Dynochem

Equations used in COVID-19 Model

The population (P) is split into groups that are susceptible (S), exposed (E), infectious (I), isolated (Is) and cured (C).

These are related by the following sum:

$$P = S + E + I + Is + C \quad (1)$$

The following ordinary differential equations track the numbers of people in each grouping other than S:

$$\frac{dE}{dt} = \beta \frac{I}{P}\frac{S}{P}P - \frac{E}{t_{inc}} = \beta \frac{IS}{P} - \frac{E}{t_{inc}} \qquad (2)$$

$$\frac{dI}{dt} = \frac{E}{t_{inc}} - \frac{I}{t_{inf}} - \frac{I}{(t_{inf} + t_{cure})} \quad (3)$$

$$\frac{dIs}{dt} = \frac{I}{t_{inf}} - \frac{Is}{t_{cure}} \quad (4)$$

$$\frac{dC}{dt} = \frac{I}{(t_{inf} + t_{cure})} + \frac{Is}{t_{cure}} \quad (5)$$

where β is the rate constant for exposure leading ultimately to infection, t_{inc} is the incubation period, t_{inf} the period during which a patient is infectious, t_{cure} the amount of time it takes to recover.

Infectious patients can recover without isolation, likely if their symptoms are minor; this is represented by the third term in equation (3). According to this term, they are assumed not to recover faster by skipping isolation.

The number of susceptible is obtained by rearranging (1):

$$S = P - (E + I + Is + C) \quad (6)$$

For the purposes of studying the effect of reduced contact between members of the population in which transmission may occur, the first term in (2) is expressed in terms of a *mixing* parameter that equals 1 for normal levels of people movement and zero when there is no movement at all:

$$\frac{dE}{dt} = \beta \frac{IS}{P} mixing - \frac{E}{t_{inc}} \qquad (2a)$$

The cumulative number of infected cases is obtained by integrating:

$$\frac{dcumul_{infected}}{dt} = \frac{E}{t_{inc}} \quad (7)$$

The cumulative number of detected cases is obtained from a similar equation, with the detection rate lagging by the period t_{test} and the detected percentage denoted $test_{frac}$:

$$\frac{dcumul_{detected}}{dt} = \frac{E|_{(t-t_{test})}}{t_{inc}} test_{frac} \quad (8)$$

The infection rate at a time t_{test} earlier is calculated using a tanks-in-series model to create a time lag of t_{test}.

The number of hospital beds required to treat patients at any time is estimated from the number of isolated cases at that time and the fraction of known cases that requires hospitalization. This fraction may be as high as 20% for all hospitalized cases or more typically 5% when calculating the number of intensive care beds needed:

$$beds_{needed} = beds_{frac} Is\ test_{frac} \quad (9)$$

The model calculates the peak day of requirements as the day on which the maximum number of beds needed was required. If there are multiple peaks, this calculation reflects the day on which the highest of all peaks was reached.

Even if the estimate for beds needed is not highly accurate, the underlying model predictions of the number of exposed, infectious and isolated cases and the timing of the peak are valuable to judge the state of the outbreak.

Most of the model parameters (on *scenarios* worksheet) are taken from literature, notably reference 1 below, based on information about the outbreak in China. The rate parameter β is fitted to known / detected case data during the initial exponential period when restrictions are not yet in place. The initial number of infectious is also estimated from those data. For most regions, this requires setting the number of days after which the first case was reported to be approximately the 10^{th} day of the outbreak.

The profile of the *mixing* parameter (reflecting people movement) is inferred by fitting to the known case data after the exponential period. This helps to see whether and to what extent the curves are flattening. To make predictions of the impact of continued restrictions, deeper restrictions or relaxation of restrictions, *mixing* profiles for subsequent weeks or months may be tried using the model (e.g. Set Parameters in Simulator), based on the inferred effectiveness of prior measures in that region.

Fitting to known case data is still the most viable option, as reliable sources of other responses are not generally available, apart from the unfortunate number of deaths. Known case data for any specific region depend on the rate of testing in that region and the criteria for testing. When the testing rate and criteria change during the outbreak, this changes the relationship between known cases and actual case numbers. Challenges experienced in sourcing test reagents and testing facilities may distort the meaning of known case data in this way.

Finally, we predict the number of reported deaths as follows. The isolation rate has a lag applied by a period t_{death}, representing the time taken from isolation until the death is reported. The fraction of isolated cases in which death is reported after that period is denoted k_d:

$$\frac{d\ deaths_{reported}}{dt} = k_d \left.\frac{I}{t_{inf}}\right|_{(t-t_{death})} \quad (10)$$

The isolation rate at a time t_{death} earlier is calculated using a tanks-in-series model to create a time lag of t_{death}. Both k_d and t_{death} are fitted to reported death figures.

In order to model the later stages of the outbreak in 2020, especially after 'reopenings' have started and testing rates have ramped up, additional variables have been included in the model. The basic equations remain the same as above. However, the following behaviour has been added for the reopening periods:

- The date on which reopening commences is t_{start}
- After t_{start}, k_d and t_{death} are replaced by k_{d2} and t_{death2}, reflecting the lower death rate in the reopening period, when the age profile of cases has shifted to younger people
- After t_{start}, $test_{frac}$ is calculated from a linear relationship that allows the detected fraction to increase as test volumes are increased (and the positive test rate declines):

$$test_{frac} = test_{frac0} + \left(test_{frac2} - test_{frac0}\right)\frac{time - t_{start}}{test_{ramp}} \quad (11)$$

Data from reference 2 have been used as the basis to fix $test_{frac2}$ at 50%; $test_{ramp}$ is taken as 30 days. Both parameters could also be fitted in principle.

These additional parameters reflect the reality of how the dynamics have changed with reopening, but also introduce additional uncertainty as their values are difficult to estimate reliably.

1. Ruiyun Li et al, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2), Science 16 Mar 2020, DOI: 10.1126/science.abb3221, available at https://science.sciencemag.org/content/early/2020/03/24/science.abb3221 (accessed 4 April 2020)
2. Steven John Phipps, Rupert Quentin Grafton, Tom Kompas, Estimating the true (population) infection rate for COVID-19: A Backcasting Approach with Monte Carlo Methods, medRxiv 2020.05.12.20098889; doi: https://doi.org/10.1101/2020.05.12.20098889, available at https://www.medrxiv.org/content/10.1101/2020.05.12.20098889v1 (accessed 23 July 2020)