



# Effects of Latency on Estimates of the COVID-19 Replication Number

Lorenzo Sadun<sup>1</sup> 

Received: 12 June 2020 / Accepted: 6 August 2020 / Published online: 20 August 2020  
© Society for Mathematical Biology 2020

## Abstract

There is continued uncertainty in how long it takes a person infected by the COVID-19 virus to become infectious. In this paper, we quantify how this uncertainty affects estimates of the basic replication number  $R_0$ , and thus estimates of the fraction of the population that would become infected in the absence of effective interventions. The analysis is general, and applies to all SEIR-based models, not only those associated with COVID-19. We find that when modeling a rapidly spreading epidemic, seemingly minor differences in how latency is treated can lead to vastly different estimates of  $R_0$ . We also derive a simple formula relating the replication number to the fraction of the population that is eventually infected. This formula is robust and applies to all compartmental models whose parameters do not depend on time.

**Keywords** Latency · SEIR · Coronavirus · Replication number · Estimation

**Mathematics Subject Classification** 92-10 · 92B99

## 1 Introduction

In any epidemic, there are two basic questions that need to be answered:

- How much does our behavior need to change to contain and eventually suppress the outbreak?
- How many people will eventually get infected if our behavior doesn't change?

Both questions boil down to estimating the basic reproduction number  $R_0$ , which describes the average number of new individuals that each infected individual goes on to infect, assuming no interventions and no depletion of the population of susceptible individuals.

---

✉ Lorenzo Sadun  
sadun@math.utexas.edu

<sup>1</sup> Department of Mathematics, University of Texas, Austin, USA

The connection of  $R_0$  to the first question is clear. If  $R_0 = 1.2$ , then blocking 20% of would-be transmissions is enough to halt and reverse the outbreak; mild interventions, such as promoting frequent hand-washing and the wearing of masks, might be enough. If  $R_0 = 5$ , then blocking 80% or more of the would-be transmissions is needed, so much more drastic actions, such as full scale lock-downs, are called for. At of July 2020, strong measures have largely halted the spread of COVID-19 in East Asia and Europe, and policy-makers there are trying to determine the extent to which these measures can be safely relaxed. In much of South Asia, Africa, and the Americas, however, where strict measures either were not imposed or were abandoned prematurely, the disease is running unchecked and policy-makers must come to grips with what measures need to be imposed or re-imposed. Throughout the world, effective public policy depends on understanding  $R_0$ .

The connection of  $R_0$  to the second question is more subtle. We will see that, in a wide variety of models, the fraction  $x$  of the population that is eventually infected is purely a function of the replication number  $R$ .  $R$ , as opposed to  $R_0$ , describes the average number of new individuals that each infected individual goes on to infect, *taking a constant level of intervention into account* but assuming no depletion of the population of susceptible individuals. Specifically, we will show that

$$1 - x = e^{-Rx}. \quad (1)$$

For instance, if  $R_0 = 5$  and our interventions only serve to bring  $R$  down to 2, then the fraction that eventually gets infected is the nonzero solution to  $1 - x = e^{-2x}$ , or about 0.797. That is, in this scenario almost 80% of the population would eventually get infected.

To the best of our knowledge, Eq. (1) has not previously appeared in the literature, and in discussions with epidemiologists, none have been aware of such a simple relationship. However, the reasoning behind it is very similar to arguments that have previously appeared in compartmental models of population dynamics, such as Gyllenberg and Hanski (1997) and Hastings (2003), particularly the calculations that appear at the very end of Hastings (2003).

Equation (1) assumes that  $R$  is constant. In reality,  $R$  varies with time as interventions are applied and relaxed, perhaps multiple times, and as the mood of the population oscillates between complacency and fear. Long-term projections in those circumstances are much more difficult, and are not attempted in this paper.

Since knowing the value of  $R_0$  is so important, it is essential to understand the accuracy of our estimates. Unfortunately, published estimates of  $R_0$  for COVID-19 have varied tremendously, from 1.3 (Du et al. 2020) to more than 6 (Tang et al. 2020), often with non-overlapping error bars. The most widely cited estimate of  $R_0 = 2.2$  is that of Li et al (2020), and (as of July 2020) much of the modeling of the future course of the pandemic has used Li's value or lower.

The basic technique for estimating  $R_0$  involves tracking the increase in cases in the early stages of an outbreak, before interventions are applied, and then computing how much transmission was needed to explain the observed growth. This requires model-building, so it's important to understand the extent to which the answer depends on the details of the model.

