Forecasting COVID-19

Simple models and misleading data

If we apply these ideas to the COVID-19 pandemic, it is easy to see why forecasting its effect is difficult. While we have a good understanding of how it works in terms of person-to-person infections, we have limited and misleading data. The current numbers of confirmed cases are known to be vastly underestimated due to the limited testing available. There are almost certainly many more cases of COVID-19 that have not been diagnosed than those that have. Also, the level of under-estimation varies enormously between countries. In a country like South Korea with a lot of testing, the numbers of confirmed cases are going to be closer to the numbers of actual cases than in the US where there has been much less testing. So we simply cannot easily model the spread of the pandemic using the data that is available.

The second problem is that the forecasts of COVID-19 can affect the thing we are trying to forecast because governments are reacting, some better than others. A simple model using the available data will be misleading unless it can incorporate the various steps being taken to slow transmission.

In summary, fitting simple models to the available data is pointless, misleading and dangerous.

The good news is that we do have good models of the spread of epidemics, and we have just enough data to use them.

Compartmental epidemiological models have been developed over nearly a century and are well tested on data from past epidemics. These models are based on modelling the actual infection process (a bit like weather forecasts model atmospheric processes). The simplest models are based on classifying living individuals in the population as Susceptible, Infectious or Recovered – hence they are called SIR models. Using

differential equations, they describe how people move between groups. More complicated variations allow for several more categories of individuals. These models allow for undiagnosed as well as diagnosed cases, so they can account for limited testing. The available data is used to estimate the various parameters in the models and then simulations are done to see what would happen under various scenarios. Some of the parameters can only be crudely estimated, but we can test what happens to the forecasts by varying the parameters within the likely range.

A good example of using this approach for COVID-19 is by a team from Imperial College London (Ferguson et al, 2020) who applied an agent-based modelling approach to the UK.

I'm yet to see a similar model for Australia, but I know there are many good people at places like PRISM and the Doherty Institute working on it.

Meanwhile, Dr Alison Hill (Harvard) has created a great Shiny app which allows exploration of the effect of the various parameters in these models.