OPTIMIZING VACCINE DEPLOYMENT FOR COVID-19

Johannes Lee

University of California, Los Angeles
Department of Electrical and Computer Engineering

ABSTRACT

Using individual-level epidemiological model similar to the SIRD model, we investigate the number of deaths resulting from different vaccination policies when the number of vaccinations per day is limited. We show that using an intelligent vaccination policy which prioritizes at-risk individuals can prevent up to 1.27 million deaths in the United States compared to a random vaccination policy.

1. INTRODUCTION

As of June 9, 2020, the United States has had over 2 million confirmed cases of COVID-19, with more than 7 million recorded cases worldwide. With the rapid and worldwide spread of COVID-19, the halting of all but essential businesses in order to slow its spread poses an enormous pressure on government and public health officials to accelerate a return to normalcy. This return is predicated on the ability to prevent a surge of cases which would in turn overwhelm the public health system and see a large number of deaths.

Vaccines have been a safe and effective method to immunize a population from disease, and countless research efforts are now being put into developing a vaccine for Sars-Cov-2, the virus which causes the disease COVID-19. If we assume that we are able to develop and produce an effective vaccine in the relatively near future, because of the constraints of the healthcare system, we would still left with an unanswered question: whom should we vaccinate first? A vaccination deployment strategy which disproportionately targets particular communities leaves other communities vulnerable, thus leading to higher strain on their healthcare system and increased deaths.

Previous studies regarding influenza (H1N1) have compared children-first and at-risk-first vaccination policies and concluded that an at-risk-first policy is more effective [1]. The CDC currently recommends prioritization of high-risk individuals and critical healthcare personnel [2]. However, the best vaccination policy for COVID-19 may be different than that for influenza.

2. A PARAMETERIZED BANDITS PROBLEM

Using per-day SIRD modeling at the individual scale [3], we seek to find the vaccination policy which minimizes the total number of disease-related deaths after a specified T_0 days. The disease in our model was given a death rate of approximately 1.8%, which is less than the estimated death rate for COVID-19 [4], and likelihood of death is generated from an individual's symptom level, which is a probabilistic function of the mortality variable. Similarly to COVID-19, individuals who are infected may manifest no noticeable symptoms. Further details of the model are given in Appendices A and B. In a population of N inhabitants with N_{inf} infected persons at the start of the first day, the vaccination policy is used to vaccinate up to λ persons each day. The vaccination policy is best characterized as an order of vaccination with which public health officials can prioritize vaccinating certain people based on M recorded variables. The variables used are the outside contact rate ρ, susceptibility σ, mortality μ, infectivity κ, symptom level χ, and household size h, as well as the maximum value of each variable for each person's household (excluding household size). This totals to $M = 11$ variables, with individual data vector $x \in \mathbb{R}^M$. The action space is then defined as

$$\mathcal{A} = \{T : \mathbb{R}^M \to \mathbb{R} | T \in \mathcal{K}\}, \tag{1}$$

i.e. the set of transformations from M dimensions to 1 dimension such that the transformation is in class $\mathcal{K}$. This makes the action equivalent to the vaccination policy. At each trial, the agent selects a vaccination policy $A \in \mathcal{A}$, and on each day, the λ persons with the highest score $A(x)$ are vaccinated. After T_0 days, the agent receives a reward r equal to $-d$, where d is the total number of disease-related deaths after T_0 days.

For simplicity, during training trials, we keep the initialization (N_{inf}, N) constant. In this case, the problem is best seen as an infinitely many-armed bandits problem, since there is an infinite number of actions in the action space [5]. To make this tractable, we parameterize the action space by a set of parameters θ. Therefore, given a function class $\mathcal{K}$, the goal of this parameterized bandits problem is to find the optimal parameters θ^* such that the vaccination policy $A_{\theta^*|\mathcal{K}}$ minimizes the total number of deaths after T_0 days. We note that by this problem formulation, the expected reward is invariant

to monotonic transformations of the action output, and take advantage of this in the case where the action transformation is constrained to be linear.

3. TRAINING

In order to reduce training time, we set $T_0 = 125$ days and use (N_{inf}, N) = (50, 2000) for all training trials. In addition, we use $\lambda = \frac{1}{200}N = 10$. Using these settings, the model is typically near equilibrium at the end of each trial.

