UNIVERSITY OF UTRECHT

BACHELOR THESIS

Prioritising COVID-19 vaccination

Author:
M.F.C. van Merrienboer
5956641

Supervisor:
Dr. M. Ruijgrok

A thesis submitted in fulfilment of the requirements for the degree of Bachelor of Science in the major

Mathematics
14 November 2021

Utrecht University

Abstract

In this thesis, we would like to explain how different vaccination strategies could result in fewer casualties during the COVID-19 pandemic. By using mathematical models that couple both social and epidemiological dynamics, the evolution of the pandemic can be simulated for a population that is divided into different age groups. Depending on the time course of the pandemic within this population, different vaccination strategies would be most effective for a minimal amount of deaths caused by COVID-19. If early vaccination is possible, prioritising vulnerable age groups would be most effective, while vaccination at a later stage would be most effective by targeting age groups that cause the most transmission.

Contents

1 Introduction

In the past year, the world has been taken captive by the coronavirus. Only after more than one and a half years, it seems like we are slowly but surely recovering from its devastating social and economical consequences, fighting the virus with vaccines. The usage of the word *we* needs to be stressed here, which is referring to rich western countries in this context, because while we hesitantly release our breath, developing countries are still holding it.

Throughout history, many mathematical methods have been developed to and model the dynamics of infectious diseases. Bernoulli implemented an epidemiological model as early as 1760 to look at smallpox [4] [12]. Often, these models include certain (finite) compartments. These compartments represent certain stages of illness. The simplest one is the SIR-model, where the compartments are made off of Susceptibles, Infectious and Recovered stages in which different parts of a population are represented. While this method captures the epidemiological dynamics, it does not include the effect that social interaction has on the spread of a virus.

With the vaccination program in full motion for a lot of countries, an important question is posed by P. Jentsch, Madhur Anand and C.T. Bauch in their article about the prioritisation of certain groups during vaccination [16]. They combine an enlarged version of the SIR-model as mentioned above, with behavioural dynamics (evolutionary game theory) to include adherence of the population to so-called NPIs. These are Non-Pharmaceutical Interventions, like social distancing, the washing of hands, and the wearing of a face mask. The main research question from the article is: to save the most lives, which age group should get the vaccine first?

The aim of their mathematical modelling study was to compare projected COVID-19 mortality under four different strategies for the prioritisation of SARS-CoV-2 vaccines in the province of Ontario, Canada. [16] A coupled social-epidemiological model was developed to model the transmission of the coronavirus in which social and epidemiological dynamics interact with each other. The social dynamics describe how individual adherence to non-pharmaceutical interventions (NPIs), relates to the amount of reported COVID-19 cases. The epidemiological dynamics include a set of governing differential equations, as well as linear algebraic calculations. The model used parameters based on data for the number of cases, seroprevalence (immunity), population mobility and demography. It assumed a vaccine with 75% efficacy against disease and transmissibility.

To determine the influence of prioritising certain age groups for vaccination, four different vaccination strategies are introduced:

- Vaccinating people who are of ages 60 and older (oldest first strategy)

- Vaccinating people aged 20 and younger (youngest first strategy)
- Vaccinating uniformly by age (uniform strategy)
- Contact-based strategy

Note that the last three strategies aim to interrupt transmission of the virus (i.e. ages 15-9 first, 20-59 second and lastly older of younger ages), whereas the first one targets a vulnerable group. Vaccination rates ranged from 0.5 to 5 % of the population per week, with as starting date the 1st of either January or September of 2021.

The number of cases, adherence to NPIs, and lockdown measures undergo successive waves and depending on the starting date (and rate) of the vaccination programme predicts the effectiveness of the four vaccination strategies. The key finding of this research is that transmission-interrupting strategies become favourable with time as herd immunity builds.

If no vaccine was made available, 72000 deaths would occur (95% credible interval 40000-122000) in Ontario from January 1st, 2021 to March 14th, 2025. Assuming a vaccination rate of 1.5% of the population per week, COVID-19 mortality would be reduced on average by 90.8% (oldest first strategy), 89.5% (uniform strategy), 88.9% (contact-based strategy) and 88.2% (youngest first strategy).

For the September 1st, 2021 to March 14th, 2025 scenario, 60000 deaths (95% credible interval 31000-108000) would occur in the absence of a vaccine. Assuming again a vaccination rate of 1.5% of the population per week, COVID-19 mortality would be reduced on average by 92.6% (contact-based strategy), 92.1% (uniform strategy), 91.0% (oldest first strategy) and 88.3% (youngest first strategy).

In conclusion, the preferred group to prioritise for COVID-19 vaccination depends on the time and development of the disease in the population. If vaccination begins at later dates, it might save more lives to choose for a transmission-interrupting vaccination strategy than to prioritise vulnerable age groups.

This can be explained by the fact that if the effective reproduction number R_{eff} (the average number of secondary infections produced by a single person) decreases towards 1 (from ≈ 2.2 in the absence of pre-existing immunity for COVID) as natural immunity rises, the benefits of transmission reducing vaccines become stronger. Take example seasonal influenza with an $R_{eff} \approx 1.5$, it is estimated that only 33% of the population needs to be immune to let transmission die out.

2 SEPAIR model

2.1 SIR model

The epidemiological dynamics of COVID-19 are configured using a (somewhat extensive) SIR model. SIR models simulate the behaviour of infectious diseases, using a set of differential equations, dividing the population into separate disease 'stages'. This can be seen mathematically as [12]:

$$\frac{dS}{dt} = \frac{-\beta IS}{N} \tag{2.1}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \tag{2.2}$$

$$\frac{dR}{dt} = \gamma I \tag{2.3}$$

With $S(t)$ being the number of Susceptible at time t, $I(t)$ the number of Infected people, $R(t)$ the number of Removed (recovered, vaccinated, and deceased) and $S(t) + I(t) + R(t) = N$. β is the probability of disease transmission per contact times the number of contacts per unit time, also called the transmission rate, this is multiplied by the infectious population at time t, since contact with non-infectious people will not result in a new infection. γ is the recovery rate of infectious people.

