CoViD-19 epidemic follows the “kinetics” of enzymes with cooperative substrate binding

Todd P. Silverstein

Chemistry Department (emeritus), Willamette University, Salem, Oregon

Correspondence
Todd P. Silverstein, Chemistry Department (emeritus), Willamette University, Salem, OR 97301. Email: tsilvers@willamette.edu

Abstract
The Hill equation, which models the cooperative ligand-receptor binding equilibrium, turns out to be useful in modeling the progression of infectious disease outbreaks such as CoViD-19. The equation fits well the data for total and daily case numbers, allows tentative predictions for the half-point and end point of the epidemic, and presents a mathematical characterization of how social interventions “flatten the curve” of the disease progression.

KEYWORDS
CoViD-19, enzyme kinetics, hill equation, ligand binding, Michaelis–Menten

1 | INTRODUCTION

Those who followed the CoViD-19 epidemic (and who did not?) watched with trepidation as the number of cases rose steadily. A common question was “when will this end and life return to normal?” The answer was invariably “who knows?” Well, it turns out, biochemists know. A glance at the plot of total number of cases vs. day showed a sigmoidal rise. This curve is familiar to biochemists as the saturation plot for a protein that binds ligand cooperatively, that is, with a Hill coefficient \( n \geq 2 \).

There are a number of sophisticated mathematical models of infectious disease progression.\(^1\)\(^-\)\(^4\) They can be stochastic or deterministic, statistical or mechanistic. The most complex models can account for variables like reproduction number (\( R_0 = \text{infectivity} \)), steady state, vaccination, social distancing, hospital capacity, etc. But wouldn’t it be nice if there was a simple, familiar mathematical model that we all had access to and could use to model the severity and timeline of a pandemic? Enter, the Hill equation.

The Hill equation describes how ligand binding to a protein receptor saturates with the concentration of free ligand, \([L]\):

\[
f_b = \frac{\text{occupied receptor}}{\text{total receptor}} = \frac{[PL_n]}{[P]_\text{tot}} = \frac{[L]^n}{K_d + [L]^n} = \frac{1}{1 + \left(\frac{[L]}{[L]_{50}}\right)^n}
\]  

(1)

where \( f_b \) is the fraction of total receptor that has bound \( n \) ligands, and \( K_d \) is the \( PL_n \) dissociation equilibrium constant; the exponent \( n \) is known as the Hill coefficient. Because \( K_d \) is the free ligand concentration that gives \( f_b = 0.5 \), also known as \([L]_{50}\), Equation (1) can be written as:

\[
[PL_n] = \frac{[P]_\text{tot}}{1 + \left(\frac{[L]}{[L]_{50}}\right)^n}
\]  

(2)

For an enzyme, the enzyme-substrate complex \((E_n = P_n)\) goes on to catalyze the reaction, so the Hill equation can be transformed to describe the dependence of initial reaction rate \((v_0)\) on initial substrate concentration, \([S]_0^0\):

\[
v_0 = \frac{V_{\text{max}}}{1 + \left(\frac{K_m}{[S]_0^0}\right)^n} = \frac{V_{\text{max}}}{1 + \left(\frac{[S]_{50}}{[S]_0^0}\right)^n}
\]  

(3)

Received: 5 May 2020 Revised: 19 May 2020 Accepted: 26 May 2020
DOI: 10.1002/bmb.21397

Biochem Mol Biol Educ. 2020;48:452–459.
where \( V_{\text{max}} \) is the asymptotic \( v_{0} \) at saturating \([S]_{0}\). A Hill coefficient \( \geq 2 \) signifies the cooperative binding of two or more substrate molecules at the same time. Equation (3) reduces to the well-known Michaelis–Menten equation when \( n = 1 \) (i.e., non-cooperative binding of a single substrate molecule); in this case, the Michaelis–Menten plot, \( v_{0} \) vs. \([S]_{0}\), saturates hyperbolically at the asymptote \( V_{\text{max}} \). For \( n \geq 2 \), cooperative \( v_{0} \) vs. \([S]_{0}\) plots saturate sigmoidally.

