Towards the Control of Epidemic Spread: Designing Reinforcement Learning Environments

Andrea Yañez[1] , Conor Hayes[1,2] , Frank Glavin[2]

[1] Data Science Institute, Insight Centre for Data Analytics, National University of Ireland Galway, Ireland.
[2] School of Computer Science, National University of Ireland Galway, Ireland.
{andrea.yanez,conor.hayes}@insight-centre.org
frank.glavin@nuigalway.ie

Abstract. Throughout history, epidemic outbreaks have led to spikes in human illness and mortality, with major challenges to communities and society in general. An epidemic situation requires decisions to be made about interventions that could reduce or contain the disease spread, taking into account all the information received and the projections of the current situation into the future. Decisions made by public health officials involve determining the best sequence of actions to perform from a set of alternatives (school closure, vaccination, isolation). In order to decide which intervention strategies to implement, decision makers need to analyse a large number of scenarios and variables. This task can be overwhelming. Reinforcement Learning (RL) optimisation strategies have been proposed in the past years to automatically find optimal intervention strategies for a disease spread in order to support decision makers. An important component in RL is the environment, which describes the task that the RL agent (solution approach) aims to optimise. This work focuses on *how to design environments* to represent the problem of epidemics and finding optimal interventions. We present different challenges that need to be addressed for environment design and provide diverse examples from the state of the art.

1 Introduction

Infectious disease spread is a persistent threat to humankind. Emergence or re-emergence of pathogens can lead to an unexpected increase in the number of infected and sick individuals in a local area (epidemic) or even a global area (pandemic). In the last decade, diseases such as Measles, AIDS, Malaria, Ebola, among others, still cause millions of deaths. The crucial question is not whether an epidemic will emerge, but when it does, *Which interventions are the most effective(e.g., social distancing, school closure, vaccination) to contain or reduce the spread?* This question poses a challenge to governments, public health officials and emergency response personnel, who must select the optimal intervention to implement from a vast number of possible alternatives.

To make analyse the possible outcomes, policy makers generally use mathematical models or simulations. In this artificial setting, different types of decisions can be made without experiencing real costs or consequences. Using a simulation, the decision maker could employ intervention strategies at discrete decision points during the evolution of an epidemic. The effect of any simulated intervention is stochastic, which increases the complexity of the decision-making process.

The space of possible interventions or combination of interventions strategies is large and finding the optimal health policy can be an overwhelming task. Generally, decision makers analyse and compare the performance of a one-time decision over a limited number of *pre-selected* intervention strategies. For example, in [21], authors compare the cost-effectiveness of a pre-defined set of health policies. Nonetheless, this approach may fail to consider the optimal intervention strategy. In addition, it does not consider that a plan of sequential interventions over time may be better than a one-shot intervention. However, finding an optimal combination of interventions over time can quickly become intractable and leads to a complex sequential decision-making problem.

A solution would be an optimisation technique that can effectively search the space of possible health policies and identify the optimal interventions to control an epidemic. In this work, we first frame the task of finding an optimal intervention strategy as a Reinforcement Learning (RL) problem. In RL, an agent (optimisation technique) interacts with an environment (problem definition) to learn and find an optimal policy in a sequential decision-making scenario. The agent and environment components can be implemented in different ways depending on the problem definition. In this work, we mainly focus on *how to design environments* to represent the problem of finding optimal interventions during an epidemic outbreak. We present different challenges that need to be addressed for environment design and provide diverse examples from the state of the art. Next, we offer a concise overview of possible solution approaches to address the optimisation problem represented by the environment. Lastly, we conclude and highlight future work.

2 Problem Formulation

We frame the task of finding an optimal intervention strategy as a Reinforcement Learning (RL) problem in Figure 1. In RL, an agent within an environment executes an action a_t over $t \in T$ discrete time steps or decision points. For each action, the agent receives a reward r_t. The goal is to find an optimal policy π^* that can indicate the best way to act for each environment state. Similarly, we assume that public health officials employ mitigation strategies at discrete decision points during the evolution of an epidemic. The mapping of RL components to our problem setting can be summarised as follows (see Figure 1 above):
Agent/Optimisation Component: the Optimisation Component aims to explore the space of possible policies to learn an optimal policy. See subsequent sections for more details.

