GAVIN PUBLISHERS

Archives of Epidemiology

McCauley JL. Arch Epidemiol 4: 138.

DOI: 10.29011/2577-2252.100038

Brief Report

Covid-19 how Complexity Defeats Modeling Illustrated by Social Distancing vs. Recoveries in the Daily Infection Rates

Joseph L McCauley*

Department of physics, University of Houston, USA

***Corresponding author:** Joseph L McCauley, Department of physics, University of Houston, USA

Citation: McCauley JL (2020) Covid-19 how Complexity Defeats Modeling illustrated by Social Distancing vs. Recoveries in the Daily Infection Rates. Arch Epidemiol 4: 138. DOI: 10.29011/2577-2252.100038

Received Date: 28 April, 2020; **Accepted Date:** 15 May, 2020; **Published Date:** 21 May, 2020

Abstract

Epidemic and pandemic modelling is generally based on estimates of unknown and unknowable parameters and therefore fails systematically. The failures of model predictions of the covid-19 pandemic are often reported in the news media. We present as the useful alternative a very simple data-based forecast of future infections and deaths: we need only the active infection and death data for today and yesterday. Instead of a government grant and massive computer time, we need only a hand calculator to calculate logarithms. We use the daily growth rates to extract both observed and predicted doubling times for the covid pandemic for five countries. The observed doubling time allows us to compare the effectiveness of the lock downs in the different countries. The predicted doubling time allows us to make a data-based forecast. The daily growth rates change in unforeseeable steps and are algorithmically complex: we can only know the rates as they unfold; modelling them in any useful way is impossible. We apply standard chemical kinetic rate equations in discrete time of one day intervals to quantify the effect of social distancing vs. recoveries in the daily data. Flattening and peaking are precisely defined. Social distancing can cause flattening, but recoveries are required in order that the active infections to peak and decay. Two models and their predictions are also analyzed.

The Observed and Predicted Doubling Times

The infection and death rates present an example of complexity: the infection rates can be observed daily, but they cannot be modelled and predicted in any useful way in advance of their occurrence. E.g., averaging past rates cannot tell us future rates. Any attempt to fix the infection and death rates in advance in order to make predictions from a model will always be defeated by the daily unfolding. You can watch the unfolding, you can take measures to reduce infections, but you cannot predict when active infections will flatten or peak, or if they will merely flatten without peaking. We state the conditions for flattening and peaking in part 2. This is not about finding a formula or computer simulation to replace a discrete set of points with daily changing chords (rates) by a smooth curve. Such an effort would distract us from our job: nature has given us the data. Our job is to understand the data. That means that no arbitrary parameters will be introduced. The only quantities discussed below are determined daily by the data.

In what follows

N=index labeling day n

P(n)=uninfected population on day n (not immune)

I(n)=total # active infections on day n

R(n)=# recoveries on day n (immune)

D(n)=total # deaths on day n

4/3/20 means April 3, 2020 or 3 April 2020.

Infections are reported on several daily data websites like worldometer [1]. There are total infections IT where

$$I_T=I+D+R \qquad (1)$$

We will focus on active infections I(n), the source of new infections, assuming that dead and recovered patients cannot transmit the disease (to zeroth order, at least). In an epidemic the doubling time N_n (the number of days required for active infections to double starting from day n), $I(n+N_n)=2I(n)$, is one way to characterize exponential growth. The doubling time may fall between two days so generally not an integer. If the daily infection rate r would be constant

$$I(n+1)=rI(n) \quad (2)$$

then after n days

$$I(n)=r^nI(1) \quad (3)$$

So r>1 is necessary for an exponential infection rate. There is a bifurcation from growth to decay at r=1; if r<1 then the contagion dies out exponentially in the population. The doubling time would be N=ln2/lnr for r>0 but the daily infection rate seen in the data, while deterministic, is not constant. We are not concerned here how accurate are the data for a given country, which is beyond our control. The better data, the better our understanding. Accuracy of the data, it suffices to say that errors in reporting will be magnified at the rate λ=lnr if r>1 and contracted at the rate λ=lnr if r<1. Error reporting in data matters a lot during growth of an epidemic but matters less as the epidemic dies out. Instead of applying a preconceived (and therefore wrong) model we will ask what can the data teach us. We will use two qualitatively related methods to extract useful information from the data. The methods apply to any epidemic or pandemic. The first method is to read off the observed doubling time. The second method is to use the predicted doubling time. The latter can be used to make a "forecast", a prediction of the future that goes beyond what we can know today.