To model the sigmoidal rise of the total number of disease cases (e.g., CoViD-19) with time, one would use this version of the Hill equation:

\[
cases(t) = \frac{\text{cases}_{\text{final}}}{1 + \left(\frac{t_{50}}{t}\right)^{n}}
\]

(4)

Here \( t \) represents days, \( t_{50} \) is the day that you reach half of the final number of cases (cases_{final}), and \( n \) controls the steepness of the sigmoidal rise. Together, \( t_{50} \) and \( n \) provide information on how long the outbreak will last.

For example, Figure 1a shows sigmoidal plots for \( n = 3, 5, \) and 8, given cases_{final} = 100 and \( t_{50} = 40 \) days. Note that a lower \( n \) flattens the curve, so that the rate of increase of cases is shallower. Although this does not decrease the final number of cases, it is advantageous in that it would decrease the number of hospital admittances on a particular day. The curve is also flattened by decreasing cases_{final} (Figure 1b) and by increasing \( t_{50} \) (Figure 1c).

Many print media stories that discussed flattening the CoViD-19 epidemic curve illustrated the effect not with total cases vs. time curve, but with the bell-shaped daily cases vs. time curve. Daily cases as a function of time is the first derivative of the total cases function:

\[
d(\text{cases}(t))/dt = \#\text{daily cases}(t) = \frac{n \cdot \text{cases}_{\text{final}} \cdot \left(t_{50}\right)^{n}}{\left(t/n + 1\right) \cdot \left[1 + \left(\frac{t_{50}}{t}\right)^{n}\right]^{2}}
\]

(5)

In Figure 2 we can see how the three fit parameters affect the shape of the first derivative plot. Lowering case_{final} (Figure 2a) decreases the amplitude, that is, the number of daily cases on each day, but leaves the relative width and time-axis position of the bell-curve unchanged. Lowering \( n \) (Figure 2b) increases the relative width of the curve, and shifts its peak time to the left (to shorter times). This has the effect of increasing the number of daily new cases early and late in the epidemic (before day 25–30, and after day 55–60), while dramatically decreasing this number at the peak of the epidemic. Finally, increasing \( t_{50} \) (Figure 2c) also increases the relative width of the curve, while shifting the peak time to the right (later times). This spreads out the number of new daily cases, decreasing it early in the epidemic, while increasing it later on.

**FIGURE 1** Sigmoidal saturation plots for various values of (a) \( n = 3, 5, 8 \) (cases_{final} = 100 and \( t_{50} = 40 \) days); (b) cases_{final} = 40, 60, 100 (\( n = 5, t_{50} = 40 \) days); and (c) \( t_{50} = 20, 40, 60 \) (\( n = 5, \) cases_{final} = 100) [Color figure can be viewed at wileyonlinelibrary.com]

From Figure 2c we can see that the curve is asymmetric—it rises faster than it falls. Furthermore, this asymmetry declines as \( n \) increases (Figure 2b, \( n = 8 \)). Another surprising observation regarding the first derivative curve is that its peak is always \( \leq t_{50} \) (vertical dotted line in Figure 2a,b). The peak approaches \( t_{50} \) as \( n \) increases, rising from about \( 0.6 \times t_{50} \) for \( n = 2 \), to \( 0.975 \times t_{50} \) for \( n = 8 \).
To explore some real epidemic data sets, Figure 3 presents a plot of the number of CoViD-19 cases in China and South Korea. Clearly, the sigmoidal fits to Equation (4) are good, although not perfect: For example, the early undershoot and late overshoot of the China curve are due to the fact that Chinese authorities changed their method of diagnosing the disease after day 24, suddenly increasing the number of cases.

Before we go any further, it is important to state up front that fitting infection data to the Hill equation is purely a statistical process. There is no theoretical model explaining why this equation fits the data, thus there is no way to connect the Hill equation parameters to viral or human biology, or to social policy implementation. On the other hand, one might hypothesize that faster viral reproduction inside the host would raise both \( n \) and \( c_{\text{final}} \); the same would be expected for anything that raises infectivity, for example, more infectious viral particles, easier viral spread from asymptomatic individuals, higher frequency of close human contact, etc.