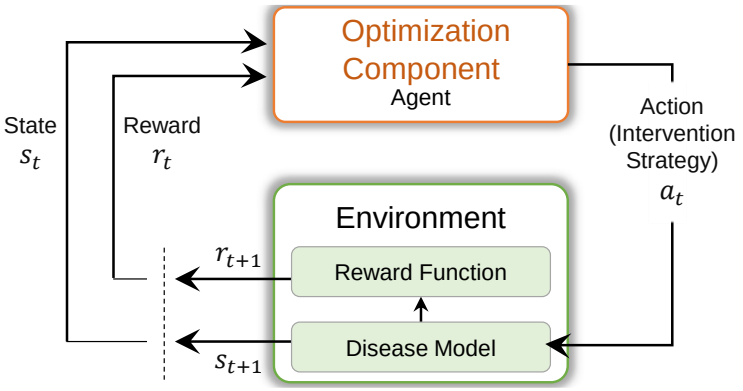

Fig. 1. Identifying Optimal Intervention Strategies as a Reinforcement Learning Task.

Environment: The environment is an abstraction of the problem that reduces it to signals that represent the option selected by the agent (interventions), the agent's new state and how the agent is performing (reward) [27]. Both the disease model and reward function will be further discussed in subsequent sections.
Reward r**:** The reward r_t is a numeric feedback signal given by the environment after an action a_t is executed. It is representative of the goals of the public health officials (e.g. reduce the number of infected individuals and costs). The reward is made up of an accumulation of positive and negative reinforcements based on the decision-making.
State s: The state of the disease model has all the information that is necessary to represent any stage of the disease, e.g., number of infected individuals.
Action/Intervention Strategy a**:** In our proposed approach, an intervention strategy is the action executed by the optimisation component at a specific state s. An intervention x is a specific mitigation measure carried out to prevent or interrupt the spread of the disease. The set $\mathcal{A} = \mathcal{P}\left(\bigcup X_n\right)$ is the power set of all possible interventions, representing all combinations including the empty set indicating "no intervention". Each element $a \in \mathcal{A}$ is an action or intervention strategy. At each state s_t, only $a \in \mathcal{A}(s_t)$ actions are possible. For instance, possible actions depend on the availability of resources (e.g., vaccine doses) and interventions that can be carried out simultaneously.
Policy π**:** The policy is a function that indicates an action $a \in \mathcal{A}$ to perform for each state $s \in \mathcal{S}$, $\pi : \mathcal{S} \rightarrow \mathcal{A}$.

The main goal of this research is to design an optimisation component that can find an *optimal policy* which can be used by a public health official, to choose an action a in each disease state s to obtain the maximum expected reward after $t \in T$ decision points, i.e. $\pi^* = \arg\max_\pi \mathbb{E}_\pi\left[\sum_{t=1}^{T} r_t\right]$.

3 Environments for Epidemic Control

The environment in RL characterises the world that the agent interacts with to learn and achieve its goal. We define two main elements for the environment: the

disease model and the reward function. The *disease model* is a mathematical or simulation model that represents the disease dynamics. At each decision point t, the optimisation component executes an action a_t to test its impact over the disease spread. The outcomes generated by the disease model is represented as a state s_t and interpreted by the *reward function*, which issues a reward r_t that the agent uses to measure its success. Lastly, other design decisions related to the definition of the environment are those related to the *intervention strategy* (what are the actions/interventions?) and the *state representation* (how are states defined and represented?). In the following sections, we will discuss in depth on how each of these components can be implemented and provide different types of examples.

Throughout this section, we will use a running example to introduce several topics. This example is called the **Boarding school scenario**. In 1978, an outbreak of the influenza virus in a boys boarding school was reported [1]. The epidemic started with one initial student infected and lasted 14 days. It confined 512 students, out of a total of 763, to bed. This example has been used in related work such as in [16, 18, 28]. The information about this use case is shared in a dataset that informs the number of infected individuals (students confined to bed) per day.

