

# Calculating the curves: from theory to practice

A W H Tsui

I ‘met’ my fellow prospective college freshers on Messenger about a month ago. I suppose that the light-hearted chatter about baking, Netflix, and raspberry gin provides an escape from the tedium of lockdown. Even so, it would be a lie to claim that we are not worried – the occasional exchange of distressed voice notes is evidence enough. Cambridge has cancelled all 2020-21 lectures. The freshers’ fair was (predictably) the next to yield. *What next*, we ask. *What next?*

Answering the ‘what next?’ question is not always straightforward, particularly in epidemiology. If we were to consider *everything* about *every* person in the community – where they go, who they meet, their physiology, etc. – it would in principle be possible to predict every transmission route, every infection, every recovery. This is, obviously, an impossibly complex task for any computer. Nevertheless, we can make good predictions of such *chaotic systems*, if the number of individuals is sufficiently large.

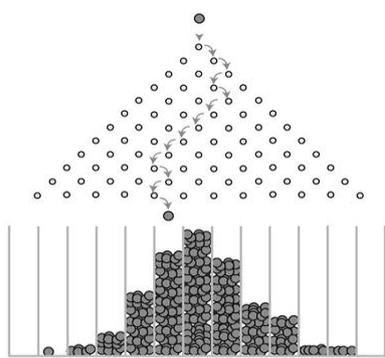


Fig. 1. The ‘Pascal’s triangle’ fairground game.

Consider, for example, this popular fairground game (Fig. 1). I start with releasing a single ball from the top. Unless I measure everything (angle and height of release, mass and radius of ball, etc.) ridiculously accurately, I have no way of consistently predicting where this one ball will land. However, a pattern emerges when I repeat this with many balls. When the ball meets a peg, the probability of it going left is roughly equal to it going right; therefore, we can reasonably predict the frequency that one ball will end up in a certain pocket. (For the aspiring mathematicians, this probability distribution is governed by the famous *Pascal’s triangle*.)

We have established that we can get a good idea of the behaviour of chaotic systems, **given that the number of trials is large enough**. To simplify the situation, every individual in our community could be considered as an independent<sup>1</sup> ‘trial’ of the system, as they follow the same rules of probability that underlie the epidemic. You might have heard of the Imperial<sup>2</sup> and Oxford<sup>3</sup> models: both mathematically elegant and sound, but offer different conclusions. The Imperial model is an *individual-based* model which, using our fairground game analogy, attempts to predict the outcome of every single ball; the Oxford model is a mathematically more simple method which considers groups of people, hence a *population-based* model. My approach more resembles the Oxford method in accounting for the population as a whole – not all of us have a supercomputer in our room!

## Where to begin

Dr. John Cullerne’s earlier Plague Pit publication<sup>4</sup> introduces the **SIR** model (Fig. 2). He visualises three ‘reservoirs’, corresponding to the *susceptible*, *infected*, and *recovered* people respectively. We also assume, like the Oxford model, that all people who recover become immune<sup>5</sup>. The progression of the epidemic will be determined by the **rate of flow** between the reservoirs.

<sup>1</sup> The assumption of ‘independence’ here might seem a little shaky, as people meeting up is likely not to be independent. However, as said, if there are enough people in the community and enough ‘meetings’, this assumption is fairly reasonable.

<sup>2</sup> Ferguson NM, Laydon D, Nedjati-Gilani G et al. *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. Imperial College London. March 2020.

<sup>3</sup> Lourenço J, Paton R, Ghafari M et al. *Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic*. University of Oxford. March 2020.

<sup>4</sup> Cullerne JP. *What is it to model an epidemic?* The Plague Pit. April 2020. <https://plaguepit.com/modelling/>

<sup>5</sup> This assumption seems reasonable enough. There is, however, not enough evidence to show that people can indeed become immune forever after infection! This uncertainty is one of the reasons why a coronavirus vaccine has never been developed successfully. This may be a subject for future discussion.

There will be some mathematical notation, although used rather sparingly. The maths is meant to elaborate on Dr Cullerne's article. For the readers who shudder at the sight of an  $x$ , you may consider skipping to wherever you are most comfortable. Don't panic!

For simplicity's sake, we have also assumed an unchanging population – a good assumption, considering that birth and death rates are roughly the same, and movement across the border has been frozen due to the global outbreaks.