An important feature of COVID-19 is the large extent to which the disease is carried by individuals who are not yet showing symptoms. While there is general agreement that some pre-symptomatic individuals can still infect others (see, e.g., Mandavilli (2020) for early reporting on this fact), there is as yet no agreement on the length of the latency period during which an infected individual is not yet infectious. Some early models treated infected individuals as immediately infectious (Du et al. 2020; Pasco et al. 2020), while others assumed that they only become infectious toward the end of the incubation period (Ferguson et al. 2020).

In this paper, we determine how the estimated value of  $R_0$  depends on the length of the latency period for three versions of the popular SEIR (Susceptible, Exposed, Infectious, Removed) model. A detailed explanation of the results will have to wait until we describe these versions, but we find, as illustrated in Fig. 2, that all three versions are very sensitive to the average length of the latency period. Even for a fixed value of this parameter, the different versions, all of which are more or less equally plausible scientifically, give very different answers. More complicated (and realistic) models are even more parameter-dependent. The upshot is that essentially all published estimates of  $R_0$  for COVID-19 should be viewed with skepticism. Until we understand latency for COVID-19 much better than we currently do, we won't understand  $R_0$ .

Variation in early data for the spread of COVID-19 leads to additional uncertainty in estimates of  $R_0$ . Initial estimates of  $R_0$  were based on the early stages of the epidemic in Wuhan, China, where Li et al. (2020) estimated a doubling time of 7.4 days. In places like Europe and the USA, where efforts had already been made to contain the spread, the actual reproduction number  $R$  should have been less than  $R_0$ . However, outbreaks in many places have grown much faster than in Wuhan. For instance, in the USA the number of confirmed COVID-19 cases grew 100-fold, from 541 to 54,856, from March 8 to March 24, 2020, corresponding to a doubling time of only 2.4 days (<https://www.worldometers.info/coronavirus>, <https://interaktiv.morgenpost.de/corona-virus-karte-infektionen-deutschland-weltweit/>). Over the same time period, the number of confirmed cases in the entire world, excluding China, grew from 29,256 to 341,356, which indicates a doubling time of 4.5 days, slower than the USA but faster than Wuhan.<sup>1</sup>

A further complication for COVID-19 is the (still unknown) extent to which the disease is spread by individuals who *never* develop symptoms; these asymptomatic carriers are very difficult to identify and isolate. The simplest way to model such individuals is to imagine that there is a long tail to the distribution of the length of the infectious period. This is the motivation behind one of the three variants of the SEIR model developed below. A more sophisticated approach, taken by most COVID-19 modelers, is to sort infected individuals into non-infectious, asymptomatic, pre-symptomatic, and symptomatic compartments and treat each compartment differently. See Pasco et al. (2020) for an example. However, this has the problem of involving

---

<sup>1</sup> These numbers are for confirmed cases. Due to severe testing inefficiencies, the actual number of infections on March 8, and likewise the actual number of infections on March 24, were probably an order of magnitude higher. However, even if testing efficiency doubled in that time and the true number of infections in the USA only grew 50-fold instead of 100-fold, the resulting doubling time of 2.8 days would still be far smaller than in Wuhan.

even more parameters than before, making the estimated value of  $R_0$  be even more dependent on the details of the model. With enough parameters, almost anything can happen. As John von Neumann said, *With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.* (Dyson 2004)

The outline of the paper is as follows. In Sect. 2 we review stage models for epidemics and develop three versions of the SEIR model. For each version we present, without proof, the formula for the estimated value of  $R_0$  in terms of the growth rate of the epidemic and the parameters of the model. This is illustrated by comparing, in Fig. 2, the different estimates for  $R_0$  for COVID-19 in the USA in March 2020. We also illustrate the consequences of Eq. (1) and briefly discuss the phenomenon of herd immunity. In Sect. 3 we derive the formulas for  $R$  presented in Sect. 2 and we prove that Eq. (1) applies, not only to the three variants already discussed, but to all compartmental models meeting some mild assumptions. Finally, in Sect. 4 we discuss the real-world implications of these results.