Hill equation fits are quite good for almost all data sets (\( R^2 = 0.9940 \)–\( 0.9997 \)). Table 1 lists Equation (4) fitted parameters for data sets from 35 countries (fitted plots for each data set can be found in the Figures S2–S38). The Hill coefficients (\( n \)) vary about two-fold (from 4 to 8), and half-times (\( t_{50} \)) vary about five-fold (from 20 to 100 days); on the other hand, cases\(_{\text{final}}\) vary over 3,000-fold, from 384 for Taiwan to 1.4 million for the United States.

Above, we speculated that anything that enhanced infection rate might raise both \( n \) and \( c_{\text{final}} \). In fact, the fitted data from 35 countries show no strong relationship between any of the three fit parameters (Figure S1). For example, the United States had a cases\(_{\text{final}}\) value almost six times higher than the next-highest country, but its \( t_{50} \) and \( n_{\text{Hill}} \) are in the middle of the range of values. In general, countries with the lowest and highest values of \( t_{50} \) span the entire range of \( n_{\text{Hill}} \); the same is true of the lowest and highest values of cases\(_{\text{final}}\). There may be a weak correlation between \( t_{50} \) and cases\(_{\text{final}}\) (Figure S1): Countries with the lowest \( t_{50} \) values (<36 days) all had low cases\(_{\text{final}}\) (<30,000); conversely, countries with the highest cases\(_{\text{final}}\) (>170,000) all had moderate to high \( t_{50} \) values (≥40 days). It seems logical that strong social intervention, including...
| Country      | # cases_{final} | t_{50} (day) | n   | R^2     | % completion |
|--------------|----------------|-------------|-----|---------|--------------|
| China        | 74,780 ± 120   | 29.76 ± 0.08| 6.03 ± 0.09 | 0.9975  | 102^b        |
| S. Korea^c   | 10,594 ± 26    | 20.58 ± 0.07| 4.85 ± 0.08 | 0.9995  | 101          |
| Taiwan^d     | 384 ± 4        | 25.90 ± 0.13| 5.06 ± 0.13 | 0.9986  | 100          |
| Hong Kong^d  | 1,054 ± 9      | 34.11 ± 0.09| 7.41 ± 0.13 | 0.9993  | 98           |
| Austria      | 15,380 ± 50    | 32.47 ± 0.08| 5.91 ± 0.07 | 0.9995  | 99           |
| Australia    | 6,709 ± 18     | 34.91 ± 0.06| 7.81 ± 0.09 | 0.9996  | 100          |
| Thailand     | 3,000 ± 15     | 30.80 ± 0.10| 4.97 ± 0.07 | 0.9993  | 98           |
| Luxembourg   | 3,847 ± 25     | 25.15 ± 0.13| 4.13 ± 0.08 | 0.9989  | 97           |
| Switzerland  | 30,700 ± 150   | 33.51 ± 0.11| 4.57 ± 0.05 | 0.9995  | 95           |
| Germany      | 171,400 ± 900  | 42.44 ± 0.12| 5.48 ± 0.06 | 0.9996  | 94           |
| Germany^e    | 174,100 ± 800  | 42.77 ± 0.11| 5.36 ± 0.06 | 0.9996  | 96           |
| Israel       | 16,760 ± 200   | 46.69 ± 0.26| 6.13 ± 0.14 | 0.9984  | 94           |