We first let $\mathcal{K}$ be the set of all linear transformations, such that $\forall A \in \mathcal{A} | A \in \mathcal{K}, \exists\ c : A(x) = c^T x + b$ for $c \in \mathbb{R}^M$ and for some scalar b. We approach this case with 3 methods:

1. **Least Squares.** Since the value of an action $Q(A)$ is invariant to scale, we draw 10000 uniform samples on the unit sphere using $c_i = \frac{u_i}{||u_i||}$, $u_i \sim \mathcal{N}(0, I_M)$ without loss of generality, and observe the reward r_i on trial i. Given the set pairs of action vectors and rewards $\{c_i, r_i\}$, we can predict r from c as

$$\hat{r} = \theta^T c + b, \tag{2}$$

where θ is the least squares predictor of r from c, and b is a bias term. Then our optimal action is equivalent to $\theta^* = \frac{\theta}{||\theta||}$.

2. **REINFORCE with Gaussian distribution and learned covariance**. In order to use REINFORCE [6], we first let $\theta = \{\mu_a, \Sigma\}$. We then use an ϵ-greedy sampling method, and sample $c \sim \mathcal{N}(\mu_a, \Sigma)$ with probability $1 - \epsilon$, and $c \sim \mathcal{N}(0, I_M)$ with probability ϵ. We initialize $\mu_a = 0, \Sigma = I_M$. Since the REINFORCE update is given by

$$\theta \leftarrow \theta + \alpha(r_i - B)\nabla \ln p(A|\theta), \tag{3}$$

our update becomes:

$$\mu_a \leftarrow \mu_a + \alpha(r_i - B)\Sigma^{-1}(c_i - \mu_a) \tag{4}$$

$$\Sigma \leftarrow \Sigma + 0.5|\Sigma|\alpha(\Sigma^{-1}(a - \mu_a)(a - \mu_a)^T\Sigma^{-1} - \Sigma^{-1}), \tag{5}$$

where $B = -15$ is a constant baseline. In the Σ update, $|\Sigma|$ is added to alleviate numerical issues where Σ becomes non-positive-semidefinite before sufficient exploration, and training is terminated when Σ would become non-positive-semidefinite. We let the step size $\alpha = \frac{1}{1000+i}$, where i is the trial number, starting at 0. This can occur in less than 500 trials while still achieving good results. The optimal action vector is given by μ_a at the end of training.

3. **REINFORCE with Gaussian distribution and fixed covariance**. Similarly to method 2, $c \sim \mathcal{N}(\mu_a, 0.1I_M)$, where the only difference is the covariance matrix. The REINFORCE update is equal to

$$\mu_a \leftarrow \mu_a + \alpha(r_i - B)(c_i - \mu_a). \tag{6}$$

Due to instability reasons, we remove the Σ^{-1} term from the learned covariance REINFORCE update. We also inherit the step size α and baseline B from method 2, and train for 10000 trials.

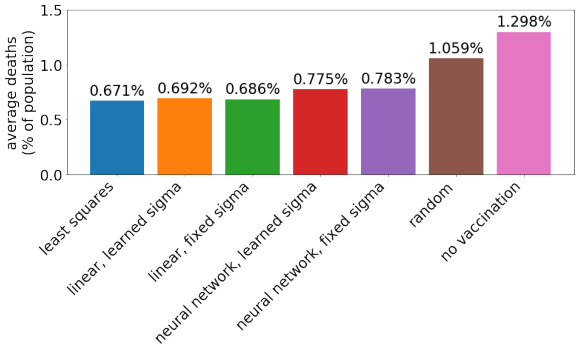

Fig. 1. Average deaths using the function found by each method for (N_{inf}, N) = (500, 20000), $\lambda = 100$. 1000 trials each.

We also use methods 2 and 3 to search for optimal parameters for a neural network. We choose the neural network to have a fixed structure with one hidden layer with 5 units and Leaky ReLU activation function with negative slope 0.1. This network has a total of 66 parameters, such that $\mu_a \in \mathbb{R}^{66}$.

