Statistically Speaking

Managing the R0 of COVID-19: mathematics fights back

"All models are wrong but some models are useful" – George Box [1]

Perhaps for the first time in history, a single statistical measure is now dictating the entirety of the UK government policy. The 'basic reproduction number', R0 value for COVID-19 is more directly determining economic and social policy than has ever the inflation rate, interest rate or exchange rate. It is encouraging to see political policy for once 'rational' but disappointing as it took a pandemic to make it so. However, is R0 an appropriate and significant measure? Like many mathematics/statistical parameters, R0 is relatively easy to explain, more complicated to understand (even graphically), and very difficult to calculate or use for modelling. Given its significance for all our lives, it is important to understand a little of its background. This article seeks to explain the issues in a non-technical way, relegating all equations (used sparingly) to appendices.

What is R0?

R0 can be described simply as the average number of secondary infections produced when one infected individual is introduced into a susceptible host population [2]. Pedants will immediately recognise that we are currently at a stage well beyond R0; that is, long after the disease 'has been introduced' so more accurately the term is R_E, the 'effective reproduction number'. Avoiding pedantry, we will use R0, R_E and R interchangeably. If one person becomes infected (patient zero) and passes it onto two others, and they in turn each to two others, and so on, the R0 value is 2. This is an example of exponential, 'Malthusian' growth, first described in 1798 (Eqn 1). We tend to view the problem in terms of our human health but in Eqn 1, the data are viewed from the perspective viral growth, akin to an animal population.

In 1847, Verhulst noticed an obvious problem with the Malthusian model: exponential growth cannot go on indefinitely. Animals will run out of food, space, etc, and even for viruses, once they have infected everyone, they have reached a limit to their theoretical population. A different, logistic equation (Eqn 2) is required [3]. In this logistic, S-shaped growth pattern the population increases at first exponentially, but then the slope of curve (reflecting the R0 value) changes from > 1 to < 1 mid-way up the curve at what is known as the 'inflection point' (Fig. 1). Hence, in this sense the R0 value may be regarded as naturally changing in a population over time. For a virus, the plateau may not represent the entire population infected but only the proportion susceptible (see below).

However, even this logistic growth equation is not the whole story. The plateau population is never likely to be entirely static, but will vary with prevailing conditions, and May's related 'logistic map' equation (Eqn 3) better reflects this [4]. In this recursive equation, the term R not only influences the rate of early exponential rise in population, but also the degree to which the pseudo-stable population later fluctuates. As the value of R increases, the population varies increasingly cyclically or periodically around a mean value at the plateau, with the notable finding that at some critical value of R, the size fluctuates chaotically over time. It is interesting in this regard that the range of reported values for COVID-19 R0 are predicted to cause such periodic or even chaotic effects by the logistic map equation [5]. An accompanying article shows how this has implications for NHS capacity planning [6]. Figure 1 plots the three equations.

Measuring R0

There are broadly two approaches to estimating R0. One is individual level modelling (ILM) where data are collected ideally from the very start of an epidemic: the contacts of an infected individual (patient zero) are traced and tested, and this continues as the disease spreads. R0 is then the average over the number of secondary (or tertiary, etc) cases of such many diagnosed individuals. A second is the population level model (PLM) which uses the change in infected numbers within the population from one day to the next, often using adjusted cumulative models. For example, if the number of cases on day x is 1000, and new cases on day x + 1 is 1300, then R0 by this method is 1.3 [7].

Individual level modelling, which can be regarded as prospective and PLM, which might be viewed as retrospective, do not lead to the same answer, as they depend on the efficiency of contact tracing, the use of test results (and their accuracy) vs. symptoms, etc. More generally, the estimate of R0 by either method is also dependent upon, among other things (a) the average number of people an infected person is exposed to each day; (b) the probability of each exposure becoming an infection; (c) duration of infectiousness, including periods of asymptomatic infectiousness; (d)

