# Mathematical Modelling of Recurrent Epidemics

### David J. D. Earn[†]

One of the most famous examples of an epidemic of an infectious disease in a human population is the Great Plague of London, which took place in 1665–1666. We know quite a lot about the progression of the Great Plague because weekly bills of mortality from that time have been retained. A photograph of such a bill is shown in Figure 1. Note that the report indicates that the number of deaths from plague (5533) was more than 37 times the number of births (146) in the week in question, and that wasn't the worst week! (As Fred Brauer notes in his article in this issue, an even worse plague occurred in the 14th century, but no detailed records of that epidemic are available.)



**Figure 1:** A photograph of a bill of mortality for the city of London, England, for the week of 26 September to 3 October 1665.

---

[†] **David J. D. Earn** is a professor in the Department of Mathematics & Statistics at McMaster University. His web site is http://www.math.mcmaster.ca/earn.

Putting together the weekly counts of plague deaths from all the relevant mortality bills, we can obtain the *epidemic curve* for the Great Plague, which I've plotted in the top left panel of Figure 2. The characteristic exponential rise, turnover and decline is precisely the pattern predicted by the classic susceptible-infective-recovered (SIR) model of Kermack and McKendrick [1] that I describe below (and Fred Brauer also discusses in his article). While this encourages us to think that mathematical modelling can help us understand epidemics, some detailed features of the epidemic curve are not predicted by the simple SIR model. For example, the model does not explain the jagged features in the plotted curve (and there would be many more small ups and downs if we had a record of daily rather than weekly deaths). However, with some considerable mathematical effort, these "fine details" can be accounted for by replacing the differential equations of Kermack and McKendrick with equations that include stochastic (i.e., random) processes [2]. We can then congratulate ourselves for our modelling success… until we look at more data.

The bottom left panel of Figure 2 shows weekly mortality from plague in London over a period of 70 years. The Great Plague is the rightmost (and highest) peak in the plot. You can see that on a longer timescale, there was a complex pattern of plague epidemics, including extinctions and re-emergences. This cannot be explained by the basic SIR model (even if we reformulate it using stochastic processes). The trouble is likely that we have left out a key biological fact: there is a reservoir of plague in rodents, so it can persist for years, unnoticed by humans, and then re-emerge suddenly and explosively. By including the rodents and aspects of spatial spread in a mathematical model, it has recently been possible to make sense of the pattern of 17th century plague epidemics in London [3]. Nevertheless, some debate continues as to whether all those plagues were really caused by the same pathogenic organism.

A less contentious example is given by epidemics of measles, which are definitely caused by a well-known virus that infects the respiratory tract in humans and is transmitted by airborne particles. Measles gives rise to characteristic red spots that are easily identifiable by physicians who have seen many cases, and parents are very likely to take their children to a doctor when such spots are noticed. Consequently, the majority of measles cases in developed countries end up in the office of a doctor (who, in many countries, is required to report observed measles cases to a central body). The result is that the quality of reported measles case data is unusually good, and it has therefore stimulated a lot of work in mathematical modelling of epidemics.
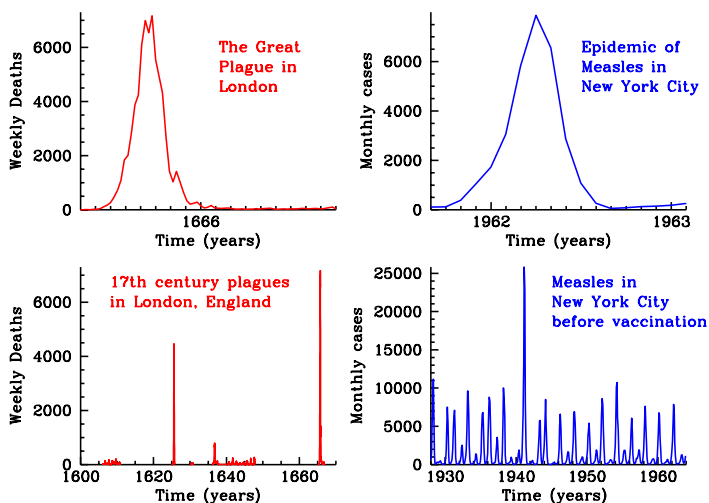
An epidemic curve for measles in New York City in 1962 is shown in the top right panel of Figure 2. The period shown is 17 months, exactly the same length of time shown for the Great Plague of London in the top left panel. The 1962 measles epidemic in New York took off more slowly and lasted longer then the Great Plague of 1665. Can mathematical models help us understand what might have caused these differences?