3.1 Disease Model

In order to understand the large-scale dynamics of a disease spread, we use a mathematical model structure. The models allow prediction of the spread of the disease through a population using a description of the infectious disease dynamics at the individual level.

What are some examples of infectious diseases and their properties? For infectious disease we understand the effective colonisation of a host by a micro-organism like bacteria, viruses, parasites or fungi, producing a subnormal functioning of the host, which can lead to disease [12]. The pathogens are transmitted directly between people or indirectly via a vector (e.g. mosquito, rat) or the environment (e.g. water, air). The symptoms of the infected individual can fluctuate from mild to severe, including death. During the course of the infectious disease, we can identify three periods (incubation, clinical symptoms and non-disease) associated with the severity of symptoms and amount of pathogen particles in the host [12]. The first period (Incubation) starts when the susceptible host is exposed to the pathogen, and this starts to multiply. During this period, the host is still not contagious. The next period (clinical symptom) is characterised by the presence of symptoms, and the host can transmit the infection to another susceptible individual. In the the last period (non-disease) the individual has either recover(ed), suffered death, is immune (Measles), is susceptible to reinfection (Chlamydia), or remains infected for life (HIV). The periods and other properties of the disease are fundamental to understand how the disease is modelled.

Transmission The propagation of a pathogen in the population depends on several factors. The World Health Organisation (WHO) suggests a list of measures to quantify the transmissibility of the disease [20] such as the number of symptomatic cases, the basic reproduction number (R0), and the clinical attack rate. A measure that is frequently used in modelling is the basic reproduction number [5], which is defined as the average number of secondary infected individuals resulting from one infected individual in a totally susceptible population.

Virulence This is the ability of the pathogen to cause damage or death in the host.The WHO list of measures for virulence are found in [20]. Pandemic Severity Assessment Frameworks [24, 22, 19] consider the severity or health impact of an epidemic by incorporating multiple measurements of transmissibility, and Clinical severity (virulence). The 2009 influenza pandemic demonstrated that the degree to which costly interventions are justified depends on the severity of the disease.

How are diseases modelled? The value of using mathematical models in epidemiology has been in existence since 1766 when Bernoulli published a model that shows the increases in people's life expectancy if a portion is inoculated against smallpox [9].

Mathematical models are a representation of the reality that describes the evolution of an infectious disease over time. The models are useful to explore the different hypothesis, analyses and assumptions on a variety of scenarios. In consequence, the model should be as complex as needed to fulfil its purpose.

The modelling of the disease's dynamics is based in the division of the population into categories or compartments that represent a specific stage of the epidemic (e.g., susceptible, infected, recovered). The transition of individuals between compartment depends on the equation that describes the system (e.g., differential equation, transition probability) that considers the characteristics of the disease (e.g., virulence, transmissibility), the population dynamics (e.g., population size, effective contact networks) and the applied intervention strategies.

The model structure should consider the distinctions of disease transmission, with examples shown in Fig.2 The simplest is SI (Susceptible Infected), the susceptible individuals have no immunity, when they get infected they remain in this compartment their entire life. It is usually used to model HIV [14, 4]. SIR (Susceptible Infected Recover), the population is classified into three compartments Susceptible(S), Infected(I) and Recovered(R). The SIR model is adequate to describe the spread of viral diseases (e.g., Influenza, Measles, Chickenpox), but not to describe bacterial infections (e.g., Encephalitis) where recovered individuals do not remain immune and may be re-infected. The SEIR structure contrasts with the SIR in the addition of a latency period, represented by the compartment exposed(E). Individuals in this category who have had contact with an infected individual, but cannot yet transmit the pathogen to others. This structure can be used in diseases that have a long incubation period such as dengue hemorrhagic fever(DHF), chickenpox.