## 2 Stage Models

### 2.1 The SEIR Model and its Variants

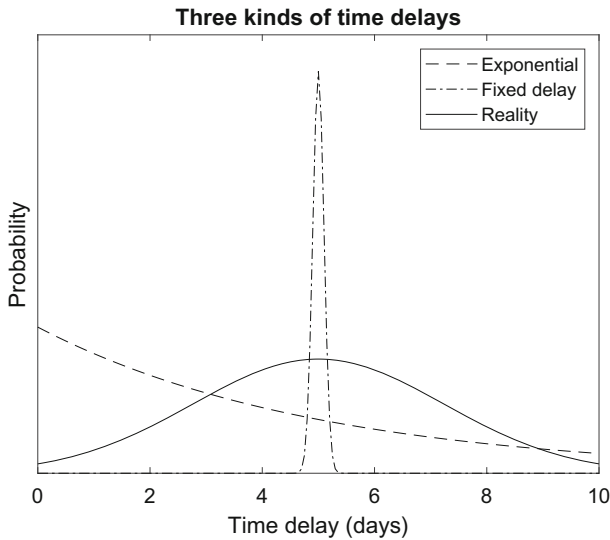
Most models of disease spread are variants of the SEIR model, which is itself an extension of the popular SIR model ([https://en.wikipedia.org/wiki/Compartmental\\_models\\_in\\_epidemiology](https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology)). In the SEIR model, people are classified as “Susceptible”, “Exposed”, “Infectious”, and “Removed” (or “Recovered”). Susceptible individuals become Exposed from contact with existing Infectious individuals, Exposed individuals become Infectious after a latency period, and Infectious individuals eventually recover, die, are quarantined, or are otherwise Removed from circulation. We let  $S(t)$ ,  $E(t)$ ,  $I(t)$  and  $R(t)$  denote the number of Susceptible, Exposed, Infectious and Removed individuals at time  $t$  and write equations to govern the evolution of these quantities.

The basic SEIR model has three continuous parameters.

1. The mean time  $t_1$  spent by patients in the uninfected “exposed” state. We call this time the latency.
2. The mean time  $t_2$  spent by patients in the infectious state.
3. The reproduction number  $R$ , which is the mean number of people that each infected person infects in turn.

The sum  $t_{\text{tot}} = t_1 + t_2$  is the total average time from exposure to the appearance of symptoms severe enough to cause an individual to isolate. We call this the extended incubation period.  $t_{\text{tot}}$  is relatively easy to measure, and for symptomatic COVID-19 patients it is believed to be about one week (Backer et al. 2020). The individual numbers  $t_1$  and  $t_2$  are much harder to determine. In this paper, we will compare different values of the latency  $t_1$ , while holding the sum  $t_{\text{tot}}$  fixed.

In addition to these three parameters, we consider several ways to treat the distribution of incubation times for individuals. The most common choice is to assume that the time each individual spends in the exposed state and the infectious state are inde-

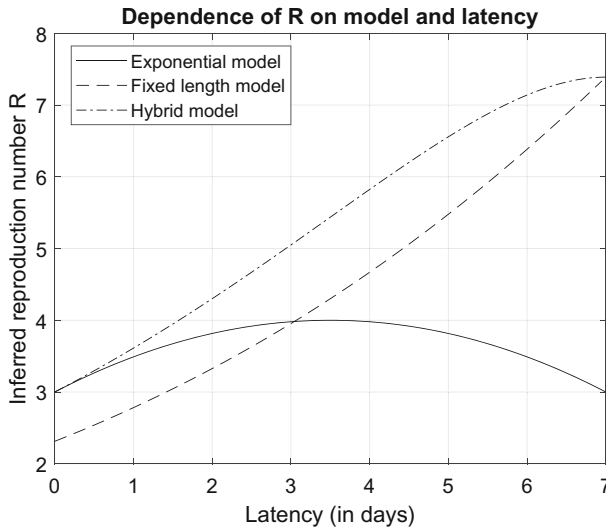


**Fig. 1** The actual distribution of latency (or infectiousness) is not well-approximated by either an exponential distribution or a fixed value

pendent exponential random variables with means  $t_1$  and  $t_2$ , respectively. This choice leads to a simple system of differential equations. Alternatively, we can assume that the time spent in the exposed state is always exactly  $t_1$ , and the time spent in the infectious state is exactly  $t_2$ . A third approach is a hybrid, keeping the time in the exposed state constant but making the time in the infectious state follow an exponential distribution.

Of course, none of the three methods captures the actual dynamics. In reality, the length of the incubation period is roughly bell-shaped, as is the length of the infectious period. See Fig. 1. Approximating this bell with an exponential distribution introduces a substantial chance of almost immediate retransmission of the disease, since the peak of the distribution is at  $t = 0$ . This is unrealistic, as there is a minimum length of time between getting infected and having a big enough viral load to infect others. However, replacing the bell with a spike at the mean length is also unrealistic, as it suppresses all variation among disease carriers. There is also the question of symptomatic versus asymptomatic carriers. This is partially addressed in the hybrid model, which incorporates a minimum retransmission time while recognizing that some people (especially those who are asymptomatic or who defy orders to isolate themselves) may be infectious for much longer than others.