© 2020 Association of Anaesthetists

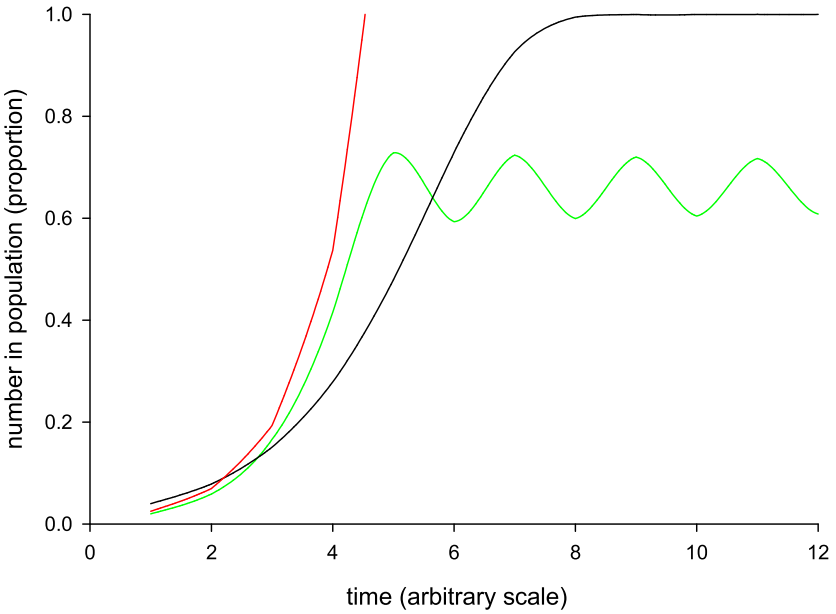

Figure 1 Three models of population (viral) growth: exponential (Eqn 1; red), logistic (Eqn 2, black) and logistic map (Eqn 3, green). Different values have been used to separate the lines. Note that whereas in exponential, R0 value is constant, it varies for the other two models depending on the time-point it is measured

population size, especially the proportion of susceptible people; and (e) rate of recovery or death. These experimental approaches, ILM vs. PLM, therefore, need to take these factors into account by combination with modelling (see below) and different models do this to different degrees in calculating R0.

Modelling R0

If models assist data analysis, then in turn, the data support modelling. A simple model is the SI model (susceptible-infected). The total population, N, consists of the sum of those susceptible (S) and infected (I). Moreover, since I transmits to S, the rate at which I increases is the inverse of the rate at which S declines. In the SI model, the rate at which the disease spreads will be proportional to S (the more susceptible people there are, the faster the disease spreads) and also to I (the more infected people there are, the faster the disease will spread), with the constraint that the sum of S and I must equal N. The transmission rate of infection (not precisely the same as R0) is the product of the proportion of susceptibles and proportion of infected (Eqn 4) [7].

However, the SI model does not take into account recovery from disease, which the SIR model does. SIR assumes that those who recover are immune, and therefore as their proportion rises, this puts a limit on spread of disease as it reduces the proportion of susceptibles. The recovery rate, by definition, increases in proportion to the rate of increase of the infected. It can therefore be shown that, in the SIR model, R0 is proportional to the ratio of the transmission rate and the recovery rate (Eqn 5). Figure 2 shows an example of an SIR modelling plot [7].

It should readily be seen that various other factors could be introduced to these SI and SIR models to complicate the equations even further. There is always a trade-off in modelling between the 'computational cost' of increasing the model's complexity, balanced against the diminishing returns on the model's accuracy in representing reality. A model that is as complicated as reality is useless as a model, regardless of accuracy. Less accurate models will be much more useful.

Confusion

The equations in the Appendix 1 reflect the fact that R0 is not a straightforward statistic: R appears as a factor in different places, in different forms of equations (although further mathematics could in fact show that each equation is logically consistent with the others). Specific methods to calculate R0 from theoretical models include: the survival function; next-generation method; Jacobean matrix eigenvalues; endemic equilibrium; metapopulation models; partial differential equation models, hierarchical Bayesian regression and constant term-polynomial methods [8]. Models take into account different numbers of factors (e.g. compare SI vs. SIR; other models take into account subgroups like age, health, etc), and they also use different types of statistics (e.g. frequentist vs. Bayesian). Therefore, estimates from these methods do not agree.

© 2020 Association of Anaesthetists

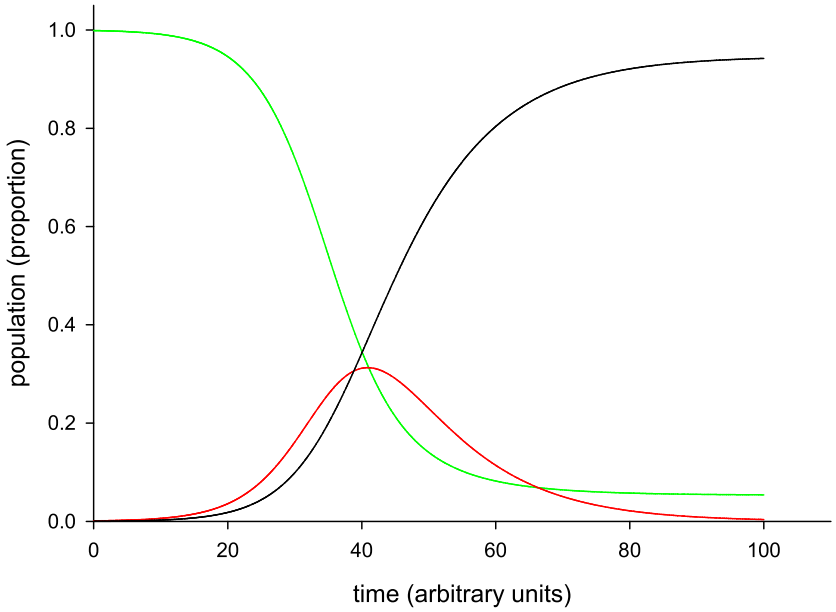

Figure 2 Example of an SIR model plot for a hypothetical infectious disease. The green curve is the susceptible population; the red curve is the infected, and the black curve is the recovered. Note the inverse relationship between the susceptible and recovered, with the infected being the balance

