**A quick guide to Covid-19 epidemiology**

BDI Pathogen Dynamics Group
Oxford University

**Reproductive number R**

There is just one mathematical truth in epidemiology: If each infected individual infects on average less than one other person (reproductive number $R<1$), the epidemic declines. If each infected person infects on average more than one person ($R>1$), the epidemic grows exponentially. Interventions designed to stop the spread of the epidemic will therefore be successful if they achieve a reduction of $R$ below 1.

**Dependencies of R**

$R$ for a given virus in a given population is influenced by many factors, among them the routes of transmission, the number of viral particles in the infected individual, the likelihood of contact between individuals, the likelihood of the virus surmounting the initial defences of the immune system and the proportion of individuals in the population which is immune due to previous exposure to the same or a similar virus. Interventions to stop the spread of an epidemic therefore aim to reduce $R$ by reducing one or more of the contributing factors.

**Reducing R by vaccination**

The easiest way to stop an epidemic from spreading is the use of an effective vaccine which increases the proportion of individuals who are immune. If a sufficient percentage of the population is immune (often between 90 and 95%), transmission chains cannot be established and imported cases will not spread. This is called herd immunity.

**Reducing R by other means**

If an effective vaccine is not available, the aim is to reduce other factors contributing to $R$. Isolating patients and quarantining their contacts aims to disrupt transmission chains by reducing the likelihood of an infected individual coming into contact with susceptible individuals. Social distancing decreases the likelihood of individuals getting in contact in general and as such also reduces the contact between infected and susceptible individuals. If contact between infected and susceptible individuals is reduced drastically, $R$ will drop below 1 and herd protection is achieved. From a public health point of view, herd protection is very similar to herd immunity, however, it is only maintained as long as contact remains reduced and will break down if contact rates increase again. It can however keep an epidemic at bay until a vaccine or effective treatment are available. The degree of contact reduction that is required depends on the $R$ at the beginning of the epidemic and on how targeted the approach of contact reduction is (targeted contact tracing affecting only the contacts that matter for disease reduction or general social distancing measures affecting the whole population).
Epidemiological models of disease

A main focus of epidemiology is the construction of mathematical models of epidemic spread that enable epidemiologists to understand how a given disease spreads and to simulate how it will develop with and without interventions. Very broadly, two types of models are used. Models are either compartment-based models which assume that individuals are members of groups and that their characteristics can be assigned from a distribution of values for the characteristic in question. Individual-based models model every individual in the population separately and are better at capturing heterogeneities between individuals that cannot be described by a mathematical distribution. These models can be informed by real-world data should this data be available.

Key model parameters

Epidemiological models use a key set of parameters that describe an epidemic. Some of these parameters are intrinsic to the virus, others are intrinsic to the interaction of the virus with a given population. At the beginning of an epidemic, these parameters can only be determined empirically and are therefore subject to corrections as more data becomes available.

Several parameters are used with two different denominators: confirmed cases and infected individuals, which is an important distinction. The case fatality rate (CFR) for example is the proportion of confirmed cases that succumb to the disease, while the infection fatality rate (IFR) is the proportion of infected individuals that die. The infection mortality rate is the more interesting of the two, but it is difficult to measure as the number of mild and asymptomatic cases is very difficult to determine, especially early on. The case fatality rate depends on the infection fatality rate, but also on the level of testing and the accuracy of the test. However, it is easy to determine and widely used.

Key parameters include:

- Growth rate (r): the increase in the number of infected individuals (or confirmed cases) in a given period of time
- Doubling time: The number of days it takes for the number of infected individuals to double
- Generation time: The average time between the time of infection of one individual and the time of infection of someone infected by this individual (sometimes symptom onset is used instead of time of infection as this is easier to determine)
- Incubation period: The average time between infection and onset of symptoms
- Symptomatic period: The average time during which infected individuals experience symptoms
- Infectious period: The period during which an infected individual can transmit the disease
Individual-based models can accommodate more complex input parameters like network or connectivity matrices.