In the early stages of an outbreak, when almost all individuals are susceptible, SEIR models (and more complicated extensions of these models) are all linear. Linearity implies that the total number of infected individuals grows exponentially as  $e^{t/\tau}$ , where  $\tau$  is the doubling time of the growth divided by  $\ln(2) \approx 0.69$ . Since all models lead to the same qualitative behavior,  $\tau$  is the only quantity that can be *directly* deduced from the data. The subsequent step of inferring  $R$  from  $\tau$  then depends on which model we choose and which parameters we choose for that model.



**Fig. 2** Estimated values of  $R$  for all three models as a function of the average latency  $t_1$ , assuming  $\tau = 3.5$  and  $t_2 = 7 - t_1$ . While  $R$  greatly increases with latency in the fixed length and hybrid models, the exponential model behaves differently. The exponential model allows a few exposed individuals to reinfect others very quickly and these rapid transmitters drive much of the growth of the outbreak

### 2.2 Estimating $R$

Here are the formulas for  $R$  as a function of  $t_1$ ,  $t_2$  and  $\tau$  for each of the three versions of the SEIR model. When the duration of the Exposed and Infectious states are both treated as exponential random variables,

$$R = 1 + \frac{t_1}{\tau} + \frac{t_2}{\tau} + \frac{t_1 t_2}{\tau^2}. \tag{2}$$

When both durations are fixed,

$$R = \frac{(t_2/\tau)e^{t_1/\tau}}{1 - e^{-t_2/\tau}}. \tag{3}$$

Finally, in the hybrid model

$$R = \left(1 + \frac{t_2}{\tau}\right) e^{t_1/\tau}. \tag{4}$$

These formulas are derived in the next section.

To illustrate how estimates of  $R$  depend on the latency and the choice of model, we consider the rapid growth of COVID-19 in the USA in March 2020. The 100-fold growth over 16 days corresponds to  $\tau = 3.5$ , and the extended incubation period of a week means that  $t_2 = 7 - t_1$ . Figure 2 gives the computed values of  $R$  under these assumptions as a function of the latency  $t_1$ .

Note how the inferred value of  $R$  depends on the average latency  $t_1$ , and also how it depends on the seemingly innocuous choice of whether to make the lengths of the Exposed and Infectious stages fixed or random. While it is easy to make a good case for each variant of the SEIR model, they give vastly different results, especially for larger values of  $t_1$ ! Without a clear scientific justification for choosing one model over the other two, we must acknowledge that all of our estimates of  $R$  are extremely uncertain.

The exponential model behaves differently from the other models because it involves some individuals becoming infectious very quickly, even when the average latency is large, and involves some of those individuals infecting others very quickly as well. These rapid transmitters do not account for the majority of the transmissions, which still take a total of 7 days on average, but they have an oversized effect on the growth rate. The number of these rapid transmitters is maximized when either  $t_1$  is small or  $t_2$  is small, which explains why the estimated value of  $R$  reaches its peak when  $t_1 = t_2 = 3.5$  days.

The lesson is that when the *minimum* time from infection to reinfection goes up, so does the estimated value of  $R$ . There is a simple explanation for this. If  $\tau = 3.5$ , then the outbreak doubles every 2.4 days, quadruples every 4.8 days, and multiplies by 8 every 7.2 days. If it takes at least 4.8 days for an exposed individual to infect somebody else, then  $R$  must be at least 4 to account for the quadrupling in that time; if the time interval is always at least 7.2 days, then  $R$  must be at least 8.

### 2.3 Extent of the Epidemic

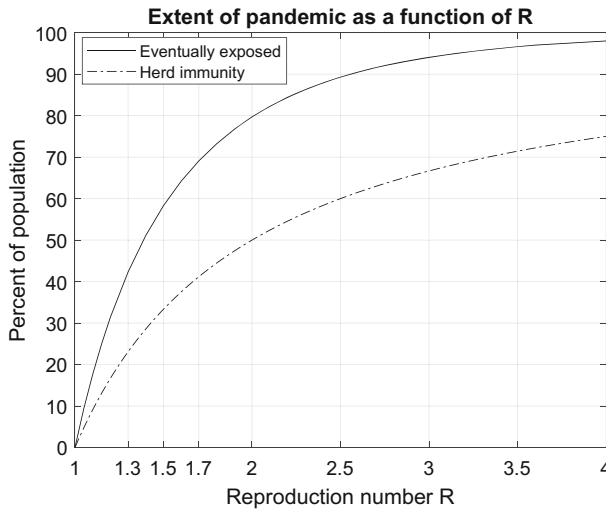
We next consider the dependence on  $R$  of the fraction  $x$  of the population that is eventually exposed to the disease. As noted earlier, and as proved in Sect. 3,  $x$  is the nonzero solution to the equation

$$1 - x = e^{-Rx}. \quad (5)$$

This equation applies to the SIR model, to all three versions of the SEIR model, and to compartmentalized extensions of the SEIR model.

