just a few examples of how the model can adapt to changing circumstances with updated initial conditions and half-lives.

Sensitivity Analysis

Returning to the original formulation, the time $t^*$ be thought of as the timing of an intervention and the half-life $HL$ as its intensity. With closed-form expressions for epidemiological parameters, one can then explore analytically the effect of $t^*$ and $HL$ on three measures of “outcome severity”: the time of maximum daily infections $t_x$, the maximum number of infections $I_{\text{max}}$, and the cumulative total $I_\infty$. We illustrate such a sensitivity analysis with baseline values $t^* = 10$ and $HL = 3.3$ obtained for China below (with $DT_0 = 2.2$). Table 1 shows the sensitivities of the three severity outcomes to variations of $\pm 1$ day for $t^*$ and $HL$.

In view of (19), it is not surprising that the sensitivity $\Delta t_x$ is equal to 1 regardless of the value of $HL$. Of greater interest is the impact of a one-day decrease in $t^*$ or $HL$ on $I_\infty$ and $I_{\text{max}}$. Table 1 shows that the peak number of infections $I_{\text{max}} = 4889$ drops to 3568 and 4023 for $\Delta t^* = -1$ and $\Delta HL = -1$ respectively. This means that interventions starting one day earlier have a greater impact on the relatively short-term $I_{\text{max}}$ metric than a one-day drop in $HL$.

The total number of infections $I_\infty = 82440$ drops to 60160 and 52237 for $\Delta t^* = -1$ and $\Delta HL = -1$, respectively. This time, a one-day reduction in $HL$ has the greatest impact on the total number, no doubt because the shorter $HL$ impacts total numbers over a long time. This simple example shows that the best intervention may depend on the goal. If we need to choose between a one-day reduction of $t^*$ and $HL$, then we choose $t^*$ if the main goal is to protect the healthcare system by “flattening the curve.” If

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

© 2020 Springer Science+Business Media, LLC, part of Springer Nature, Volume 42, Number 3, 2020
Table 1. Sensitivity analysis of three outcome severities (the time of maximum daily infections \(t_0\), the maximum number of infections \(i_{\text{max}}\), and the cumulative total \(I(t)\)) to changes in \(HL\) and \(r^*\). Results in the central shaded area are the (baseline) severities for the estimated values \(HL = 4.3\) and \(r^* = 10\) obtained for China (\(DT_0\) of 2.2). The sensitivity analysis is for \(\pm 1\) day variations for both \(HL\) and \(r^*\).

| \(HL\) | 9   | 10  | 11  |
|-------|-----|-----|-----|
| \(t_0\): | 10.9 | 11.9 | 12.9 |
| \(i_{\text{max}}\): | 2951 | 4043 | 5541 |
| \(I_\infty\): | 38185 | 52237 | 71707 |
| \(t_0\): | 13.2 | 14.2 | 15.2 |
| \(i_{\text{max}}\): | 3568 | 4889 | 6699 |
| \(I_\infty\): | 60160 | 82440 | 112971 |
| \(t_0\): | 15.7 | 16.7 | 17.7 |
| \(i_{\text{max}}\): | 4560 | 6249 | 8563 |
| \(I_\infty\): | 94779 | 129880 | 177982 |

the goal is to minimize the total number of infecteds, then a shorter \(HL\) has the greater impact.

Many other types of sensitivity analyses can shed light on optimal intervention strategies. For example, one can investigate what combinations of \((r^*, HL)\) values bring about a given reduction in either \(i_{\text{max}}\) or \(I_\infty\).

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.

Applications

China: A Completed Epidemic with Stringent “Nonpharmaceutical Interventions”

We plotted in panel A of Figure 2 the doubling times of numbers of infected from January 22 to May 11, 2020, for mainland China (black dots). This panel is analogous to the one in Figure 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. Fitting the data by hand yielded parameter values \(r^*, DT_0, \dot{I}(0),\) and \(HL = \ln 2/r_0\) satisfying these conditions (they are given in Figure 2). 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 \(r^* = 10\) 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 A shows that a good fit for total numbers \(I(t)\) comes at the cost of a modeled \(DT\) that is in agreement with observed doubling times only between the 10th and approximately 55th days. Panel B shows that this was a month and a half during which much of the spread took place. After the 55th day, observed doubling times drift downward, 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 has essentially ended, as can be seen from the low daily numbers in panel B at that time.