Using the same notation as Fred Brauer uses in his article in this issue, the basic SIR model is

$$\frac{dS}{dt} = -\beta SI, \tag{1}$$

$$\frac{dI}{dt} = \beta SI - \gamma I. \tag{2}$$

Here, $S$ and $I$ denote the numbers of individuals that are susceptible and infectious, respectively. The derivatives $dS/dt$ and $dI/dt$ denote the rates of change of $S$ and $I$ with

**Figure 2:** Epidemic curves for plague in London (left panels) and measles in New York City (right panels). For plague, the (red) curves show the number of deaths reported each week. For measles, the (blue) curves show the number of cases reported each month. In the top panels, the small ticks on the time axis occur at monthly intervals.

respect to time. The *mean transmission rate* is $\beta$ and the *mean recovery rate* is $\gamma$ (so the *mean infectious period* is $1/\gamma$). As Fred Brauer discusses, if the total population size is $N$, and everyone is initially susceptible ($S(0) = N$), then a newly introduced infected individual can be expected to infect $\mathcal{R}_0 = \beta N/\gamma$ individuals (the *basic reproduction number* $\mathcal{R}_0$ is also discussed at length by Marjorie Wonham in her article in this issue). You can find a discussion of the SIR model together with the mathematical ideas it is based on in some introductory calculus textbooks (see, for example, [4]).

As Fred Brauer notes, we cannot solve the SIR equations and obtain formulae for the functions $S(t)$ and $I(t)$. Yet the epidemic curves that we are trying to explain are essentially given by $I(t)$, so it is hard to proceed without it! Fortunately, computers come to our rescue. Rather than seeking an explicit formula for $I(t)$, we can instead obtain a numerical approximation of the solution. One simple approach is Euler's method, which we can implement as follows (using a spreadsheet or any standard programming language).

The derivative $dS/dt$ is defined as the ratio of the change in $S$ in a given short time interval $dt$, divided by that time interval, in the limit that $dt$ approaches zero. Dealing with that limit is tricky, but at any time $t$ we can approximate the derivative by writing $dS = S(t + dt) - S(t)$ and solving for the number of susceptibles at a time $t + dt$ in the future,

$$S(t + dt) = S(t) - \beta S(t)I(t)dt. \tag{3}$$

Similarly, we can approximate the number of infectives at time $t + dt$ as

$$I(t + dt) = I(t) + \beta S(t)I(t)dt - \gamma I(t)dt. \tag{4}$$

Equations (3) and (4) together provide a scheme for approximating solutions of the basic SIR model. To implement this scheme on a computer, you need to decide on a suitable small time interval $dt$. If you want to try this, I'd suggest taking $dt$ to be one tenth of a day. I should point out that I am being extremely cavalier in suggesting the above method. Do try this at home, but be forewarned that you can easily generate garbage using this simple approach if you're not careful. (To avoid potential confusion, include a line in your program

that checks that $S(t) \geq 0$ and $I(t) \geq 0$ at all times. Another important check is to repeat your calculations using a much smaller $dt$ and make sure your results don't change.)

In order for your computer to carry out the calculations specified by equations (3) and (4), you need to tell it the parameter values ($\beta$ and $\gamma$, or $\mathcal{R}_0$, $N$ and $\gamma$) and initial conditions ($S(0)$ and $I(0)$). For measles, estimates that are independent of the case report data that we're trying to explain indicate that the mean infectious period is $1/\gamma \sim 5$ days and the basic reproduction number is $\mathcal{R}_0 \sim 18$ [5]. The population of New York City in 1960 was $N = 7\,781\,984$. If we now assume one infectious individual came to New York before the epidemic of 1962 ($I(0) = 1$), and that everyone in the city was susceptible ($S(0) = N$), then we have enough information to let the computer calculate $I(t)$. Doing so yields the epidemic curve shown in the top panel of Figure 3, which does *not* look much like the real data for the 1962 epidemic in New York. So is there something wrong with our model?