The dependence of  $x$  on  $R$  is shown in Fig. 3. When  $R$  is as small as 1.15,  $x$  is already at 25%, and when  $R = 1.7$ ,  $x$  is already about 70%. When studies such as (Murray 2020) said in March 2020 that 25–70% of the population would be infected in the absence of control measures, they implicitly assumed that  $R_0$  was only between 1.15 and 1.7. Under that assumption, moderate efforts should have been enough to bring  $R$  below 1 and suppress the outbreak. However, if  $R_0$  is actually larger, and if our best long-term social distancing efforts don't bring  $R$  below 2, then the vast majority of the population will be infected in the first full wave of the pandemic. In that case the resulting herd immunity would be impressive, but almost no vulnerable people would be left to benefit from it.

Finally, the value of  $R$  affects the possibility of having multiple waves of the pandemic. Suppose that extreme interventions are sufficient to contain the first wave of the pandemic, during which a fraction  $x$  of the population is infected, and then society re-opens and  $R$  returns to its pre-lockdown value. Unless



**Fig. 3** The percentage of the population that eventually becomes infected as a function of the reproduction number  $R$ . Note that this percentage is already over 40% when  $R = 1.3$ , and rapidly approaches 100% as  $R$  increases. The lower curve is  $x = 1 - \frac{1}{R}$ , the threshold to achieve herd immunity and avoid successive waves of infection

$$x > 1 - \frac{1}{R}, \tag{6}$$

successive outbreaks are extremely likely to occur.

### 3 Mathematical Derivations

All three versions of the SEIR model (with no further compartmentalization) have the same general structure.

1. Susceptibles turn into Exposed at a rate equal to  $\beta S(t)I(t)$ , where  $\beta$  is a transmission coefficient. In the early stages of the pandemic,  $S(t)$  is approximately the entire population  $T$ , so the number of Susceptibles who become Exposed, per unit time, is approximately  $\beta TI(t)$ .
2. Exposed individuals become Infectious in an average of  $t_1$  days. In the exponential version, the number of Exposed who become Infectious per unit time is  $\alpha E(t)$ , where  $\alpha = 1/t_1$ . In the fixed-length version, the rate at which Exposed become Infectious today is exactly equal to the rate at which they became Exposed  $t_1$  days ago, namely  $\beta TI(t - t_1)$ .
3. Infectious individuals are Removed in an average of  $t_2$  days. In the exponential version and the hybrid version, the number of Infectious who are Removed per unit time is  $\gamma I(t)$ , where  $\gamma = 1/t_2$ . In the fixed-length version, the rate at which Infectious are Removed today is exactly equal to the rate at which they became Infectious  $t_2$  days ago, which in turn is the rate at which they were Exposed  $t_1 + t_2$  days ago, namely  $\beta TI(t - t_1 - t_2)$ .



4. The reproduction ratio is  $R = t_2\beta T$ , so we can replace  $\beta T$  with  $R/t_2$  or  $R\gamma$ .

### 3.1 Exponential Incubation Times

In the purely exponential model, these considerations yield the following system of differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta S(t)I(t), \\ \frac{dE}{dt} &= \beta S(t)I(t) - \alpha E(t), \\ \frac{dI}{dt} &= \alpha E(t) - \gamma I(t). \end{aligned} \tag{7}$$

At the beginning of the outbreak  $\beta S(t) \approx \beta T = R\gamma$ , resulting in the linear system of differential equations

$$\frac{d}{dt} \begin{pmatrix} E(t) \\ I(t) \end{pmatrix} = \begin{pmatrix} -\alpha & R\gamma \\ \alpha & -\gamma \end{pmatrix} \begin{pmatrix} E(t) \\ I(t) \end{pmatrix}. \tag{8}$$

The exponential rate of growth ( $1/\tau$ ) is the positive eigenvalue of the matrix on the right-hand side. Since the sum of the two eigenvalues is the trace, namely  $-(\alpha + \gamma)$ , the other eigenvalue must be  $-(\frac{1}{\tau} + \alpha + \gamma)$ . The determinant of the matrix is  $-(R - 1)\alpha\gamma$ . Since the determinant is the product of the two eigenvalues, we have

$$\begin{aligned} -(R - 1)\alpha\gamma &= \frac{-1}{\tau} \left( \frac{1}{\tau} + \alpha + \gamma \right) \\ R - 1 &= \frac{1}{\alpha\gamma\tau^2} + \frac{1}{\tau} \left( \frac{1}{\alpha} + \frac{1}{\gamma} \right) \\ &= \frac{t_1t_2}{\tau^2} + \frac{t_1 + t_2}{\tau}, \end{aligned} \tag{9}$$

which is equivalent to Eq. (2).

