Communicating the models

Dr. Vandana Revathi Venkateswaran (MPI Evolutionary Biology, Germany)
Dr. Madhav Ranganathan (IIT Kanpur, India)

Introduction

In an event like an epidemic, policymakers are keen to know how the disease will spread - for instance, they might be keen to know how many people are likely to be infected in future. This will help them make decisions and allocate resources towards disease control. For instance, this knowledge can give information about the number of ICUs or ventilators required in a region. These and several other predictions concerning the spread of disease are accomplished using mathematical models of epidemiology.

Mathematical models help us make our mental models more quantitative. Models are not reality; however, studies across the natural and physical sciences have shown the importance of models in understanding nature. Say, we need to send a spacecraft to the Moon. To find out how much velocity a spacecraft needs for it to escape the Earth, we would not design hundreds of spacecrafts and launch them at different speeds to see which one reaches the Moon, right? Instead, we rely on mathematical equations that clearly predict the velocity and all other features that a spacecraft should have to reach the Moon.

Models come with limitations. To extend the analogy, our spacecraft is designed to reach the Moon, not Jupiter and our model would need some tweaks to get it there. It is important to understand the assumptions behind a model and its scope before using it to make predictions and policies.

Why do we need mathematical models for CoVID-19?

Models are used to predict the future of a population. In the case of epidemics too, we need mathematical modelling to understand how the disease is most likely to spread, and where it is more likely to spread. This will help avoid risky trial experiments or random guesses with real populations. It can also be viewed as a shortcut, instead of implementing many guesses about how to deal with the spread of a disease we can see what implementing each of these guesses would mean, using some nifty equations, and take more well-informed decisions. Even as you read this, mathematical modelling has been at the heart of several policy decisions worldwide regarding the response to CoVID-19.

However, there can be several possible models. So, the question is, how are models developed and used? Typically, models are constructed based on some reasonable hypotheses. They are then validated using available data. For

instance, if we want to use a new model to predict the number of CoVID-19 infections in Chennai in June 2020, we first validate it using the data on infections until now. In other words, we see if the model is able to explain the daily number of infections until today (April 4, 2020). Once the model is validated, it can be used to predict future behaviour and suggest new experiments to study the population. As days go on and new data becomes available, it is possible to test the model predictions. In some cases, the model is improved/refined as more data becomes available and the cycle continues.