Exo-r Modeling of Deaths

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 a lagged relationship between infections and deaths could justify using the Exo-r approach to model deaths as well as infections.

To explore this possibility, we let \(D(t)\) and \(\dot{D}(t)\) represent the total and daily numbers of deaths on day \(t\). We assume that the true number of infections is \(\alpha_1\) times the reported number modeled as \(R(t)\) and \(\dot{R}(t)\) above; the probability of death due to Covid-19 is \(\alpha_2\); and the time between infection and death (when it occurs) is a constant \(\alpha_3\). Then the number of deaths \(\dot{D}(t)\) on day \(t\) is the “real”

![Figure 2](image-url)
number \(a_1I(t-a_3)\) of infections \(a_3\) days earlier multiplied by the probability of death \(a_2\):

\[
\dot{D}(t) = a_1a_2I(t-a_3).
\] (22)

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

We begin by replacing all the \(I\)’s of the section titled “Results” with \(D\)’s. The expressions are recalled here for convenience. Total deaths are modeled as

\[
D(t) = \begin{cases} 
D(0)\exp(rs) & \text{if } t < t^*, \\
D(0)\exp\left(r_0\exp(s(t-t^*)) - 1 + st^*\right) & \text{if } t \geq t^*, 
\end{cases}
\] (23a)

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

\[
\dot{D}(t) = \begin{cases} 
D(0)\exp(rs) & \text{if } t < t^*, \\
D(0)\exp\left(r_0\exp(s(t-t^*)) - 1 + st^*\right) & \text{if } t \geq t^*, 
\end{cases}
\] (23b)

We will apply this approach to Covid-19 deaths in the United States, a country where the epidemic consists of myriad superimposed outbreaks of 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?

**United States: An Ongoing Epidemic with Some NPIs**

We examined the doubling times of the total number of deaths and total and daily numbers of deaths between March 4 and May 11 in the United States. Panel A of Figure 3 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 squares approach to obtain the linear trend for \(\ln(D(t))\) during these first 28 days. The resulting intercept \(D(0)\) and doubling time \(DT_0\) are given in Figure 3.

We chose values of \(HL\) equal to 11 and 14, which provide plausible high and low trajectories of deaths beyond the 28th day. The goal is to assess, however crudely, the level of uncertainty concerning the final number of deaths when the daily incidence has just passed its peak (panel B of Figure 3). The final numbers are reported in panel A.

With a relatively short half-life of 11 days, the final death toll is 97,000. With a half-life of 14 days (a 27% increase), the toll jumps to 259,000, i.e., a 367% increase. This example shows that considerable uncertainties remain even in projecting a mature epidemic that has peaked. This is because small changes to growth rates applied over several months make a big difference to the ultimate outcome, in the same way that a retirement fund at the end of a working life will be extremely sensitive to the (average) rate of interest.

**Discussion**

**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 to estimate the model’s parameters.

Applying the model to the United States 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.

---

\(^5\)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 that makes expected deaths \(D(t)\) really a moving average of past infections.
May 11. We do not know how long it will take for potential contacts to either make themselves scarce through depletion/social distancing, or become immunized following infection and recovery. Not knowing whether or 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 infection of intervention measures into a plausible pattern of decrease in 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) at the University of Washington projected in March a range from 38,000 to 162,000 U.S. deaths. Our figures are even consistent with the White House’s massaged range of 100,000 to 240,000 deaths, perhaps based on the IHME modeling [4, 6].