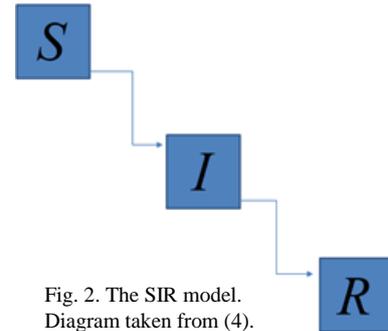


Fig. 2. The SIR model. Diagram taken from (4).

The rate at which  $S$  flows into  $I$ , i.e. the *infection rate*, will be dependent on both the number of susceptibles and the number of infectives. It would be reasonable to assume that the chance of an infection happening would be higher if a) there are more people to infect, and b) there are more people who can infect! As this is a one-way system, we can express the rate of change of  $S$  over time as follows:

$$\frac{dS}{dt} = -\beta SI \quad (\text{i})$$

The  $d$  essentially indicates a difference, so  $\frac{dS}{dt}$  is the change in the number of susceptibles over time.  $\beta$  is just some numerical constant. The rate is negative because it represents a *loss* of susceptibles.

The rate at which  $I$  flows into  $R$ , i.e. the *recovery rate*, will be dependent on the number of infectives. Surely more people would recover if there were more people who *can* recover! Following the assumption that all recovered people stay immune, i.e. another one-way system, we can write the following expression:

$$\frac{dR}{dt} = \alpha I \quad (\text{ii})$$

As above,  $\frac{dR}{dt}$  is the change in the number of recovered people over time,  $\alpha$  is another constant, and the rate is positive because the amount of recovered people can only increase (there is no way back!).

$I$  receives people from  $S$  and loses people to  $R$ . Logically, the rate at which  $S$  loses people to  $I$  must equal the rate at which  $I$  gains people from  $S$ , and the same can be said for  $I$  and  $R$ . Therefore, this statement should not be controversial:

$$\frac{dI}{dt} = \beta SI - \alpha I \quad (\text{iii})$$

If the number of infectives increases, then  $\frac{dI}{dt}$  at that moment in time must be positive. If we have a negative  $\frac{dI}{dt}$ , the number of infectives drop, signalling that the epidemic is past its peak.

Something interesting happens when we manipulate (iii):

$$\frac{dI}{dt} = (\beta S - \alpha)I$$

For  $\frac{dI}{dt}$  to be positive,  $(\beta S - \alpha)$  must be greater than 0. The *threshold* condition, is therefore  $\frac{\beta S}{\alpha} = 1$ . Above that, the epidemic soars; below that, the epidemic dies down. If we consider what this means:

- $\beta$  is the infection rate, or *the probability of infection per unit time*;
- $\frac{1}{\alpha}$  is actually the mean lifetime of an infective (i.e. how long it takes for an infective to recover);
- $S$  is just the number of susceptibles.

Therefore  $\frac{\beta S}{\alpha}$  is essentially ‘probability of infection per time  $\times$  time  $\times$  number’, i.e. **the expected number of new infections in that time!** This is the famous  $R_0$  that you will have heard about: the average number of people an infective can spread the disease to. The subscript  $0$  indicates that the value here is for the beginning of the epidemic at  $t = 0$ .

If we consider this at the very start of an epidemic ( $t = 0$ ), something subtle emerges. As shown, for an epidemic to take off,  $\frac{\beta S}{\alpha}$  must be larger than 1. Therefore, for an epidemic to even start in the first place, **enough susceptibles must be present at the start** (that threshold is of course  $\frac{\alpha}{\beta}$ !) This is one of the targets of social distancing: by creating physical barriers between groups of people, the effective susceptible population in an area is reduced, reducing the chance of an outbreak starting.

We can now calculate a variety of epidemic curves, based on these three equations (i-iii). (I view  $R_0$  as merely a useful by-product of the mathematics, and not what we base our models on.) These curves can then be fitted to existing data. The ‘training wheels’ of our modelling journeys have often been the data from the 1666 Black Death outbreak in Eyam, which features in Dr. Cullerne’s article. (I also reference his original Eyam paper here<sup>6</sup>.) How will our model fare against the COVID-19 outbreaks?

### The Pearl passes the test

Armed with the experience of SARS (severe acute respiratory syndrome; caused by SARS-CoV) in 2003, Hongkongers were quick to react when word of the Wuhan outbreak got out. The medical system was already alerted in early January, despite Chinese officials claiming that there was ‘no evidence’ of human-to-human transmission. Every citizen, remembering the costly battle with its predecessor, was wearing a facemask by Chinese New Year<sup>7</sup>; schools were already closed before then, fearing the worst. On 25<sup>th</sup> January, three days after the first imported case from Wuhan, the government declared the highest level of emergency. The decisive measures taken by the government and hospitals, combined with the discipline of citizens, were able to suppress the epidemic effectively; the first wave<sup>8</sup> was stalled at only 73 active (imported and local cases combined) cases at its peak<sup>9</sup>.

