# Predicting the Unpredictable

Ana Cascon & William F. Shadwick

1 April 2022

## 1 Modelling the course of Covid-19

> ...you know there's certain mathematics to the outbreak of infectious disease. - Bruce Aylward[1]

The use of mathematics to describe processes in the physical world is one of the triumphs of human invention. Mathematical models allow us to see into the future–to predict the trajectory of a Mars lander or, with much less accuracy, the landfall of a hurricane.

There is, indeed, a *certain mathematics* to the outbreak of infectious diseases. But that mathematics is not what has driven the response to Covid-19. Instead, wildly inaccurate predictions of severity and misplaced confidence in unprecedented interventions with no scientific basis have led almost every western democracy to follow the same disastrous path.

Guided by data from past epidemics and from Covid-19, we have encoded that *certain mathematics* in a simple model. The Extended Gompertz Function Model exemplifies the power of mathematics to make forecasts. It predicts the course of Covid-19 waves with enough accuracy to manage healthcare demand–the most critical aspect of any epidemic. It allows us to see that the interventions made no difference to outcomes. In particular, we can use it to show that the success of the Chinese strategy that was recommended by the WHO and adopted all over the world was a mirage. Covid-19 infections in China had almost certainly peaked before any of their 'brutal but effective'[2] measures were imposed.

If you're wondering where this mathematical model was in February and March of 2020, the answer is that it had yet to be discovered. It can be found in a paper that we posted to MedRXiv on 26 December 2021[3]. The present article is a less technical version which outlines how the model works and how it can be used not only to predict the course of Covid-19 waves but to draw inferences about the nature of the epidemic process that drives them.

It is the latter that tells us that the Chinese lockdown–the prototype and inspiration for unprecedented restrictive measures the world over–could not have had the effect that the WHO assumed it did in that fateful press conference of 24 February 2020.

**If you want to predict the future, you need start by understanding the past** Without data about a phenomenon that can be used to validate it, a mathematical model is a purely intellectual 'toy'. It may well have uses–such as testing hypotheses about what happens in the *model world* when inputs are varied, interventions imposed, behaviours change, vaccination occurs and so on, but it is *not* a tool for making predictions about the world we live in.[4]

So the first thing we need to do is to look at epidemic data from the past. When we do, we see striking similarities in the behaviour of epidemics over the past 170 years. These are illustrated by Figure 1, the cumulative deaths from a Cholera outbreak in Denmark in 1853, Figure 2, cumulative influenza cases in Portugal in the 'flu season' of 2018-2019 and Figure 3, cumulative Covid-19 hospital admissions in London on 28 April 2020. In the plots the blue dots record the daily totals and the red curve is the graph of the Gompertz Function[5] that best fits the data.

Exactly the same observation can be made for numerous other epidemics from the past, including Ebola and SARS as can be seen in our paper.The past tells us that cumulative epidemic events, be they

[1] WHO-China Joint Mission on Covid-19, Press Conference on 24 February 2020, Transcript page 10

[2] The Guardian 19 Mar 2020 China's coronavirus lockdown strategy: brutal but effective. Emma Graham-Harrison and Lily Kuo.

[3] Ana Cascon and William F. Shadwick 2021 Predicting the course of Covid-19 and other epidemic and endemic disease. `https://www.medrxiv.org/content/10.1101/2021.12.26.21268419v1`

[4] As Graham Medley, chairman of the UK's SPI-M modelling group, has been at pains to explain, providing scenarios is the, often misunderstood, purpose of the Covid-19 modelling presented to the UK SAGE committee.`https://www.spectator.co.uk/article/what-the-media-gets-wrong-about-sage-s-models` Unfortunately, many other modellers have presented their outputs as firm predictions of the future.

[5] The Gompertz Function is defined by $Gompertz(t, a, b, N) = Ne^{-e^{-(at+b)}}$.