Moreover, all models face the problem that diseases that have R0 > 1 do not always become epidemics (and can die out), and conversely, those with R0 < 1 can persist. Even at any one time-point, R0 does not have a single value in the population, but different values for various differently susceptible subgroups or those that live in clusters or partial isolation (e.g. higher in care homes) [9].

Noting these deficiencies, most commentators nevertheless agree that R0 is all we really have. A recent text proposed that healthcare staff should march on the Department of Health with placards shouting "*Give us confidence intervals! Give us standard deviations!*" [10]. The NHS has traditionally been averse to error bars in its publications and this is also true of its presentation of R0: some indication of the variance associated with the estimates being used would be welcome. An independent group of senior scientists (led by a former Chief Scientific Advisor) has made similar criticisms to those above, of the government's presentation of R0 and argued that R0 should be presented with the confidence intervals of different estimates and caveats on how it was measured [11].

Hope

There is some consensus that offers hope across all the models. One is that, generally, diseases can be tackled effectively by reducing the R0, and this can be achieved by reducing the contacts of an infected person. Commentators may differ in emphasis, but it is agreed that a combination of social distancing, rigorous contact tracing and isolation of contacts and the vulnerable should be effective. Second, is that exponential equations work both ways. Whereas infections rates can rise at alarmingly high rates (Fig. 1), reducing the exponent only modestly will have dramatic effects in the other direction. With a starting value of 21,000, reducing R0 by just 9% from say, from 1.15 to 1.05 reduces the projected infections (e.g. after 61 days) from > 100 million to just over 400,000. Third, the proportion of the population required be immune to achieve herd immunity (whether through past infection or vaccination) is dependent upon the R0, and particularly sensitive in the range R0 1–2 (Fig. 3). If R0 can be kept to < 2, then just < 50% of the population needs to be immune [12]. In other words, a vaccine does not have to be perfectly effective to manage the disease, if combined with social distancing and other measures – it only needs to be 'good enough'. Despite the complexity and confusion, this consensus should offer some promise.

Postscripts

Thomas Malthus (1766–1834) was an English polymath – cleric, mathematician, economist and founding Fellow of the (Royal) Statistical Society – who wrote an early treatise on population growth (*An Essay on the Principle of Population*, 1798; Eqn 1). His ideas remain controversial. The notion that exponential population growth outstrips

© 2020 Association of Anaesthetists

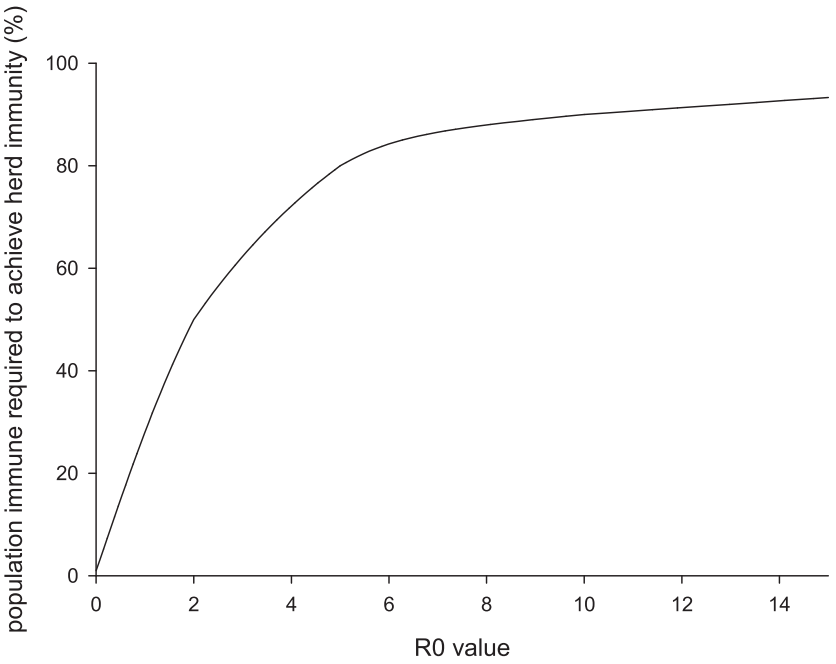

Figure 3 Estimate of the proportion of population needed to be immune to achieve herd immunity, as a function of R0. Note that for R0 < 2, the relationship is very steep with < 50% of population needing immunity,