So susceptible individuals progress to the infected compartment at rate β (hence the minus term from the equation(1) and the plus term in equation(3)) and infected individuals progress from the infected to recovered compartment at rate γ. $\frac{S(t)}{N}$ is the fraction of the population that is susceptible and $\frac{I(t)}{N}$ the fraction that is infectious. At the infection free state $I = R = 0$, and $S = N$.

Note that equations (2.1) and (2.3) correspond to an exponentially distributed 'waiting time' per disease compartment. For (2.1) we have $\frac{dS}{dt} = \frac{-\beta IS}{N}$, then the function $P(t) = e^{\frac{-\beta It}{N}}$ correspond to the fraction that is still in compartment S at time t, which is 1 at $t = 0$, since then $I = 0$. For (2.3) we have $\frac{dR}{dt} = \gamma I$, which means after time t, $P(t) = e^{-\gamma I(t)}$ will represent the amount of individuals that is still present in compartment I. A typical virus progression will start with an infectious curve near zero, and quickly reach a peak before dying out as a function of time. The number of susceptibles will always decrease, but it will stay positive when $t \to \infty$ as can be seen in figure 2.1. This has to do with the fact that the pandemic will already die out if an infectious individual will infect no more than one other individual.

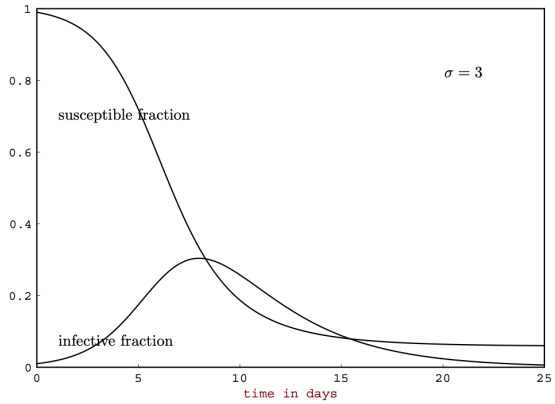

FIGURE 2.1: Solutions of the classic SIR epidemic model with contact number $\sigma = 3$ and average infectious period $\frac{1}{\gamma} = 3$ days. figure and caption from [12]

To explain this in a bit more detail, let us introduce a few parameters (note that we use the same argumentation as in section 2.3 of [12]):

- R_0.
  The first parameter is the basic reproduction number R_0 which is defined as the average number of new cases of infection caused by one typical infected individual in a population that consists completely of susceptibles [6] [12].

- σ.
  The next parameter is the number of adequate contacts that one typical infectious individual has during its infectious period. Note that adequate in this sense means that contact between a susceptible and infectious individual will result in transmission of the virus. In [6] this parameter is represented as σ.

- $\bar{R}$.
  The last parameter is the replacement number $\bar{R}$, which represents the typical number of new infections one infectious individual produces during its period of infectiousness.

The difference between these parameters can be seen as follows, R_0 only determines the number of new infections at the beginning and is a constant, whereas $\bar{R}$ will change over time as it represents the actual new infections produced. At $t = 0$ they are equal. Also note that the typical number of adequate contacts of an infected individual that will result in a new infected individual σ will be equal to both $\bar{R}$ and R_0 at $t = 0$, since all contacts of the infected individual will be susceptible at the beginning of the pandemic.

The replacement number $\bar{R}$ will be smaller than the contact number after a population is invaded, as the fraction of susceptibles will no longer be 1, so not all contacts that infected individuals encounter will result in a new infection: $\sigma \geq \bar{R}$. Furthermore, R_0 will almost always be equal to σ, but in certain special cases σ will become lower than R_0 after a virus entered a population, as new classes of infected individuals appear which have a lower transmission rate. In conclusion: $R_0 \geq \sigma \geq \bar{R}$.

Now, if we divide all of the equations (2.1), (2.2) and (2.3) by the total number of the population N, we get the following equations ([12], equation (2.2)):

$$\frac{ds}{dt} = -\beta is, \qquad s(0) = s_0 \geq 0, \tag{2.4}$$

$$\frac{di}{dt} = \beta is - \gamma i, \qquad i(0) = i_0 \geq 0, \tag{2.5}$$

With $r(t) = 1 - s(t) - i(t)$ and $s(t), i(t)$ and $r(t)$ the fractions of the population that are present in the classes susceptibles, infectious or removed. Now, one can determine the phase plane of s and i, and show that the triangle T, which is defined as ([12], equation (2.3)):

$$T = \{(s,i) | s \geq 0, i \geq 0, s + i \leq 1\} \tag{2.6}$$

is positively invariant, i.e. solutions that start in T will stay in T over time. Furthermore, unique solutions exist in T for all $t \geq 0$ [13].

Now, let the contact number σ be defined by the transmission rate β divided by the rate at which individuals recover, γ. Note that this indeed gives the number of adequate contacts that one typical infectious individual has during its infectious period, since β is the probability of disease transmission per contact times the number of contacts per unit time and γ denotes the amount of time an individual stays in its infectious state. Furthermore, let the replacement number $\bar{R}$ at $t = 0$ be defined by σs_0, i.e. the contact number times the initial fraction of susceptible which will be 1 at the beginning of the pandemic (note that only the initial invader will not be included in s_0). We can now introduce the following theorem to back our claims made regarding the epidemiological dynamics as portrayed in figure 2.1.