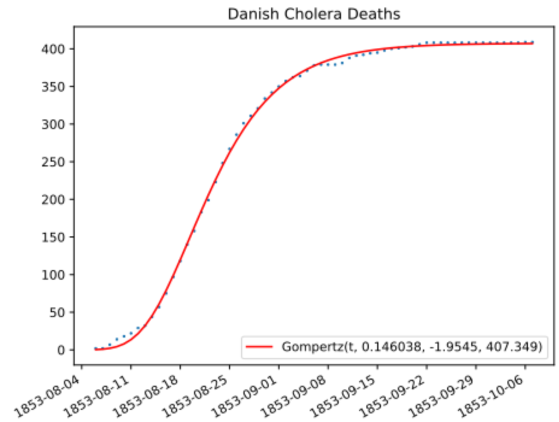



Figure 1: Deaths from the 1853 Cholera epidemic in Aalborg Denmark.

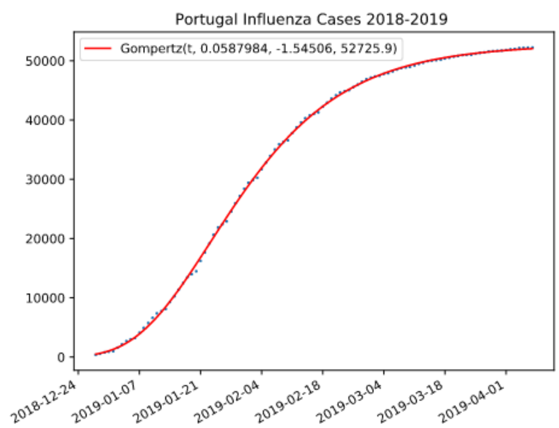



Figure 2: Influenza Cases in Portugal 28 Dec 2018-7 Apr 2919.

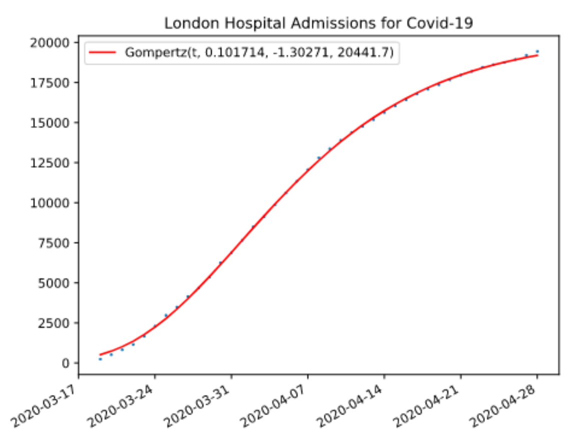



Figure 3: Covid-19 Hospital Admissions in London 19 Mar-28 Apr 2020.

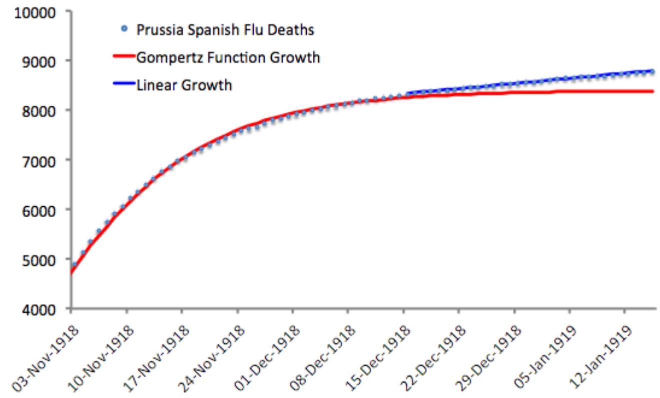



Figure 4: Spanish Flu Deaths in Prussia transition to linear growth December 1918.

cases, hospital admissions or deaths, all look remarkably similar. Whether it's the deaths in a horrific Cholera epidemic (where over 50% of those infected died), a routine winter outbreak of endemic disease or the hospital admissions in London during the first Covid-19 wave, they all follow a characteristic Gompertz Function curve.

