ETH zürich

A Compartmental Model Utilising Reinforcement Learning to Guide Country-Specific Vaccination Prioritisation

Class report for
Complex Social Systems: Modeling Agents, Learning and Games

Artem Anchikov, Nicola Brunello, Claire Antoinette Dick, Héctor Adrian Ramírez Lara & Ross Straughan

Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland
{aanchikov,nbrunello,dickcl,hramirez,rstraughan}@student.ethz.ch

Supervisors:
Prof. Dr. Dirk Helbing
Dr. Nino Antulov-Fantulin
Thomas Asikis

Novemeber 15, 2021

Abstract

The global pandemic caused by the highly infectious disease COVID-19 has led to millions of deaths world-wide, whilst also inflicting huge economic damage internationally. In this paper, a compartmental model was generated that models the dynamics of the disease for various countries to show the effectiveness of various vaccination strategies. In an attempt to aid in vaccination policies for countries with various age demographics, a reinforcement learning algorithm was created that grants vaccinations to various age groups to limit deaths whilst also preserving the economy. Although the reinforcement learning algorithm failed to create a more effective alternative to current vaccination strategy, it was shown to attempt to vaccinate some younger individuals to preserve the economy. To the knowledge of the authors, this is the first attempt to create a computational model that utilises both compartmental modelling and reinforcement learning in order to generate vaccination programmes that target specific age groups for various countries of different age demographics.

Contents

List of Figures

1 Introduction

As a highly contagious disease, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has lead to a global pandemic infecting millions with the disease COVID-19. Due to its high infection rates, COVID-19 led to indeterminable amounts of illness, hospitalisations and the death of over 5 million people as of November 2021 [1]. Indirectly, the pandemic led to a multitude of problems including social isolation [2], severe economic stagnation [3], in addition to many direct and indirect deaths; evident with Switzerland experiencing a 8.8% increase in mortality during 2020 [4]. Globally, there has been a drive to vaccinate the general population, based on the principle of targeting the oldest of society first [5–7]. In an attempt to validate this rationale, offer new insights and to guide future vaccination programmes, various computational models have analysed prioritising individuals based on age [8–13], sociability [8, 9, 12], serostatus [10], sociodemographics [11] and occupations [11, 14]. Not all are congruent with the status-quo of current vaccination strategies, with some demonstrating that vaccination prioritising the eldest of society is not always optimal [8, 9, 12]. The alternative approach of indirectly protecting the most vulnerable in society by vaccinating others may not appear intuitive at first, however modelling has shown that indirect protection by vaccinating younger, more social individuals is superior when transmission is low [8] and vaccine effectiveness is high [12].

Much of the computational modelling research that has been conducted on the COVID-19 pandemic has a bias towards more economically developed countries, such as those in North America and Europe [9–14]. Conclusions made from data in such countries may not be directly translatable to countries with vastly different demographics. The age distribution between countries is often linked to the economic activity of a country, with poorer countries having a relatively larger younger population [15]. Accordingly, it is necessary to validate whether the vaccination strategies that have been deployed in countries with an older population are optimal for an age demographic skewed to a younger average age. As 77% of Ethiopian, Malawi, Nigerian and Ugandan households have seen a decrease in economic income since the start of the pandemic, 30% of whom cannot now consistently access medicine and staple foods, there is a need to tailor vaccination roll-out based on country-specific demographics especially as the socioeconomic impacts are more drastic [16]. Consequently, as there is still a limited understanding of how age demographics impact the effectiveness of current vaccination programmes, there needs to be further research into tailoring vaccination strategies to mitigate economic damage and limit mortality.

In this work, we propose a compartmental model that derives the most optimal vaccination policy based on age through the means of reinforced learning. This approach is seemingly the first computational model to investigate how vaccination policy should be tailored based on the age demographics of a country determined by reinforced learning. Our reinforcement model aims to reduce deaths and economic burden, two highly important variables for COVID-19 and potential future pandemics. The models are representative of real world demographics, and as such, offer insight on how vaccinations should be prioritised based on age. The compartmental model created is able to simulate pandemic scenarios and simulate how various vaccination strategies impact the number of infections, deaths and economic output. Utilising reinforcement learning, an optimised vaccination strategy was able to be developed for specific countries based on their respective age demographics. Undoubtedly, future work will need to be performed to correct inaccuracies, nevertheless, the computational model offers a foundation for simulating an often unaddressed question of how vaccination policy should be tailored to specific population demographics.

2 Related Work

Compartmental modelling is often utilised in mathematical modelling of infectious diseases with the population being split according to labels such as SEIR with the letters referring to an individual being susceptible, infectious, exposed or recovered respectively [9, 12, 17–20]. Some models, have attempted to vaccinate based on local infections inside a network, which is unlikely to work effectively in the reality due to the complexity in tracking true positive cases and the rights of an individual to decline the vaccine [17, 21]. A more useful result from such a model would be to instead target specific demographics, reducing the overhead required to track, communicate and the eventual immunisation of the individuals, significantly reducing the logistics of the problem. Such models have already investigated the effectiveness of various vaccination strategies [12] and determined that vaccination prioritising the eldest results in more deaths when vaccines have effectiveness of 50% or higher [12]. In a model that accounts for age-specific sociability, Li *et al.* [9] highlighted that targeting the most social in society could reduce infections by 10% versus targeting the most elderly first [9]. The two aforementioned papers both account for the increased sociability of younger individuals and thus the increased likelihood to spread COVID to others. In these papers, various vaccination strategies were manually selected which is unlikely to create the most optimal strategies unlike a strategy that utilised reinforcement learning.

In reality, pandemic modelling is inherently stochastic in nature; an extreme example being so called superspreaders [22]. It is possible to implement both a deterministic and stochastic approach to compartmental modelling, with the former allowing for the use of ordinary differential equations to benefit from dynamical systems. However, the rigid structure of a deterministic approach does not allow for a broader generalisation of the problem and reflect the non-linearity of infections and will accordingly lack more realistic infectability profiles [23]. Furthermore, they are often not representative of epidemiological events such as extinction and gradual fade out [24]. Often, these factors are deemed insignificant or ignored, however there has been an increasing amount of literature that accounts for these with the changing of states being based stochastically rather than deterministically by utilising stochastic differential equations rather than ordinary differential equations. [14, 18, 19].