| Czechia      | 8,070 ± 50     | 35.72 ± 0.15| 4.57 ± 0.06 | 0.9995  | 92           |
| Norway       | 8,295 ± 70     | 34.72 ± 0.23| 3.82 ± 0.07 | 0.9990  | 92           |
| Spain        | (2.51 ± 0.04) × 10^5 | 43.8 ± 0.4  | 4.91 ± 0.12 | 0.9980  | 92           |
| France       | 191,200 ± 2,100 | 46.90 ± 0.25| 5.84 ± 0.11 | 0.9989  | 90           |
| Italy        | (2.26 ± 0.03) × 10^5 | 48.9 ± 0.3  | 4.44 ± 0.08 | 0.9989  | 89           |
| Malaysia     | 6,690 ± 60     | 41.94 ± 0.23| 4.30 ± 0.06 | 0.9994  | 87           |
| Portugal     | 28,200 ± 400   | 40.8 ± 0.3  | 4.49 ± 0.09 | 0.9990  | 85           |
| NY state     | 366,000 ± 5,000| 40.0 ± 0.4  | 4.18 ± 0.07 | 0.9991  | 82           |
| NY state^e   | 378,000 ± 4,000| 40.84 ± 0.3 | 4.06 ± 0.06 | 0.9992  | 80           |
| Iran         | 115,000 ± 2,500| 50.5 ± 0.7  | 3.92 ± 0.10 | 0.9978  | 81           |
| SF Bay Area  | 6,010 ± 60     | 42.92 ± 0.24| 4.08 ± 0.04 | 0.9997  | 81           |
| Turkey       | 149,900 ± 1,100| 39.15 ± 0.15| 4.51 ± 0.04 | 0.9998  | 80           |
| USA          | (1.29 ± 0.03) × 10^6 | 56.8 ± 0.5  | 5.84 ± 0.13 | 0.9986  | 79           |
| USA^e        | (1.48 ± 0.04) × 10^6 | 59.9 ± 0.6  | 5.35 ± 0.13 | 0.9983  | 79           |
| UK           | 210,000 ± 2,500| 59.05 ± 0.26| 6.16 ± 0.07 | 0.9997  | 77           |
| UK^e         | 236,000 ± 4,000| 61.6 ± 0.4  | 5.72 ± 0.09 | 0.9993  | 79           |
| Belgium      | 60,800 ± 1,000 | 47.6 ± 0.4  | 4.48 ± 0.07 | 0.9993  | 77           |
| Netherlands  | 51,450 ± 800   | 48.2 ± 0.4  | 4.04 ± 0.06 | 0.9994  | 74           |
| Canada       | 72,200 ± 2,700 | 64.1 ± 0.9  | 5.66 ± 0.14 | 0.9984  | 67           |
| Ireland      | 29,900 ± 1,400 | 52.21 ± 1.1 | 4.84 ± 0.16 | 0.9973  | 66           |
| Japan        | 21,350 ± 1,400 | 74.5 ± 1.5  | 6.85 ± 0.29 | 0.994   | 64           |
| Calif. state | 88,200 ± 7,500 | 54.8 ± 2.7  | 3.32 ± 0.13 | 0.9970  | 52           |
| Calif. state^e | 114,100 ± 9,000| 62.9 ± 2.9  | 3.08 ± 0.10 | 0.9978  | 49           |
| Russia       | 211,000 ± 8,000| 60.9 ± 0.6  | 7.11 ± 0.09 | 0.9997  | 44           |
| Russia^e     | (3.48 ± 0.24) × 10^5 | 69.1 ± 1.3  | 6.26 ± 0.12 | 0.9992  | 42           |
| Sweden       | 44,600 ± 2,500 | 69.7 ± 2.0  | 3.38 ± 0.06 | 0.9993  | 42           |
| India        | 70,050 ± 3,300 | 64.8 ± 1.0  | 5.51 ± 0.08 | 0.9996  | 42           |
| India^e      | 117,000 ± 8,000| 75.6 ± 1.8  | 4.97 ± 0.09 | 0.9994  | 36           |
| Peru         | (9.0 ± 1.7) × 10^4 | 64 ± 4      | 5.14 ± 0.23 | 0.9966  | 32           |
| Peru^e       | (6.3 ± 5.8) × 10^5 | 109 ± 26    | 4.44 ± 0.16 | 0.9979  | 7            |
| Brazil       | (4.4 ± 1.5) × 10^5 | 89 ± 10     | 4.05 ± 0.12 | 0.9984  | 15           |
| Brazil^e     | ≈2 × 10^6      | ≈130        | 4.02 ± 0.12 | 0.9984  | ≈6           |