The Gompertz Function is the first component of our 'Extended Gompertz Function Model' for epidemics. The fact that the actual numbers of events in Figures 1 to 3 are so close to the outputs of the model is what allows us to say that the epidemic events 'follow' the Gompertz Function.

This statement is exactly analogous to Galileo's assertion that projectiles follow parabolic paths (which he demonstrated long before Newton explained why they *must* do so). In neither case does this phenomenological model explain why the path is followed. But in both cases the knowledge that it *is* followed allows one to make predictions about the future based on observations of the past. The knowledge that a projectile follows a parabola allows us to make inferences about the nature of the motion. For example, we can see what a 'buzzer beater' basketball shot would look like if it were made on the moon, at least if we ignore the constraints of a spacesuit and a few other details. (See Box Galileo: Basketball on the Moon...).

In exactly the same way, knowing that epidemics follow Gompertz Functions allows us to make inferences about them as well. For example, the difference between successive Gompertz Function values is a model for the number of daily events. For a Gompertz Function, these daily differences rise to a peak and then become smaller every day. Eventually the number of daily events shrinks to zero. At this point the model epidemic is over. So if an epidemic continues to follow Gompertz Function growth, then the mathematics tells us that it eventually dies out. This was the case with the Cholera epidemic shown in Figure 1.

But that's not the only possible outcome. The alternative is that cumulative events continue to grow linearly–a feature of endemic disease.

Our analysis of several years of daily influenza case records from Portugal's national health service showed that while each outbreak follows a Gompertz Function, cases never die out completely. Instead the growth of cases becomes linear. We also found the same behaviour at the end of each of the first two waves of the Spanish Flu in England and at the end of the 1918 Spanish Flu outbreak in Prussia as Figure 4 shows.[6]

And we saw exactly the same thing in all the initial Covid-19 outbreaks. Figure 5 illustrates this for Covid-19 hospital admissions in London in April 2020.Thus the second component of our Extended Gompertz Function model is linear growth. The model predicts that, just like endemic influenza, Covid-19 waves consist of alternating phases of Gompertz Function growth and linear growth.

**Using the Extended Gompertz Function Model to make predictions** When, as in the case of cumulative Spanish Flu deaths in Prussia, observations are very well approximated by a line, the 'line of best fit' to the data is determined by linear regression. This method for finding the line, an error minimising process, uses a simple formula that has been known since the end of the 18th Century. In the linear growth phase, we can make predictions simply by finding the line of best fit and extrapolating it forward.

Finding *curves* that fit the data, rather than lines, is also done by minimising errors. But this time

[6]See Section 4 of Predicting the course of Covid-19 and other epidemic and endemic disease.

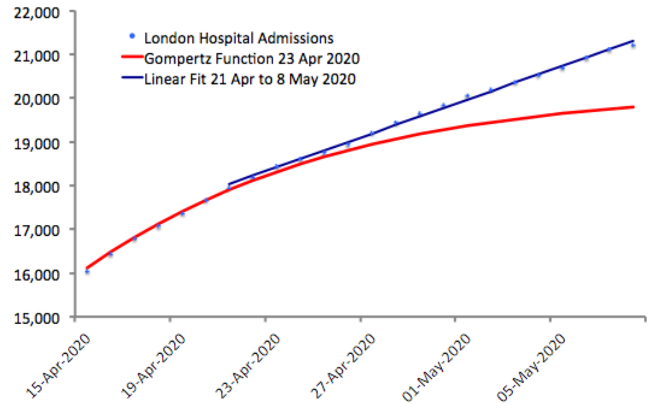



Figure 5: Covid-19 Hospital Admissions in London, transition to linear growth Apr 2020.