Theorem 2.1 ([12], Theorem 2.1)
Let $(s(t), i(t))$ be a solution of ((2.4), (2.5)) within triangle T. If $\sigma s_0 \leq 1$, then $i(t)$ decreases to zero as $t \to \infty$. If $\sigma s_0 \geq 1$, then i(t) first increases up to a maximum value $i_{max} = i_0 + s_0 - 1/\sigma - [ln(\sigma s_0)]/\sigma$ and then decreases to zero as $t \to \infty$. The susceptible fraction $s(t)$ is a decreasing function and the limiting value s_∞ is the unique root in $(0, 1/\sigma)$ of the equation

$$i_0 + s_0 - s_\infty + ln(s_\infty/s_0)/\sigma = 0 \tag{2.7}$$

All though we will not provide a proof of this theorem in this thesis (for this, we would like to refer the reader to [12]), a couple of remarks should be made. Note that the number of infectious individuals in the population starts at (very small) I_0 and as time passes reaches a peak before dying out to zero.

The fraction of susceptibles will start at 1, denoting the pre-pandemic situation, and strictly decreases with time before attaining its 'end-value' of s_∞. Note that s_∞ is strictly positive, since the epidemic will die out when the replacement number $\bar{R} = \sigma s_t \leq 1$, i.e. when $s_t \leq 1/\sigma$. We can see this clearly in figure 2.1, where the infectious curve

starts to decrease as the susceptible fraction gets below $1/\sigma = 1/3$.

One last note remains to be made about the epidemiological dynamics that are introduced previously in this chapter. Since an epidemic affects a population within a relatively short time span, the birth rate at which individuals enter compartment S and the death rate at which individuals leave the compartments are set equal and we hence neglect these terms in all dynamics used in this thesis.

2.2 SEPAIR model

We have now explained the general dynamics of a relatively straightforward virus. But as the title of this section might already have given away; the model that is used for COVID-19 is somewhat more complex, including also an E, P, and an A term, referring to the Exposed, Pre-symptomatic and Asymptomatic compartments. The I term in this model represents the number of infectious individuals, which are divided into symptomatic (I_{s_i}) and asymptomatic (I_{a_i}) cases. N_j denotes the total number of individuals in age group j. A more detailed description of the different stages of the disease will be provided after we have introduced the governing epidemiological dynamics later on in this chapter.

2.2.1 Closure of public places

Since the beginning of the COVID-19 pandemic, government shutdown policies have been a very important factor in the decline of reported cases. The extent to which public places such as schools and workplaces are being closed is represented by the variable $D(t)$ in the equation $dD(t)/dt = \Omega(D(t))$. It represents the reaction of the public health authorities to the number of confirmed cases in the following way:

$$\Omega(D(t)) = \begin{cases} k_1(1 - D(t)) & \text{if } \sum_{i=1}^{16} \alpha_i(I_{s_i} + I_{a_i}) > T \\ -k_2 D(t) & \text{if } \sum_{i=1}^{16} \alpha_i(I_{s_i} + I_{a_i}) \leq T \end{cases} \tag{2.8}$$

With $D(t) \in [0, 1]$, i.e. the closure level of public places. The variable α_i denotes the ascertainment rate, i.e. the fraction of total infectious cases in age group i ($I_{s_i} + I_{a_i}$) that are reported. k_1 and k_2 denote the workplace shutdown rate and the opening rate, which are fitted to available data.

Note that the equation changes when the number of ascertained cases crosses threshold T (shutdown threshold). If the number is larger than T, the exponent will be positive as $k_1 \geq 0$ and $(1 - D(t)) \geq 0$. This implies that closure of work and school places is triggered and increases exponentially. When the number of ascertained cases is equal to or lower than threshold T, the exponent will be negative, and an exponential decrease in the amount of closed work and school places will kick in until it completely dies out to zero and all work and school places are open.

2.2.2 Epidemiological dynamics

We will now have a look at the general epidemiological dynamics that are used to model the virus spread of COVID-19. The model is structured in a way that divides the individuals of the population into age classes $i \in [1, 16]$ which consist of five year increments each, and the last group consisting of people aged 75 and over. We will first present the governing equations as they are used in the article [16]. Subsequently, we will try to explain what the equations represent and refer to some further chapters for certain components.

$$\frac{dS_i^1}{dt} = -r\rho_i[1 + s\sin(\frac{2\pi}{365}(t-\phi) - \frac{\pi}{2})]S_i^1 \sum_{j=1}^{16} C_{ij}(t)\frac{I_{s_j} + I_{a_j} + P_j}{N_j} - \tau S_i^1 \tag{2.9}$$

$$\frac{dS_i^2}{dt} = -r\rho_i[1 + s\sin(\frac{2\pi}{365}(t-\phi) - \frac{\pi}{2})]S_i^2 \sum_{j=1}^{16} C_{ij}(t)\frac{I_{s_j} + I_{a_j} + P_j}{N_j} - \tau S_i^2 \tag{2.10}$$

$$\frac{dE_i}{dt} = r_i[1 + s\sin(\frac{2\pi}{365}(t-\phi) - \frac{\pi}{2})](S_i^1 + S_i^2) \sum_{j=1}^{16} C_{ij}(t)\frac{I_{s_j} + I_{a_j} + P_j}{N_j} - \sigma_0 E_i + \tau(S_i^1 + S_i^2) \tag{2.11}$$

$$\frac{dP_i}{dt} = \sigma_0 E_i - \sigma_1 P_i \tag{2.12}$$

$$\frac{dI_{a_i}}{dt} = \eta\sigma_1 E_i - \gamma_a I_{a_i} \tag{2.13}$$

$$\frac{dI_{s_i}}{dt} = (1-\eta)\sigma_1 E_i - \gamma_s I_{s_i} \tag{2.14}$$

$$\frac{dR_i}{dt} = \gamma_a I_{a_i} + \gamma_s I_{s_i} \tag{2.15}$$

$$\frac{dD_i}{dt} = \Omega(D(t)) \tag{2.16}$$

Following these equations, we can classify the stages of disease for COVID-19 in the following way:

- Susceptibles.
  The S represents the number of susceptible individuals in the population, the same way they did in the SIR-model. They are divided into two groups: S_i^1 and S_i^2. S_i^1 being the number of the susceptible and not vaccinated individuals in age group i. If vaccination leads to immunisation, the individuals move to the R_i compartment. However, since we assume a vaccine with 75% efficacy, it will remain possible for individuals to be infected and transmit the disease. These individuals will be present in S_i^2. So more concretely: S_i^2 is the number of susceptibles in age group i who had a standard two-dose vaccination but were not immunised. In the S-equations a sinusoidal term is present, which indicates a periodical effect that influences the number of susceptibles at time t. This periodical effect is caused by the variation in susceptibility due to seasonal changes.

- Exposed.
  The E_i term represents the exposed individuals. These individuals have been exposed to an infectious individual, resulting in transmission of the virus. However, they are not yet capable of producing new cases through contact with susceptible individuals. Only after the latent period, denoted by $1/\sigma_0$, individuals 'evolve' from being exposed to being pre-symptomatic. The number of individuals in the E_i compartment therefore decreases exponentially by the number of individuals that are exposed at time t divided by the latent period, which is represented by the minus term in (2.11).

- Pre-symptomatic.
  The P_i term represents the pre-symptomatic individuals, which are able to infect susceptible individuals but do not display symptoms (yet). After a pre-symptomatic latent period of $1/\sigma_1$, individuals move to either I_{s_i} or I_{a_i}. The number of individuals in the P_i compartment therefore decreases exponentially by the number of individuals that are exposed at time t divided by the latent period, which is presented by the minus term in (2.12).

- Infectious.
  The I terms represent the number of infectious individuals, which are divided into categories I_{s_i} and I_{a_i}, denoting symptomatic and asymptotic as explained previously in this section.

- Removed.
  R_i denotes the number of removed individuals in compartment i, i.e. recovered, vaccinated (with resulting immunity) and deceased.

To clarify these terms; when a person is infected by another person with the virus, it enters the exposed compartment, where it is not yet infectious. After the latent period, the person transfers to the pre-symptomatic compartment from where it either proceeds to the symptomatic or asymptomatic infectious compartment. Lastly, after infectiousness ends, the person enters the removed compartment. The latent periods are crucial for the dynamics regarding virus spread, since they give information about the infectiousness of an individual at a certain time.

It is important to note the difference between pre-symptomatic and asymptomatic. An asymptomatic individual will be tested positive for the virus but not present any COVID symptoms during its infectious period, whereas a pre-symptomatic person will initially present itself as asymptomatic but can develop symptoms days or even weeks after testing positive [20].

An additional note needs to be made about the parameter τ that is used in equation (2.9)-(2.11) to describe a certain constant flow from the susceptible compartments to the exposed compartments. Since human interaction and increased susceptibility due to age and seasonality factors are already accounted for, we can assume τ to be the rate at which people get infected in ways that are unrelated to human contact, e.g. touching

or eating a product in the supermarket that has been contaminated.

To take into account certain factors that are specific to this COVID-pandemic, such as the adherence to NPIs and the closure of work- and school places, the SEPAIR model is equipped with a lot of extra parameters. A short description of most of these parameters is provided in table 2.1.

TABLE 2.1: Parameter definitions

Parameter	Meaning
r	probability of transmission per contact
ρ_i	susceptibility-factor for age group i
s sin terms	variation in seasonal susceptibility
ϕ	phase of seasonal susceptibility
C_{ij}	contact rate between age group i and j
σ_0	inverse of latent period for exposed individuals
σ_1	inverse of latent period for pre-symptomatic individuals
γ_a	inverse of infectious period for asymptomatic individuals
γ_s	inverse of infectious period for symptomatic individuals

Most of these parameters are calibrated using the available data in Ontario [16], except for the probability of transmission per contact r and the contact rate between age group i and j, C_{ij}. It is important to note that the epidemiological dynamics together with the calibrated parameters can not model the pandemic without the crucial information provided by r and C_{ij}, since the probability of transmissibility and the rate at which individuals can transmit the virus by being in contact are arguably the most important factors in an epidemic.

We shall focus first on the method to obtain C_{ij}, and continue with the calculation of r in chapter 6. To calculate C_{ij}, which denotes the contact rate between age group i and j, variables that depend on government shutdown policies and individual adherence to Non-Pharmaceutical Interventions (NPIs) are used. The choice to adhere to NPIs depends on the possible cost or gain of practicing them. This 'game' in which individuals will opt for the highest possible payoff will be the main subject of the next chapter. After examining the dynamics behind NPI adherence, we will return to the contact rate C_{ij}.

3 NPIs

During the emerge of a new virus, certain Non-Pharmaceutical Interventions (hereafter abbreviated as NPI) are often the first response that is taken to minimize spread [1]. NPIs include for example social distancing, the wearing of face masks, and the disinfecting of our hands. For COVID-19, we see that the strict adherence of the population to these NPIs reduced the number of covid-cases effectively, resulting in a 'flattening of the curve'[19]. Over time however, we see that the adherence to NPIs decreases, depending on the number of reported infected people. This is seen as a 'free-rider problem'[1]; People who choose not to practice NPIs benefit from people who do.

In our article, the fraction of the population that adheres to the NPIs is modelled using evolutionary game theory. This is supported by the fact that an individual's choice to adopt a certain strategy depends on the choice of others, weighing the risks and benefits of switching to another strategy.