### 3.2 Fixed Incubation Times

In the model with fixed incubation times, Exposed become Infected at rate  $RI(t - t_1)/t_2$ , while Infected are Removed at rate  $RI(t - t_1 - t_2)/t_2$ , so

$$\frac{dI(t)}{dt} = \frac{R}{t_2} (I(t - t_1) - I(t - t_1 - t_2)). \tag{10}$$

Plugging in  $I(t) = Ce^{t/\tau}$ , where  $C$  is an unknown constant, we obtain

$$\frac{C}{\tau} e^{t/\tau} = \frac{CR e^{t/\tau}}{t_2} \left( e^{-t_1/\tau} - e^{-(t_1+t_2)/\tau} \right)$$

$$\begin{aligned} \frac{1}{\tau} &= \frac{R}{t_2} \left( e^{-t_1/\tau} - e^{-(t_1+t_2)/\tau} \right) \\ R &= \frac{t_2}{\tau \left( e^{-t_1/\tau} - e^{-(t_1+t_2)/\tau} \right)} \\ &= \frac{(t_2/\tau)e^{t_1/\tau}}{1 - e^{-t_2/\tau}}, \end{aligned} \tag{11}$$

which is Eq. (3).

### 3.3 The Hybrid Model

Here we treat the length of the exposed stage as fixed and the length of the infectious stage as random. The rate at which Exposed become Infectious is  $R\gamma I(t - t_1) = RI(t - t_1)/t_2$ , but the rate at which Infectious individuals are removed is  $\gamma I(t) = I(t)/t_2$ . This gives the differential equation

$$\frac{dI(t)}{dt} = \frac{R}{t_2} I(t - t_1) - \frac{I(t)}{t_2}. \tag{12}$$

Plugging in  $I(t) = Ce^{t/\tau}$  as before, and dividing both sides by  $Ce^{t/\tau}$ , gives

$$\begin{aligned} \frac{1}{\tau} &= \frac{R}{t_2} e^{-t_1/\tau} - \frac{1}{t_2} \\ \frac{t_2}{\tau} + 1 &= Re^{-t_1/\tau} \\ R &= \left( \frac{t_2}{\tau} + 1 \right) e^{t_1/\tau}, \end{aligned} \tag{13}$$

which is Eq. (4).

### 3.4 How Many People Will Eventually Get Sick?

Here we consider what fraction of people will get sick before the pandemic ends. We will derive Eq. (1) twice, first for the three versions of the SEIR model that we previously considered, and then for more general compartmental models.

In the SEIR model, the pandemic only subsides when it starts to run out of Susceptibles to infect, so for this part of the calculation we do *not* make the simplifying approximation that  $S(t) \approx T$ . Instead we look at the exact equation for  $S$ , which is the same in all three versions of the model.

$$\frac{dS(t)}{dt} = -\beta S(t)I(t). \tag{14}$$

Dividing both sides by  $S(t)$  gives

$$\frac{d \ln(S(t))}{dt} = -\beta I(t). \tag{15}$$

Integrating then gives

$$\ln \left( \frac{S(\infty)}{S(0)} \right) = -\beta \int_0^\infty I(t) dt. \tag{16}$$

Let  $x$  be the fraction of people who eventually get sick. Since  $S(0)$  is essentially  $T$  (minus a few infected individuals who seed the outbreak), and since  $S(\infty) = (1 - x)T$ , the left-hand side of Eq. (16) is  $\ln(1 - x)$ . Meanwhile, the right-hand side is  $-\beta t_2 T x$ , since the number of people who get sick is  $xT$ , and since each sick person is infectious for an average of  $t_2$  days. However,  $\beta t_2 T = \beta T / \gamma = R$ , so we have

$$\begin{aligned} \ln(1 - x) &= -Rx, \\ 1 - x &= e^{-Rx}, \end{aligned} \tag{17}$$

which is Eq. (1).

Next we consider a general compartmentalized model, with individuals who have been infected divided into  $n$  categories  $I_1, \dots, I_n$ , with infectivities  $\beta_1, \dots, \beta_n$ , with average durations  $s_1, \dots, s_n$ , and with probabilities  $p_1, \dots, p_n$  of an infected individual eventually landing in each compartment. The average durations  $s_1, \dots, s_n$  are fixed, but we make no additional assumptions about the distribution of times spent in each compartment. Some individuals may spend time in more than one compartment (e.g., pre-symptomatic followed by symptomatic), so the sum of the  $p_i$ 's may be greater than 1.