If r_n varies from day to day then when is the rate from day 1 to day n exponentially increasing? We can write

$$I(n)=r_nr_{n-1}\ldots r_1I(1) \quad (3b)$$

Where the effective daily rate is

$$r_{en}=[r_1\ldots r_n]^{1/n}. \quad (3c)$$

As simple mathematical but non-data examples, if r=2 then the infections would double daily. If r=1+ε/n then the infections would plateau as n increases. Many mathematical models can be made but we are only interested in understanding the data, not in model building. The effective rate is not the rate on any given day, it's simply a number that directly connects the states of active infections on days 1 and n,

$$I(n)=r_{en}{}^nI(1). \quad (3d)$$

In other words, the states are path-independent: any two states I(0) and I(n) are connected globally by a single exponential (3d) with a constant growth rate r_{en}. Furthermore, the sequence of steps doesn't matter, I(3)/I(1)=r1r2=r2r1, which has implications for understanding flattening. If for each day we have $r_n>1$ then $r_{en}>1$ and the infections (3d) are exponentially increasing. But this condition is only sufficient and is not necessary: if $r_k<1$ for some values of k=1,…,n but the overall product (3c) is greater than unity then the process (3d) is exponentially increasing through day n. Infections may decay, as in Austria after 4/3/20 (Figure 1), or they may oscillate, as for Finland after 4/1/20 (Figure 2). Any variation in daily slope above, equal to, or below unity is consistent with eqns (2) and (3b).

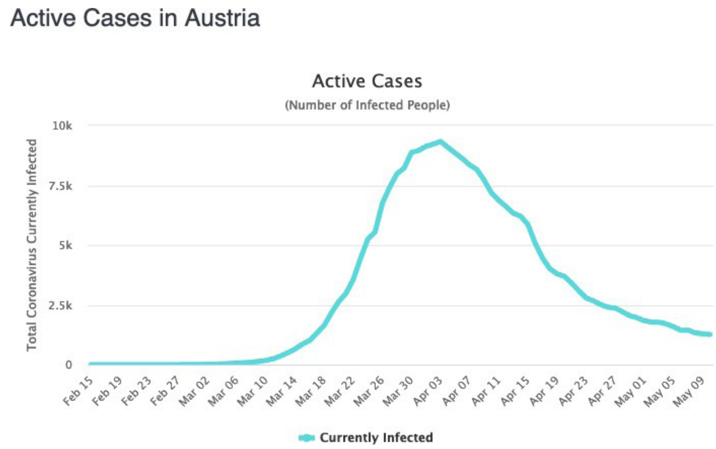

Figure 1: Austria peaked on 4/3/20 and reopened small businesses on 4/14/20.

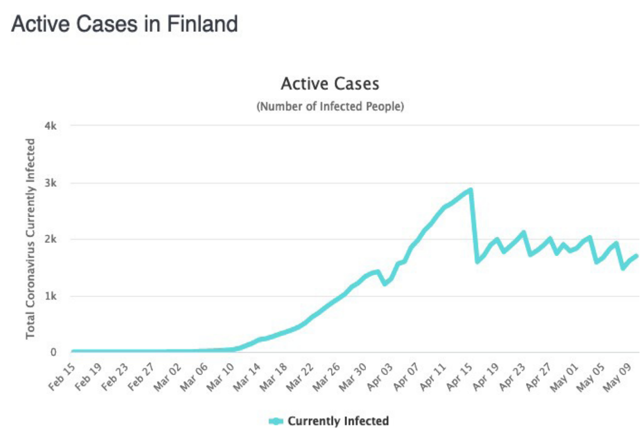

Figure 2: Finland's oscillating active infections after peaking on 4/1/20.

This is described by daily exponential behavior. Any two points on the graph are connected by a global exponential with a single effective rate. The exact doubling time starting from day n is then $N_n=n\ln2/\ln(r_1\ldots r_n)$. We can also define the predicted doubling time $T_n=\ln2/\ln r_n$ starting from day n where $r_n=I(n+1)/I(n)$. This will turn out to be a decent forecast of the future if. The daily rates do not deviate too much from r_n over the next T_n days. As a forecast, it's the best one possible and it's the only forecast that we will consider or advocate. The predicted doubling time provides a much better forecast of the future than do models based on memory or computer simulations (see part 5). We can similarly write a predicted doubling time for the death rate, and the two rates are not the same. Deaths lag infections by about a month.

The observed doubling time N_n is read directly from the data with a sliding window. One starts at day n in the data and slides the window forward until $I(n+N_n)=2I(n))$. This method is limited because when you reach day n where N_n is larger than the number of days left in the data set then you must stop. This is a useful method so long as you're not at that limit. The resulting effective infection rate is then $r_{ne}=2^{1/Nn}$ and depends on n. We illustrate this method in table 1. E.g., from 3/27 to 3/31 in the U.S. $r\approx 2^{(2/11)}$ with a doubling time of 5.5 days. From 3/2 to 3/27 the doubling time increased in jumps from 2 to 5 days. One can also start at the last data point and read backward until the infections halve, but we will leave this to the reader. When in a decay of infections, r<1 then we can speak of the halving time $N=-\ln 2/\ln r$, satisfying $I(n)=I(0)/2$. Before the 3/16/20 lockdowns in Austria, Germany and US the respective doubling times respectively were 2-3 days, 2.5-3 days, and 2.5 days for the three doubling intervals immediately preceding lock down (roughly 3/7-3/16). Here are the results that I have read from the data (Table 1).