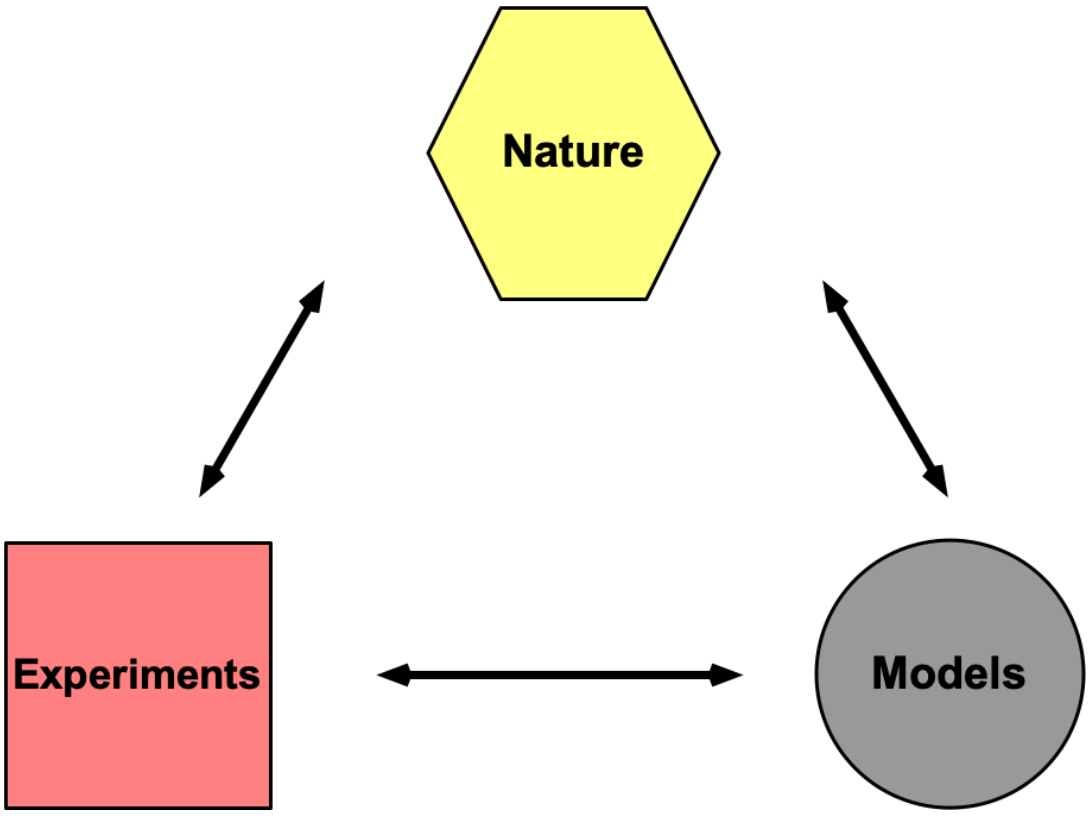

Fig 1: Experiments and models are tools that work together to understand nature

SIR and SEIR Models of Infectious Diseases

SIR models are commonly used to study the number of people having an infectious disease in a population. The model categorizes each individual in the population into one of the following three groups :

1. *Susceptible* (**S**) - people who have not yet been infected and could potentially catch the infection.
2. *Infectious* (**I**) - people who are currently infected (active cases) and could potentially infect others they come in contact with.
3. *Recovered* (**R**) - people who have recovered (or have died) from the disease and are thereby immune to further infections.

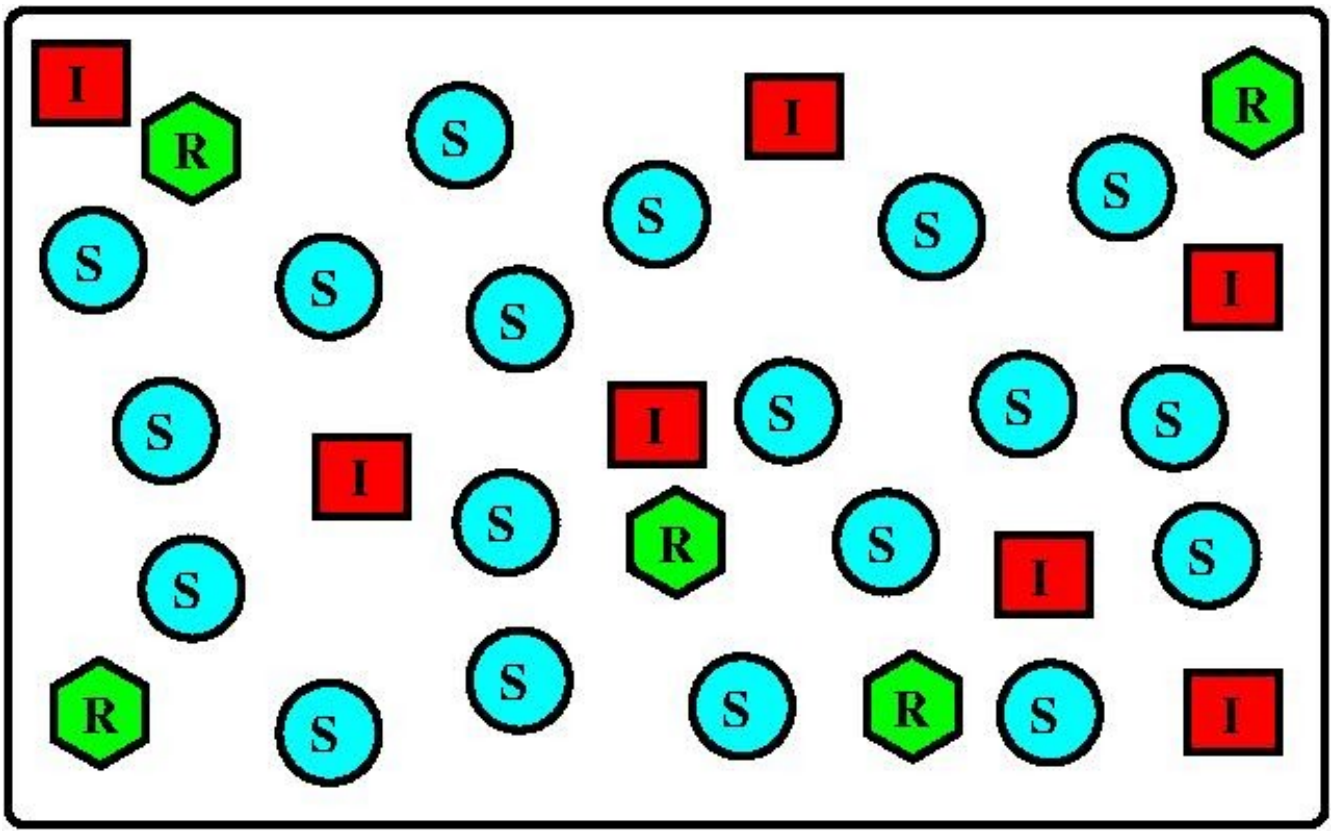

*Fig 2: Cartoon showing individuals in a population categorized as **S**, **I**, **R**.*

These *compartments* contain a certain number of people on each day. However, that number changes from day to day, as individuals move from one compartment to another. For instance, individuals in compartment **S** will move to the compartment **I**, if they are infected. Similarly, infected people, **I** will move to the recovered **R** compartment once they recover or die from the disease.