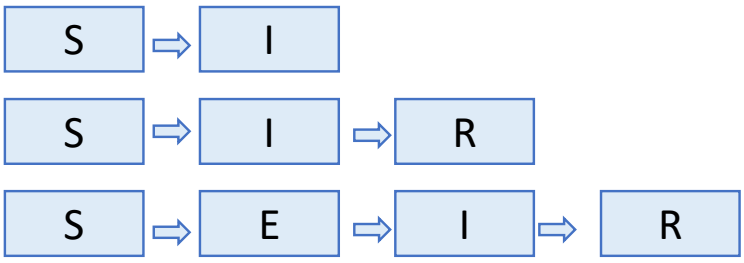

Fig. 2. Disease Model Structure. In the Figure: 'S' stands for Susceptible, 'E' stands for Exposed, 'I' stands for Infected and 'R' stands for Recovered.

The SIR model has been used to describe the Boarding school use case. Since we only have a dataset about the disease spread, different methods can be used to implement the SIR model to build a simulation, for instance in [28] authors used a Markov chain and in [16] authors used ordinary differential equations (ODE).

Generally, epidemic models can be classified into two categories: *deterministic* and *stochastic* models. In a *deterministic* model, the outcome is fully determined by initial conditions, parameter values and underlying equations. Deterministic model simplify the dynamics of the disease spread and the population composition enabling relatively fast computation of model outputs. However, these models do not capture the complexity of human interactions and behaviours. A *stochastic* model has some randomness in that the same initial conditions and parameter values leads to a different result each time the model is executed. Infectious disease transmission is a stochastic process. To create more realistic models, it is necessary to incorporate random elements at some level. For this reason, stochastic models are preferable to study a small population such as households [30]. In local social networks (e.g., schools, works places, households), the social structure is very important in how the disease is transmitted because it is based on the interactions between the individuals. Disadvantages of stochastic models are the high computational cost of simulating a disease spread in a large population. There are many types of stochastic epidemic model: network [13, 23] ,agent based [25] ,probabilistic [28, 11] among others.

Different approaches to model the Boarding school scenario can lead to a deterministic or a stochastic epidemic model. On the one hand, the Markov chain used in [28] has stochastic transition probabilities between states. On the other hand, using differential equations the authors of [16] arrived to a deterministic epidemic model.

Note that using a deterministic or stochastic model will make the environment deterministic or stochastic, which in turn affects the way an agent should be evaluated. If the environment is deterministic, the interaction results between

the agent and environment will always provide the same results. If the environment is stochastic, then it is necessary to run the interaction between the agent and the environment several times to evaluate how the agent performs in expectation.

3.2 Intervention Strategy

Interventions are a set of measures that aim to mitigate the impact of a disease spread by preventing disease propagation, by reducing the severity or duration of an existing disease, or by restoring functions lost as a result of the disease.

When an epidemic arises, the attention of decision makers shifts to questions such as: What should be the target population of the intervention be? What material and human resources are needed, and how much? What are the indicators of success (e.g. vaccine coverage, number of people in treatment, number of people infected)? In this way, the decision makers start analysing which mitigation measures may be more effective. In the research literature, some authors divide the interventions into pharmaceutical interventions (PHI) and non-pharmaceutical interventions (NPI). PHI includes treatment, vaccines and antiviral drugs. NPI include quarantine, isolation, travel restrictions and school and workplace closure

Another classification of interventions can be found in [15]:

- Preventive interventions: these type of interventions avoid new cases of infected individuals by interrupting the transmission of pathogens to susceptible human hosts, increasing their resistance to infection, or detecting the symptoms at the earliest possible stage such as immunisation, prophylaxis, screening.
- Treatment of disease: these type of interventions restore the normal health state of the infected individual and remove the infection.
- Reduce-transmission interventions: these type of interventions limit contact in order to prevent transmission, they may be applied to both individuals and groups such as quarantine, isolation, social distancing, personal protection and hygiene measures like hand washing, using masks or coughing into sleeves.

In an epidemic model, when interventions are added they are implemented to reduce the number of susceptible individuals (aka preventive interventions), reduce the length of time an individual is infectious (aka treatment of disease) or reduce the transmissibility by limiting contacts or introducing sanitary measures (aka reduce-transmission interventions), among other. In general, the application of an intervention implies a change to the epidemic model parameter values. For example, reduce-transmission interventions such as isolation (restricts the movement of infected individuals and separates them from those who are healthy) decrease the transmission rate parameter in the model. Note that, depending on the level of granularity of the model, several interventions of the same type can be applied to the epidemic model by modifying the same parameter. For

instance, in the Boarding school scenario which is described by the SIR model, modifying the transmission rate parameter can have several meanings in terms of interventions such as isolation of individuals, school closure or wearing masks.

