My research:
Mathematical modelling of infectious diseases

Robyn Stuart

I started working on infectious diseases more or less by accident. I'd just finished a PhD in a much more pure-maths topic (Markov chains + mixing times), and I'd accepted a postdoc position in the medical department mostly because I wanted to try something new. But within an hour of starting, I felt like I'd been given a whole new lease on life. It was almost inconceivable to me that I'd be able to make useful contributions to something that I found so exciting. It felt too good to be true, as though I'd randomly decided to walk down a particular street and then had found a hundred dollars on the road. What are the odds that the decision to study maths could have led me to something so cool! (As it turns out, the odds are pretty high - there are lots of cool jobs that you can do with a math degree. This is definitely something that I think we could emphasise more to our students!)

There's a fascinating history of mathematical modelling for infectious diseases. For me, a major part of the appeal is the fact that women have been contributing for a relatively long time to the field. If you ask epidemiologists how the field was founded, many will know the story of how the physician John Snow traced the cause of the 1854 London cholera outbreak to a particular water pump. But at the exact same time, Florence Nightingale was doing similarly important work with her comprehensive statistical studies in Turkey and India, which demonstrated that poor sanitary practices were the main cause of high mortality. And if we fast-forward by half a century, to the next major breakthrough, we again find a woman at the helm: the mathematician Hilda Hudson was the brains behind many of the advances in outbreak analysis led by the much more famous Ronald Ross. Had they not been stalled by the first World War, Ross had hoped that their work would eventually tackle "questions connected with statistics, demography, public health, the theory of evolution, and even commerce, politics and statesmanship". This lofty ambition is one that modern epidemiological modellers are still pursuing.

Like so many other fields, the pace of development in the world of infectious disease modelling grew astronomically with the advent of affordable high-performance computers. Mathematical models of infectious disease spread can now offer real-time predictive analyses of disease control and treatment strategies. The emergence of COVID-19 has highlighted the critical importance of such work. But the faster that computers get, the more we challenge them with ever more complex problems and models. As a result, a major part of my research is to develop the statistical and mathematical tools necessary for increasingly sophisticated models to be rigorously tested and validated.

In my daily work, I'm very much at the intersection of mathematical modelling and public health. Much of my research aims to inform disease control policy-making, and is done in collaboration with governments and non-governmental organisations like the World Health Organization, the World Bank, or the Bill and Melinda Gates Foundation. Often, I make user-friendly, open-source modelling software to support better decision-making in health, or work directly with governments to use such software.

A key practical focus is the analysis and optimisation of strategies aimed at reducing transmission or disease burden. For example: let's suppose that you're in charge of a country's HIV program, and you need to decide how to spend your budget for the year. With a fixed quantity of money, how much should you spend on treatment versus prevention strategies? Should you spend more money in areas that

already have higher numbers of people with HIV, or should you target it more towards areas where the number of new infections is increasing most rapidly? And how would the answers to these questions change depending on whether you were most concerned about decreasing the number of infections, decreasing the number of deaths, or maximising the equity of healthcare services? These questions might sound abstract, but mathematical modelling can help here; let's see how.

For such a question, we can start with what's called a compartmental model. Compartmental models divide the population into compartments based on useful properties. The classic version divides people into Susceptible, Infected, and Recovered, leading to the so-called SIR model. Even with a simple model like this, there's a lot of cool math you can do. But for a disease like HIV where people don't recover, we might instead set up an "SIT" model, where the T stands for "treated". People who are treated are less infectious and less likely to die than people who are infected but not receiving treatment. So one strategy – the one advocated by the UN – is to provide treatment to as many people as possible, since this will reduce transmission and deaths. But it's not so simple as that: before people are given treatment, they need to be diagnosed, and there's still a lot of stigma around getting tested for HIV in many parts of the world. Therefore, there's still an important role for interventions that can be used whether people know they have HIV or not, like prophylactics. Moreover, both treatment and testing programs can be very expensive to operate. Much of my work has looked into the question of how countries can optimise the allocation of their funding for HIV, as well as for TB and malaria, across different interventions, geographies, and people.

Over the last year, infectious disease modellers have gotten more press than they'd ever expected or indeed hoped for. In the wake of the COVID-19 pandemic, researchers around the world moved with incredible speed to develop statistical and mathematical models to support worldwide responses to the pandemic. Many of these models were compartmental models, like the one described above. Others, including the one that I worked on, fell into a different category, of agent-based models. In an agent-based model, you don't just track broad categories of people, but instead track individuals. You create a simulation in which representative agents, each with their own properties like age and health, interact with other agents over networks of contacts. These contact networks represent settings like households, schools, workplaces, and communities. When a person becomes infected, they might transmit the virus to other people in their contact networks. If the person is wearing a mask, the probability of transmission is lower. If they have been tested and know that they are infected, they might inform the other members of their household, as well as their colleagues/classmates and friends, and then these people might also get tested and/or stay at home until they know they are no longer infected. These individual-level details are super important for tracking the spread of viruses especially in their early stages, when the difference between whether a new disease evolves into an outbreak or not might come down to random stochastic details, such as whether one particular person went to work one day or not.

With mathematical modelling of infectious diseases, we can understand how the actions and properties of individuals translate to population-level outcomes. This simple statement provides a line of continuity from the earliest stages of my PhD research to now: I've always been fascinated by how macro-level patterns form from seemingly random micro-level behaviour. Thanks to mathematical modelling, we can investigate such questions in theory before putting them into practice in the real world.