(Continues)
wide-spread testing, contact tracing, and quarantining, would lower both $t_{50}$ and cases$_{final}$.

Only three data sets were not well-fit by Equation (4), and their deviations allow some interesting speculation on the effectiveness of social intervention. The Taiwan and Hong Kong (Figure 4) data sets had to be fit with a non-zero intercept. This could be because their interventions, after the first scores of cases were identified, were so effective that these initial cases did not lead to a sigmoidal rise; only after the first 120 cases appeared in Hong Kong did the disease spread as an epidemic, presumably via community transmission.

The South Korea data set showed a sigmoidal rise in cases until day 25, after which the cases numbers rose more slowly (Figure 5). One could speculate that South Korea, which was the second nation to experience an outbreak of CoViD-19, managed to mount an intensive social intervention around day 25, including for example, extensive testing, contact tracing, quarantining, etc. This would be another example of “flattening the curve”. Note that although the curve was indeed flattened from day 25–55, the number of cases$_{final}$ was unchanged by this intervention: 10,594 ± 26 for the blue dashed curve (i.e., the normal sigmoidal rise without the intervention) vs. 10,550 ± 110 for the red solid curve. In this case, flattening the curve did not decrease the eventual number of cases, it just slowed the rate at which those cases appeared every day, allowing hospitals to accommodate the influx of patients.

As mentioned above, in the midst of a pandemic, one question on everyone’s mind is “when will it be over?” The nature of an asymptote is that you never actually reach it. However, one can ask how long it would take to get close to the end, for example, within 90 or 95% of the maximum value. Manipulating Equation (4), we can show that if $f$ is the fraction of the total you are interested in (e.g., 0.9 or 0.95), then the time to reach that value ($t_f$) can be calculated:

$$t_f = \frac{t_{50}}{(\frac{1}{f} - 1)^{\frac{1}{n}}}$$

TABLE 1 (Continued)

| # cases$_{final}$   | $t_{50}$ (day) | $n$ | $R^2$  | % completion$^a$ |
|--------------------|---------------|-----|--------|------------------|
| Mexico (1.9 ± 1.4) x 10^5 | 100 ± 22      | 3.97 ± 0.13 | 0.9983 | 8                |
| Mexico$^e$ (1.2 ± 0.3) x 10^5 | 84 ± 8        | 4.23 ± 0.13 | 0.9983 | 19               |

Note: Fitted curves for each country can be found in the Figures S2–S38. Unless otherwise specified, all fits are performed by nonlinear regression, using the Kaleidagraph software program, and the last reported data point is from April 27, 2020.

$^a$Last reported total number of cases = cases$_{final}$ × 100%.

$^b$The final reported value exceeds cases$_{final}$ (>100% completion) due to case influx from abroad.

$^c$Days 25–55 are omitted from the fitting process, as discussed below.

$^d$Hong Kong, Taiwan data were fit with non-zero y-intercept (120 and 35 cases, respectively).

$^e$The final reported value is from May 3, 2020.

FIGURE 4 Hong Kong CoViD-19 cases fit to Equation (4) with a zero y-intercept (red curve), and with a non-zero intercept (black curve) of 122 ± 4 cases [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 5 South Korea CoViD-19 cases fit to Equation (4) with all points (red solid curve), and with days 25–55 omitted (blue dashed curve) [Color figure can be viewed at wileyonlinelibrary.com]
We can see from Equation (6) that, unsurprisingly, $t_1$ increases linearly with $t_{50}$ but it also decreases with $n$ (larger $n$ gives a steeper rise, which gives a shorter $t_1$). For example, if $n = 3$ and $t_{50} = 20$ days, then $t_{50} = 42$ days, but for $n = 4$, $t_{50} = 35$ days. Again, decreasing $n$ flattens the curve, which spreads out the duration of the epidemic (Figure 2b). Using Equation (6) and fit parameters from Table 1, $t_{50}$ for the United States and New York state come out to 92 and 72 days, respectively; that would be May 24 (2020) for the United States as a whole, and May 6 for New York state.

A better way to judge if you are close enough to the end of an epidemic to ease social restrictions is to identify a target maximum number of new daily cases. The World Health Organization has stated that an outbreak is deemed “contained” when < 10% of tests for the disease are positive; to be even safer, we can aim for < 1%. Harvard University’s Global Health Institute has opined that the “bare minimum” of testing capacity in the United States should be 500,000 per day (0.15% of the population), so 1% would be 5,000 positive tests per day. Taking the fit parameters from Table 1 and using the first derivative equation (Equation 5), we can predict that the outbreak could be deemed contained in the U.S. by day 99 (May 31). For New York, the target would be < 300 new daily cases (0.015% × 20 million), which should occur by day 106 (June 9).