Reinforcement learning has been used extensively to determine optimal strategies to help mitigate the effects of pandemic [17–20, 25]. In such techniques, a reward system drives the simulation to create solutions that can limit a number of variables including deaths [17, 20, 25], economic burden [17, 20, 25], infection [19, 25] and active cases [20]. A challenge often faced is the balancing of the various goals. Both Kerrigan [17] and Ohi *et al.* [20] implemented thresholds for the largest amount of disease spread at which point economic consideration is not considered and the goal is simply to reduce the number of deaths [17, 20]. However, this approach can be viewed as unsatisfactory when one regards the time to reach immunity from vaccinations, often after at least two vaccinations with a spacing of 4-12 weeks between them [26], and thus the approach could be seen as reactionary rather than preventative. Meanwhile other studies have modelled a particular vaccination rate based arbitrarily [9] or on real world data [12] which mirrors the continuous vaccination administration we have seen in reality. Studies utilising compartmental models thus far have mainly focused on limiting the spread of diseases through lockdown measures [20, 25], travel policy [25] and school closure [18]. There are multiple possible algorithms that can be used for reinforcement learning, however Proximal Policy Optimization (PPO) has already proven to be effective with a number of authors illustrating how it can be used to optimise strategies to reduce the impact of COVID-19 [17–19]. There has seemingly been little to no research on using reinforcement and compartmental modelling to determine vaccination policy based on allocating vaccinations according to age.

3 Aims

Based on the review of literature, there are a number of areas where current modelling of pandemics could be further improved. It is the aim of our project to introduce a method to create a compartmental model that determines the optimum vaccination strategy based on a country's age demographics with the aid of the reinforcement learning. This includes:

1. Development of a virtual compartmental model that simulates the spread of COVID-19 in a population.
2. Comparing the common approach of vaccinating the most elderly in society first to an optimised approach created with reinforcement learning.
3. Illustrating how the vaccination programme of individual nations may need to be tailored according to their respective age demographics.

4 Materials and Methods

To gain some insights on a more efficient vaccination policy to contain the spreading of the pandemic, we define an age-structured probabilistic network model. This allows the observation of what effect the age of each individual has on the definition of the vaccination policy, particularly taking into account the different risk of death that is associated with each age group. To do so, a network graph static over the time steps was employed, where the nodes have age as a property.
The network model is drawn as a regular undirected graph with nodes with a degree of $k = 3$ to compare with common metrics on other literature [27] with the use of the Python module *Graphtool*. Individuals are nodes, and the connections between them are edges. The use of *Graphtool* allows efficient handling of the changes at each time step, as the characteristics of each individual and neighbor filtering can be easily iterated over using built-in methods. In particular, the characteristics of each node are the state of the node, the age group to which it belongs, and the economic value that the individual brings to the community. The economic value is set to either 1 or 0, depending on whether the individual is of working or not. In the model, the individuals are considered to be contributing to the economy from 20 to 75 years of age.

4.1 The Model States

One of the most common methods for epidemic simulation is the SIRS (Susceptible-Infected-Recovered-Susceptible) model. By employing this model, it is possible to observe the dynamics of epidemic spreading, having the nodes in one of the three states: Susceptible (S), Infected (I), Recovered (R). The transfer from a state to the other is regulated by probability rates, namely the infection rate β, the recovery rate θ, and the rate at which recovered people transition back to susceptible.

The model we decided to employ was developed from the SIRS model but does not correspond to it entirely. Contrary to the typical SIRS model, our model does not include a rate for the subject to become susceptible again after recovery, as we deemed it inconsequential for the time scale observed. To model vaccination status which leads to modified infection and recovery rates, this was added as two additional states: the Vaccinated + Susceptible (VS) state and the Vaccinated + Infected (VI) state. Furthermore, we added a Dead (D) state. These states make the interaction between nodes more articulated, as illustrated in figure 1.

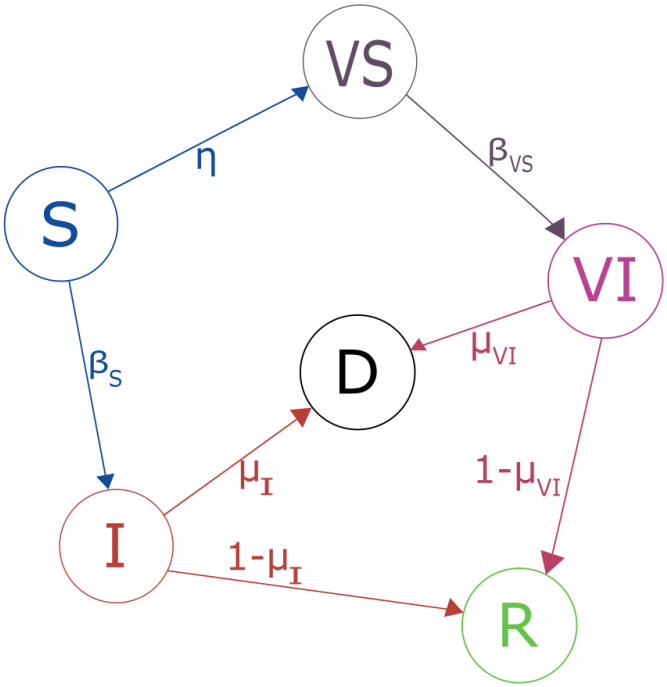

Figure 1: The possible change in states of a node, with the respective rates. The rates are written with reference to the state prior to the change.

As a result, there were also new rates included: the rate of vaccination η, and the rate of death μ. On the other hand, in the model, there is no real rate of recovery, as the infected individual will either die or recover, taking into account the duration of the time steps that we implemented. Each of the rates was referred to the initial state the node was in, and they are defined to be age-dependent. Moreover, in the model, it was decided to exclude all possible deaths unrelated to the pandemic that we are observing.

4.2 Model Evolution

Traditional models of spreading dynamics on networks are memory-less Markovian, as an individual contracting the virus and recovering from it is viewed as a Poisson process [28]. Although social interactions exhibit correlation and memory (non-Markovian effects) and therefore are not described by exponential time distributions [29], in some certain conditions Markovian dynamics are equivalent [30]. This work simplifies the dynamics into a Bernoulli memory-less process (which corresponds to a Poisson process in discrete time arrivals [31]) to focus on the initial feasibility of an agent discriminating vaccination by age. Further work can be done to shift towards exponential inter event-time distributions. Gillespie and kinetic Monte Carlo could be used to provide these dynamics to the death and infection event times [27]. Initial attempts were preform to obtain the time evolution of death and infection through KMC. However time evaluations of different age groups would conflict due to their different infection rates and weather or not a vaccine protects them. Further research could be done to determine weather the average of the time steps or if a competition for the next time step would provide a suitable model, but this effort falls outside the scope of this work.