4. RESULTS

In order to test our results, we use (N_{inf}, N) = (500, 20000), $\lambda = 100$, and $T_0 = 150$ days, and compare the average number of deaths over 1000 trials resulting from using each of the five methods policies with the number of deaths from random vaccination and no vaccination (Figure 1). Compared to random vaccination, the optimal linear vaccination policy found from the least squares method results in 0.388% of the population less deaths. For the United States, this percentage would amount to 1.27 million fewer deaths on average. We can make this extrapolation if we assume that the total population is composed of many distinct subpopulations, better known as cities. The fact that the least squares method performs best indicates that with this model and sampling scheme, the expected number of deaths is well modeled by the linear parameterization. Interestingly, even though each neural network is able to implement any linear function, and thus the best performance from a neural network is at least as good as the performance of any linear function, we find that our parameterized neural networks perform worse than the linear functions, both in training and in testing. This may

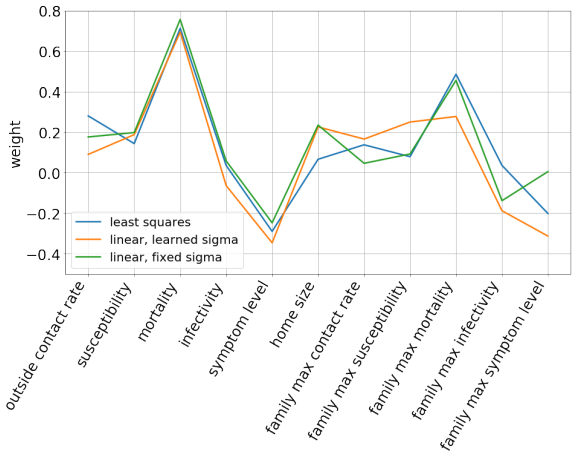

Fig. 2. Weights of data variables when normalized to unit variance for linear vaccination policies.

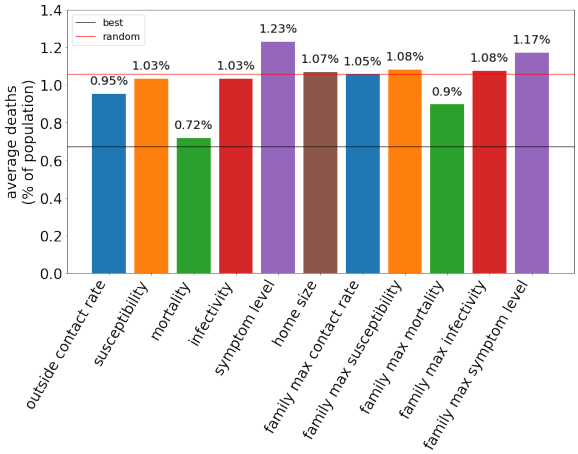

Fig. 3. Average deaths when vaccinating based on each variable. (N_{inf}, N) = (500, 20000), $\lambda = 100$, with 1000 trials each.

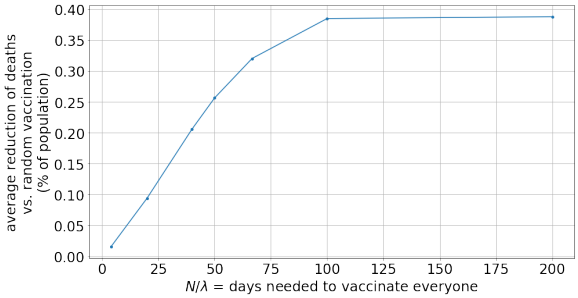

Fig. 4. Advantage of using an informed vaccination policy vs. random vaccination for varying N/λ. (N_{inf}, N) = (500, 20000), with 1000 trials each point.

be due to non-convexity of outputs with respect to both inputs and parameters.

In order to examine the relative importance of each variable, the recorded variables were normalized to unit variance before the training phase. We then plot the unit-normalized weights of the 11 variables (Figure 2). A variable with large positive weight is deemed to be important, with individuals with high values being prioritized, while a variable with large negative weight has inverse importance, where an individual having a high value would be deprioritized. In our case, we notice that the dominant positive factors are mortality and family maximum mortality, indicating that those who are vulnerable or live with those who are vulnerable should be vaccinated first. On the other hand, the largest negative factor is the symptom level, indicating that those who have had noticeable COVID-19-like symptoms should not be vaccinated first, presumably because the chance of already being infected is higher than those without symptoms.