3.3 State Representation

The state S is defined as the set of all possible states $\{s_1, s_2, \ldots, s_m\}$ where the size of the state space is $|S| = m$. The state completely encapsulates the status of the system and includes all the information necessary for the agent to make a decision, discarding other irrelevant aspects of the data. In a chess game, a state can be all possible chessboard configurations.

In [7] the properties of a good state representation is described. The authors note that a state is good if it has Markovian properties, that is, it is able to represent the true values of the current policy well enough for the agent to improve the policy. It is therefore capable of generalising the learned value-function to unseen states with similar futures, and has few dimensions for efficient estimation.

We can find examples of state representation for disease dynamics in the following works: In [8], the authors use a deterministic disease model divided into four compartments: susceptible (S), HIV-infected (I), AIDS (A), and dead (D). The state is represented by a three-dimensional vector indicating the number of individuals in compartment S, I, and A at period t. The state $s = [S_t, I_t, A_t]$. Because the assumption of the population size is fixed, the number of individuals in dead (D) can be calculated. In [29] the authors uses a disease model with six compartments, and the population has two age groups (adults, children). The authors used, as a state representation, a record of the stream of past actions and observations that they refer to as history.

3.4 Reward Function

Imagine that you are teaching your dog to do tricks. You will provide treats as a reward every time the dog does what you are teaching it. But If it fails to do the trick, the punishment is not giving any treats. Therefore, your dog will figure out the action that made it receive treats and repeat that action.

In the RL environment, the term reward is used to denote any type of feedback presented to the agent, which may be positive (reward) or negative (punishment). This feedback provides the agent with an idea of the actual value of being in a state. Therefore, the agent which receives the reward/punishment will improve itself.

The reward function in our problem represents the decision maker's goals. It can be single or multi-objective, and the reward function structure should reflect this. If a conflict between the objectives is generated, we need to prioritise the components of the function. The reward function punishes bad performance which is commonly considered in term of costs. It provides positive reinforcement for good performance which we will call benefit.

To understand how a reward function is defined, let's use simple examples with the boarding school scenario. The goal is to minimise the number of individuals infected during the epidemic; a simple reward is utilising the number of new infected in each time step. $r_t = NewInfected$. The problem with this reward approach arises when the agent learns that an intervention that reduces the infection, without evaluating the social and economic costs, is always the best.

Another goal can be to minimise the total expenditure cost during the epidemic. The reward function is represented as the sum of all cost incurred during each time step. $r_t = CostSick + CostIntervention + CostHospitalization$.

In this case, the challenge is to determine the value of each cost. In interventions such as school closure it is difficult to calculate the cost because it involves multiple variables such as tutors absent from work. An example using benefit is the maximisation of the health of the population. The reward function can utilise health units such as QALY (Quality-Adjusted Life Year), DALY(Disability-Adjusted Life Year) to represent the gain of health in the population. Thus, a reward function can be a combination of these cost and benefits components. It is widely used in health economic evaluation.

In [17, 3], the authors use a health outcome and economic cost performing incremental cost-effectiveness ratio (ICER) analysis to evaluate the impact of interventions strategies. In [29, 10], the net health benefit (NHB) approach[3] was adopted to evaluate the cost-effectiveness of the interventions. The NHB is defined by $gain - (cost/wtp)$, where $gain$ is the gain of QALY, $cost$ is the cost of the intervention and wtp is the willing to pay parameter indicating how much the decision maker is willing to pay for a QALY gain. An intervention is cost-effective if NHB expressed as QALYs is positive.

4 Agents for Epidemic Control

There are different methods to tackle the optimisation problem that is represented by the environment (previous section). Although we have focused on environment design up to now, in this section we offer a concise overview of different solution approaches, with a focus on RL-based solutions.