The updating function passes through each node in random order, and determines its state change according to the transition rates. For each transition, a random number is drawn using *numpy.random*. If the random number is lower than the age group specific transition rate for that transition, the state changes. A node can perform maximum one transition per time step, it can not e.g. go directly from susceptible to recovered. This also means that a node who becomes vaccinated will not become infected within the same time step. For determining whether a 'Susceptible' node becomes infected, we iterate over all neighbours of the node, defined by the edges in *Graphtool*. For each of these neighbours, a random number is drawn and if at least one is below the infection rate threshold, the node moves from the 'Susceptible' state to 'Infected'. The 'Infected' state can transition either to the 'Dead' state or the 'Recovered' state. In the model, the transition to the 'Recovered' state is implemented as a consequence of the fact that the node does not transition into the 'Dead' state. so after one time step, each 'Infected' node will have transitioned either to 'Dead' or 'Recovered.The model is initialized with all individuals in the 'Susceptible' state. One random node is 'Infected' to start the pandemic. In order to not have the pandemic end after just one time evolution, this initial 'Infected' node experiences no state transition in the first time step. Otherwise the 'Infected' node would immediately recover or die, ending the pandemic.

4.3 Timescale

In order to have time steps that allow us to observe any possible change between states in just a single step, we decided to assign each of them a duration of 20 days. In particular, this was done by considering the time to death to be of 18 days on average, as stated in [32] and that people who do not die will eventually recover. In this way, during one time step, each infected individual either recovers or dies, according to the probability of death of the age group the individual belongs to. The duration of the simulation is 400 days, corresponding to 20 time steps. This duration allows to infer some conclusions about the development of the pandemic with a reasonably long timescale, it being longer than one year.

4.4 Choice of Rates

As the model was age-stratified, it was decided to have different transition rates for different age groups. These rates, indicated in figure 1, are required to run the simulation. The rationale behind the stratification is that different age groups will have different lifestyles, as well as immune defences, and therefore the probabilities related to the evolution of the infection and subsequent eventual recovery would have to be different to reflect it. In the model, the rates were defined as constants for the full duration of the simulation. The exception to this is the rate of vaccination, which is influenced by the choices of the agent.

The infection rate β is defined by real-world data taken from Davies *et al.* [33] and Choi and Shim [34], respectively for non-vaccinated individuals and vaccinated individuals. The specific rates are illustrated in the table 1 and kept constant throughout the simulation. In the model, the possibility of infection of a node does not depend, in itself, on the amount of neighboring infected nodes. Each of the connections with nodes in the 'Infected' state of a node in 'Susceptible' or 'Vaccinated Susceptible' states is considered individually and has its probability. It is worthy of notice that the rate of infection depends solely on the receiver node, and not on the ones that are infecting it. This consideration was made to generalize the model, and to reduce ulterior complications. For example, the kind of connections that two nodes share is not defined, so there is no specific information on the duration of the contact, or if it was in a closed space or in the open air.

Initial state of the node	Infection rates β																			
Susceptible	0.4	0.38	0.38	0.79	0.79	0.86	0.86	0.8	0.8	0.82	0.82	0.88	0.88	0.74	0.74	0.74	0.74	0.74	0.74	0.74
Vaccinated Susceptible	0.08	0.076	0.076	0.158	0.158	0.172	0.172	0.16	0.16	0.164	0.164	0.176	0.176	0.148	0.148	0.148	0.148	0.148	0.148	0.148

Table 1: Infection rates across age groups

The death rate μ was obtained from real-world data as well and differentiated between the case in which the individual is vaccinated [35] and the one in which they are not [36]. The specific values can be seen in table 2. An assumption of the model is that every node in the "Infected" state can transition to either the "Dead" or the "Recovered" states over the course of a time step (corresponding to 20 days). Even in the eventuality that the recovery time was longer, it was assumed that the infection was identified during that period, and that the individual went into isolation, therefore preventing further infection of other nodes.

Initial state of the node	Death rates μ																			
Infected	0.003	0.003	0.003	0.003	0.003	0.003	0.005	0.005	0.011	0.011	0.03	0.03	0.095	0.095	0.228	0.228	0.296	0.296	0.296	0.296
Vaccinated Infected	0	0	0	0	0	0	0	0	0.006	0.00132	0.0011	0.003	0.003	0.0095	0.0095	0.0228	0.0228	0.0296	0.0296	0.0296

Table 2: Death rates across age groups

The vaccination rate η is the probability with which an individual belonging to an age group will get vaccinated. This depends, in turn, on two parameters: acceptance rate of the vaccine, and availability of the vaccine itself. While the former is defined as constant [6], see table 3, the latter is defined as a policy which the user can fix or allow the reinforcement learning model to decide at each time step, based on the reward from the previous experience that the model has collected.

Initial state of the node	Vaccine acceptance rates η																			
Susceptible	0	0	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.517	0.517	0.613	0.613	0.774	0.774	0.834	0.834	0.834	0.834

Table 3: Vaccine acceptance rates across age groups

4.5 Age Demographics

The age distribution of Switzerland was obtained with data from the United Nations [37]. Ages were categorised into 20 age groups containing 5 years each and represented as a percentage of the total population. In order to compare vaccination strategies based on age demographics, two more countries were chosen. Japan and Niger were selected as they represent the oldest and youngest populations in the world [37]. The age distributions can be seen in figure 2.

4.6 Assumptions

Despite the reference to real-world data, the model has some limitations. In the definition of its structure, in order to reduce the complexity, some assumptions were made that limit its resemblance to the real world. With respect to the compartmental model that was implemented, it was assumed to be static with a fixed amount of nodes at the start, and the only variation accepted with regards to it is death or recovery due to the pandemic, at which point a node loses all its connections. As the model is static, there is no formation of new connections over the time steps. This would be unrealistic in a real-world scenario, as it would imply not having any interaction with new individuals for the full duration of the simulation. On the other hand, assuming that there are no newborns during any iteration of the simulation, no deaths unrelated to the pandemic, or contact with individuals outside of the community limits the resemblance that the model can have with the real world. In particular, the last point prevents