Day n	Austria	Germany	Italy	UK	USA
3/16	3	2.5	6	3	2.5
3/17	3	2.5	6.5	3	1.5
3/18	3.5	4	7.5	3.5	2
3/19	3.5	5.5	8	4	2.5
3/20	4	6.5	11	4	2.5
3/21	4.5	6	14.5	4	2.5
3/22	4.5	6	15	4	3
3/23	7	9	20	3.5	3.5
3/24	>11	9	25	3.5	3.5
3/25	Peaked 4/3	9.5	Peaked 4/21	4.5	3.5
3/26		27		5	4.5
3/27		Peaked 4/6		5	5
3/28				5.5	5.5
3/29				5.5	5.5
3/30				6	5.5
3/31				6	6.5
4/1				7.5	8

Table 1: Observed doubling times for five countries.

Germany, Austria and US locked down on 3/16/20, Italy on 3/9/20, and UK on 3/23/20. Where a doubling time fell between 2 days we labeled it as half a day. Before lock down the doubling time was 2-3 days. The doubling time reflects the effectiveness (or lack of same) of a lock down. The US and UK lock downs clearly were relatively ineffective.

Next are the one dimensional discrete point sets from which the data were taken. In the worldometer link [1] you can see the numbers by using your computer mouse to place the cursor on each data point in a curve on their website. The date and number then appear. The doubling times systematically increase as social distancing takes effect. Early in the pandemic, social distancing is the only factor that can reduce the infection rate. Recoveries come into play later (Figure 3-6).

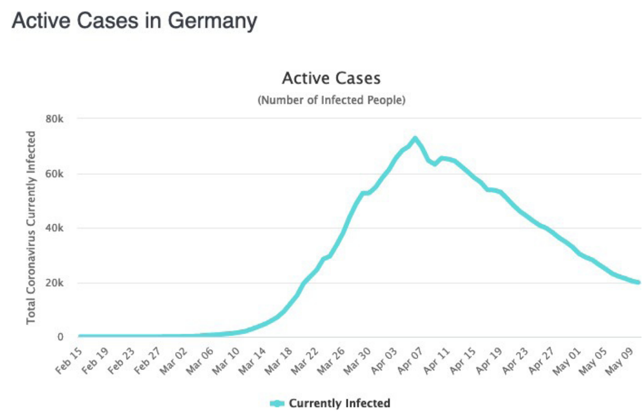

Figure 3: Germany peaked on 4/6/20 and partly reopened in April. The borders with Austria remained closed except for work commuting.

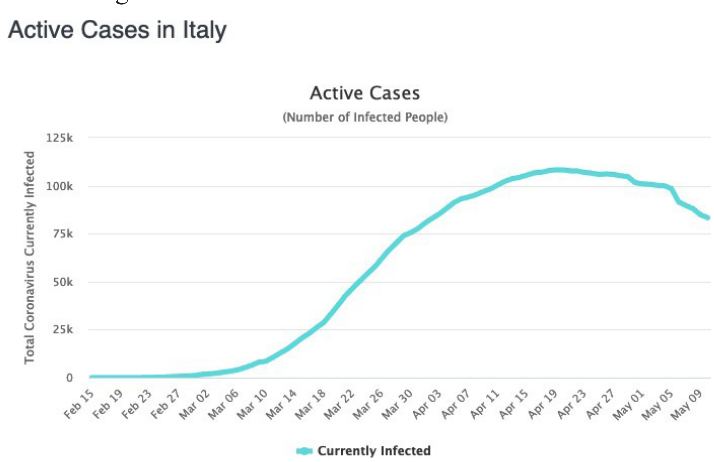

Figure 4: Italy flattened for 3 weeks without peaking, then peaked on 4/20/20 with $r\approx 1.004$.

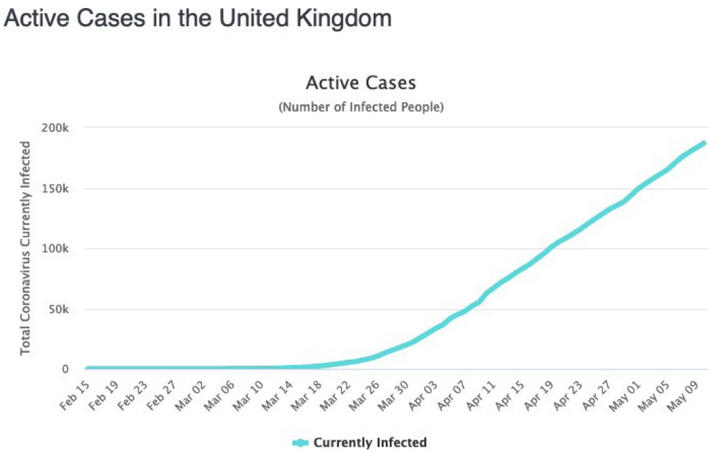

Figure 5: UK, like the US, shows increasing active infections through 5/10/20.