A possible solution approach is a brute force exhaustive search. This approach is computationally infeasible due to the large solution space (discussed above). Testing the cost-effectiveness of a pre-defined set of health policies, as in [21], could miss the optimal solution. Exploration-Exploitation methods could provide better solutions. These initially carry out a broader search and slowly focus on exploring policies with the most potential of being optimal. In [2] the authors compare the use of a Multi-Armed Bandit and a Genetic Algorithm to explore actions in a single state. The approach aims to find a one-shot policy recommendation for a five year period. In contrast, Dynamic Programming can consider the sequential value of actions (as in [28]). This method assumes the

[3] The NHB approach is further explained in [26].

optimisation component has access to a complete Markov Decision Process representation of the disease. This is usually not the case for more complex models and existing simulators. Finally, RL techniques can represent a more realistic scenario where only the actions and states are known. The RL agent learns the expected total reward from the current state by trial and error. Example approaches include Q-Learning and TD-Learning. To our knowledge this area has been under-explored.

5 Conclusion

In this paper, we described the different components needed to build environments to support epidemic control. First, we framed the problem of finding optimal policies as a reinforcement learning problem. Next, we focused on how to design environments, which are the components that characterise the epidemic problem setting in RL. We describe environments using the following components: disease model, intervention strategy, reward function, and state representation. Also, we mentioned challenges that need to be tackled for environment design and provided some examples.

We have shown that each of the environment components can be defined in a variety of ways. Components can also be combined in different ways to express different types of problems. Concretely specifying a component requires setting values to its parameters. This flexibility allows us to represent numerous epidemic scenarios but also brings a challenge related to reproducibility. If all the details of how the environment was built is not shared, it is very hard to replicate evaluation results. However, it is common for published works to miss certain details of their environment design as there are many details that need to be shared to entirely describe environments. A possible solution, that to our knowledge has not been explored in the case of epidemic control, is to share the exact environment definition in a platform such as OpenAI Gym [6].

As a next step, we aim to review the different types of agents that can be developed to solve our problem of interest. Our overall goal is to propose a new solution to find optimal health policies that can be used by decision makers. The work described in this paper is an attempt to define the environmental model suitable for representing an epidemic outbreak and the potential interventions that are available to decision makers. We hope that this work is useful to others that are also working on the same field.

Acknowledgement. This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2, co-funded by the European Regional Development Fund.

References

1. Anonymous: News and notes. British Medical Journal 1, 587 (1978), `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1603269/`