No, but there is something very wrong with our initial conditions. The bottom right panel of Figure 2 shows reported measles cases in New York City for a 36 year period, the end of which includes the 1962 epidemic. Evidently, measles epidemics had been occurring in New York for decades with no sign of extinction of the virus. In late 1961, most of New York's population had already had measles and was already immune, and the epidemic certainly didn't start because one infectious individual came to the city. The assumptions that $I(0) = 1$ and $S(0) = N$ are ridiculous. If, instead, we take $I(0) = 123 \cdot (5/30)$ (the number of reported cases in September 1961 times the infectious period as a proportion of the length of the month) and $S(0) = 0.065N$, then we obtain the epidemic curve plotted in the middle panel of Figure 3, which is much more like the observed epidemic curve of Figure 2 (top right panel). This is progress—we have a model that can explain a single measles epidemic in New York City—but the model cannot explain the recurrent epidemics observed in the bottom right panel of Figure 2. This is not because we still don't have exactly the right parameter values and initial conditions: no parameter values or initial conditions lead to recurrent epidemics in this simple model. So, it would seem, there must be some essential biological mechanism that we have not included in our model. What might that be?
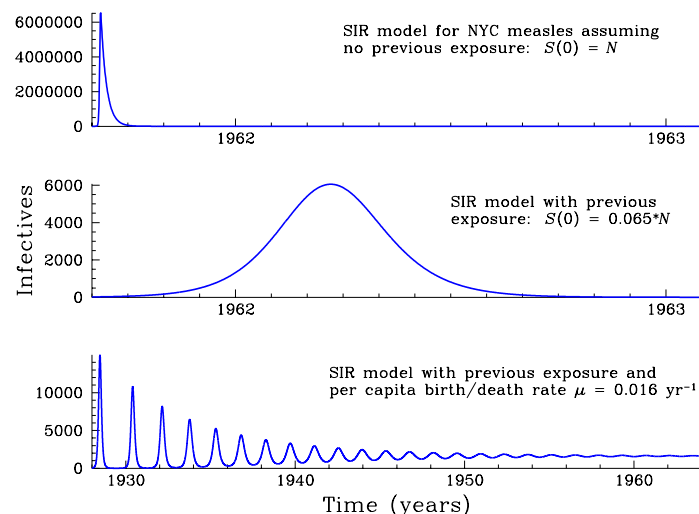


**Figure 3:** Epidemic curves for measles in New York City, generated by the basic SIR model. The curves show the number of infectives $I(t)$ at time $t$. In the top two panels, the small ticks on the time axis occur at monthly intervals.

Let's think about why a second epidemic cannot occur in the model we've discussed so far. The characteristic turnover and decline of an epidemic curve occurs because the pathogen is running out of susceptible individuals to infect. To stimulate a second epidemic, there must be a source of susceptible individuals. For measles, that source cannot be previously infected people, because recovered individuals retain lifelong immunity to the virus. Newborns typically acquire immunity from their mothers, but this wanes after a few months. So births can provide the source we're looking for.

If we expand the SIR model to include $B$ births per unit time and a natural mortality rate $\mu$ (per capita), then our equations become

$$\frac{dS}{dt} = B - \beta SI - \mu S \,, \tag{5}$$

$$\frac{dI}{dt} = \beta SI - \gamma I - \mu I \,. \tag{6}$$

The timescale for substantial changes in birth rates (decades) is generally much longer than a measles epidemic (a few months), so we'll assume that the population size is constant (thus $B = \mu N$, so there is really only one new parameter in the above equations; we can take it to be $B$). As before, we can use Euler's trick to convert the equations above into a scheme that enables a computer to generate approximate solutions. An example is shown in the bottom panel of Figure 3, where I have taken the birth rate to be $B = 126\ 372$ per year (the number of births in New York City in 1928, the first year for which we have data). The rest of the parameters and initial conditions are as in the middle panel of the figure.

Again we seem to be making progress. We are now getting recurrent epidemics, but the oscillations in the numbers of cases over time damp out, eventually reaching an equilibrium. While the graph is just an approximate solution for a single set of initial conditions, it can actually be proved that all initial conditions with $I(0) > 0$ yield solutions that converge onto this equilibrium. So we still don't have a model that can explain the real oscillations in measles incidence from 1928 to 1964, which showed no evidence of damping out. Back to the drawing board?