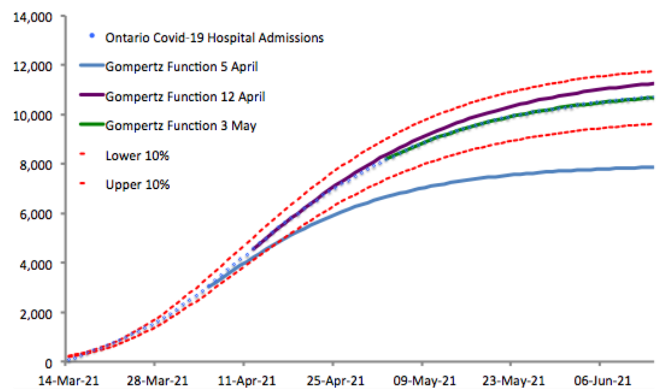



Figure 6: Successive Gompertz Function Fits for Ontario Hospital Admissions.

there's no formula. Instead 'fitting the curve' is done by searching through the set of curve parameters in order to find the minimiser. (See Box: Galileo, Basketball on the Moon and Nonlinear Regression for an example of fitting a parabola.)

This is non-linear regression–a simple task for a computer which can be done by numerous off the shelf programs. That's how we find the Gompertz Function that best fits the cumulative event data in all of the examples shown here and in our paper.

As time progresses, and we have more data to use in making the fit, the Gompertz Function changes, gradually converging to the one we find when all the data is in. Each time we find a new Gompertz Function we use it to make predictions of the future by extrapolating forward in time along that curve. Figure 6 illustrates how successive applications of this process produce predictions which become increasingly accurate.

The two phases of our model are very different and the transition between them is a matter of observation rather than prediction. Obviously it's advantageous to be able identify the transition early–especially the onset of a new Gompertz Function growth phase.

For example, in November 2021, Covid-19 hospital admissions in England were growing linearly. Despite the panic over Omicron, our observation of this phase predicted, correctly, that we would see only about half as many admissions as in the previous winter's outbreak.

We detected the transition to Gompertz Function growth on 6 December. It continued until 7 February 2022 when the next linear phase began. On 6 March 2022 the Omicron variant BA.2 began driving yet another Gompertz Function phase.

## 2 The model allows us to make inferences about epidemics

The great utility of mathematical models is that they can provide insights beyond their original purpose. Our goal was to predict the course of Covid-19–but the model can do much more than that. The fact

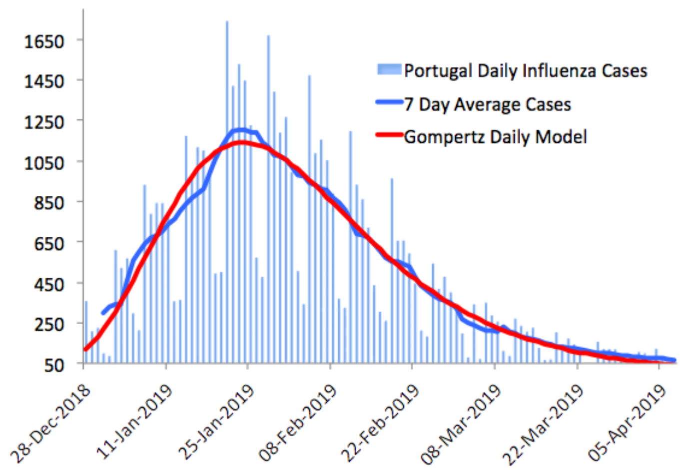



Figure 7: Influenza Cases in Portugal 28 Dec 2018–7 Apr 2019 with Gompertz Function Daily model.

that cumulative events data follow the model so closely means we should expect that the epidemic process evolves in the same way the model process does. Important features of the model should correspond to important features of the disease so we can use the model to make inferences about epidemics.

For example, linear growth is the hallmark of endemic disease where each infected person, on average, infects only one other so the Reproduction Number $R$ is equal to 1. In this case there will be the same number of infections next week as there were this week and total infections will grow linearly. Conversely, if infections are growing linearly, the same number must be occurring in each week–so $R$ must be equal to 1. The same conclusions follow, approximately, with growth being approximately linear if and only if $R$ is approximately equal to 1.