food availability has been used to support population control (in the poor). Pierre François Verhulst (1804–1849) was a Belgian mathematician (Eqn 2) with a passionate social conscience. When visiting Rome, he was moved to submit a democratic constitution for the Papal State, and was promptly banished from the city [3]. He died young. Carl Gustav Jacob Jacobi (1804–1851), who described the 'matrix' named after him now used to model R0 in some methods, was a German pure mathematician – a giant of the subject. He had no idea of the future applications of his work. He also died young, of smallpox, in one of the many waves of epidemics that killed up to 400,000 people annually in Europe in the mid-19th century. The population biologist Robert May (1936-2020), Fellow of Merton College, was former Chief Scientific Advisor to the UK Government (1995-2000), and President of the Royal Society (2000-2005). His logistic map (Eqn 3) has been described as one of the most beautiful equations in science [13]. He died in an Oxford care home of dementia, on 28 April at the height of the COVID-19 pandemic. It is not known what he would have thought of the central role of R0 in current government policy.

Acknowledgements

JP is an elected member of the Council of the Royal College of Anaesthetists and Fellow of the Royal Statistical Society (the society founded by Malthus). The views expressed are personal and not of these organisations. No other competing interests declared.

J. J. Pandit
Consultant
Nuffield Department of Anaesthetics, Oxford University Hospitals Foundation Trust,
Oxford, UK
Email: jaideep.pandit@dpag.ox.ac.uk

References

1. Box GEP. Science and statistics. *Journal of the American Statistical Association* 1976; **71**: 791–9.
2. May RM, Anderson B. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press, 1992.
3. Cramer JS. The early origins of the logit model. *Studies in History and Philosophy of Science* 2004; **35**: 613–26.
4. May RM. Simple mathematical models with very complicated dynamics. *Nature* 1976; **261**: 459–67.
5. Alimohamadi Y, Taghdir M, Sepandi M. The estimate of the basic reproduction number for novel coronavirus disease (COVID-19): a systematic review and meta-analysis. *Journal of Preventive Medicine and Public Health* 2020; **53**: 151–7.
6. Pandit JJ. Demand-capacity modelling and Covid-19 disease: identifying themes for future NHS planning. *Anaesthesia* 2020; **75**: 1278–83.
7. Breban R, Vardavas R, Blower S. Theory versus data: how to calculate R0? *PLoS One* 2007; **2**: e282.
8. Harko T, Lobo FSN, Mak MK. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation* 2014; **236**: 184–94.

 © 2020 Association of Anaesthetists

9. Li J, Blakely D, Smith RJ. The failure of R0. *Computational and Mathematical Methods in Medicine* 2011; **2011**: 1–17.
10. Pandit JJ. *Practical Operating Theatre Management*. Cambridge: Cambridge University Press, 2019.
11. The Independent Scientific Advisory Group for Emergencies (SAGE). The Independent SAGE Report. May 2020. www.independentsage.org/wp-content/uploads/2020/05/The-Independent-SAGE-Report.pdf (accessed 11/05/2020).
12. Fine P, Eames K, Heymann DL. "Herd immunity": a rough guide. *Clinical Infectious Diseases* 2011; **52**: 911–6.
13. Farmelo G. *It Must Be Beautiful: Great Equations of Modern Science*. London: Granta, 2003.

Appendix

Equation 1: simple exponential equation where P(t) is the population at time t, P_0 is the initial population, R is the rate of growth and t is the time:

$$P(t) = P_0 e^{R.t} \quad (1)$$

Equation 2: logistic growth equation, where symbols as in Eqn 1, with K being the maximum population achievable for the given conditions:

$$P(t) = \frac{K}{1 + A.e^{-R.t}} \text{ where } A = \frac{K - P_0}{P_0} \quad (2)$$

Equation 3: logistic map equation: symbols as in previous equations, with P_{n+1} being the population at day n + 1, P_n the population at day n. The equation is recursive:

$$P_{n+1} = R.P_n(1 - P_n) \quad (3)$$

Equation 4: SI model, where β is the transmission rate, and P_1 and P_S are the proportions of susceptibles and infected at any given time:

$$\text{Rate of infection} = \beta.P_I.P_S \quad (4)$$

Equation 5: SIR model; symbols as above, R and R0 interchangeable, where γ is the recovery rate:

$$R0 = \frac{\beta}{\gamma} \quad (5)$$

© 2020 Association of Anaesthetists