Don't give up. We've nearly cracked it. So far, we have been assuming implicitly that the transmission rate $\beta$ (or, equivalently, the basic reproduction number $\mathcal{R}_0$) is simply a constant and, in particular, that it does not change in time. Let's think about that assumption. The transmission rate is really the product of the rate of contact among individuals and the probability that a susceptible individual who is contacted by an infectious individual will become infected. But the contact rate is *not* constant throughout the year. To see that, consider the fact that in the absence of vaccination, the average age at which a person is infected with measles is about five years [5]; hence most susceptibles are children. Children are in closer contact when school is in session, so the transmission rate varies seasonally. A crude approximation of this seasonality is to assume that $\beta$ varies sinusoidally,

$$\beta(t) = \beta_0(1 + \alpha \cos 2\pi t) \,. \tag{7}$$

Here, $\beta_0$ is the mean transmission rate, $\alpha$ is the *amplitude* of seasonal variation and the time $t$ is assumed to be measured in years. If, as above, $\beta$ is assumed to be a periodic function (with a period of one year) then the SIR model is said to be *seasonally forced*. We can still use Euler's trick to solve the equations approximately, and I encourage you to do that using a computer for various values of the seasonal amplitude $\alpha$ (you must have $0 \leq \alpha \leq 1$: why?).

You might think that seasonal forcing is just a minor tweak of the model, but in fact this forcing has an enormous impact on the epidemic dynamics that the model predicts. If you've taken Physics and studied the forced pendulum, then you might already have some intuition for this. A pendulum with some friction will exhibit damped oscillations and settle down to an equilibrium. But if you tap the pendulum with a hammer periodically then it will never settle down and it can exhibit quite an exotic range of behaviours including chaotic dynamics [6] (oscillations that look random). Similarly complex dynamics can occur in the seasonally forced SIR model.

Most importantly, with seasonal forcing, the SIR model displays *un*damped oscillations similar to the patterns seen in the real measles case reports. But we are left with another puzzle. If you look carefully at the New York City measles reports in the bottom right panel of Figure 2 you'll see that before about 1945 the epidemics were fairly irregular, whereas after 1945 they followed an almost perfect two-year cycle. While the SIR model can generate both irregular dynamics and two-year cycles, this happens for different parameter values, not for a single solution of the equations. How can we explain *changes over time* in the *pattern* of measles epidemics?

Once again, the missing ingredient in the model is a changing parameter value. This time it is the birth rate $B$, which is not really constant. Birth rates fluctuate seasonally, but to such a small extent that this effect is negligible. What turns out to be more important is the much slower changes that occur in the average birth rate over decades. For example, in New York City the birth rate was much lower during the 1930s (the "Great Depression") than after 1945 (the "baby boom") and this difference accounts for the very different patterns of measles epidemics in New York City during these two time periods [7].

A little more analysis of the SIR model is very useful. It is possible to prove that changes in the birth rate have exactly the same effect on disease dynamics as changes of the same relative magnitude in the transmission rate or the proportion of the population that is vaccinated [7]. This equivalence makes it possible to explain historical case report data for a variety of infectious diseases in many different cities [8].

One thing that you may have picked up from this article is that successful mathematical modelling of biological systems tends to proceed in steps. We begin with the simplest sensible model and try to discover everything we can about it. If the simplest model cannot explain the phenomenon we're trying to understand, then we add more biological detail to the model, and it's best to do this in steps because we are then more likely to be able to determine which biological features have the greatest impact on the behaviour of the model.

In the particular case of mathematical epidemiology, we are lucky that medical and public health personnel have painstakingly conducted surveillance of infectious diseases for centuries. This has created an enormous wealth of valuable data with which to test hypotheses about disease spread using mathematical models, making this a very exciting subject for research in applied mathematics.

## References

[1] Kermack, W. O. and A. G. McKendrick, 1927: A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London, Series A*, **115**, 700–721.

[2] Bailey, N. T. J., 1975: *The mathematical theory of infectious diseases and its application.* Griffin, London, 2nd edition.

[3] Keeling, M. J. and C. A. Gilligan, 2000: Metapopulation dynamics of bubonic plague. *Nature*, **407**, 903–906.

[4] Hughes-Hallett D. et al, 2002: *Applied Calculus.* John Wiley & Sons, Toronto, 2nd edition.

[5] Anderson, R. A. and R. M. May, 1991: *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press, Oxford.

[6] Gleick, J., 1987: *Chaos.* Abacus, London.

[7] Earn, D. J. D., P. Rohani, B. M. Bolker and B. T. Grenfell, 2000: A simple model for complex dynamical transitions in epidemics. *Science*, **287**, 667–670.

[8] Bauch, C. T. and D. J. D. Earn, 2003: Transients and attractors in epidemics. *Proceedings of the Royal Society of London, Series B*, **270**, 1573–1578.

## Practical Further Reading Suggestions from the Editors

*The SIR Model for Spread of Disease* web site at http://www.math.duke.edu/education/ccp/materials/diffcalc/sir/contents.html includes Maple, Mathematica and Matlab files.