We assume that there are only two strategies: either to adopt the NPIs or not. This is described mathematically in [1], by the imitation dynamic. In [1], the imitation dynamic is used to determine the strategy of individuals to either get vaccinated or not. This same method can however also be used to determine the strategy to practice NPIs, as will be explained in the following paragraphs.

The imitation dynamic assumes that people sample other people at some constant rate s and that a person switches to the other person's strategy if it has a higher payoff with a probability proportional to the expected gain in payoff ΔE [8][15][3]. For the following equations, the same notation is used as in [1] with a short description, but the reader is kindly referred to this article for a more extensive analysis.

The payoff to practice NPIs will be taken as:

$$E_v = B - c_v \tag{3.1}$$

Where B denotes the baseline payoff that represents a state of perfect health and c_v denotes the payoff penalty to practice NPIs. For COVID, one can think of this as the 'cost' to reduce social activities and money spent to maintain better hygiene, e.g. face masks and soap.
The payoff to *not* practice NPIs will be taken as:

$$E_n = B - c_i m L \tag{3.2}$$

With again, B being the baseline payoff in state of perfect health and c_i the penalty for getting infected. m is the probability of disease transmission and L is the number of

cases at time t. Note that $L = \sum_{i=1}^{16} \alpha_i (I_{s_i} + I_{a_i})$ in our case.

We can subtract equation (3.1) from (3.2) to obtain the profit for an individual that changes from NPI-practitioner to non-practitioner:

$$\Delta E_{nv} = E_n - E_v = c_v - c_i m L \tag{3.3}$$

Since individuals sample others at sampling rate s, this can be combined in the following equations. Let x be the fraction of the total population that is practicing NPIs. Equation 3.4 gives the rate at which individuals switch from adhering NPIs to not adhering. Note that this rate is given by the probability to encounter a non-practitioner times the payoff gain. Since the payoff difference will be negative if $c_v - c_i m L = \Delta E_{nv} \leq 0$ the rate to change from NPI adherer to non-NPI adherer will be zero, i.e. individuals will only change strategy if there is an improvement in expected payoff.

$$\begin{cases} sx(1-x)\theta(c_v - c_i m L) & \text{when } c_v - c_i m L = \Delta E_{nv} > 0 \\ 0 & \text{when } c_v - c_i m L = \Delta E_{nv} \leq 0 \end{cases} \tag{3.4}$$

A similar equation can be found for the rate at which people decide to start practicing NPIs, also depending on the different payoffs for each strategy. The total rate of change in the fraction of the population adhering to NPIs; given by the rate at which non-adherers switch to adhering minus the rate at which adherers switch to non-adhering, is given by:

$$\frac{dx}{dt} = s\theta x(1-x)(c_v - c_i m L) \tag{3.5}$$

Rewriting this equation and substituting the following parameters:

$$s\theta c_i m = \kappa \tag{3.6}$$

$$\frac{c_v}{m c_i} = \omega \tag{3.7}$$

Gives us the behavioural model with social learning feedback:

$$\frac{dx}{dt} = \kappa x(1-x)(L - \omega) \tag{3.8}$$

Recall that L is the total number of cases present at time t, so the fraction of total cases in the population is given by $\frac{\sum_{i=1}^{16} \alpha_i (I_{s_i} + I_{a_i})}{\sum_{i=1}^{16} N_i}$. Furthermore, ω depends on the penalty for practicing NPIs, which changes over time and is also depending on the fraction of the population x that already practices NPIs. We can replace ω therefore with cx. The evolution of the proportion of individuals who practice NPIs is given by[16]:

$$\frac{dx}{dt} = \kappa x(1-x) \left(\frac{\sum_{i=1}^{16} \alpha_i (I_{s_i} + I_{a_i})}{\sum_{i=1}^{16} N_i} - cx \right) + p_{ul}(1-2x) \tag{3.9}$$

All the terms in this equation are explained previously in this section, except for the phenomenological term p_{ul}. To comprehend its function, let us take a closer look at the dynamics of equation (3.8).

Let L and ω both be constant. In reality, both of these variables are changing with time, but we view them as constants for the sake of understanding the role of p_{ul}. We can distinguish two cases:

$$(L - \omega) > 0 \tag{3.10}$$

If $(L - \omega) > 0$ then the solution of (3.8) will increase until it reaches 1 (note that $k > 0$).

$$(L - \omega) < 0 \tag{3.11}$$

If $(L - \omega) < 0$, then the solution for (3.8) will decrease until it reaches 0.

This can be interpreted as follows: if the number of infected COVID-19 L individuals is small in comparison to the number of individuals practicing NPIs $\omega = cx$, x will decrease. Since there are not a lot of infected individuals, the cost to practice NPIs will not outweigh the payoff, and $\omega = cx$ will decrease. At a certain point however, this decrease will reach a turning point when $(L - \omega) > 0$, i.e. the number of infected individuals is large in comparison to the number of NPI practitioners. The payoff to practice NPIs will then be larger than the cost to practice them, resulting in an increase of x.

Now, if we had $p_{ul} = 0$, as in equation (3.8), the solutions $x = 0$ and $x = 1$ would be fixed points. This means that if the system would reach a state where the whole population decides to either practice NPIs or to *not* practice NPIs, it will stay in that state. No one will ever start practicing NPIs if $x = 0$ or stop practicing NPIs if $x = 1$. To avoid this, p_{ul} is introduced.

Let $p_{ul} > 0$. Then $(1 - 2x) = 1 > 0$ if $x = 0$. This means that if the whole population is not practicing NPIs while the number of infections L might be extremely high, the system will not stay in the state where no one will start to adhere to the NPIs but x will increase.