Furthermore, we may wish to compare our results to heuristic vaccination policies, wherein we prioritize individuals to be vaccinated based on a single variable (Figure 3). Notably, in this paradigm, prioritization based on mortality alone achieves a death percentage within 0.05% of the optimal linear vaccination policy (164,000 persons for US population). This indicates that even without perfect access to the presented variables, intelligent vaccination policies should result in fewer deaths caused by COVID-19.

Finally, we would also like to know how much the advantage of using an informed vaccination policy compared to random vaccination varies as the rate of vaccination λ is changed (Figure 4). The vaccination policy used is the one found using the least squares method. Here we find that, as expected, as the number of days needed to vaccinate everyone increases up to 200 days, the advantage of using an informed vaccination policy increases as well. This informs us that as long as the number of vaccinations administered per day is limited, an informed vaccination policy should be used to minimize the expected number of deaths.

5. DISCUSSION

In practice, some of the variables use are opaque and difficult to determine with certainty. This does not mean that these results are not interpretable, however. For example, for COVID-19, mortality is closely linked with age, and outside contact rate is fairly reasonable to estimate. Practically, public health officials may wish to categorize individuals into different vaccination priority groups based on prominent characteristics using a tree or flowchart structure. Using the results from this study, a possible prioritization flowchart is given in Figure 5.

Because this study uses an epidemiological model, it is important to note the weaknesses of the model in order to accurately recognize the usefulness of our results. In particular, the model takes advantage of several assumptions:

1. Individuals are contagious for every day that they are infected.

2. Transmission is only through contact, which corresponds to being in the same area at the same time or close in time.

3. Expectation of contact rate is constant across all days.

4. Individuals who have been infected and have recovered cannot be reinfected.

5. There is no data available from COVID-19 testing.

6. Vaccinations are single-dose and 100% effective after τ days if an individual has not been infected τ days after vaccination.

7. Vaccination rate is constant beginning from the first day.

8. Initial infected percentage was not varied during training and testing.

9. Each life/death is weighted equally.

Perhaps least realistic are assumptions 3 and 4. Our model assumes a rate of transmission higher than is recorded for COVID-19 in light of face coverings and social distancing measures, and also assumes that individuals do not change their contact rate in response to having COVID-19 symptoms or being tested for COVID-19, which is not modeled. Taking these into account would decrease the rate of spread of the disease and consequently decrease the average number of deaths on any trial. Based on these differences, we expect and hope that the total number of deaths in real life is less than the estimates presented here. As a result, we can interpret our results as an upper bound for the efficacy of vaccination policies. Nevertheless, our results show that if we adopt a vaccination policy that first prioritizes the most vulnerable individuals and their families, and then those who come into contact with many people on a regular basis, such as essential workers, we can expect to save many thousands of lives compared to random vaccination.

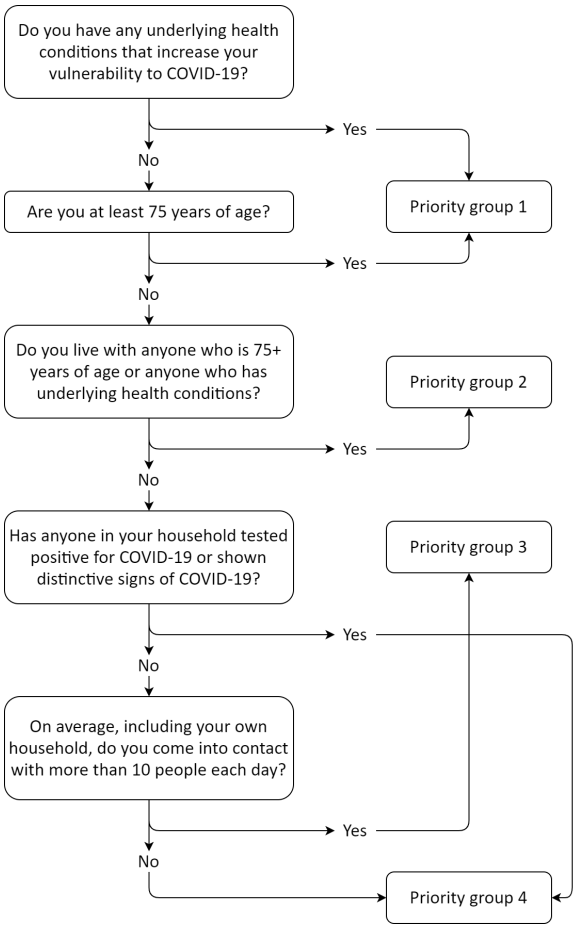

Fig. 5. Possible prioritization flowchart representing a vaccination policy.