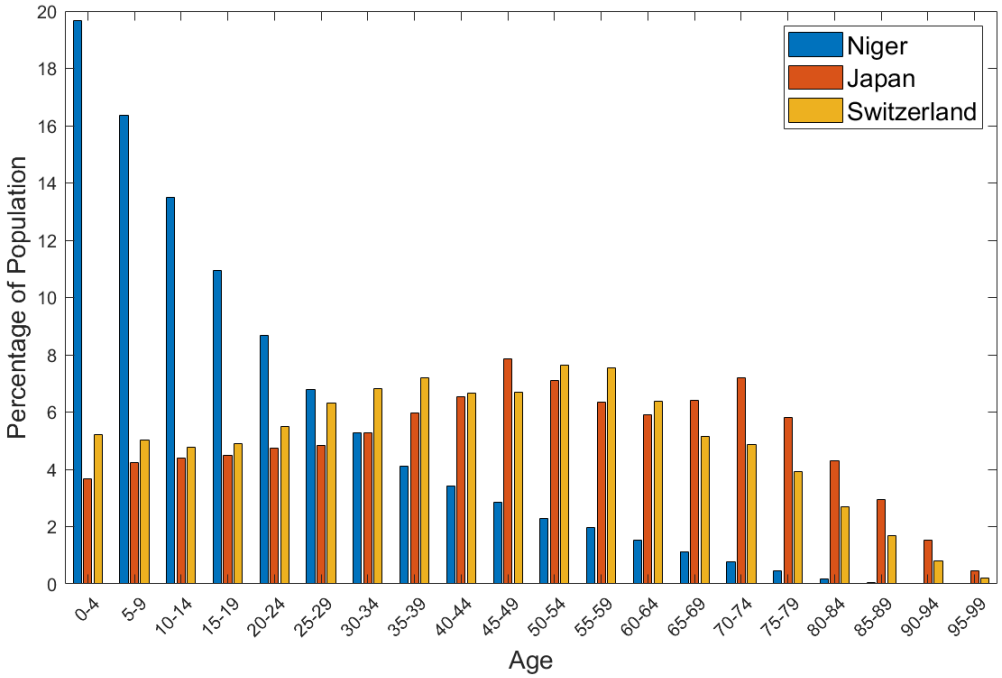

Figure 2: Normalized age demographics of Niger, Japan and Switzerland.

the possibility of having new nodes in the graph that were not there initially, and that could be infected or a new origin for infection.

Looking at the states and their interaction, the model analyses each of the nodes belonging to the same age groups in the same way. Assuming that the individuals belonging to the same age group are identical reduces the complexity of the model greatly, and standardizes the treatment of each node in the update function. Other assumptions that were made are that the nodes either die or recover in 20 days or less, that there are no spontaneous infections, and that the recovered nodes cannot transition to the infected state anymore. The rationale supporting the first is that nodes that are in the "Infected" state will eventually die or recover, and that the average time to death is of 18 days [32]. By having time steps corresponding to 20 days, nodes that are not in the "Dead" state at the end of this time period will eventually recover, so they transition to the "Recovered" state, as explained in the subsection 4.4. The absence of spontaneous infection is supported by the closed environment that is defined by the graph. Finally, the choice of not giving the possibility to nodes in the "Recovered" state to transition to the "Infected" one is supported by the heightened resistance to the infection that is developed during the infection period. Despite being a simplification, as recovered individuals can still get infected, the model is focused on allocating vaccines, and so the degree of resistance given by the recovery from the infection is considered enough to not be susceptible anymore.

The last assumptions were made on the nodes and the evolution of the model. In particular, the supply of vaccines per time step was defined as a constant throughout the simulation, and all nodes in the work force contribute equally to the economy. Both of these assumptions greatly reduce the complexity of the model itself. Defining the supply of vaccines constant at each time step, decouples vaccination roll-out from the economy of the community that is under scrutiny. In real life, a community with greater economic power would be able to access a larger supply of vaccines. Furthermore, when the majority of the population is either recovered or vaccinated, a large amount of vaccines per time step is not needed anymore. On top of not diversifying the economic output between different working age groups, it is also fixed for all individuals within an age group. Our model also does not consider one individual changing their economic output, which would happen by changing jobs for example, but instead only changes the economic output for when an individual becomes sick or dies. This does not correspond to a real-world scenario, but once again helps limit the complexity of the model.

4.7 Reinforcement Learning

OpenAI demonstrated that their new developed PPO algorithm was superior to many other algorithms due to the need of lower sample sizes, increased efficiency in learning for each time step and the algorithm being less prone to noise in data [38]. Indeed, as previously stated, PPO has been used effectively in reinforcement learning for compartmental models of COVID-19 [17–19]. Due to ease of implementation and previously shown effectiveness, PPO was chosen as the reinforcement algorithm to utilise.

The reward function of this model is based on both the economy and the number of deaths. We deemed the number of infections to not be an important parameter as this does not align with the ultimate goal of our model and already indirectly contributes through the economic status. The reward function R is defined as:

$$R = \frac{E_t}{P} - \frac{2D_t}{P} \tag{1}$$

Where:

1. P is the total number of the population i.e. the number of nodes
2. E_t is an indication of the status of the economy at each time step, representing the total number of workers that are not infected and are therefore able to contribute towards the economy. More workers results in a higher reward. Workers are classed as anyone between the ages of 20 to 65 years.
3. D_t is the number of deaths at each time step. Fewer dead results in a higher reward.

It was decided to weight the number of deaths more than the number of workers as ultimately it is desirable for our model to reduce the deaths of all individuals, including those that are not of a working age. However, incorporating the economy was deemed to be an important factor, especially for weaker economies where individuals can struggle to access food and medicine [16]. This arbitrary formulation could be altered to have the desired output. The reinforcement learning acts at the end of every time step. An epoch state is defined when the pandemic reaches a steady state, meaning there are no new infections being generated. A new random pandemic scenario is then generated for the reinforcement learning algorithm to learn from. If steady state is not reached, the epoch is defined as 50 time steps. In total, the reinforcement learning algorithm learns from 0.7 million pandemic simulations.

4.8 Simulating the Models

A series of tests were performed to evaluate the effectiveness of the compartmental model and the reinforcement learning. Initially two separate scenarios were simulated with the compartmental model, in the absence of reinforcement learning. These aforementioned scenarios were simulating the various compartmental states, as well as the economic activity, of the populations of Niger, Switzerland and Japan for a scenario in which no vaccinations are administered and when vaccinations are offered to the eldest in society first. There are various similar policies to vaccinate the eldest in society first, however, in this model we aim to match the approach given by the Swiss government in which the first age group to be given prioritisation were over the age of 70 years old [39]. For the purposes of this report, this will be known as the *conventional* vaccination policy. Subsequent simulations were then performed utilising the reinforcement learning, in order to compare the optimised vaccination strategy created to the one used in Switzerland. For each simulation scenario, 100 simulations were performed due to the inherent randomness that each simulation has. Data used in the results section is an average of the 100 simulations performed for each scenario.

5 Results

As previously stated in section 4.8, simulations were performed for policies in which no one is vaccinated (figure 3), the conventional vaccination policy in which the eldest are first vaccinated (figure 4) and then vaccinated according to the policy developed by reinforcement learning (figure 5). As all three countries saw a decrease in the average number of deaths with the conventional vaccination policy versus no vaccination policy (refer to table 4), it can be deduced that the compartmental model was able to effectively model the health benefits of the vaccine. As seen in table 4, the older populations of Japan and Switzerland experienced far higher deaths than the younger population of Niger, highlighting how pandemic responses ought to take into account the country-specific age demographics.

No Vaccination	Average Deaths	Average Infections
Niger	9.16 ±3.23	848.09 ±14.43
Switzerland	50.39 ±7.78	918.19 ±92.79
Japan	74.46±8.32	929.53 ±10.68

Conventional Vaccination Policy	Average Deaths	Average Infections
Niger	6.01 ±2.41	830.98 ±119.44
Switzerland	23.80 ±5.61	829.20 ±84.52
Japan	32.71 ±7.89	791.56 ±81.41

RL Vaccination Policy	Average Deaths	Average Infections
Japan	56.33 ±12.53	801.47 ±141.94

Table 4: Average number of deaths and infections with corresponding standard deviations for each simulation.

