The field of epidemiological modelling has grown steadily in size over the past several decades. Its value manifests in several ways: first and foremost, in its ability to unify disparate sources of data and knowledge across the various subfields of public health. Using modelling, researchers can combine data from genomics, policy research, and epidemiological studies to model phenomena that could not be accurately understood otherwise.

Modelling also allows for other types of research to be validated and expanded in scope to broaden its applicability. For instance, small-scale public health studies that test a new model of community health care in a small town or province are lacking in size to serve as effective evidence for rollout of the same program in a larger context. Additionally, funders and governments are generally unwilling to undertake medium-size studies to build on the results of a small-scale study, given the financial burden involved and the risk that the results of the smaller study will not scale well. In these contexts, modelling excels as a predictive tool to enhance the impact of smaller-scale studies. The modeller can adapt the study as best as possible to represent its systems programatically, and then run the model with parameters drawn from the study. Lastly, models can quantify uncertainties in decision making in a way that empirical studies can struggle to, since stochastic models can be run many times to produce a range of model outputs, and Bayesian techniques can allow for the representation of uncertainty in model parameters.

One of the most well-studied and oldest models in epidemiology is the SIR (Susceptible-Infected-Recovered) model. As a compartmental model, it classifies every member in a population of fixed size as either susceptible to infection, actively infected and infective, or recovered and immune. The number of individuals in each compartment changes according to the rate at which individuals transition from susceptible to infected, and the rate at which they transition from infected to recovered. The rate of transition from susceptible to infected is dependent on the number of infected individuals, since a population with more sick people means it is more likely to become infected.
Because of its simplicity, the SIR model is one of a handful of compartmental epidemiological models which has an analytic solution. However, the convenience and simplicity of the model belies several shortcomings. The compartmental nature of the model makes it such that individuals are nothing but quantities which are shuffled between compartments. In real life, individuals have differential risks of infection according to many factors, such as age, sex, income, geographic location, history of prior disease, inoculation, and many others. Individuals also have differential rates of recovery from disease, which vary according to access to healthcare, the individual’s immune response, and a whole host of other factors which can make the difference between a speedy recovery and lifelong illness. Additionally, given the same parameter set, the SIR model will produce the same outcomes, which fails to represent the reality that epidemics are often very stochastic and difficult to predict.

Agent-based stochastic modelling has started to take hold as a strategy to overcome many of the innate shortcomings of traditional epidemiological models like SIR. In an agent-based stochastic model, each individual has a full set of traits, including infection status, and a set of events that can occur over the lifetime of that individual, such as marriage, infection, death, enrolment in a treatment program, migration, etc. Events schedule other events, and the behaviour of a particular event and its scheduling of future events is generally guided by a random variable. Thus, every run of the model produces a different outcome, and the model can be run many times to produce a distribution on each output.

**Deliverable 1.** Produce a reference implementation of an agent-based SIR model. This model should be able to export data, exist within a flexible and well-designed framework so as to enable development of more complicated agent-based models, be able to run in parallel, and be performant under population sizes in excess of 1,000,000.

Complex epidemiological models inevitably have large sets of parameters to describe various event probabilities, quantities, and other values of interest to the simulation. However, it is exceedingly common that a subset of the parameters cannot be observed or are not known with confidence in the literature. For instance, medication compliance can be difficult to quantify because compliance varies with medication, side effects, burden of medication administration,
and other factors, and because it is difficult to directly observe patients taking medication. Thus, most models will contain a subset of parameters whose values must be estimated through some method. These methods have been well-characterized for compartmental epidemiological models, but there is little research on how to calibrate an agent-based model.

**Deliverable 2.** Implement and analyze a maximum-likelihood, linear regression approach to the calibration of the aforementioned agent-based SIR. The SIR model has two parameters – the force of infection, and the duration of infection – and thus is a relatively easy model to calibrate. Therefore, an agent-based version of the model is an excellent candidate for examining the behaviour of a calibration method in the agent-based context. The analysis of the calibration routine will include evaluating goodness-of-fit between model output and historical epidemic data across a range of parameters of the model and a range of parameters for the calibration routine. A written report of the results will be authored.

**Deliverable 3.** The lab I do research with primarily focuses on modelling HIV and tuberculosis (TB) coepidemics in a southern African setting with the goal of evaluating the utility of various interventional strategies and public health policies. One of the lab’s projects is evaluating the efficacy of two anti-TB interventions, “second-degree isoniazid preventative therapy” and “active case finding for previously treated patients.” Building an agent-based model for this scenario is much more complicated than for an SIR model, but builds upon techniques and frameworks developed to support the agent-based SIR model. TB infection and HIV infection have many different disease states and act in tandem to affect the health and infectivity of the individual, and the agent-based mindset is well-suited to dealing with this additional complexity. This deliverable comprises implementing the agent-based model, calibrating it using the technique honed in Deliverable 2, and visualizing model output.