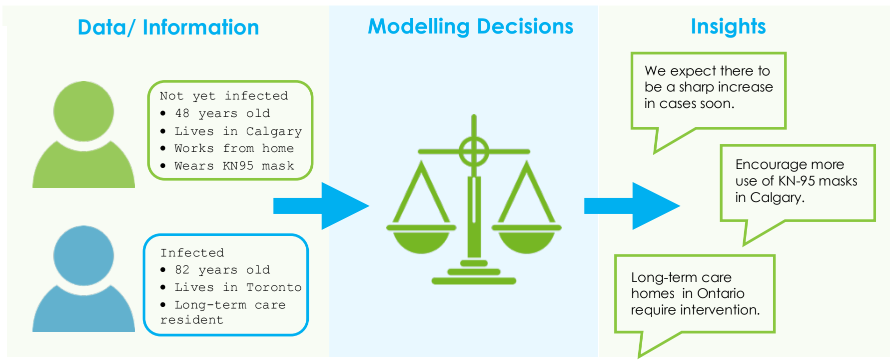
STATISTICS OF INFECTIOUS DISEASE TRANSMISSION

STATISTICAL SIGNIFICANCE

Emerging infectious diseases, such as COVID-19, carry a lot of uncertainty. People are unsure of the best way to respond, weighing risk management against life disruptions. Policymakers need to decide which interventions would be most appropriate, and when to implement them. Statistical models can help inform these types of decisions. Statisticians are continuously working to make infectious disease models more accurate and informative, while working within the constraints of data availability and other challenges.

Disease Modelling

From the first days of data availability for the COVID-19 pandemic, statisticians all over the world have been thinking about the best ways to harness that data into models. In fact, searching a database for scientific articles about statistical COVID-19 models yields around 250,000 results for articles published in 2020 alone. These statistical models are useful in different ways and provide unique insights we wouldn't be able to obtain from only raw data or lab studies.



Statisticians take data and combine it with expert knowledge to weigh decisions about what components need to be included in a model in order to translate to accurate and interpretable insights about an infectious disease epidemic.

Statistics at Work

There are many ways statistical models can help deepen people's understanding of how infectious diseases spread and inform decision making. When considering a new intervention – say, running a public health campaign about effective handwashing and its benefits – a statistical model can be used to project the potential impact on future disease case numbers under various scenarios (such as if 10, 25, or 50% of people improve their handwashing as a result). After implementing an intervention, models can also be used to evaluate how effective it actually was. More generally, over time as more data becomes available models can help illuminate what is most important in the spread of the disease – is there a strong effect of age? Household size? Geographical location?

Modelling Challenges

At the beginning of a new epidemic (where disease is spreading within a region/ country) or pandemic (where disease is spreading across multiple countries), there is often very little information available. Data collection systems are not in place yet, so records might be sparse, inconsistent, or unreliable. It takes time to gather information and learn enough about a disease to effectively represent it as a model. Sometimes this leads to the same data seemingly showing very different results in different studies, as statisticians navigate the many decisions that go into making a model. Models are a work in progress that improve as more data comes in.

Data availability is a major limiting factor in how complex statistical models can be. While many aspects impacting disease transmission could theoretically be translated into a mathematical equation, these equations are limited in their usefulness without data to inform them. Models could be considered at a very fine granularity for example, accounting for factors such as individual people's daily interactions or genetic predispositions to disease. While a model like this might give a very accurate picture of how the disease is spreading, this level of data is unlikely to ever be available on the scale that would be needed to be useful in modelling an epidemic.

From Data to Insight

Given the challenges at hand, statisticians need to make trade-offs. They combine their mathematical knowledge with the knowledge of other experts such as epidemiologists, physicians, and infectious disease researchers to develop models that ensure the data that *is* available is being used to the fullest extent possible. In turn, they recognize that there will always be some gaps in our knowledge that even the most clever models can't fill. This is why statistical models provide a range of plausible scenarios, rather than a single guess about what's at play.

While statistics alone can't put an end to an infectious disease, models can certainly help with risk evaluation and exploring possible future outcomes. The work of statisticians is critical in taking what can seem like unintelligible information and turning it into insights about what is impacting the spread of a disease and what course of action is most likely to lead to minimizing negative impacts.♦

By: Madeline Ward
Written for the 2023 JSM ASA SPAAC Statistical Significance Competition