As it can be seen in figures 3 and 4, the average age of death is significantly reduced for the conventional vaccination policy, highlighting that the vaccinations are effective at protecting the eldest in society with the conventional vaccination policy. As observed in the real world, with no vaccination policy, the most vulnerable people die at a higher rate. For all of the simulations, the transition from states with respect to time are similar. The number of people in the *susceptible* state decreases with time as people are infected, die, recover or transition to *susceptible whilst vaccinated*. Prominent in all simulations is the initial dip in the economy where a proportion of the work force has to isolate due to being infected. This decrease in productivity then rebounds after individuals recover. Niger, in contrast to Switzerland and Japan, has a relatively young population, as seen in the distinct skew in figure 2. For this reason, one can expect that the deaths per population to be lower than a country like Japan. This relationship has effectively been portrayed in the results with Niger consistently having a lower death count that that of Switzerland and Japan.

As highlighted in figure 4, the amount of individuals that are in the vaccinated state is relatively low in comparison to reality. The pandemic ends with less than the 20% of the population vaccinated, likely as the transitions equations are not perfectly fine tuned. In both the vaccinated and non-vaccinated scenarios, it can be concluded that the pandemic ends due to an immunity gained mainly from prior infection rather than immunity acquired via vaccination. Another somewhat unexpected result is the fact that the number of infections for both the simulation with and without vaccination is similar. With the conventional vaccination policy, there was only a 2.0%, 10.1% and 16.0% decrease in infections for Niger, Switzerland and Japan respectively. In reality, inoculation with vaccines produced by PfizerBioNTech and Moderna has been shown to reduce infection by up to 90% [40].

The reinforcement learning develops a vaccination policy that differs from the conventional vaccination policy. As seen in figure 5, the economic output is virtually the same as seen in figure 4 for Japan. In figure 6, the distribution of vaccines according to various age groups is illustrated with respect to time, illustrating that the RL model more frequently gives younger individuals vaccinations in order to attempt to preserve the economy. As a consequence of this decision, it can be seen in figure 5 that more of elderly people population are left exposed to the virus leading to higher deaths. Nevertheless, the algorithm recognises that the eldest in society should be prioritised more than younger people in order to avoid excessive deaths. In spite of that, the RL model develops a higher average death count, leading to the deaths of 56.33 on average, versus the conventional policy that results in 32.71 deaths on average, whilst also generating a poorer performing economy. Without doubt, if such a vaccination policy were to be deployed in the real world, it would be deemed unsatisfactory.

No Vaccination Policy

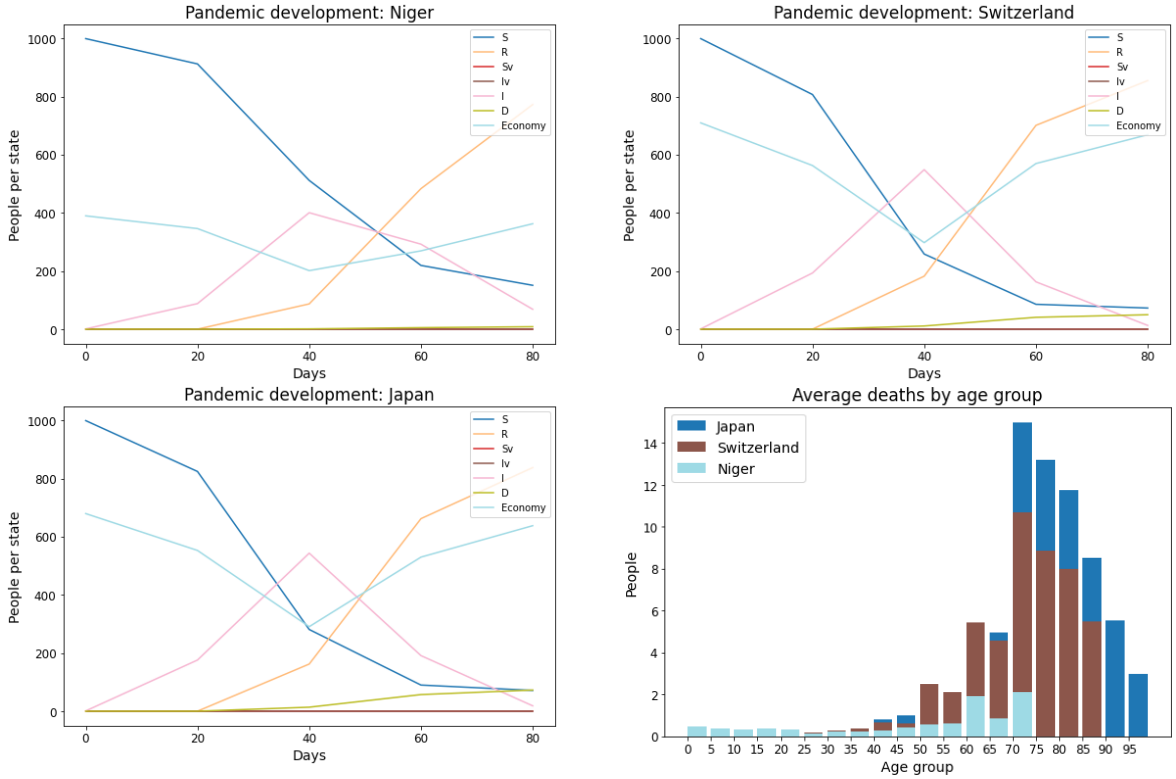

Figure 3: Simulation of pandemic development for a simulation with no vaccines administered, where S=susceptible, Sv=susceptible and vaccinated I=infected, Iv=infected and vaccinated, R=recovered, D=dead and *Economy* represents the number of working adults.