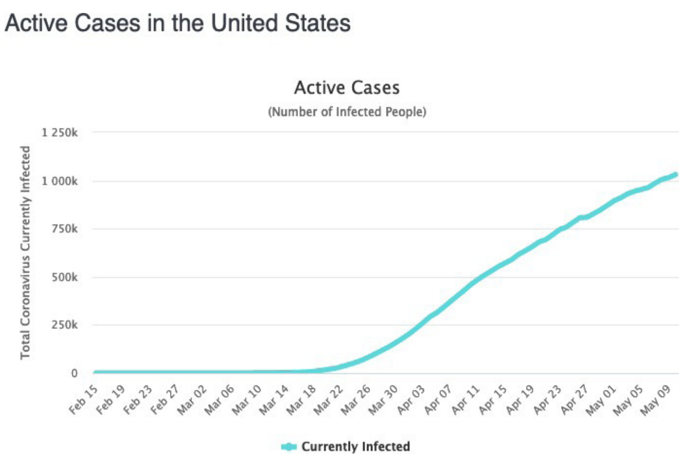

Figure 6: Over a month after Austria and Germany peaked the US still has not peaked, in part due to nonuniform social distancing compliance.

Finally, we compare the observed doubling time N_n with the daily predicted doubling time T_p=ln2/lnr_n but leaving off the subscript n. The latter will provide a fairly decent prediction if. The daily rate r_n doesn't vary very much over the doubling time T_p that is predicted. The point of the next table is not to show that the predicted doubling time is a good forecast but rather to show that it should be taken with a grain of salt. It is, in any case, the best that we can do until the data unfold. The next table illustrates our ability to forecast (Table 2).

Day n	UK T_n	UK N_n	US T_n	US N_n
3/16	3	3	2.5	2.5
3/17	2	3	1.5	1.5
3/18	1.5	3.5	2	2
3/19	3.5	4	2	2.5
3/20	3	4	3	2.5
3/21	6	4	2.5	2.5
3/22	4.5	4	3	3
3/23	3.5	3.5	3	3.5
3/24	4	3.5	3	3.5
3/25	3.5	4.5	3	3.5
3/26	3	5	3.5	4.5
3/27	4.5	5	4	5
3/28	5.5	5.5	5	5.5
3/29	5.5	5.5	5	5.5
3/30	6	6	5.5	5.5
3/31	5	7	5.5	6.5
4/1	4.6	7.5	5.5	7
The time N_n is exact to within half a day, but when we approach the last date for which data is available then T_n is the only estimate at our disposal.				

Table 2: Shows the observed doubling time N_n vs. the predicted doubling time T_n calculated from the daily infection rate for the UK and US.

There is no guarantee that a long predicted doubling time will be reached. Using the unrealistic but simple plateauing model where $r_n=1+\varepsilon/n$ we get $T_n=\ln2/(\ln(1+\varepsilon/n)\approx n\ln2/\varepsilon>>n$ for n large and ε small. Real data might peak and decay before such a limit is reached. This is to illustrate that the predicted doubling time is merely the best and only possible data-based forecast, it need not be what we actually observe after a time T_n. E.g., the day before Germany peaked r_n=1.04 and T_n=16 days. There was nothing in the active infections to suggest that Germany would peak on 4/6/20 and then decay on the next day, although I am sure that people who use technical analysis in the stock market could also find equally meaningless patterns in the covid data.

Discrete kinetics of Exponential Growth and Decay

Any process that grows or dies like $I(n+1)/I(n)=r(n)$ is exponential. We use the standard method of chemical kinetics [2] with the observed daily varying rate coefficients to describe how a growth process can flatten and may or may not jump from r>1 to r<1 on one day and then decay. The process is time-reversible: the map $I(n+1)=rnI(n)=$ with r given by the data can be iterated either forward or backward in time n with the unique inverse $I(n)=r-1I(n+1)$. Time reversibility is the signature of determinism in dynamics. Randomness (noise) at each time step would erase the past, as in the stock market. Pure exponential growth is not only time reversible, the daily rates with 0<r<∞ form an Abelian group under multiplication. Reversible dynamics doesn't mean you can make the physical system run backward, it means that you can run the mathematical equations backward in time (e.g., the earth-sun Kepler problem is time- reversible).

Total infections include active infections, recoveries and deaths $I_T=I+R+D$ so $\Delta R=\Delta I_T-\Delta I-\Delta D$. The world has $P_T\approx 7.6$ billion people with 81 million births/yr., so because the world population change is relatively small over a few weeks or a month we can take $P_T\approx$ constant=P+I+R. This is a number conservation law,

$$\Delta P+\Delta I+\Delta R=0 \quad (4)$$

Where $\Delta P<0$ is the daily change in the uninfected, not-immune part of the population. We can apply (4) to one country so long as deaths are small compared with recoveries. If deaths would rise to the order of magnitude as recoveries then the required conservation law would be

$$\Delta P+\Delta I+\Delta R-\Delta D=0 \quad (4b)$$

An infection peak occurs on day n when, on day n-1, we find $-\Delta P>\Delta R$ and then on day n we find $-\Delta P<\Delta R$. Social distancing and vaccines reduce ΔP while recoveries are increased by a good health care system or a good immune system, or both. We will see the competing effects of both terms in the graphs. Next, consider standard chemical reaction kinetics using [2].