A spreadsheet program that is easy to use is available from http://ugrad.math.ubc.ca:8099/mathsheet/index.html.

## MATH jokes
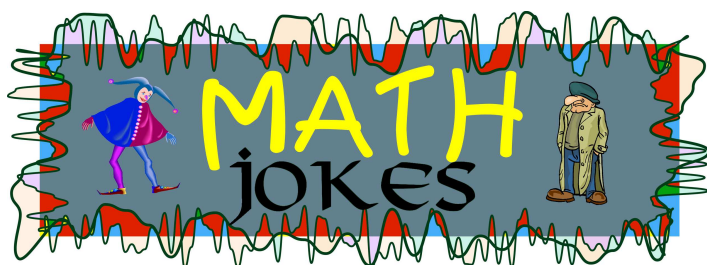
Q: What is a mathematician's pick when faced with the choice between poutine and eternal bliss in the afterlife?
A: Poutine! Because nothing is better than eternal bliss in the afterlife, and poutine is better than nothing.

Q: What is a topologist?
A: A person who cannot tell a doughnut from a coffee mug.

Q: Why did the mathematician have trouble computing $A^{-1}A$?
A: Because he was having an identity crisis.
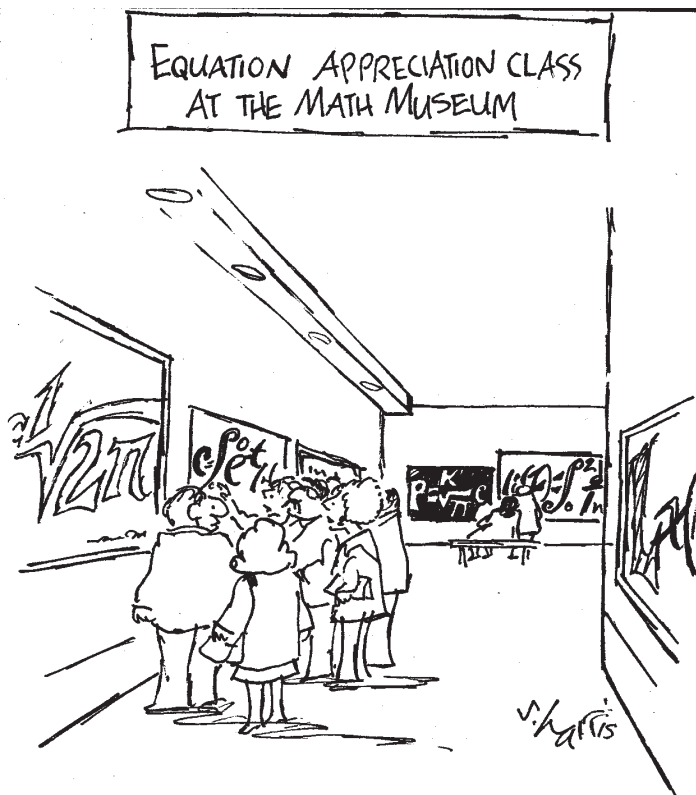
Absurdity of zero: there is no such a thing as nothing.

Q: What is yellow, sour, and equivalent to the axiom of choice?
A: Zorn's lemon....
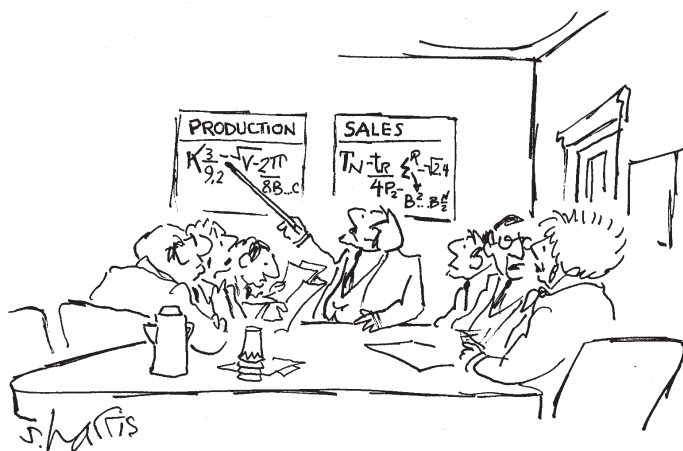


"HE NEVER LETS US FORGET THAT HE WAS A MATH MAJOR."

Q: What is normed, complete, and yellow?
A: A Bananach space....