6. REFERENCES

[1] Lee, B. Y., Brown, S. T., Korch, G. W., Cooley, P. C., Zimmerman, R. K., Wheaton, W. D., ... & Burke, D. S. (2010). A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 H1N1 influenza pandemic. Vaccine, 28(31), 4875-4879.

[2] CDC, Center for Disease Control (2018). Iterim updated planning guidance on allocating and targeting pandemic influenza vaccine during an influenza pandemic.

[3] Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, 115(772), 700-721.

[4] Onder, G., Rezza, G., & Brusaferro, S. (2020). Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. Jama, 323(18), 1775-1776.

[5] Agrawal, R. (1995). The continuum-armed bandit problem. SIAM journal on control and optimization, 33(6), 1926-1951.

[6] Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine learning, 8(3-4), 229-256.

7. APPENDIX A: INDIVIDUAL-LEVEL EPIDEMIOLOGICAL MODELING

Our model is based heavily on the continuous-time Susceptible-Infected-Recovered-Deceased (SIRD) model, which adds the deceased category to the SIR model (Kermack and McKendrick, 1927). The SIRD model assumes 4 possible states for each person with zero chance of reinfection (a transition from R to S or I). Its dynamics are characterized by the following system of differential equations:

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \tag{7}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I - \mu I \tag{8}$$

$$\frac{dR}{dt} = \gamma I \tag{9}$$

$$\frac{dD}{dt} = \mu I, \tag{10}$$

where β is the infection rate, γ is the recovery rate, and μ is the mortality rate. Under this model, the number of susceptible and infected individuals both asymptotically converge to 0, so that every individual is either recovered (and thus immune) or deceased.

Our model models each individual instead of only the 4 categories, and uses sparse computations using Scipy. Modeling at the individual level allows us to characterize COVID-19 in a more realistic manner which is more suitable for reinforcement learning. Our model inherits the 4 states from the SIRD model, but also adds degrees of infectivity and individualized mortality rates due to predefined risk factors, among other additions. We let $\mathbf{S}_t, \mathbf{I}_t, \mathbf{R}_t$ and $\mathbf{D}_t \in \mathbb{R}^N$ with components $\in \{0, 1\}$ be discrete vectors representing which of the 4 states the N individuals are in. We model the symptom level χ as ranging from 0 to 5, as it can represent how confident we are that a person has been infected. Days since vaccination is given by $\mathbf{U}_t$, and is 0 by default. We then define the susceptibility of an individual to infection as $\sigma_t \geq 0$ and the infectivity to be $\kappa_t \geq 0$, keeping also the mortality μ_t. We therefore model the probability that an uninfected individual i is infected by an individual j due to a single contact as

$$Pr(i \Leftarrow j) = \alpha \sigma^i \kappa^j \mathbf{I}^j \tag{11}$$

for some universal infection multiplier $\alpha > 0$.

In this work, we also introduce a more realistic model for contacts between individuals. We define the (sparse) contact matrix $C(t)$ at time t as:

$$C(t) = C_{household} + C_{random}(t) \tag{12}$$

with $C_{household}$ being a block diagonal matrix representing contact with members of the same household with each block a square matrix of all 1's except along the diagonal. Each block has size h, which is the size of each household. $C_{random}(t)$ represents random contact with strangers. We assume that all contact is symmetric, so $C_{household}$ and C_{random} are symmetric. We typically model contact as binary, however, for computational convenience using sparse matrices, we let $Pr(C_{ij} \notin \{0, 1\}) > 0$. The number of contacts an individual makes on day t is equal to $\sum_{j=1}^{N} C(t)_{ij}$. We therefore define the outside contact rate of individual i as:

$$\rho^i = \mathbb{E}_t \left[\sum_{j=1}^{N} C_{random}(t)_{ij} \right]. \tag{13}$$