Normally, chemical kinetics is formulated with differential equations because the reaction processes are fast on the time scale of one second. Our processes change on a time scale of the order of a day so that our kinetic equations are discrete. With concentrations of two species n1, n2 with reaction rate k then $n_1+n_2=n=$constant so that $\Delta n_1+\Delta n_2=0$. Assuming independence then $\Delta n_1=kn_1n_2=-\Delta n_1$. With three different populations it's similar unless one population doesn't react but simply grows directly from either population, say population 2. In that case we have the number (or mass) conservation law

$$n_1+n_2+n_3=n=\text{constant} \quad (5)$$

So that

$$\Delta n_1+\Delta n_2+\Delta n_3=0. \quad (6)$$

If populations 1 and 2 react while

$$\Delta n_3=mn_2 \quad (7)$$

Then

$$\Delta n_1=-\,kn_1n_2=-\Delta n_2-\Delta n_3 \quad (8)$$

and so

$$\Delta n_2=kn_1n_2-mn_2. \quad (9)$$

In the chemical kinetics of binding of ions in an electrolyte [2] to form neutral atoms (or for the creation and annihilation of quantized vortices in a thin superfluid film [3]) then we would have $n=n_1=n_2$ while $\Delta n_3=-\Delta n=kn^2-pn_3$, where k is the association/binding rate while p is the dissociation/unbinding rate.

In our population notation defined above we have

$$\Delta P=P(n+1)-P(n)=-bP(n)I(n)$$

$$\Delta I=I(n+1)-I(n)=[bP(n)-g]I \quad (10)$$

$$\Delta R=R(n+1)-R(n)=gI(n).$$

This is called an SIR model in the literature [4] but this should not be regarded as a model: it's simply the standard set of kinetic equations that any technically competent chemistry or physics graduate student should be able write down. A nonequilibrium thermodynamic application in continuous time to weak electrolytes is given in ref [2].

We can extract the rate r=1+bP-g daily. We obtain the recovery coefficient

$$g=(\Delta I_T-\Delta D)/I-(r-1) \quad (11)$$

from (4). If we know the data for days n-1 and n then we can describe that transition. Were the rate coefficients constants fixed by thermal equilibrium (or by any equilibrium or steady state) then the equations would have predictive power. But, unlike chemical kinetics, there is no equilibrium condition; the rate coefficients may jump daily. Any discrete time SIR or SEIR model is immediately falsified by the data if one takes the rates to be constant. But feeding the observed rates into the kinetic equations allows us to organize and understand the data. Continuous time SIR and SEIR models are useless mathematical exercises. The iterated map (10) using daily rate data is the only method that sheds light on infection growth. The smaller is r-1, the easier will a small daily jump in recoveries in (11) make r<1.

The variable rate coefficient bP reflects the effect of social distancing (or lack of same) while g is the variable recovery rate. A reduction in bP early in the data (before recoveries become significant) tells that social distancing is working. The condition for peaking is $b_{n-1}P>g_{n-1}$ with $b_nP<gn$. The condition to be near a possible peak is $r\approx 1$, but being near the peak doesn't imply crossing it: we still need a jump that gives us g>bP. You can't get over the hump without feeding enough immune people into the population. Without recoveries (without immunity developed) the population simply dies out very, very slowly with r approaching unity from above. So a small positive growth rate r can lead either to (i) a continued epidemic with a long doubling time (r>1 with r-1 small, flattening without peaking), or (ii) a sudden jump on a single day from r>1 to exponential decay (r<1). Early in a contagion the rate r is only reduced by social distancing. The daily rate coefficient separates these two important effects. As Onsager once commented, a good theory helps you to understand the data.

On that count we have a good theory.

A contagion with a long doubling time can be approximated over shorter times by linear contagion growth. If

$r^n=(1+\varepsilon)^n$ with $\varepsilon<<1$ (12)

then

$I(n)/I(0)=r(n)=(1+\varepsilon)n=e^{n\ln(1+\varepsilon)}\approx 1+n\ln(1+\varepsilon)\approx 1+n\varepsilon$ (13)

This approximation can be seen in the data for a number of days $\Delta n<<N$. Approximately linear data over a few days do not reflect flattening, they merely reflect a longer doubling time. By flattening I mean a definite slope break with sudden change to a smaller slope. This can only be seen after it has occurred, flattening is not predictable. The oft-read news statement 'the data seem to be flattening' or 'the data are expected to flatten' is meaningless.

Flattening and/or Peaking

We will use the daily growth rate r to obtain the predicted daily doubling time T=ln2/lnr for the most recent data. It's instructive to see how the peaks were approached in Austria and Germany (Table 3).

Day (n)	r_n	Predicted T_n(days)
3/25	1.22	3.5
3/26	1.1	7.2
3/27	1.08	9
3/29	1.03	
3/30	1.08	
3/31	1.01	69
4/1	1.02	
4/2	1.01	69
4/3	.98 (peaked)	
4/4	0.97	

Table 3: Austria's daily growth rates r=I(n+1)/I(n) read by a 2-window sliding window before and after the peak on 4/3/20. Predicted doubling times each day are T= ln2/lnr. r<1 is the condition for exponential decay. We can then speak of a 'halving-time'.