The fact that Covid-19 growth changed to the linear phase after the initial outbreaks indicates that it became endemic at that point–exactly as influenza does each year when the current viral variant fades out. And, as with influenza, new variants emerge in a series of waves of Covid-19.

Similarly, we can make inferences about the epidemic during its Gompertz Function growth phase. Once we reach the end of this phase the fits have converged to the ones that we've illustrated in the examples. The only way this can happen is if, day by day through the epidemic, the way the data evolves is approximated very well by the way the final Gompertz Function evolves in tracing out its path.

So the speed of epidemic growth must match that of the Gompertz Function. While this is rapid at first, a Gompertz Function *never* grows exponentially and this means that epidemics don't either.[7]

Because the Gompertz Function describes the daily cumulative events, the sequence of differences between its values on consecutive days is a model for daily events, which we call the Gompertz Function Daily Model. Daily data is typically very 'noisy', as Figure 7 shows, due to the effect of weekends, holidays and other extraneous factors that are not representative of the disease process. The Daily model is a bonus. It gives us a smooth 'de-noised' version of daily events. In particular we have a systematic way of determining the date on which those events peaked.[8]That is a critical part of our analysis of China's lockdowns.

**Reassessing China's lockdown** The only possible test of efficacy of a lockdown is whether it caused a decline in infections. If lockdowns (or any other interventions) occurred *after* infections had already peaked they cannot have been the cause of the decline. It follows that they could not have been the cause of a subsequent decline in cases, hospital admissions, deaths and so on.

Simon Wood has combined studies estimating the time from infection to the appearance of symptoms of Covid-19 with studies estimating the time from the onset of symptoms to death in fatal cases. This gives an estimate (as a probability distribution) of the time between infection and death.[9] He used

[7]See Predicting the course of Covid-19 and other epidemic and endemic disease, Section 5.1 The 'Exponential Growth' Phase.

[8]The difference between Gompertz Function values on successive days is an extremely good approximation to the Gompertz Function's derivative–so the peak in daily differences occurs at the same time as the peak speed.

[9]Simon Wood (2021) Inferring UK COVID-19 fatal infection trajectories from daily mortality data: Were infections already in decline before the UK lockdowns? Biometrics `https://onlinelibrary.wiley.com/doi/10.1111/biom.13462`

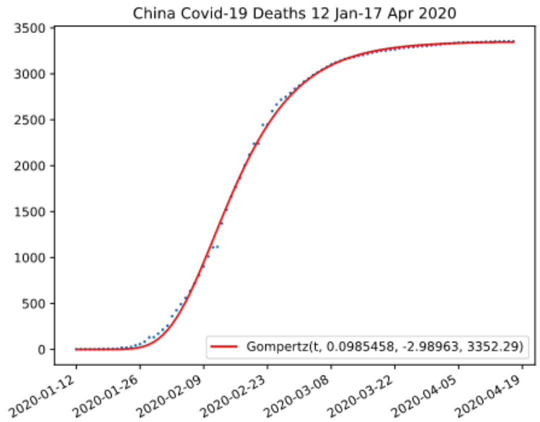



Figure 8: Covid-19 Deaths in China 12 Jan to 17 Apr 2020. Data source WHO

this approach to show that the English lockdowns almost certainly came too late to have influenced the decline in infections. Therefore they were *not* effective and no analysis of costs and benefits is necessary in England. There were no benefits.

Our model for Covid-19 also provides a model for infections, from which we can determine the time at which they must have peaked.[10] As a result, we can use our model to provide independent confirmation of Wood's conclusions (which we will discuss further in Part 2 of this article). Most importantly, our model for infections allows us to refute the WHO's claim that China's lockdown policy was what stopped the initial Covid-19 outbreak.

**China's lockdown came after infections had already peaked** From the initial report on 12 January 2020 until 17 April 2020, Figure 8 shows that Covid-19 deaths data followed a Gompertz Function to an extremely good approximation.[11]

The day on which the Gompertz Function model for daily deaths peaked was 12 February 2020. It is overwhelmingly likely that the time from infection to death is at least 22 days–so infections peaked no later than 21 January 2020–two days *before* the Wuhan train station and airports were closed. 'Non-essential' businesses were not closed until 13 February.