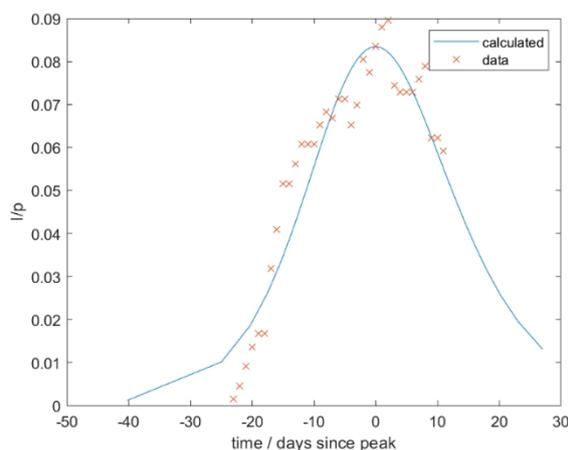


Fig. 3. The curve of best fit (blue) to the Hong Kong first wave data (orange). Note that a) the time axis is shifted so that  $t = 0$  at the chosen peak value, and b) the infectives on the y-axis are in units of  $\rho$ , the threshold number of susceptibles (which also turns out to be  $\alpha/\beta$ , as shown earlier).

The Hong Kong first wave data, like the Eyam data, was very well documented. Through rigorous testing, hospitals managed to track down every single case; patients were closely monitored in quarantine, so their times of recovery were reliable (they have to test negative twice!). This means that we can calculate, with confidence, the number of *active* cases on any day. I will spare you the maths, but here is a curve fitted to the Hong Kong data (Fig. 3). Note that I have actually subtracted 14 after calculating the number of active cases for every day, as the first 14 cases were all imports from Wuhan<sup>10</sup>, and so not governed by the mathematical rules of the local

<sup>6</sup> French A, Kanchanasakdichai O, Cullerne JP. *The pedagogical power of context: iterative calculus methods and the epidemiology of Eyam*. Phys. Educ. 2019.

<sup>7</sup> Taken from conversations over the phone with family and friends who were in Hong Kong at the time.

<sup>8</sup> The ‘second wave’ from March to May consisted almost only of imported cases, mainly from the UK and US.

<sup>9</sup> <https://www.statista.com/statistics/1105425/hong-kong-novel-coronavirus-covid19-confirmed-death-recovered-trend/>

N.B. Note that the above data is cumulative. The number of *active* cases on any day must be calculated using the recoveries.

<sup>10</sup> <https://chp-dashboard.geodata.gov.hk/covid-19/en.html>

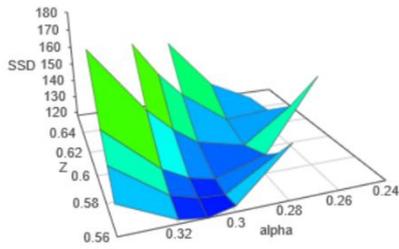


Fig. 4. The search for the combination of epidemic constants (here  $Z$ , the proportion of susceptibles who succumb, and  $\alpha$ ) that minimises the *sum of squared differences* (SSD) between the calculated and recorded time-values. SSD is a measure of how closely a curve fits the data points. The lowest point on this surface, i.e. when SSD is a minimum, corresponds to  $Z = 0.56$  and  $\alpha = 0.31$ .

transmission. (Afterwards, the number of imported cases was very small when compared to the local cases as the borders closed.)

A family of curves were calculated, each based on different peak values and epidemic constants; out of all these curves, we pick the one that is closest to the real data points (Fig. 4). This best-fit curve corresponds to an  $R_0$  of about 1.5, a threshold susceptible population\* of 659, and a peak value of 55 active local cases. It is worth noting that our model relies on choosing a peak value for the epidemic; if the data we are using has not peaked yet, then we will have to guess the position of the peak, which is a very unreliable approach.

\* Recall that the threshold susceptible population is  $\frac{\alpha}{\beta}$ , as  $\frac{\beta S}{\alpha} > 1$  is required for the epidemic to take off.