On the other hand, $(1 - 2x) = -1 < 0$ if $x = 1$. This means that the state wherein the whole population is practicing (costly) NPIs while the number of infections might be extremely low, will not hold but still decrease in time.

In [16], p_{ul} is described as the term that encompasses the social heterogeneity and influence from external populations, which follows naturally from the fact that p_{ul} can be seen as the few individuals that 'go against the trend' of imitating an individual's strategy if it has a higher payoff. We do like to note that since these are only a few individuals, p_{ul} is often very small.

As one may have noticed already, the term representing the fraction of the population that adheres to NPIs, x, is not present anywhere in the governing epidemiological

differential equations that are introduced in 2. This is however not completely true, since they are implicitly present through the contact matrix term, C_{ij}, which is calculated using x. The exact relation between these variables will be explained in the following chapter.

4 Contact matrix

To further examine the social dynamics of the COVID-19 spread, a closer look has to be taken at contact patterns between individuals of certain age groups. Data is available for the amount of time that different age groups are in contact [23]. This data is used and altered to fit the interaction change due to individual adoption of NPIs and government policies to close workplaces and schools. The term $C_{ij}(t,x)$, introduced in 2.1 as the contact rate between age group i and j, can be defined more concretely as the average number of contacts per day at workplaces, schools, households, and other locations, as a function of NPI adherence and time:

$$C_{ij}(t,x) = C_{ij}^{W}(t) + C_{ij}^{S}(t) + (1-\epsilon_P x(t))(\overline{C}_{ij}^{O} + \overline{C}_{ij}^{H}) \tag{4.1}$$

With the superscripts, W, S, H, O, representing the different places i.e. work, school, house and other. The contact rate at work develops in the following way:

$$C_{ij}^{W}(t) = \begin{cases} (1-\epsilon_W)\overline{C}_{ij}^{W} & \text{if } t > t_{close}^{w}, t < t_{open}^{w} \\ \overline{C}_{ij}^{W} & \text{if } t < t_{close}^{w} \\ (1-D(t)(1-\epsilon_W))\overline{C}_{ij}^{W} & \text{if } t > t_{open}^{w} \end{cases} \tag{4.2}$$

The first term will be put into operation if $t > t_{close}^{w}, t < t_{open}^{w}$, so after workplaces close due to government policies. The $\overline{C}_{ij}^{W}$ represents the pre-corona ('normal') contact hours per day between individuals of age i and j and $0 < \epsilon_W < 1$ is the workplace closure effectivity which is less than perfect due to unauthorised occasional use of workplaces, the inability to close essential workplaces like supermarkets, hospitals, etc. Note that if $\epsilon_W \to 0$, i.e. workplace closure efficacy approaches zero, the contact rate will approach its pre-pandemic value $C_{ij}^{W}(t) \to \overline{C}_{ij}^{W}$. If $t < t_{close}^{w}$, the workplaces will remain open and the contact hours will be identical to those in the non-pandemic situation. For $t > t_{open}^{w}$, the reduced contact rate if active, which depends on the workplace closer effectiveness and the closure level $D(t)$.

For schools we have a similar equation, except for the assumption of perfect closure effectiveness for school places:

$$C_{ij}^{S}(t) = \begin{cases} 0 & \text{if } t > t_{close}^{s}, t < t_{open}^{s} \\ \overline{C}_{ij}^{S} & \text{if } t < t_{close}^{s} \\ (1-D(t))\overline{C}_{ij}^{S} & \text{if } t > t_{open}^{s} \end{cases} \tag{4.3}$$

An important note to make here is that we did not include the delay between an individual's decision to start practicing NPIs (and hence cease contact in case of space

closure) and its impact on transmission. The correct notation is therefore $t = t - t_{delay}$. Also, contact hours spent at home are expected to rise as work and school places close, but since there rapid saturation of infection will occur due to repeated contacts at home, this will have a negligible effect.

Since the government shutdown policies do not have total control on social interaction within households and public places other than houses and schools, the COVID influence on contact rates $\overline{C}_{ij}^{O} + \overline{C}_{ij}^{H}$ is given by the fraction of the population that is practicing NPIs at that time, corrected by the efficacy ϵ_P of NPI adoption. One could argue that NPI adherence does not have the same efficacy at home as at other places. However, the contacts at home are repeated far more frequently which leads to rapid saturation and infection risk reduction, which in combination with the lack of practicing NPIs at home can be set at equal efficacy as the adoption of NPIs at other places where a lot of new contacts are made. Additionally, since no data was provided about the location of infection (e.g. at home or at public spaces), a location-specific efficacy cannot be determined.

The sum of the contact rate at all of the places mentioned above will result in the total number of contacts an individual has per day: $C_{ij}(t, x)$. This term is present *directly* in the epidemiological differential equations in section 2 to determine how many susceptibles from the S_i^1 and S_i^2 compartments move to the exposed E_i compartment due to interaction with an individual from the infectious compartments $I_{s_j} + I_{a_j} + P_j$. However, the contact rate is also present *indirectly* through the transmission rate r on which we shall elaborate more in the next chapter.

5 Next generation matrix and reproduction number

[4] [6] [12] [21] As the covid-pandemic evolves over time, there is one term in particular that keeps returning to describe the status of the spread of the virus at a certain point in time: the basic reproduction number R_0. R_0 is defined as the average number of new cases of an infection caused by one typical infected individual in a population that exists completely of susceptibles [6] [12]. Exponential growth in the number of COVID cases is indicated with an R_0 larger than one, meaning that one infected individual results in more than one new infection, whilst an R_0 smaller than one will let the disease die out [4].