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 [2] for Imperial’s early modeling efforts, [6] and [4] 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?” 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 United States can be reduced by an order of magnitude (to low hundreds of thousands) with aggressive mitigation efforts. Crucially, we must hope that this is 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; for example, Imperial College provides reassurances on this issue. Still, it does not bear thinking about 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.

### Complex Versus 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 ambition 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.

In its extreme simplicity, the Exo-r approach has several advantages. First, it has few moving parts that can go wrong. Indeed, it is incontrovertible that the growth rate of infections $I/I$ is a (positive) quantity $r(t)$ that changes with time. The only moving part is the nature of this change. Second, the model provides closed-form expressions for important epidemiological parameters. Sensitivity analyses can then shed light on the timing and effects of different intervention strategies. Third, the model can be updated to explore in a simple and transparent manner the consequences of changing epidemiological conditions, such as an early reopening of the economy or the “second wave” feared by many. Being able to update projections easily is particularly useful at the local level, e.g., if one wishes to capture the dynamics of the 50 different epidemics across 50 U.S. states, each with its own timing, social distancing rules, and “exit strategy.”

### Final Remarks

We come full circle by first recalling 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” [5]. These words of wisdom were in sharp contrast to his earlier skeptical remarks. This apparent contradiction in utterances a few days apart could result from out-of-context quotation, 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 significantly from week to week” [1]. In an ideal world, public health authorities

---

6Just 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 1980s and 1990s had the opposite problem: they had to wait years, if not decades, to see whether their projections panned out.

7The 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. See https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html.
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.

Addendum
As this article goes to press in July 2020, many countries are seeing a resurgence of Covid-19 cases, even though they had successfully brought early outbreaks under control. In view of this situation, the suggestion made earlier that the Exo-r approach be used or extended to update projections takes on new meaning. The simplest update can be obtained with the Excel spreadsheet provided in the Supplementary Material (see below). It allows users to explore the implications of a growth rate that continues to decrease exponentially, but at a different rate (typically with a longer half-life). Other variants are possible; the effect of reopened economies and relaxed social distancing measures can be captured through a growth rate that ceases to decrease, thus bringing about a period of exponential growth, a period that can be followed by a renewed decrease in the growth rate if control measures are tightened. Projecting these cycles hinges on our understanding of the (empirical) relationships between the various control measures and the growth rates.

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 an interactive Google spreadsheet that can be used to explore and update projections of U.S. deaths. The user can enter his/her own daily data on infections or deaths, and seek parameter values that provide the best fit. There is also a link to an Excel spreadsheet used to project a “mature” epidemic, i.e., one that has passed its exponential phase, with the Exo-r model assumption that growth rates (of either cases or deaths) are decreasing exponentially.

ACKNOWLEDGMENTS
The author wishes to thank an anonymous referee and the managing editor, David Kramer, for useful comments that significantly improved the paper.

Marc Artzrouni
Department of Mathematics (CNRS-UMR 5142)
University of Pau and Pays de l’Adour
64000 Pau
France
e-mail: marc.artzrouni@univ-pau.fr

REFERENCES
[1] N. Aizenman and S. MacMinn. How to make sense of all the COVID-19 projections? A new model combines them. May 13, 2020. Available online at the National Public Radio website. Retrieved May 14, 2020.
[2] N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, Imperial College COVID-19 Response Team. https://doi.org/10.25561/77482, 2020.
[3] P. Rucker and W. Wan. Trump projects up to 240,000 coronavirus deaths in U.S., even with mitigation efforts. Washington Post, March 31, 2020.
[4] W. Wan, J. Dawsey, A. Parker, and J. Achenbach. Experts and Trump’s advisers doubt White House’s 240,000 coronavirus deaths estimate. Washington Post, April 2, 2020.
[5] W. Wan and C. Y. Johnson. America’s most influential coronavirus model just revised its estimates downward. But not every model agrees. Washington Post, April 8, 2020.
[6] K. Wu, D. Darcet, O. Wang, and D. Sornette, D. Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and the rest of the world. MedRxiv, https://doi.org/10.1101/2020.03.11.2003436, 2020.

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