The daily rate predicts a doubling time $T_n=\ln 2/\ln r_n$. But that doubling time will be reflected later in the data only if the data should become approximately linear on a time scale that is large compared with the doubling time. The data for Austria approached the peak linearly and peaked on 4/3/20. A rough estimate for 3/12/20 gives $bP\approx r-1\approx .34$ when $g\approx 0$. But while $P_n\approx P$ is constant b changes with n due to social distancing. We next use the daily data on to find

$bp=(\Delta I_T-\Delta D)/1$ (14)

and

$g=bp-(r-1)$ (15)

On 4/2/20 we find bP=.025 and r-1=.01 or g=.015. On the next day g>bP and r<0.

Austria was under a uniform and very effective 'lockdown' after the borders were closed midnight 3/16/20. One wore a mask to grocery shop or visit a pharmacy, travel beyond the village was forbidden, distance between pedestrians had to be maintained; in order to go for a walk only household members were allowed to come along. All sports were banned (hiking, biking, etc). The total lock down was very effective. As of April 14 small businesses were allowed to reopen. Biking and hiking was permitted, but only alone or with household members. Masks must still be worn in stores. During the lock down one could not drive outside the local region without a note from an M.D. stating that travel is necessary for health reasons, or some other necessary reason. As of 5/8/20 the border with Germany was still closed except for commuting back and forth to work. According to the data Austria had a most effective lockdown: it peaked first and has had the fastest recovery (infection decay rate). Therefore we should pay attention to the effect of lock down for that case. The initial doubling time on 2/29 was N=3 days with I=10 active infections. Austria peaked 44 days later on 4/3/20 with I=9334. Recoveries have little effect on r until distancing reduces the rate bP, and bP remains roughly constant without social distancing. So without social distancing we could have expected I=9334 active infections only 29 days later on 3/29/20 (instead of the I=8223 reported), and on 4/3/20 we should have seen I=29702, three times the number reported under lock down. Austria had a model lock down.

Lockdowns in both Italy and Germany (less effective but imposed at the same time) reduced the doubling rates in those countries significantly. Germany and Austria closed their borders and locked down at midnight 3/16/20. Italy only locked down very late on 3/9/20 when the number of infections were already 23000 compared with 7000 in Germany (Italy's population is ¾ Germany's). The US and UK hesitated, both first considered no lockdown at all, dilly-dallied, and then locked down half-heartedly. The English PM who (like the US President) had first considered a free market solution (let the virus run with no lock down) caught the virus and was released on 4/9/20 from ICU (Table 4).

Day (n)	r_n	Predicted T_n (days)
4/2	1.07	10.2
4/3	1.04	
4/4	1.03	23.3
4/5	1.04	18
4/6	.96 (peaked)	

Table 4: Germany's daily infection rates near the peak on 4/6/20.

Germany peaked on 4/6/20. We get bP=.34 and r-1=.04 from the early data near 3/16/20. From 4/3/20 to 4/5/20 bP=.04 while r-1=.04 and g≈0, so social distancing brought Germany to the peak, but the transition to r<0 on 4/6/20 was caused by recoveries.

Day (n)	r_n	Predicted T_n (days)
4/5	1.03	23.3
4/6	1.02	
4/7	1.01	69
4/8	1.02	

Table 5: The daily infection rate in Italy is stuck, and is approximately linear with T large.

by to try to see why Italy did not peak 4/6/20 we compare bP with g for recent data. On 3/6 when g≈0 we get bP≈r-1=.28. On 4/6/20 we find bP=.04 and g=.03. There has been no jump to r<1. Instead, Italy is in a long, slow increase in infections with T=69 days (Table 5). The daily rate is probably stuck at r≈.02 because the recovery rate coefficient g is a bit too small. Social distancing has been very strict in Italy, but the doctors and hospitals were overloaded and had to turn people away to die.

For the UK we get from 3/16/20 that bP=r-1≈.27 (perhaps reflecting ineffective social distancing) but on 4/9/20 g≈.001. On 4/9/20 the UK is in the pandemic with a doubling time of 9 days with no peak in sight (Table 6).

Day (n)	r_n	Predicted T_n (days)
4/5	1.08	9
4/6	1.06	
4/7	1.09	
4/8	1.06	
4/9	1.13	

Table 6: The unfavorable daily infection rate in the UK.

For the US, on 3/16 we have bP≈r-1= .39 while on 4/9/20 bP≈.14 and g≈.004. Social distancing has been too weak and the recovery rate, as in the UK, is much too low for significant flattening, much less for a peak, to be in sight (Table 7).

Day (n)	r_n	Predicted T_n (days)
4/4	1.07	10
4/5	1.09	8
4/6	1.085	
4/7	1.07	10
4/8	1.07	
4/9	1.07	10
4/10		

Table 7: The unfavorable daily infection rate in the U.S.A.