Of course, it is easy to fit an asymptotic curve when your data approach the asymptote, as they do in Figure 3. But what if you are in the middle of the epidemic, and you want to predict the end? This is much more difficult. And here the function and limitations of data-fitting must be considered. By fitting an entire data set to a mathematical equation, we can then use the fit parameters to predict the value of the dependent variable ($y$) for a value of the independent variable ($x$) that has not been measured, but lies within the range of measured values. Furthermore, the $R^2$-value gives you a notion of the quality of the fit, and if we know the uncertainties in the fit parameters, we can calculate the uncertainty in the predicted $y$-value.

However, that is not the situation in the middle of an epidemic. Here we are collecting data points one day at a time, and trying to use the fit equation to predict $y$-values appertaining to $x$-values that lie far beyond any that we have measured so far. In other words, we are using nonlinear regression to attempt to fit an incomplete data set. In this case, the $R^2$ value and the fit parameter uncertainties do not mean the same thing that they do when working with a complete data set.

Consider the countries that were below 70% completion as of April 27, 2020, that is, the last ten countries in Table 1. One can see (Figure 6) that the relative uncertainties of the $cases_{final}$ and $t_{50}$ parameters are substantially higher than for the countries closer to the end of their epidemic. On the other hand, $n_{Hill}$ relative uncertainties are only slightly higher for these bottom 10 countries, compared to the other countries (Figure 6). Table 2 affords a closer look at this effect.

It turns out that even the high uncertainties for the low-completion data sets are woefully underestimated. And it goes almost without saying that an $R^2$ value above 0.99 for fitting a data set that is less than 50% complete is meaningless. One way to examine this important issue is to take a complete data set, and remove points from the end, examining how this affects the fit. For example, Germany’s data set was very well-fit by Equation (4), with $R^2 = 0.9996$, and its final reported number of total cases (from April 27) was 94% of the fitted $cases_{final}$ value. Thus, its fit parameters are extremely well-determined: $cases_{final} = 171,400 ± 900; t_{50} = 42.44 ± 0.12$ days; and $n_{Hill} = 5.48 ± 0.06$. Removing points from the end of the data set allows us to examine the fit parameters at 90% completion, 80%, and so on.

![Figure 6](https://example.com/figure6.png)

**FIGURE 6** Relative uncertainties for $cases_{final}$ (red), $t_{50}$ (blue), and $n_{Hill}$ (black), plotted against the percentage completion of the case load for all countries listed in Table 1 [Color figure can be viewed at wileyonlinelibrary.com]

| Completion | ≥90% | 70–89% | <70% |
|------------|------|--------|------|
| $cases_{final}$ | 0.7% | 1.5% | 4% for Canada (67% complete) to 74% for Mexico (8% complete) |
| $t_{50}$ | 0.4% | 0.8% | 1.4% for Canada (67% complete) to 22% for Mexico (8% complete) |
| $n_{Hill}$ | 1.7% | 1.7% | 1.3% for Russia (44% complete) to 4.5% for Peru (32% complete) |

*Note: Relative uncertainties are highest for the lowest completion % data sets.*
In Figure 7 we plot the fitted values for \( t_{90} \) at completion percentages of 94\% down to 14.5\% (below 14\% the fitting process failed). The first thing to notice is that the uncertainties in \( t_{90} \) returned by nonlinear regression, plotted as vertical error bars, do not reach the value obtained for the full data set (dotted line). From this we learn that when using nonlinear regression to fit an incomplete data set, the reported uncertainties are greatly underestimated. The second observation is that below 30\% completion, the \( t_{90} \) value is far too low, by 2- to three-fold. If your set of reported total data can be viewed at wileyonlinelibrary.com

Another way to look at this problem is to examine \( t_{90} \) predictions at various points of completion \%. Figure 8 plots \( t_{90} \) (calculated from the Equation (4) fitted parameters, using Equation (6)) for the Germany data set ending on day 31 (14.5\% completion), day 32 (17\% completion), etc., up to the full 68-day data set (94\% completion). Note that \( t_{90} \) rises dramatically from day 31 to day 39 (39\% completion), after which it continues to rise more slowly all the way to the end of the data set on day 69 (94\% completion). This receding “goal line” is another example of the dangers of trying to fit an incomplete data set.