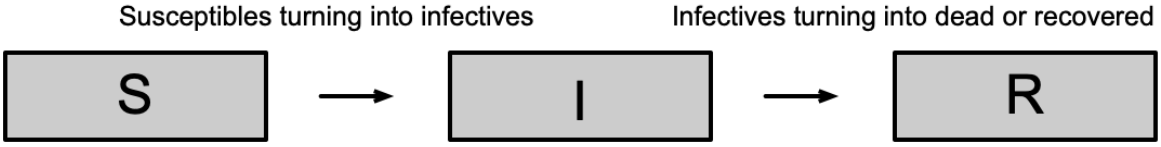

*Fig 3: Population divided into compartments, **S**, **I** and **R** whose numbers change with time. The total population (the sum of the populations in **S, I, R**) remains the same at all times.*

The total population across the three compartments **(S+I+R)** is assumed to remain the same at all times. This is just the total population of the country (or state/region) we are considering. This means that everyone exists in one of these 3 compartments. This ignores the fact that in the natural course of things (epidemic or not), births and deaths continue to happen in the country. But for short epidemics that last a few months, this is a reasonable assumption to make! For modelling other diseases like childhood infectious diseases such as measles that recur regularly, natural birth and death rates of the population will also have to be taken into account.

As in the current epidemic, from media reports each day, one can find the numbers of active cases (**I**) and the number of recovered or dead (**R**)**.** The media

also reports the total number of infected people to date, which if one thinks about it is nothing but the sum **I+R.**

Our goal is to find out how the number of people in each compartment changes with time. In order to do that we make two simple hypotheses on what drives the movement of people between these compartments.

The first hypothesis: Let us suppose you have not been infected at this point in time. So, you would belong to the **S** compartment. You can be exposed to the virus only when you come in contact with an infected person. The greater the number of infected people in the general population, the higher the chance that you will come in contact with an infected individual. This same principle which applies to you, applies equally to every other susceptible individual in the population. Therefore, the rate at which susceptible people become infected, i.e., the rate at which people are transferred from the **S** to the **I** compartments on a given day is proportional to the size of the **I** compartment as well as to the size of the **S** compartment on that day.

The second hypothesis: Infected people will either recover or die of the disease. On each day, a certain fraction of infected individuals will recover or die. This fraction is taken to be a constant, independent of the number of susceptible, infected, or recovered individuals on that given day. This fraction is somehow "intrinsic" to the specific pathogen and captures the average human body's recovery time for that particular disease.

What mathematical modellers do is to write the above hypothesis in terms of mathematical equations which tell you how the number of susceptible, infected and recovered individuals change with time. In the language of mathematics, such equations are referred to as *differential* equations. These equations are solved by a process called *integration*, and these solutions will allow us to calculate, for example, the number of infected people for any time in future.

For diseases such as CoVID-19, we need to consider another compartment called 'Exposed' (**E**). This consists of individuals who might have the virus (due to travel, direct/indirect with an already positively tested person), but do not show any symptoms. For example, if your cousin travelled to Wuhan and came back she is more susceptible than you - because she has been around the virus. In other words, they are between the susceptible and infected compartments. However, despite not showing any symptoms, these (asymptomatic) individuals can still transmit the disease to susceptible individuals. The modelling proceeds in the same way as in the previous case and the solution allows us to calculate the number of infectious people at any future time.

Disease Transmission and Containment

Models enable the quantification of the spread of diseases. The rate of spread of infections in a certain population is governed by a quantity R0, which is the basic reproduction number. The R0 value can be looked at as the intensity of the

infectious disease outbreak. Higher the R0 value of a disease, the faster the disease would spread among the population. In simple terms, the value of R0 is equal to the number of newly infected cases, on average, an infected person will cause. The R0 for measles ranges from 12–18, depending on factors like population density and life expectancy. This shows measles that is a highly infectious disease. If one person gets it, then about 18 will follow. Compared to measles, the novel coronavirus virus is less contagious. As this virus is new, we are not conclusive, but from the evidence we have, R0 ranges from 2.2–2.6. Several biological and social factors come into play in determining the R0. The incubation period, host density, modes of transmission — all affect the R0.