We looked further at the approximately linear data for 4/10-13 to see we could extract systematic changes in g and bP to extrapolate and predict a peak or crossover, but that didn't work. Scandinavia presents a very good case for lock down study because the five Scandinavian countries are culturally very similar with similar health systems and governments, even if Finland (and Estonia) has a non-Nordic ethnic origin. Iceland, Finland and Denmark all peaked before or by mid-April whereas the active infections were still growing in Norway and Sweden on 4/27/20. Sweden did not lock down. The death rate has been 15/100, twice as high as in the US. Finland oscillated with large magnitude about r≈1 from 4/16/20 through 5/10/20. It's clear that strong regulations imposed early have worked very well in reducing the pandemic in Austria and Germany. it's equally clear that waiting too late and/or imposing half-measures half-heartedly does not work well (UK, USA).

What about the data provided for China? (Figure 7).

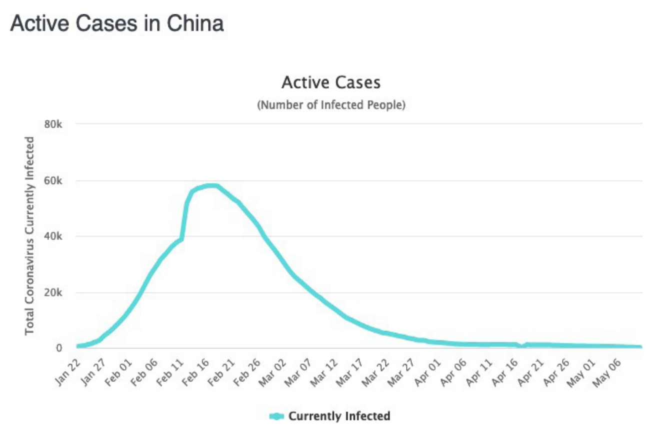

Figure 7: Data from China.

If correct, then China went rapidly from nearly infinite slope to a very smooth, flat transition across r=1. I wonder if the data were generated by a computer program (see Figure 1 in [5]). The day before the peak r=.1.001, ten times smaller than a rate from the German or Austrian data. Those data peaked 1 ½ months later than China and show a very sharp transition with a big slope break from r>1 to r<1 with r=1.02-1.04. Italy showed r≈1.008 before peaking smoothly, and the worldometer data for Iran also shows a smooth flattening with r≈1.02. Skepticism persists because we've learned from documentary TV programs that local officials in China are under enormous pressure to provide political and economic results that Beijing expects. We also know that there was the melamine milk scandal in 2008 [6]. The data provided by Iran to worldometer shows 4000 recoveries on a single day, from 4/10/20 to 4/11/20. How did that work?.

Comments on Model Calculations

There are very many papers by groups attempting to model the infection rate in a way that will convince other readers and grant agencies. We focus here only on two, the first because it's mentioned so often in the US news, and the second because it's easy to understand and use the model: as with this paper, a hand calculator along with the daily data are enough. We do not have access to the IHME model [7] whose predictions have been reported and revised all too often in the news media. That model predicted on 4/20/20 that US deaths would peak on 4/21/20 and then decay. Instead, on 4/20/20 D=42853 and on 4/26/20 D=55413 with a predicted doubling time T=34 days. Neither peaking nor even flattening were indicated by the data when the prediction was made. Long range predictions are impossible, but this did not prevent IHME from predicting that D=67640 will remain constant from 7/4/20 onward. The reality is that D=55413 on 4/26/20 with T≈146 days. There is no point in providing grant money for computer simulations when better predictions can be made with a hand calculator.

Consider next a simple model where all calculations can be made by hand. The model [8] replaces the observed daily rate by an average of daily rates over the preceding m+1 days. The authors state that we need to keep the infection growth rate below 5% (r-1<.05) in order to plateau. But the data do not reflect 'a highly nonlinear process' as is claimed [8], we've seen that the data reflect an iterated linear map. The model ansatz is based on storing m+1 states of initial data as memory

$$x_{n+1}/x_n = (x_{n-1}/x_{n-2}+ x_{n-2}/x_{n-3} +\ldots+ x_{n-m}/x_{n-m-1})/m, \qquad (16)$$

where the rhs replaces the daily rate $r_n=I_{n+1}/In$ by the quantity

$$R_n= (x_{n-1}/x_{n-2}+ x_{n-2}/x_{n-3} +\ldots+ x_{n-m}/x_{n-m-1})/m= x_{n+1}/x_n \qquad (17)$$

So that we are then studying the terms $x_{n+1}/x_n =R_n$ rather than the observed infection rate data $I_{n+1}/In=r_n$. Rather than the single initial condition required to iterate the map (2) one needs m+1 initial conditions to iterate the map (16). The initial conditions may or may not be taken from real data. Even if the m+1 initial conditions are taken from real data then the iterates for n≥m+1 will not be the real data. In any case, can we find a peak and decay in the model's future iterates? This has nothing to do with data analysis or epidemic prediction but it is an interesting mathematical game. With

$$x_{n+2}/x_{n+1}=(mR_n+x_n/x_{n-1}-x_{n-m}/x_{n-m-1})/m \qquad (18)$$

a negative term appears. In the next iteration

$$x_{n+3}/x_{n+2}=(mR_n+x_{n+1}/x_n+x_n/x_{n-1}-x_{n+1-m}/x_{n-m}-x_{n-m}/x_{n-m-1})/m \qquad (19)$$

