Agent-Based Epidemiological Modelling: Towards Novel Strategies for Optimizing Public Health Interventions

Marcus Russi
Yale University
marcus.russi@yale.edu

ABSTRACT
Over the past few decades, modelling has emerged as an essential tool for the design, calibration, and analysis of public health interventions. Its ability to extrapolate results of small-scale intervention studies and predict outcomes across large spaces of parameters allows epidemiologists and policy analysts to perform low-cost experiments, informing the public health community about the predicted impact of policy. This has lead to a proliferation over the last several decades of modelling techniques to describe the spread of disease and strategies to contain it. Traditionally, these models have built upon systems of analytically solved differential equations, as well as Markov Chain Monte Carlo, in order to construct models that are simple to reason about and quick to solve. However, these models tend to simplify populations of individuals into a small number of groups, and thus impose limits on the types and scope of questions researchers can ask using the model. Agent-based modelling presents an alternative to these techniques by explicitly modelling every individual in a population, with the life-course of an individual being delineated by discrete events that happen at particular times over the course of a simulation. Because each individual can be associated with a rich set of characteristics and history, the model can capture phenomena that are heavily influenced by factors that are difficult to stratify into small numbers of groups. To explore this technique and leverage, an agent-based implementation of the classic SIR model was developed and analyzed to explore the utility of a previously developed framework for agent-based modeling. Then, an agent-based demographic model was developed and analyzed as the basis for a future HIV-Tuberculosis modeling project. Good performance and relative ease-of-use was observed, boding well for future work.

1. INTRODUCTION
Agent-based modelling techniques have increased in popularity throughout the last several decades, broadly as a result of increased computational power, development of useful frameworks for optimizing the models, and intuitiveness of design and use. An agent-based model is, broadly defined, a system of agents evolving over time through their interactions with each other and with their overarching environment. Interactions can be mediated through spatial characteristics, through participation in simulated auctions, and indirectly through the modification of state variables local to some environment. Generally, the simulation can be thought of as a series of scheduled events capable of scheduling other events at future points in time.

Developed in 1927, the Susceptible-Infected-Recovered model is a famous model that describes the sizes of three groups of individuals over time: people who are susceptible to some example disease, people who have been infected by it and thus are capable of infecting other individuals, and people who have recovered from their illness and now have 100% immunity to it. These three “compartments” are connected to each other through differential equations, the rates depending on two easily-derived constants and the size of each compartment.

A brief examination of the Susceptible-Infected-Recovered model reveals how three of its shortcomings can be addressed by an agent-based approach. A first problem is that the SIR model, and others like it, are solved over the real numbers. This has the effect of making component sizes unrealistic, i.e. resulting in populations that contain fractions of a person. Most significantly, this implies that the number of infected individuals may never actually reach zero, implying that the disease is always present in the population being studied, even if every member of the population, save for a fraction of an individual, has recovered. While propensity to never reach zero is acceptable in the case where the population has an infinite number of individuals and the probability of reaching zero is infinitely small, in reality, we often need integer valued quantities since we are modelling a system whose fundamental unit is the person. Because agent-based models are generally implemented as discrete-time simulations
which are not implemented using real-valued functions, this limitation is averted.

A second issue with the SIR model is its determinism: the idea that given some set of parameters to the model, the same output will always be produced. In an idealized system this is true; in the real world, our epidemic models should reflect uncertainty. For instance, if one individual is infected at the beginning of a simulation, it is entirely possible that they will manage the course of the illness, and become “Recovered” without having infected any “Susceptible” individuals. The traditional SIR model cannot account for this. Fortunately, agent-based models provided a convenient alternative: because they are simulation models, nondeterminism is welcomed and distributions on the output space can be approximated through examining many runs of the model.