Using the epidemiological dynamics that are described by the differential equations in chapter 2, one can calculate R_0 using the 'next-generation matrix'. This method is presented in [6]. One important thing to note is that the calculation of R_0 depends on the probability of transmission per contact r, and we assume that the basic reproduction number R_0 is known (it can be calibrated using additional data) while the probability of transmission r is not. To clarify this, we will use $\overline{R_0}$ to represent the reproduction number that is *calibrated*, using the data provided by the Treasury Board Secretariat of Ontario [18] and the google mobility data[7]. For the reproduction number that is *calculated* using the epidemiological dynamics, we use R_0.

Since we now have two ways to obtain the reproduction number, one involving the unknown baseline transmission rate r, we can extract r by assuming $\overline{R_0} = R_0$. Before we can do this, we first need to show how we can obtain R_0 using the *transmission* and *transition* dynamics.

5.1 Calculation R_0 using next-generation matrix

As stated in the previous paragraphs, we use the (next-generation) method from [6] to calculate R_0 and afterwards extract r (and hence do not assume r to be a known constant). The next-generation matrix can be defined for mathematical models that describe the evolution of infectious diseases in a population which can be divided into a finite number of discrete categories that influence the behaviour of the virus [6]. Examples of these categories could be for examples gender or species. Note that in our case the population is divided by finitely many age groups, and in finitely many stages of the disease.

The name of the 'next generation' matrix, usually denoted by **M**, stems from the fact that from a mathematical perspective, infection transmission through contact is very similar to a demographic process where giving birth is analogous to causing a new infection through transmission[6]. This process then leads to new 'generations' of infected individuals, for which the size can increase rapidly generated by a large growth factor that can be seen in a natural way as R_0 [21][6]. Or more explicitly: R_0 relates the numbers of newly infected individuals in the various categories to the numbers of the 'next generation'.

The next-generation matrix calculates how people in a certain state give rise to other infections and includes the amount of time people stay in one compartment, e.g. in one age-group or in one specific sickness stage (susceptible, exposed, infectious, recovered). In the SEPAIR-model that is used in this thesis, all states change over time, but this would be different if we were to include certain genetic characteristics for example, which would remain constant.

To calculate R_0 we only need to include the stages of the virus in which individuals are infected and the first step is therefore to review the differential equations that cause new infections and those who describe the change in the infectious compartments. These equations will be referred to as the *infected subsystem* and we will be linearizing these around the disease-free equilibrium, which is assumed to exist. What we end up with is a Jacobian matrix that is derived by linearization of the original (nonlinear) differential equations.

This Jacobian matrix, derived from the epidemiological dynamics can be decomposed in a *transmission* part F, and a *transition* part V. Next we consider the dominant eigenvalue (or spectral radius ρ) of FV^{-1}. To prove that this decomposition is possible, we use the method described in [21]. This article considers R_0 as a threshold parameter, where if $R_0 < 1$, a locally asymptotically stable disease free equilibrium (DFE) is present, and if $R_0 > 1$, the DFE is unstable and an epidemic is always possible. We will not get into detail about the specifics regarding the DFE in this thesis, but we will examine the method that is used to justify the decomposition of the next generation matrix. Note that the decomposition step is necessary to obtain the *calculated* R_0 from the epidemiological dynamics.

Let $x = (x_1, ..., x_n)^t$, with each $x_i \geq 0$, be the number of individuals in each compartment. Let the first m compartments denote those that correspond to the infected individuals (i.e. the *infected subsystem*). We can then define $\mathbf{X}_s$ to be the set of all disease free states. We can write:

$$\mathbf{X}_s = \{x \geq 0 | x_i = 0, i = 1, ..., m\}. \tag{5.1}$$

Since we have individuals moving in and out of compartments due to being newly infected, or due to other changes for example getting older, it is important to distinguish between both transitions. We define the following:

- $\mathscr{F}_i(x)$ is the rate at which new infections appear in compartment i.

- $\mathscr{V}_i^+(x)$ is the rate of transfer of individuals into compartment i by all other means (e.g. they turn a certain age where they enter a different age class).
- $\mathscr{V}_i^-(x)$ is the rate of transfer of individuals out of compartment i.

The disease transmission model has non-negative initial conditions and evolves according the following system of equations:

$$\dot{x}_i = f_i(x) = \mathscr{F}_i(x) - \mathscr{V}_i(x), i = 1, ..., n, \tag{5.2}$$

With $\mathscr{V}_i(x) = \mathscr{V}_i^-(x) - \mathscr{V}_i^+(x)$. We assume every function to be continuous and twice continuously differentiable in each variable. Also each function satisfies the following assumptions:

(A1) if $x \geq 0$, then $\mathscr{F}_i(x), \mathscr{V}_i^+(x), \mathscr{V}_i^-(x) \geq 0$ for $i = 1, ..., n$.

(A2) if $x_i = 0$ then $\mathscr{V}_i^-(x) = 0$. In particular, if $x \in \mathbf{X}_s$, then $\mathscr{V}_i^-(x) = 0$ for all $i = 1, ..., m$.

(A3) $\mathscr{F}_i(x) = 0$ if $i > m$.

(A4) if $x \in \mathbf{X}_s$ then $\mathscr{F}_i$ and $\mathscr{V}_i^+(x) = 0$ for $i = 1, ..., m$.

(A5) if $\mathscr{F}(x)$ is set to zero, then all eigenvalues of $Df(x_0)$ have negative real parts.

(A1) follows from the fact that each function represents directed transfer of individuals, so they are all greater or equal than zero (there is no negative transfer).

(A2) accounts for the fact that if a compartment is empty, there can be no transfer out of the compartment. Since we defined $\mathbf{X}_s$ as all the disease free states, these are all empty in our system and thus is our outflow, $\mathscr{V}_i^-(x)$, equal to zero.