2. Bent, O., Remy, S.L., et. al.: Novel Exploration Techniques (NETs) for Malaria Policy Interventions. In 30th Conf. on Innovative Applications of Artificial Intelligence (Dec 2018)
3. de Boer, P.T., Kelso, J.K., Halder, N., Nguyen, T.P.L., Moyes, J., Cohen, C., Barr, I.G., Postma, M.J., Milne, G.J.: The cost-effectiveness of trivalent and quadrivalent influenza vaccination in communities in South Africa, Vietnam and Australia. Vaccine 36(7), 997–1007 (Feb 2018), http://www.sciencedirect.com/science/article/pii/S0264410X17318364
4. Bozkurt, F., Peker, F.: Mathematical modelling of hiv epidemic and stability analysis. Advances in Difference Equations 2014(1), 95 (2014)
5. Brauer, F., Castillo-Chavez, C., Castillo-Chavez, C.: Mathematical models in population biology and epidemiology, vol. 40. Springer (2001)
6. Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., Zaremba, W.: Openai gym. arXiv preprint arXiv:1606.01540 (2016)
7. Böhmer, W., Springenberg, J.T., Boedecker, J., Riedmiller, M., Obermayer, K.: Autonomous Learning of State Representations for Control: An Emerging Field Aims to Autonomously Learn State Representations for Reinforcement Learning Agents from Their Real-World Sensor Observations. KI - Künstliche Intelligenz 29(4), 353–362 (Nov 2015), https://doi.org/10.1007/s13218-015-0356-1
8. Cosgun, , Esra Büyüktahtakın, : Stochastic dynamic resource allocation for HIV prevention and treatment: An approximate dynamic programming approach. Computers & Industrial Engineering 118, 423–439 (Apr 2018), http://www.sciencedirect.com/science/article/pii/S0360835218300251
9. Dietz, K., Heesterbeek, J.: Daniel bernoulli's epidemiological model revisited. Mathematical biosciences 180(1-2), 1–21 (2002)
10. Edwards, C.H., Tomba, G.S., Kristiansen, I.S., White, R., de Blasio, B.F.: Evaluating costs and health consequences of sick leave strategies against pandemic and seasonal influenza in norway using a dynamic model. BMJ open 9(4), e027832 (2019)
11. FROST, W.H.: Some conceptions of epidemics in general. American Journal of Epidemiology 103(2), 141–151 (1976)
12. Giesecke, J.: Modern Infectious Disease Epidemiology (2017), https://www.crcpress.com/Modern-Infectious-Disease-Epidemiology/Giesecke/p/book/9781444180022
13. Gross, T., D'Lima, C.J.D., Blasius, B.: Epidemic dynamics on an adaptive network. Physical review letters 96(20), 208701 (2006)
14. López-Cruz, R.: Structured SI epidemic models with applications to HIV epidemic. Arizona State University (2006)
15. Magnusson, R.: Advancing the right to health: the vital role of law. Advancing the Right to Health: The Vital Role of Law, World Health Organization, Switzerland (2017)
16. Martcheva, M.: An introduction to mathematical epidemiology, vol. 61. Springer
17. Milne, G.J., Halder, N., Kelso, J.K.: The cost effectiveness of pandemic influenza interventions: a pandemic severity based analysis. PloS One 8(4), e61504 (2013)
18. Murray, J.D.: Mathematical biology. I, volume 17 of Interdisciplinary Applied Mathematics. Springer-Verlag, New York (2002)
19. Napoli, C., Fabiani, M., Rizzo, C., Barral, M., Oxford, J., Cohen, J., Niddam, L., Goryński, P., Pistol, A., Lionis, C., et al.: Assessment of human influenza pandemic scenarios in europe. Eurosurveillance 20(7), 21038 (2015)

20. Organization, W.H., et al.: Pandemic influenza risk management: a who guide to inform and harmonize national and international pandemic preparedness and response. Tech. rep., World Health Organization (2017)
21. Perlroth, D.J., Glass, R.J., et. al.: Health outcomes and costs of community mitigation strategies for an influenza pandemic in the united states. Clinical Infectious Diseases 50(2), 165–174 (2010), http://dx.doi.org/10.1086/649867
22. Qualls, N., Levitt, A., Kanade, N., Wright-Jegede, N., Dopson, S., Biggerstaff, M., Reed, C., Uzicanin, A., Group, C.C.M.G.W., Group, C.C.M.G.W., et al.: Community mitigation guidelines to prevent pandemic influenza—united states, 2017. MMWR Recommendations and Reports 66(1), 1 (2017)
23. Rahmandad, H., Sterman, J.: Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. Management Science 54(5), 998–1014 (2008)
24. Reed, C., Biggerstaff, M., Finelli, L., Koonin, L.M., Beauvais, D., Uzicanin, A., Plummer, A., Bresee, J., Redd, S.C., Jernigan, D.B.: Novel framework for assessing epidemiologic effects of influenza epidemics and pandemics. Emerging infectious diseases 19(1), 85 (2013)
25. Siettos, C., Anastassopoulou, C., Russo, L., Grigoras, C., Mylonakis, E.: Modeling the 2014 ebola virus epidemic–agent-based simulations, temporal analysis and future predictions for liberia and sierra leone. PLoS currents 7 (2015)
26. Stinnett, A.A., Mullahy, J.: Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Medical decision making 18(2_suppl), S68–S80 (1998)
27. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT press (2018)
28. Yaesoubi, R., Cohen, T.: Dynamic health policies for controlling the spread of emerging infections: influenza as an example. PloS one 6(9), e24043 (2011)
29. Yaesoubi, R., Cohen, T.: Identifying cost-effective dynamic policies to control epidemics. Statistics in Medicine 35(28), 5189–5209 (Dec 2016)
30. Yarmand, H., Ivy, J.S.: Optimal intervention strategies for an epidemic: A household view. Simulation 89(12), 1505–1522 (2013)