The lockdown was *not* what caused the decline in deaths on the graphs that Bruce Aylward exhibited in the 24 February press conference.The Guardian was half correct. The Chinese measures that the WHO endorsed and recommended to the world (and which China continues to employ) were brutal but almost certainly not effective.The same conclusion is reached again and again as we examine the timing of lockdowns and the likely peak of infections they were supposed to be controlling.

## 3 What happens next?

From now on no government anywhere should ever, in the face of new Covid-19 waves, a severe influenza season or the next viral pandemic, be uncertain about the demands the disease will put on their hospital system.This is now completely predictable.

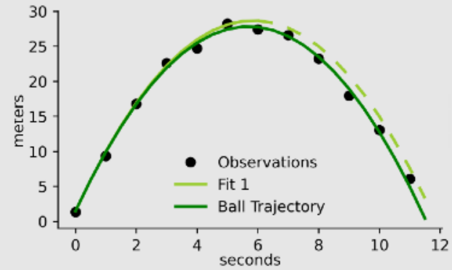
We have been predicting Covid-19 outbreaks in real time since November 2021.The Extended Gompertz Function Model accurately predicted English hospital admissions during the Omicron wave (also correctly predicting that it would be half the size of the previous winter's outbreak). Most recently, it identified a post Omicron outbreak in England at the beginning of March this year which again is evolving exactly as predicted.

Our model is in the public domain and is easily implemented. It eliminates the uncertainty about health care demands that has driven extreme measures in country after country.

[10]This is because if one process, $X$ follows a Gompertz Function then another process $Y$ whose events are a fixed proportion of the ones for $X$, perhaps beginning at a different time, must also follow a Gompertz Function that is completely determined by the one for $X$, the proportion and the time shift.See Section 3.2 of Predicting the course of Covid-19 and other epidemic and endemic disease. While we don't know the proportion of infections that result in deaths precisely, we don't need it to find the time at which infections peaked. For that we only need the time shift, in other words the time from infection to death and the Gompertz Function model for deaths.

[11]Covid-19 deaths data from China became unusable on 18 April when a single data entry increased the total deaths by almost 40%.

***Galileo: Basketball on the Moon and Nonlinear Regression*** A basketball shot on the moon, where there is no air resistance, would follow a perfect parabola. This is a curve $s(t) = \alpha + \beta t + \gamma t^2$ that depends on three constants $\alpha$, $\beta$ and $\gamma$. If we make observations of the height of the centre of the ball above the ground every second then a plot of those points would, aside from measurement errors, lie on the parabola. The force due to gravity on the moon is only about one sixth of what it is on Earth. So the ball will travel for about 6 times as long as it would on Earth giving us time to make a series of observations.




Using non-linear regression, a computer program can rapidly search over a huge range of values for the 3 parameters of the parabola to find the numbers $\alpha$, $\beta$ and $\gamma$ that minimise the average of the squared errors between the observation and the curve. In the figure, Fit 1 represents the result of a non-linear regression using the first 7 observations. The dashed line, extrapolating along the parabola determined by that fit, is the subsequent prediction. Units of height are meters and units of time are seconds. The ball reaches a height of just under 28 meters and travels for about 12 seconds.

It also removes any reason for believing that lockdown policies (or other interventions) ever worked anywhere (as we'll show in many examples in Part 2.) In all future Covid-19 waves, in severe influenza outbreaks and in any future viral pandemics, this model can be used to deal with the disease in conjunction with the public health practices that have always succeeded in the past.

**Acknowledgement** B.A. Shadwick kindly provided the Python program we have used to fit the Gompertz Functions shown in this article and in our paper.

**Part 2 of Predicting the Unpredictable *Covid-19 Interventions:The Illusion of Control*, will continue our analysis of the efficacy of lockdowns and other interventions.**