The probability that individuals are infected on a given day is equal to:

$$Pr(inf_t) = \alpha \boldsymbol{\sigma}_t \odot (C(t)(\boldsymbol{\kappa}_t \odot \mathbf{I}_t)) \tag{14}$$

where $\odot$ is the Hadamard (element-wise) product. We model vaccination by modulating $\boldsymbol{\sigma}$. An individual i is moved from susceptible to recovered τ^i time steps after being vaccinated. The probability of recovery for an infected person is modeled as:

$$Pr(recovery_t) = 0.003 \boldsymbol{\gamma}_t \odot \mathbf{I}_t \tag{15}$$

while the probability of death is:

$$Pr(death_t) = (3\text{e}{-6})(\exp(2\chi) - 1) \odot \mathbf{I}'_t, \tag{16}$$

where $\mathbf{I}'_t$ is an intermediate state representing the infected people immediately before deaths are calculated.

A successful vaccine must both prevent an individual from being infected as well as prevent him from infecting others. This also applies the recovered and the deceased. We model this by reducing σ as U increases, until reaching 0.

For convenience, we calculate deaths at the end of each time step. From this formulation, we choose a single time step of the environment to be:

1. $\Delta \mathbf{I}_{t+1} \sim Pr(inf_t)$
2. $\Delta \mathbf{I}_{t+1} \leftarrow min(\Delta \mathbf{I}_{t+1} - \mathbf{I}_t, 0)$
3. $\Delta \mathbf{R}_{t+1} \sim Pr(recovery_t)$
4. $\mathbf{I}'_t \leftarrow \mathbf{I}_t + \Delta \mathbf{I}_{t+1} - \Delta \mathbf{R}_{t+1}$
5. $\mathbf{R}_{t+1} \leftarrow \mathbf{R}_t + \Delta \mathbf{R}_{t+1}$
6. $\mathbf{S}_{t+1} \leftarrow \mathbf{S}_t - \Delta \mathbf{I}_{t+1}$
7. $\Delta \mathbf{D}_{t+1} \sim Pr(death_t)$
8. $\mathbf{I}_{t+1} \leftarrow \mathbf{I}'_t - \Delta \mathbf{D}_{t+1}$
9. $\mathbf{D}_{t+1} \leftarrow \mathbf{D}_t + \Delta \mathbf{D}_{t+1}$
10. $\mathbf{S}_{t+1} \leftarrow \mathbf{S}_{t+1} - \mathbb{I}(\mathbf{U}_t = \boldsymbol{\tau})$
11. $\mathbf{R}_{t+1} \leftarrow \mathbf{R}_{t+1} + \mathbb{I}(\mathbf{U}_t = \boldsymbol{\tau})$
12. $\mathbf{U}_{t+1} \leftarrow \mathbf{U}_t + \mathbb{I}(\mathbf{U}_t \geq 0) + \pi_t(\boldsymbol{\gamma}, \boldsymbol{\kappa}, \boldsymbol{\mu}, \boldsymbol{\rho}, \boldsymbol{\sigma}, \boldsymbol{\chi}, \mathbf{D}, \mathbf{U})$

8. APPENDIX B: MODEL PARAMETERS

For our model parameters, we use the following distributions:

$$
\begin{aligned}
h &\sim \text{True US household size distribution (2019)} \\
\rho &\sim Poisson(2) \\
\left(\sum_{j=1}^{N} C_{random}(t)_{ij}\right) &\sim Poisson(\rho^i) \\
\alpha &= 0.03 \\
\kappa &\sim \max(\mathcal{N}(1, 0.25), 0) \\
\sigma_0 &\sim \max(\mathcal{N}(1, 0.25), 0) \\
\sigma &= \min\left(1 - \frac{U-7}{\tau-7}, 1\right)\sigma_0 \\
\mu &\sim \min(\max(\mathcal{N}(1, 0.25), 0), 4) \\
\tau &\sim \max(Poisson(14), 8) \\
\chi|(I = false) &\sim B(5, 0.05) \\
\chi|(I = true) &\sim B(5, \mu/4) \\
\gamma &= \min(\text{days since infection}, 21)\frac{1}{1 + 0.2\chi}
\end{aligned}
$$

Code has been made available at `https://github.com/johannes-lee/SIRD-vaccination`.