Current Vaccination Policy

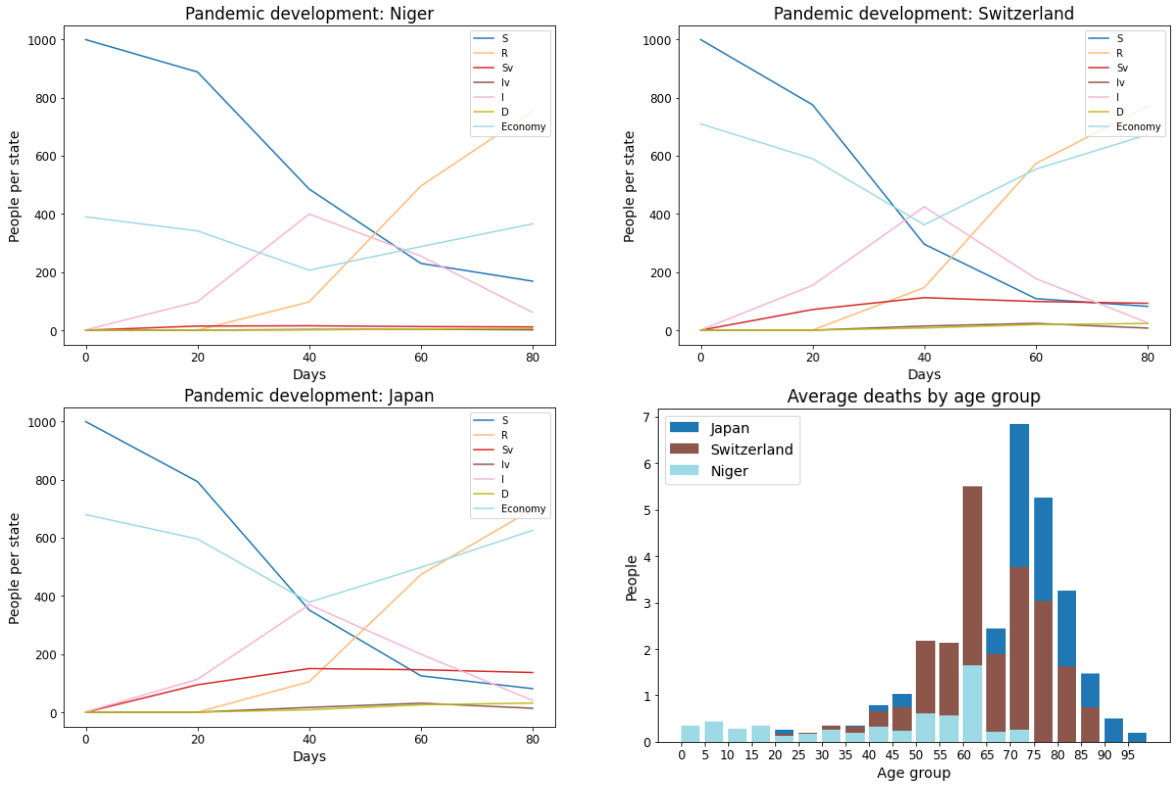

Figure 4: Simulation of pandemic development for a simulation with the conventional vaccination policy, where S=susceptible, Sv=susceptible and vaccinated I=infected, Iv=infected and vaccinated, R=recovered, D=dead and *Economy* represents the number of working adults.

RL Optimised Vaccination Policy

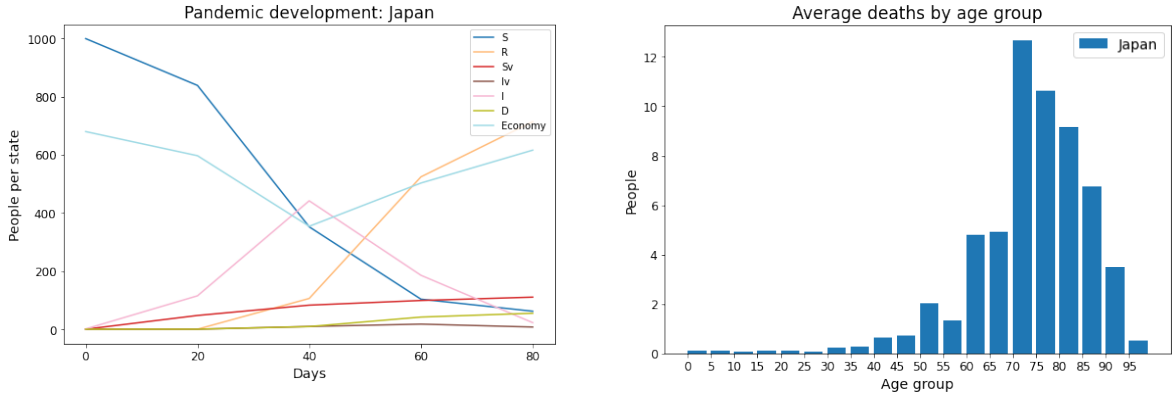

Figure 5: Simulation of pandemic development for the RL optimised vaccination policy, where S=susceptible, Sv=susceptible and vaccinated I=infected, Iv=infected and vaccinated, R=recovered, D=dead and *Economy* represents the number of working adults.

Vaccination Distribution with Time

Conventional Vaccination Policy **RL Vaccination Policy**

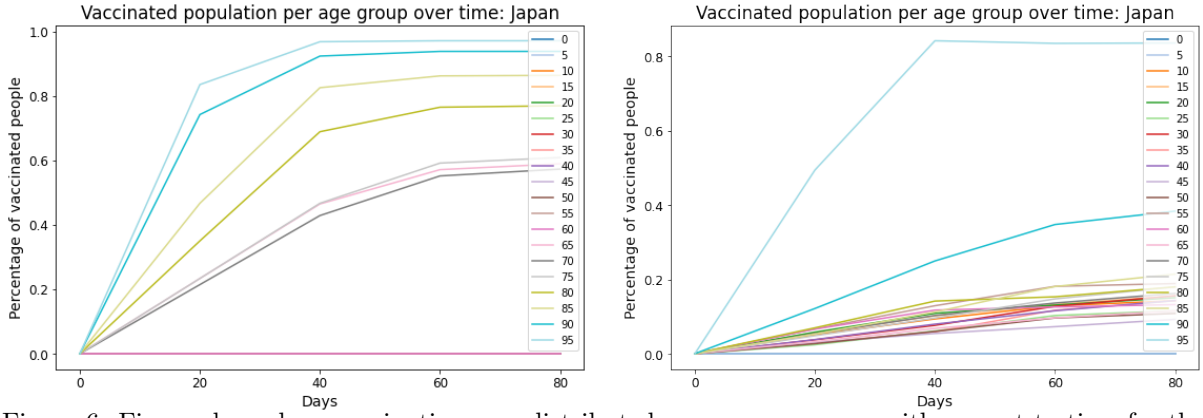

Figure 6: Figure shows how vaccinations are distributed among age groups with respect to time for the conventional vaccination policy (left) and the vaccination policy determined by reinforcement learning (right).