Lastly, many factors in modelling epidemics fail to stratify well into small numbers of groups. For instance, an individual’s CD4+ count is often used as a measure of immune strength; in turn, this count can influence an individual’s risk of contracting an infection. Ideally, a researcher would be able to model risk of infection as a function of CD4+ count. However, a traditional SIR approach does conceptualize an “individual,” and in general, compartmental models are very limited in their ability to groups of individuals beyond small numbers of traits which take on a small set of values. An agent-based model circumvents this by simply representing every individual in the population as an object with any given number of traits.

As such, agent-based modelling tools are a logical next step for epidemiological modelling. Towards this end, a framework was developed for supporting the construction of ABMs. This framework was tested using an “Agent-based SIR” model. Once tested, the framework was then used to construct a demographic model of population in South Africa, which is the basis for an ongoing project of modelling Tuberculosis and HIV in the country. Performance profiles and preliminary data were collected and analyzed.

Figure 1. Event structure for demographic model. Initial events are coloured red
3. APPROACH

A library named SimulationLib was developed in C++14 to support large agent-based models. It includes functionality such as data recording, data export, parameter import, model calibration, and event queue management. As this library would form the core of future ABMs, performance was critical, given the resource-intensive nature of ABMs.

Following this, a library named SIRLib was developed. SIRLib is an event-based, agent-based SIR model. It has two individual-based events (Infection, Recovery) and one regular event (ForceOfInfectionUpdateEvent), which, at a regular time-step, determines which Susceptible individuals will become Infected. Because SIRLib is a nondeterministic simulation, it demonstrates the heterogeneities previously discussed, including the property that many epidemics fail to “take hold” because the first individual infected fails to transmit. Performance profiles and multiple trajectories were collected from runs of SIRLib. In particular, we were interested to see if runtime would be linear on the number of individuals in the population: since future models would be more complex, it was crucial to verify that the modelling framework could scale to handle populations with millions of individuals.

With SIRLib developed, an agent-based demographic model was developed to simulate the evolution of a population in South Africa. This model is in support of further modelling work related to HIV and tuberculosis. Since both diseases are complex, chronic conditions whose within-host and inter-host properties are mediated by a variety of lifestyle and demographic factors, it is important to have an underlying model which describes, for each individual, properties relevant to the disease course. Parameters from this model were sourced from the South Africa Department of Health and other sources internal to the Yale School of Public Health.

The model structure consists of several events that interact with each other throughout time. Two initial events generate an initial population, drawn from census data, and execute a matchmaking process whereby single individuals are paired up to be married. All other events follow from these initial two; they take each individual through a complete life-course. Model runs were analyzed for face validity and for performance.

4. RESULTS

SIRLib exhibited linear performance on the size of the input and accurately reproduced the dynamics of a traditional SIR model. Importantly, depending on model parameters, many of these model trajectories did not result in epidemic.
The demographic model only allows death due to “Natural causes,” that is, deaths not attributable to AIDS or tuberculosis. Therefore, the death rate is much lower than in reality, and the population exhibits dramatic exponential growth. Over the course of a 10-year simulation on 40,000 individuals, the distribution of household size stays relatively constant, a positive indicator that the model is behaving as expected. Furthermore, running time scaled linearly on the size of the population.

5. CONCLUSION AND FURTHER DIRECTIONS
Agent-based modelling is a promising technique for exploring complex phenomena involved in disease transmission and life-course. It eliminates many of the limitations seen in compartmental epidemiological modelling and allows for exploration of data which would be very difficult to generate in other types of models. Key to successful agent-based modelling is a keen eye for performance and systems design such that the model scales to the size of the population of interest. SimulationLib, SIRlib, and the demographic model exhibit excellent performance characteristics and demonstrate the promise of ABMs. Future work will involve validating a calibration method for parameter estimation in SIRlib, and development and calibration of a HIV and tuberculosis model built on top of the demographic model.

Figure 4. (Top) Birth, death, and population size for 40,000 individuals over 10 years. (Bottom) Histogram of household size at 0 years and 10 years (bottommost). Little change is seen, as desired.