The equations for  $S$  and  $\ln(S)$  are then

$$\frac{dS(t)}{dt} = -S(t) \sum_{i=1}^n \beta_i I_i(t) \quad \frac{d \ln(S(t))}{dt} = - \sum_{i=1}^n \beta_i I_i(t). \tag{18}$$

Integrating as before, we have

$$\ln \left( \frac{S(\infty)}{S(0)} \right) = - \sum_{i=1}^n \beta_i \int I_i(t) dt. \tag{19}$$

Setting  $S(\infty) = (1 - x)T$ , making the approximation  $S(0) \approx T$  and exponentiating then gives

$$1 - x = \exp \left( - \sum_{i=1}^n \beta_i \int_0^\infty I_i(t) dt \right). \tag{20}$$

Meanwhile, the total number of people who get infected is  $xT$ , of which a fraction  $p_i$  wind up in compartment  $i$ , and spend an average time of  $s_i$  there, so

$$\int_0^\infty I_i(t) dt = p_i s_i x T. \tag{21}$$

This then gives

$$1 - x = \exp \left( -x \sum_{i=1}^n p_i s_i \beta_i T \right). \quad (22)$$

We then compute  $R$ , which is defined as the average transmission per infected individual toward the beginning of the outbreak, when  $S \approx T$ . Each person in compartment  $i$  infects others at a rate  $\beta_i T$  for an average time  $s_i$ , for an average of  $\beta_i s_i T$  newly infected people in all. Since each infected person has a probability  $p_i$  of landing in compartment  $i$ , each infected person infects an average of

$$R = \sum_{i=1}^n p_i s_i \beta_i T \quad (23)$$

others. But this is precisely the expression appearing on the right-hand side of Eq. (22), so

$$1 - x = e^{-Rx}, \quad (24)$$

exactly as before. In fact, our results for the SEIR model can be viewed as a special case of this with just two compartments, namely the Exposed and Infectious states, with  $\beta_1 = 0$ ,  $s_1 = t_1$  and  $p_1 = 1$  for the Exposed state and  $\beta_2 = \beta$ ,  $s_2 = t_2$  and  $p_2 = 1$  for the Infectious state.

### 3.5 Multiple Waves and Herd Immunity

To understand Eq. (6), imagine that  $R = 4$ , so each infected person has contact, of the sort that spreads the virus, with an average of 4 other people. If more than 3/4 of the population has already had the virus, then fewer than one of those four people (on average) will get infected. Since each infected individual actually infects fewer than one new person (on average), any local outbreak will quickly peter out.

More generally, if a fraction  $x$  of the population is already immune and a fraction  $1-x$  is still Susceptible, the average number of new people that each infected individual actually infects is only  $R(1-x)$ , not  $R$ . As long as  $R(1-x) < 1$ , or equivalently  $x > 1 - \frac{1}{R}$ , local outbreaks cannot get any traction and cannot spread into the population at large. This well-known phenomenon, where immunity in part of the population protects the rest, is called “herd immunity”.

## 4 Conclusions

All models for understanding the future trajectory of the COVID-19 pandemic depend on knowing how contagious the disease is, that is, on estimating  $R_0$  or  $R$ . This is true both for complicated models involving many different sorts of people and many different stages of disease progression, as well as for simple models such as SEIR. Unfortunately, there is no direct way to measure  $R$ . All we can do is measure the time scale  $\tau$  of the exponential growth of the pandemic and try to estimate  $R$  from  $\tau$ .

Such estimates depend strongly on the details of the model used. Eventually, the correct parameters and the most accurate models will be revealed through clinical studies of the actual distribution of the different stages of infection and especially by tracing individual contacts to determine when infectivity starts. Until those studies are completed for COVID-19, every projection will involve choosing parameters through educated guesswork.

Even in the simple SEIR models considered in this paper, changing the parameters of the model can change the estimated value of  $R$  dramatically. This was illustrated in the example with  $\tau = 3.5$  days and  $t_1 + t_2 = 7$  days. In the fixed length model the estimate for  $R$  ranged from 2.3 to 7.4 and the hybrid model was almost as sensitive. In all three models, having a latency of just one day ( $t_1 = 1$ ) increased  $R$  by 0.5 or more compared to having no latency at all ( $t_1 = 0$ ). Working with more realistic (and more complicated) models only makes the problem of model dependence worse. The more parameters there are, the more the final answers depend on the assumptions.

Most of the highly cited early estimates of  $R$  have been based either on models without latency, on data from Wuhan that shows much slower growth than seen in Europe and the USA, or both. As such, these estimates are likely to have substantially underestimated the true value of  $R$  in the USA and perhaps elsewhere.