The key insight is if R0 is less than 1, then the epidemic will die out. Thus, our goal is to reduce R0. We can reduce R0 by physical distancing, quarantining, vaccinating, etc. Studies show that the novel coronavirus can travel only about a meter in the air as compared to the 100 meters range for an airborne disease like measles. Second, the measles-causing germs can live outside the host for hours. In contrast, the novel coronavirus can survive only for 3 hours in the air. This R0 value, however, is only an average estimate. For some still mysterious reason, some infected persons, called super-spreaders, can infect a lot more. A woman in South Korea, who belonged to a religious sect, attended services in a church. (ironically, the service was held to obtain god's grace to protect them from coronavirus) 5016 people connected to that church have been affected until March 18th. This also why public gatherings are forbidden; we do not want to even accidentally trigger the hidden super-spreaders.

Fig 4: Super spreader in South Korea

Flattening the Curve

In the early stages, the disease spreads rapidly through the population via a phenomenon called *exponential growth*. The spread of infections slows down as more and more infected people recover and become immune. Eventually, the number of infected people does not increase with time, a phenomenon referred to as *flattening of the curve*. After this, the number of infections drops and eventually goes to zero as the population recovers and becomes immune. Thus, mathematical modelling can give us the maximum number of people who are infected by the disease at any given time and the exact time when we see this number of infected people. This gives an idea into the number of hospital beds/ICUs required for the population.

Now say, we have a vaccine against an epidemic. This will reduce R0 since the number of susceptible people getting infected will decrease. As more people are vaccinated, the disease will come under control. Two Scottish mathematicians, Kermack and McKendrick showed (using mathematics, of course!) that we do not have to vaccinate the entire population for an epidemic to get over. Vaccinating only a fraction of the population is enough and this fraction depends on the R0. This fraction for the novel coronavirus causing COVID-19 has been found to be equal to 60%. This result is another example to show how mathematical modelling is extremely useful.

Individual or Agent-based models

In a SIR/SEIR model, many people fall into the susceptible compartment but not every susceptible individual has the same chance of encountering an infected individual. Healthcare workers, for instance, have more chances of getting infected. Individuals belonging to the same network (social, religious) have varying chances of getting infected depending on their network(s). For example, a shopkeeper meets hundreds of customers a day and therefore his chance of being exposed is much higher.

This is where agent-based models (ABMs), also known as individual-based models or IBMs, become useful. These are models that simulate the behaviour of autonomous agents. While modelling disease, the agents are usually individual people. This contrasts the previous model which only kept track of how the number of susceptible, exposed, infectious and recovered patients varies with the progression of time.

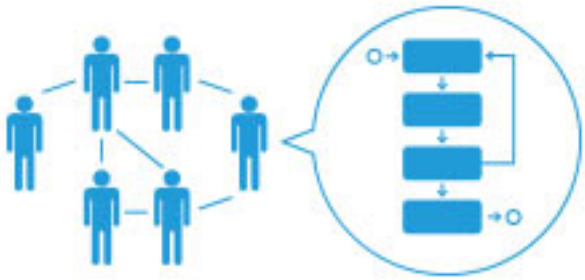

Fig 5: Agent-based models [Might be under copyright, let's make an image?] (https://encrypted-tbn0.gstatic.com/images?q=tbn%3AANd9GcQciXX3BEIo3B8foZ0nCYDa2u-MQ2KV68QHMxSHwkVnq-cDtXHq&usqp=CAU)

The agents behave according to a prescribed set of rules that determine how they interact with other agents and with their environment. These agents are capable of perceiving their environment and the other agents around them and taking decisions, leading to actions, based on their perceptions. Further, the agents may be capable of learning from the outcomes of their past actions. This kind of simulation, in which individual agents decide what to do in each step overcomes certain drawbacks of the SIR models and its derivatives, such as the assumption that the population is homogenous. It is possible, for instance, to have different types of agents that represent members of different age groups or of different professions and incorporate facts such as the greater exposure of healthcare workers to infected individuals, which in turn increases their risk of infection. They serve as "bottom-up" models, in which the emergent outcome is determined by the behaviour of the individuals in the population and require no assumptions regarding the behaviour of the system as a whole.

ABMs can thus incorporate data regarding heterogeneity in the population, social networks and individual interactions, which leads to a more realistic method of modelling populations. They also result in more realistic

contact patterns of individuals and incorporate some of the stochasticity (randomness) that exists in the real world. So, ABMs are capable of predicting interesting and unexpected emergent phenomena.

Such models have previously been used to model diseases at multiple spatial scales, from within a city to across an entire nation. This can be done by modifying the numbers and behaviours of the agents and the nature of the environment. ABMs have successfully been used to model various epidemics including H1N1, various strains of influenza, and Ebola.

One major drawback to ABMs is that as the number of agents increases, so does the computational power required to run the simulation. Large-scale agent-based models tend to require high-performance computing environments for their implementation.