Using these parameters, other characteristics of the disease can be estimated:

- Reproductive number $R$: The number of individuals infected by one infected individual
- Reproductive number $R_0$: The number of individuals infected by the first individual to contract the disease (or more often $R$ at the very beginning of the epidemic)
- Infection mortality rate: Proportion of infected individuals who succumb to the disease
- Case mortality rate: Proportion of confirmed cases who succumb to the disease
- Health economic parameters like disability-adjusted life years (DALYS), i.e the number of years lost due to ill-health, disability or early death
- Economic measures like loss of % GDP

All models can be used to simulate

- The course of the disease given different input parameters
- The course of the disease assuming different interventions

**Key parameters for Covid-19 effect on public health and the success of containment**

Several parameters for Covid-19 make it a public health problem:

- The relatively high infectioness with $R \sim 2$, which leads to a high growth rate and a fast doubling time (influenza typically has an $R$ of 1.1 to 1.5)
- The relatively high infection mortality rate of $\sim 0.7$, leading to a case fatality rate of 1-3% with current rates of testing in Europe (influenza typically has a infection mortality rate of less than 0.1%)
- The high proportion of infected individuals requiring mechanical ventilation, currently estimated to be $\sim 3\%$ of infected individuals and $\sim 7\%$ of cases (influenza caused $\sim 3000$ admissions to ICU units during the season of 2018/19). This means that in the UK and elsewhere, ICU bed and hospital bed capacity will be exhausted soon after the start of a widespread epidemic.

Several other parameters make it difficult (but not impossible) to contain:

- The infectious period starts before the symptomatic period: The fact that infected individuals are infectious several days before symptoms become apparent which makes it difficult to separate infected individuals from the rest of the population. Models suggest that at least 40% of transmissions occur during this pre-symptomatic period. This means that a large testing capacity is required for successful contact tracing.
- The presence of asymptomatic carriers and lack of data on their infectiousness: An estimated 20-40% of infected individuals experience no or very mild symptoms, among
them the majority of children that become infected. Rigorous contact tracing in Singapore suggests that asymptomatic adults are not responsible for a large number of onward transfections, however, the contact tracing did not include any children. While schools substantially facilitate the transmissions of some viruses, e.g. influenza, this is not the case for other viruses. More research in this area is urgently needed to determine the effect of school closures on onward transmission.

Benefits of epidemiological modelling

Mathematical models have increased dramatically in sophistication, and can simulate very granular interventions. Models have great utility in evaluating combined interventions (choosing from the menu above), aiding the refinement of intervention packages. Epidemiological modelling can help to estimate the number of infected individuals (rather than confirmed cases), the infectiousness of the disease, the contributions of different routes of infection, the likely loss of lives and the likely social and economic damage given different scenarios and interventions and other parameters required for evidence-based public health decision making.

Limitations of epidemiological modelling

Like all models, epidemiological models depend on accurate input parameters to produce accurate results. Especially during the early phase of an epidemic, important input parameters like growth rate, generation time and incubation period have to be determined empirically. However, the predictive power of mathematical models remains limited, typically with predictive error of ~10-20% per generation of infection, leading to rapid decay of accuracy over time. This is likely because of adaptive, non-linear, and assortative feedback loops in human behaviour that are not well understood or modelled. Microsimulations have not dramatically improved long-term forecasting accuracy, and sometimes obscure key unknowns. For example, no amount of detail in a simulation can remedy our lack of data on the infectiousness of children asymptomatically infected with COVID.

Using epidemiological modelling to design a mobile app to stop the spread of the disease

It has been suggested to stop the spread of Covid-19 by using a mobile app based on an epidemiological model. The app would register which other individuals the owner of a phone has been in contact with. Should the owner of a phone test positive for Covid-19, all individuals that have been in close contact with the patient in the past five days would automatically be assigned a risk score and asked to either call 111 to get tested or given advice on what to do, e.g. to avoid contact with vulnerable individuals like the elderly and to immediately contact 111 should they themselves show symptoms of Covid-19 disease. In order to work, the app would require an ethical and legal framework, sufficient user buy-in, and a sufficiently high testing capacity.

The epidemiological model behind the app would identify which individuals are most at risk of having been affected and combine instant and efficient contact tracing at scale with an algorithm to determine who is most at risk of having been infected and who is most at risk of passing on
the disease. The algorithm could be adjusted to reflect maximum testing capacity (though an increase in UK testing capacity would initially be needed to stem the spread of disease). The algorithm could also be adjusted to work from different base lines (interested users download the app themselves / the app is automatically installed on all phones through a system update / the app is needed to avoid social distances measures that are implemented more generally.

An app-based approach is generally credited with enabling health authorities to stop the spread of SARS-Cov-2 in China as well as in South Korea. While it is not desirable to adopt the Chinese or the South Korean approach for multifaceted reasons, the algorithm behind the suggested app for the UK would likely emulate the success of the Chinese and Korean apps while being compatible with norms of individual liberty, privacy and personal data usage in a Western democracy.