The success of any suppression strategy boils down to one thing: *getting  $R$  below 1*. Strategies that could work if  $R_0$  is between 1.3 and 2.5 might well fail if the actual value of  $R_0$  is larger. Meanwhile, mitigation strategies depend on isolating the most vulnerable members of society until the first wave of the pandemic has subsided. The amount of herd immunity required to make that work also depends on  $R$ . It won't be safe for the vulnerable to go out until the first wave of the pandemic has subsided *and* a fraction  $x > 1 - \frac{1}{R}$  of the population has developed immunity through exposure or vaccination.

If we cannot get  $R$  below 1, then a substantial fraction of the population will get infected. That fraction  $x$  is easily determined from the simple formula  $1 - x = e^{-Rx}$ . Unlike the estimates for  $R$ , this formula is robust, and applies to complicated compartmental models as well as to simple SIR and SEIR models. Given the usual assumptions implicit in compartmental models, this formula applies whenever

- The underlying dynamics of transmission and recovery do not change over time. That is, the parameters  $\beta_1, \dots, \beta_n, s_1, \dots, s_n, p_1, \dots, p_n$  of the system are constant, and
- The only nonlinearity in the dynamics comes from the transition from Susceptible to some of the other compartments, which happen at a rate proportional to  $S$  times a weighted sum of the  $I_i$ 's.

**Acknowledgements** I thank Tony Moody, Stan Schein, Alfredo Sadun, Anita Sadun, Rebecca Sadun and Rina Sadun for their enormous help in making this paper comprehensible to a broad audience. This work is supported by the University of Texas at Austin, but not through any external grants. There are no conflicts of interest. All data is available in the manuscript or the supplementary materials. A preliminary version of this paper was posted to medRxiv and can be found at <https://www.medrxiv.org/content/10.1101/2020.04.07.20056960v1>.

## References

- Backer JA, Klinkenberg D, Wallinga J (2020) Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020, *Eurosurveillance* 25(5), February 6, 2020 <https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062>
- Day-by-day statistics for the world, individual countries, and individual states may be found at <https://www.worldometers.info/coronavirus>
- Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA (2020) The serial interval of COVID-19 from publicly reported confirmed cases, *MedRxiv*
- Dyson F (2004) John von Neumann reportedly told this to Enrico Fermi, who told Freeman Dyson, who described it in an article in *Nature* 427:297
- Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, Bhatia S, Boonyasiri A, Cucunubá Z, Cuomo-Dannenburg G, Dighe A, Dorigatti I, Fu H, Gaythorpe K, Green W, Hamlet A, Hinsley W, Okell LC, van Elsland S, Thompson H, Verity R, Volz E, Wang H, Wang Y, Walker PGT, Walters C, Winskill P, Whittaker C, Donnelly CA, Riley S, Ghani AC (2020) Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. <https://doi.org/10.25561/77482>
- Finer subdivisions, such as regions of China, may be found at <https://interaktiv.morgenpost.de/coronavirus-karte-infektionen-deutschland-weltweit/>
- For a very clear introduction to the SIR model and its variants, including SEIR, see [https://en.wikipedia.org/wiki/Compartmental\\_models\\_in\\_epidemiology](https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology)
- Gyllenberg M, Hanski I (1997) Habitat deterioration, habitat destruction, and metapopulation persistence in a heterogenous landscape. *Theor Popul Biol* 52:198–215
- Hastings A (2003) Metapopulation persistence with age-dependent disturbance or succession. *Science* 301:1525–1526
- Li Q, Guan X, Peng W, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing X, Xiang N et al (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 382:1199–1207. <https://doi.org/10.1056/NEJMoa2001316>
- Mandavilli A (2020) Infected but feeling fine. The unwitting coronavirus spreaders, *New York Times*, March 31, 2020
- Murray LCJ (2020) IHME COVID-19 health service utilization forecasting team and Christopher JL Murray, *Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by US state in the next 4 months*, <https://doi.org/10.1101/2020.03.27.20043752>
- Pasco R, Wang X, Petty M, Du Z, Fox SJ, Pignone M, Johnston C, Ancel ML (2020) COVID-19 Healthcare demand projections: Austin, Texas, presentation to the Austin City Council, March 24, 2020. [https://sites.cns.utexas.edu/sites/default/files/cid/files/covid-19\\_analysis\\_for\\_austin\\_march2020.pdf](https://sites.cns.utexas.edu/sites/default/files/cid/files/covid-19_analysis_for_austin_march2020.pdf)
- Tang B, Bragazzi N, Li Q, Tang S, Xiao Y, Wu J (2019) Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J Clin Med* 9(2):E462

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