### What next?

But despite seeming a decent fit on paper, exactly how good is our curve? One thing that the SIR model does not account for is the *incubation period*. In the case of COVID-19, symptoms manifest on average 5-6 days after exposure to the virus, but this period can be anywhere from 2 to 14 days<sup>11</sup>. One way to solve this is to add another compartment to our S-I-R flow:

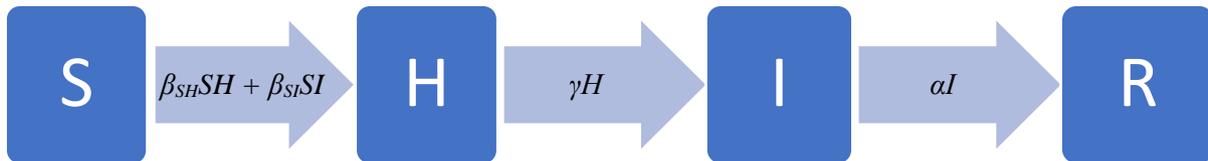
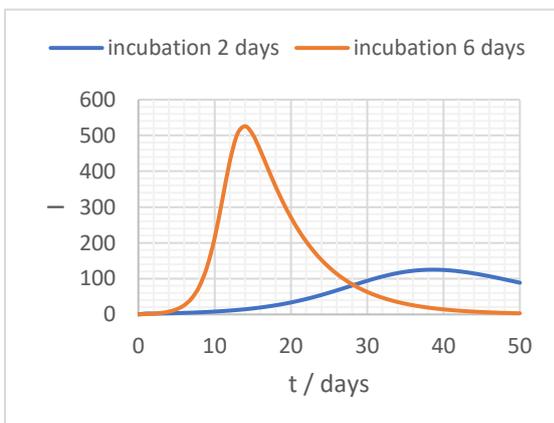


Fig. 5. A flow diagram for the SHIR model – essentially the SIR but with an extra ‘hidden’ compartment, indicated by H.

Here I have inserted the compartment H for ‘hidden infectives’ (Fig. 5). There are now two  $\beta$ s for the two ways a susceptible can be infected (by an invisible or symptomatic infective). Assuming a fixed probability for the decay of H into I, we introduce another constant  $\gamma$  (the probability of decay;  $1/\gamma$  is the mean time to decay) to govern this process. We can also expect that  $\beta_{SH} > \beta_{SI}$ , as asymptomatic carriers are still infective, and they might infect more people on average than symptomatic patients (who, upon suspecting that they are infected, will isolate themselves!). The difference  $\gamma$  makes is huge:



Here  $S_0$  is set at 1000,  $\alpha$  at 0.03,  $\beta_{SH}$  at  $3 \times 10^{-4}$ ,  $\beta_{SI}$  at  $3 \times 10^{-5}$ . Note that  $\gamma$  is the reciprocal of the incubation time.

The incubation time can also be interpreted as the time to detection. With widespread community testing, this time will drop, as carriers are identified and isolated quicker; the result is that the infectives curve is ‘flattened’ (i.e. the blue curve in Fig. 5). If the hidden infectives are left to spread the disease among the population, the curve is not flattened (i.e. the orange curve in Fig. 5).

These attempts to ‘flatten the curve’ are important so the medical system is not overloaded around the peak – the nightmare scenario that New York had to face,

<sup>11</sup> Coronavirus disease 2019 (COVID-19) situation report 73. World Health Organisation. 2 April 2020.

when even the morgues could not cope with the drastic increase in deaths. Fortunately, the NHS seems to be doing rather well thanks to the forced lockdown, but remember that social distancing only works if everyone does it!

Answering the question of ‘what next?’ in epidemiology would be to predict the progression of the epidemic, and use that to influence our plans in the near future. However, hindsight is equally as important, if not more; the curve and its associated epidemic constants provide us with a foundation from which we can build our analyses. It is also worth noting that the Hong Kong data is unnatural, due to a) the exceptionally quick identification and isolation of infectives, and b) the effective barrier to transmission that widespread mask-wearing and good hand hygiene provides.

The logical extension to this work is to compute the best-fit curves for other countries; in doing so, we can collect and compare their epidemic constants ( $R_0$  etc.). This might prove useful in hindsight as we review the approaches of different countries, and attempt to find things in common between them that correlate with successful control. It will, however, not be as straightforward as it seems; the lack of recovery data in the UK will present more mathematical challenges.

So what *is* next? Time, and the collection of more (hopefully reliable) data, will tell. However, we can be sure of one thing: the war is not over yet, and like the First World War, may not be truly over ‘by Christmas’. Until then, cover your face, stay socially distanced, and I’ll see you on the other side.