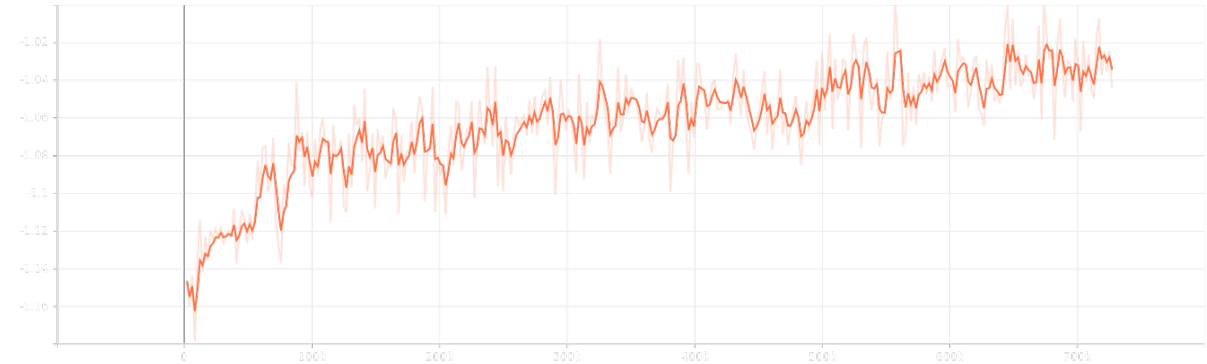

Figure 7: Graph shows how the reinforcement learning algorithm increases the reward it earns with time.

6 Discussion

6.1 Current Model

The compartmental model that was created demonstrates the key characteristics of a pandemic such as COVID-19. With all simulations, there is an initial increase in the amount of infections, which then start to level off as people recover. With the introduction of vaccinations, the rate of deaths is reduced significantly. There are however a number of ways in which the simulation is not realistic in modelling COVID-19. The first being that the pandemic only lasts for a few months even without vaccinations and, needless to say, this does not conform with the reality of the pandemic. Similarly, the number of vaccinations that are administered are far fewer that what are being given to in many economically developed countries. Indeed, to reach herd immunity, the number of individuals who are needed to be immune varies drastically. However, it has been observed that in a confined environment, COVID-19 would stop spreading when 70% of individuals were infected [41, 42]. As no inoculation provides 100% immunity, it could be expected that the number of vaccination would need to be higher in our model. The mechanism for the pandemic ending in the the simulations relies on a large proportion of the population being infected, meaning that they are not susceptible to being reinfected. The addition of vaccines does decrease the number of deaths, but very insignificantly reduces the number of infections. Therefore, it can be concluded that our model does not model the end of the pandemic due to vaccination, but rather due to immunity developed from prior infection.

The reinforcement learning algorithm produced poorer results that the conventional vaccination policy. In its current state, the RL model is not able to provide any new insight into how to produce a vaccination policy for pandemics. However, with the high standard deviation in deaths and infections and as the model was also in the process of learning further, refer to figure 7, it may be possible that if model was left to be trained further it could have produced better results.

6.2 Outlook

The model that we have presented offers a first step in modelling age based vaccination strategies, however a number of adjustments could significantly improve the accuracy. As stated in section 4.6 our model has a number of assumptions. A constant vaccine supply, whereby each day the same number of individuals are offered the vaccine, is an unrealistic scenario. Recovery time of individuals can vary drastically as shown by Liu *et al.* [43], meanwhile our model assumes this takes 20 days for every individual. In a similar vein, the time to death is identical for all individuals. On average, the time has been recorded to be 18 days, not too dissimilar to our estimate of 20 days, however Marschner [32] shows how the time to death would be more accurately described with a probability distribution with respect to time. A significant limitation of our current model is that the nodal connections are static. In reality, these relationships are ever changing and should be modelled as a dynamic model; Meirom *et al.* [19] have proven how such an approach is effective in pandemic modelling. Incorporating such networks would replicate the ever changing social interactions that people face in reality.

The current reward mechanism as stated is somewhat arbitrary. One could argue that these parameters could be more refined or have a more tangible goal such as maintaining a certain amount of workers in industry or limiting the number of deaths by a certain amount. The linear reward system we used is unlikely to be the most optimal, however, it was enough to demonstrate that the reinforcement learning could aid in generating a personalised vaccination program based on certain objectives.
In particular, the reward function that was implemented for this model will give more relevance to the economic contribution rather than the deaths, due to their natural scale. As such, the model focused more on vaccinating the nodes that could contribute to the economy of the community rather than the ones with an higher risk of dying.

A major factor missing from the current model is that younger people are not regarded as more sociable. With the computational modelling work performed by Li *et al.* [9], it was concluded that pandemics may be further reduced by targeting the most sociable in society. As our model regards every individual as equally sociable, this could mean that the model misses critical detail. Sociability could either be represented as increased amounts of edges for nodes representing younger individuals, or the reproductive number, R, could be increased for the younger population.

6.3 Conclusion

In this report, a compartmental model was developed that was effectively able to model pandemic spread. It was seen that the addition of vaccines was able to reduce the number of deaths and reduce economic impact. Reinforcement learning was utilised in an attempt to create an optimised vaccination policy. The steps taken in this report are a small step to create a computational models that are modelled for country-specific demographics. The compartmental pandemic modelling without reinforcement learning showed how the outcome of the COVID-19 pandemic can vary considerably according to varying age demographics of different countries. It also showed, as expected, that a fixed vaccination policy lowers the death toll and total number of infected people. The initial reinforcement learning algorithm that was developed did not outperform the fixed vaccination policy, however it was able to recognise that the eldest, most vulnerable in society required high levels of inoculation. Considering that our model makes large assumptions and can be regarded as simplistic, the results produced in this report are a modest step in developing tools that allow for vaccination policies that are tailored to specific countries.

References

[1] W. H. Organisation. "Who coronavirus (covid-19) dashboard." (), [Online]. Available: `https://covid19.who.int/`. (accessed: 17.11.2021).

[2] R. Clair, M. Gordon, M. Kroon, and C. Reilly, "The effects of social isolation on well-being and life satisfaction during pandemic," *Humanities and Social Sciences Communications*, vol. 8, no. 1, pp. 1–6, 2021.

[3] A. Pak, O. A. Adegboye, A. I. Adekunle, K. M. Rahman, E. S. McBryde, and D. P. Eisen, "Economic consequences of the covid-19 outbreak: The need for epidemic preparedness," *Frontiers in public health*, vol. 8, p. 241, 2020.

[4] I. Locatelli and V. Rousson, "A first analysis of excess mortality in switzerland in 2020," *Plos one*, vol. 16, no. 6, e0253505, 2021.

[5] E. National Academies of Sciences, Medicine, *et al.*, "Framework for equitable allocation of covid-19 vaccine," 2020.

[6] E. C. for Disease Prevention and Control, "Equitable allocation of covid-19 vaccines in the united states," pp. 1–26, 2021.

[7] H. Schmidt *et al.*, "Equitable allocation of covid-19 vaccines in the united states," *Nature Medicine*, pp. 1–10, 2021.

[8] S. Bansal, B. Pourbohloul, and L. A. Meyers, "A comparative analysis of influenza vaccination programs," *PLoS medicine*, vol. 3, no. 10, e387, 2006.

[9] R. Li, O. N. Bjørnstad, and N. C. Stenseth, "Prioritizing vaccination by age and social activity to advance societal health benefits in norway: A modelling study," *The Lancet Regional Health-Europe*, p. 100 200, 2021.