We pick up two negative terms, and so on. The iterate n+N has N-1 differences in the rate. But there is no such relationship between a past average and future rate jumps in real infection data. The difference/error in r-R will be magnified exponentially as the maps (2) and (17) are iterated. If we use as initial data the eight infection rates from Germany for 3/25/20 through 4/1/20, where the data are only five days short of peaking on 4/6/20, then we find from (18) and (19) that the predicted infection rates pass through the date 4/6/20 at 1.05 and then increase to 1.06 on 4/7. In the real data r=1.04 on 4/25/20 but then r=.96<1 on 4/6/20. In the model, the infections continue to increase without peaking. Germany and Austria peaked sharply without flattening. Real data may also flatten without peaking. Flattening should not be imagined as a precursor to peaking.

If the m+1 initial conditions are chosen carefully so that the successive slopes are decreasing (as in Figure 1 of [8] which looks nothing like the worldometer pandemic data) then the earlier negative terms may eventually win over the later positive ones and R_n, causing peaking and decay. But since the real, observed data do not obey (16) this ansatz unfortunately does not inform us about epidemics. We do not know any correct rule where the observed rate rn is given by any combination of preceding rates r_{n-2}, r_{n-3}, … . The rule (17) does not generate the covid data, so the authors' 5% rule [8] follows from choosing initial conditions carefully in the numerical game, not from an epidemic data analysis. Presumably, one must choose initial data with r<1.05 in order to flatten and peak in the model. But past initial conditions do not cause flattening or peaking of a real epidemic, only social distancing and recoveries can do that. There is a 1-2 week time lag between becoming infected and showing illness, but that is not the 1-2 week time lag in the model's memory above. There is no way aside from a direct corona virus test to know who is infected until the person becomes ill.

Summarizing, one can choose initial conditions in the model that are favorable for peaking and decay. In an epidemic we have no control over initial conditions, we must deal with bad initial conditions that evolve exponentially with r≈1.2-1.3 and then find a way to bring down the infection rate. Eqn. (2) describes an epidemic and is deterministic and Markovian: I(n+1) depends only on I(n) and not on the past history of active infections. This is realistic. Unlike (16), we have on hand no mathematical rule that tells us how the rates r_n evolve. The infection rates are algorithmically complex [9]: the shortest computer program that can generate any n daily rates is simply to write down the n observed rates. It's obvious that there can be no simple hidden rule to tell us the rates because the change in r depends on how people respond to social distancing, and we don't know that in advance (simply compare Austria, Germany, US, and Finland, all of whom locked down to different degrees).

Summary

Infection rates under complete or partial lock down are not predictable in advance, they can only be discovered as they unfold daily. Modelling the rates artificially does not lead either to insight or predictability. The reality of the unfolding reflects the complexity of the underlying process that determines the coefficients in the kinetic rate equations. The kinetic equations are simple but the process that determines the rates is complex. Definitions of complexity are discussed in ch. 11 of ref. [9]. If we know the infections from today and yesterday then we can use eqn. (2) to estimate the infections tomorrow and a few days in advance if the predicted doubling time is large compared with one day. This is the only correct forecast based on real, up to date data. More complicated models are not only unnecessary, they are not useful.

We have quantified the effect of social distancing, we can extract that effect from the daily data. We have defined flattening mathematically precisely. Peaking and decay of infections only possible if the rate at which immune people are fed back into the population is greater than the rate at which not-immune people become infected; both rates are extracted from the daily data by using number conservation for the populations in the kinetic equations. The implications for public health are simple: first, we see that strong lockdowns work very well while half-hearted ones do not. Recoveries are aided by a strong health system. Second, epidemic model building should not be funded because models that try to guess the infection rates cannot predict anything of use for official policy, and are even misleading.

Acknowledgement

Thanks to Cornelia Küffner, Dyt Schiller, Kevin Bassler, Bernhard Meister, Royce Zia for very helpful comments and criticism. This work was done from late March through late April in Tirol during the Austrian lock down.

References

1. Italy, Coronavirus cases. Worldometer.
2. Onsager L (2004) Deviations from Ohm's Law in Weak Electrolytes. J Chem Phys 2: 599.
3. McCauley JL, Allen CW (1982) Interactions of vortices with boundaries in superfluid films. Phys Rev 25: 5680.
4. A discrete SIR infectious disease model.
5. Chen YC, Lu PE, Chang CS, Liu TH (2020) A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons.
6. 2008 Chinese milk scandal.
7. COVID-19 Projections, United States of America.
8. Perc M, Miksić NG, Slavinec M, Stozer A (2020) Forecasting COVID-19. Frontiers in Physics 8: 127.
9. McCauley JL (2009) Dynamics of Markets: The New Financial Economics. ch. 11, Cambridge University.