Consider the Germany data set ending on day 34, which is 22\% complete. From fitting this incomplete data set, one would expect \( t_{90} \) to be day 47 (Figure 8), and the final number of total cases to end up around 75,000 (Figure 7). But moving forward 13 days to day 47, \( t_{90} \) is still 12 days away (talk about running and getting nowhere!), and the predicted final number of cases has more than doubled, to 155,000. Here is another example of this problem: An earlier draft of this paper included data up to March 31 (2020). At that point (day 39), the U.S. case load was 213,000, and fitting suggested \( t_{90} = 51 \) and \( \text{cases}_{\text{final}} = 440,000 \pm 14,000 \) and \( t_{90} = 51 \) As I write this on May 1 (day 69), the current U.S. case load is 1.1 million, and \( t_{90} = 83! \)

The bottom line is that one has to take these fitting results with a grain of salt, especially when conditions such as testing frequency and social intervention change during the course of the epidemic. Having said that, the predictions do become fairly reliable above 60\% completion. Of course there is an endless loop problem here, because one does not know the actual \( t_{90} \) until the data set is complete, so your calculation of % completion will have an uncertain denominator. However, Figure 7 allows us to calculate that this error should be less than 10\%.

Figure 7 allows us to calculate that this problem should be less than 10\%.

A word is in order concerning the use of diagnosed cases, as opposed to hospitalizations or deaths. I have chosen to use reported cases because they are a current indicator of the epidemic, whereas deaths and hospitalizations are trailing indicators. However, it is definitely true that increasing testing during the course of the epidemic skews the fits. This explains, at least partially, the modest rise in \( t_{90} \) seen in Figure 7 after 40\% completion. The skewing effect of increased testing can also be seen in the slight upward tilt of the California total cases curve after day 48 (Figure S31), and in the increase in \( \text{cases}_{\text{final}} \) and \( t_{90} \), and decrease in \( n_{\text{Hill}} \) and completion % from April 27 to May 3 for almost all of the countries that were re-fit on May 3 (Table 1). The good news is that hospitalization and death data sets follow the same trend as total cases (only delayed by 1 or 3 weeks, respectively), and can also be fit to the Hill equation (total cases, Equation 4) and its first derivative (daily cases, Equation 5).

Let us close with an example of the usefulness of this fitting procedure. On April 8 (2020), the governor of
New York announced that, based on the models his experts were using, he expected the peak of the COViD-19 outbreak would be reached in mid-June. Assuming that by “peak” he meant the maximum of the new daily cases (first derivative) curve, simple plotting and fitting would have shown him that New York state had in fact already passed its peak of new infections earlier that week, on April 4!

Some experts predict that the world will have to deal with a resurgence of CoViD-19 in the fall of 2020, and perhaps even in years afterward. Even if not, we are only one interspecies virus hop and genetic recombination away from the next epidemic. In the meantime, biochemistry students can be regaled with the unexpected usefulness of the Hill equation in predicting the progression of epidemics.

ACKNOWLEDGMENTS
I wish to thank my colleagues J. Charles Williamson for his data-fitting insight, and Mark Janeba for his help with equation differentiation, and my partner Rosalie Schele for her help with data entry.

ORCID
Todd P. Silverstein https://orcid.org/0000-0003-3349-2876

REFERENCES
1. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc London Ser A Contain Papers Math Phys Character. 1927;115:700–721.
2. Anderson RM, Anderson B, May RM. Infectious diseases of humans: dynamics and control. Oxford, UK: Oxford University Press, 1992.
3. Brauer F, Castillo-Chavez C, Castillo-Chavez C. Mathematical models in population biology and epidemiology. New York, NY: Springer, 2012.
4. Anonymous. Mathematical modelling of infectious disease. Wikipedia. https://en.wikipedia.org/wiki/Mathematical_modelling_of_infectious_disease. Accessed June 2020.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Silverstein TP. CoViD-19 epidemic follows the “kinetics” of enzymes with cooperative substrate binding. Biochem Mol Biol Educ. 2020;48:452–459, https://doi.org/10.1002/bmb.21397