[10] K. M. Bubar *et al.*, "Model-informed covid-19 vaccine prioritization strategies by age and serostatus," *Science*, vol. 371, no. 6532, pp. 916–921, 2021.

[11] M. Ferranna, D. Cadarette, and D. E. Bloom, "Covid-19 vaccine allocation: Modeling health outcomes and equity implications of alternative strategies," *Engineering*, 2021.

[12] L. Matrajt, J. Eaton, T. Leung, and E. R. Brown, "Vaccine optimization for covid-19: Who to vaccinate first?" *Science Advances*, vol. 7, no. 6, eabf1374, 2021.

[13] S. Moore, E. M. Hill, L. Dyson, M. J. Tildesley, and M. J. Keeling, "Modelling optimal vaccination strategy for sars-cov-2 in the uk," *PLoS computational biology*, vol. 17, no. 5, e1008849, 2021.

[14] J. H. Buckner, G. Chowell, and M. R. Springborn, "Dynamic prioritization of covid-19 vaccines when social distancing is limited for essential workers," *Proceedings of the National Academy of Sciences*, vol. 118, no. 16, 2021.

[15] R. Gómez and P. H. De Cos, "The importance of being mature: The effect of demographic maturation on global per capita gdp," *Journal of population economics*, vol. 21, no. 3, pp. 589–608, 2008.

[16] A. Josephson, T. Kilic, and J. D. Michler, "Socioeconomic impacts of covid-19 in low-income countries," *Nature Human Behaviour*, vol. 5, no. 5, pp. 557–565, 2021.

[17] A. H. Kerrigan, "Reinforcement learning for optimal control of network epidemic processes," 2019.

[18] P. J. K. Libin *et al.*, "Deep reinforcement learning for large-scale epidemic control," in *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track*, Y. Dong, G. Ifrim, D. Mladenić, C. Saunders, and S. Van Hoecke, Eds., Cham: Springer International Publishing, 2021, pp. 155–170.

[19] E. Meirom, H. Maron, S. Mannor, and G. Chechik, "Controlling graph dynamics with reinforcement learning and graph neural networks," in *International Conference on Machine Learning*, PMLR, 2021, pp. 7565–7577.

[20] A. Q. Ohi, M. Mridha, M. M. Monowar, and M. A. Hamid, "Exploring optimal control of epidemic spread using reinforcement learning," *Scientific reports*, vol. 10, no. 1, pp. 1–19, 2020.

[21] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. Pappas, "Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks," in *52nd IEEE conference on decision and control*, IEEE, 2013, pp. 7486–7491.

[22] J. Mushanyu, W. Chukwu, F. Nyabadza, and G. Muchatibaya, "Modelling the potential role of super spreaders on covid-19 transmission dynamics," *medRxiv*, 2021.

[23] M. Roberts, V. Andreasen, A. Lloyd, and L. Pellis, "Nine challenges for deterministic epidemic models," *Epidemics*, vol. 10, pp. 49–53, 2015.

[24] K. Rock, S. Brand, J. Moir, and M. J. Keeling, "Dynamics of infectious diseases," *Reports on Progress in Physics*, vol. 77, no. 2, p. 026 602, 2014.

[25] G. H. Kwak, L. Ling, and P. Hui, "Deep reinforcement learning approaches for global public health strategies for covid-19 pandemic," *Plos one*, vol. 16, no. 5, e0251550, 2021.

[26] M. Voysey *et al.*, "Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of chadox1 ncov-19 (azd1222) vaccine: A pooled analysis of four randomised trials," *The Lancet*, vol. 397, no. 10277, pp. 881–891, 2021.

[27] L. Böttcher and N. Antulov-Fantulin, "Unifying continuous, discrete, and hybrid susceptible-infected-recovered processes on networks," *Physical Review Research*, vol. 2, no. 3, p. 033 121, 2020.

[28] M. Feng, S.-M. Cai, M. Tang, and Y.-C. Lai, "Equivalence and its invalidation between non-markovian and markovian spreading dynamics on complex networks," *Nature communications*, vol. 10, no. 1, pp. 1–10, 2019.

[29] A.-L. Barabasi, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, no. 7039, pp. 207–211, 2005.

[30] M. Starnini, J. P. Gleeson, and M. Boguñá, "Equivalence between non-markovian and markovian dynamics in epidemic spreading processes," *Physical review letters*, vol. 118, no. 12, p. 128 301, 2017.

[31] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to probability*. Athena Scientific, 2008.

[32] I. C. Marschner, "Estimating age-specific covid-19 fatality risk and time to death by comparing population diagnosis and death patterns: Australian data," *BMC medical research methodology*, vol. 21, no. 1, pp. 1–10, 2021.

[33] N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, and R. M. Eggo, "Age-dependent effects in the transmission and control of covid-19 epidemics," *Nature medicine*, vol. 26, no. 8, pp. 1205–1211, 2020.

[34] W. Choi and E. Shim, "Vaccine effects on susceptibility and symptomatology can change the optimal allocation of covid-19 vaccines: South korea as an example," *Journal of clinical medicine*, vol. 10, no. 13, p. 2813, 2021.

[35] A. Sheikh, C. Robertson, and B. Taylor, "Bnt162b2 and chadox1 ncov-19 vaccine effectiveness against death from the delta variant," *New England Journal of Medicine*, 2021.

[36] C. Bonanad *et al.*, "The effect of age on mortality in patients with covid-19: A meta-analysis with 611,583 subjects," *Journal of the American Medical Directors Association*, vol. 21, no. 7, pp. 915–918, 2020.

[37] *United nations: World population prospects 2019*, `https://population.un.org/wpp/DataQuery/`, Accessed: 2021-12-03.

[38] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[39] Swiss National COVID-19 Task Force, *Covid-19 vaccines: Process to determine priority and allocation national and international responsibilities for access*, 2021.

[40] A. Fowlkes *et al.*, "Effectiveness of covid-19 vaccines in preventing sars-cov-2 infection among frontline workers before and during b. 1.617. 2 (delta) variant predominance—eight us locations, december 2020–august 2021," *Morbidity and Mortality Weekly Report*, vol. 70, no. 34, p. 1167, 2021.

[41] I. Yadegari, M. Omidi, and S. R. Smith, "The herd-immunity threshold must be updated for multi-vaccine strategies and multiple variants," *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.

[42] A. Fontanet and S. Cauchemez, "Covid-19 herd immunity: Where are we?" *Nature Reviews Immunology*, vol. 20, no. 10, pp. 583–584, 2020.

[43] B. Liu *et al.*, "Whole of population-based cohort study of recovery time from covid-19 in new south wales australia," *The Lancet Regional Health-Western Pacific*, vol. 12, p. 100 193, 2021.