(A3) follows from the fact that compartments $i > m$ denote the uninfected compartments, no individuals will enter these compartments through infection: $\mathscr{F}_i(x) = 0$ if $i > m$.

(A4) described the assumption that if a population is free of the disease, that it will remain uninfected. So if an infected compartment $x_i, i \leq m$ has no individuals left ($x_i = 0$), then the rate at which new infections appear in compartment i and the rate at which individuals transfer into compartment i through other ways are both equal to zero.

(A5) regards the derivatives of $f = (f_1, ..., f_n)$ near a disease free equilibrium, i.e. $Df(x_0)$ is the derivative $[\frac{\partial f_i}{\partial x_j}]$ evaluated at the DFE (x_0) or also called the Jacobian matrix. We consider (5.2) restricted to $\mathbf{X}_s$ and a population near the DFE x_0. If the introduction of a few infected individuals to this system will not result in an epidemic, the system will return to the DFE according to the following system:

$$\dot{x} = Df(x_0)(x - x_0) \tag{5.3}$$

Now, assumption (A5) ensures that our analysis is restricted to systems where the absence of new infections will result in a stable DFE. Thus, when $\mathscr{F}(x)$ is set to zero, we demand a stable disease free equilibrium. A necessary condition for this to be true is that the real part of the eigenvalues of $Df(x_0)$ must be negative.

The conditions provided by (A1)-(A5) allow us to partition the Jacobian Matrix $Df(x_0)$ by the following lemma:

Lemma 6.1 ([21], lemma 1)
If x_0 is a DFE of (5.2) and $f_i(x) = \mathscr{F}_i(x) - \mathscr{V}_i(x)$ satisfies (A1)-(A5), *then the derivatives $D\mathscr{F}(x_0)$ and $D\mathscr{V}(x_0)$ are partitioned in n x n matrices as*

$$D\mathscr{F}(x_0) = \begin{pmatrix} F & 0 \\ 0 & 0 \end{pmatrix} \text{ and } D\mathscr{V}(x_0) = \begin{pmatrix} V & 0 \\ J_3 & J_4 \end{pmatrix},$$

where F and V are m x m matrices, i.e. consisting of only the rate of change in transmission and transition of individuals in the infectious compartments, defined by

$$F = \left[\frac{\partial \mathscr{F}_i}{\partial x_j}(x_0)\right] \text{ and } V = \left[\frac{\partial \mathscr{V}_i}{\partial x_j}(x_0)\right] \text{ with } 1 \leq i, j, \leq m.$$

In addition, F is non-negative, V is a non-singular M-matrix, meaning that the eigenvalues are non-negative and non-negative and its diagonal elements are positive, and all eigenvalues of J_4 have a positive real part.

Proof. Consider matrix (5.4). Let the upper left square A be the entries where a change in *transition* and *transmission* is considered within the *infected subsystem*, so A is a mxm matrix. Then the upper right square B is a $(mx(n-m))$ matrix, C a $((n-m)xm)$ matrix and D and $((n-m)x(n-m))$ matrix.

Let $x_0 \in \mathbf{X}_s$ be a DFE. By (A3) and (A4), $F = (\partial \mathscr{F}_i/\partial x_j)(x_0) = 0$ if either $i > m$, since there is no 'inflow' of infected individuals in these compartments, or $j > m$, since this refers to the change of infected individuals in the uninfected compartments and is equal to zero, hence the only nonzero entries will occur in the upper left square A from 1 to m if we were to look at it from the structure described by matrix (5.4). We use a similar argument for $D\mathscr{V}(x_0)$. Let $x_0 \in \mathbf{X}_s$ be a DFE. By (A2) and (A4), $V = (\partial \mathscr{V}_i/\partial x_j)(x_0) = 0$ if $i \leq m$, since $\mathscr{V}_i^+ = \mathscr{V}_i^- = 0$ if $i \leq m$ and $j > m$, which refers to the upper right square in the matrix. Furthermore, by (A1) and (A4) we can conclude that F is non-negative.

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{5.4}$$

Note that we only included the proof of the first part of the lemma, since it relies heavily on the assumptions made in (A1)-(A5). For the second part of the proof regarding the claim that V is a non-singular M-matrix, the reader is referred to the original article [21] and the background information that can be found in [2].

The sign of the eigenvalues relates to the fact if the DFE is locally stable or not, which in turn relates to R_0 in the way that an unstable DFE relates to a $R_0 > 1$ and an unstable DFE relates to a $R_0 < 1$. Since we have decomposed the original system, it is important to check its effect on the sign of the eigenvalues. In conclusion: we need to check if the conditions where $R_0 > 1$ and $R_0 < 1$ are invariant under decomposition. This will be explained (briefly) in the following paragraph, but due to the technical details and the lack of relevance to this particular thesis, we will not treat it extensively and follow the reasoning from [21] very closely. Again, the interested reader is referred to the original article [21].

If we introduce a 'typical' infected individual into a population where reinfection is turned off, the dynamics are governed by ([21], eq. 3):

$$\dot{x} = -D\mathcal{V}(x_0)(x - x_0) \tag{5.5}$$

Note that by (A5), the DFE is locally stable in this system and we can use this equation to determine what happens if a small number of infected individuals enters a disease free population. Let $\psi_i(0)$ the number of infected individuals at time $t = 0$ and let $\psi(t) = (\psi_1(t), ..., \psi_m(t))^t$ represent the number of infected individuals that are still present in the infected compartments after a time t has passed. We only consider the infected ($i = 1, ..., m$) compartments, so the upper left square of the matrix defined in lemma 1. Due to the partition of the derivatives, we know that the solution for $\psi(t)$ has to satisfy $\psi(t)' = -V\psi(t)$ (the minus sign originating from (5.2)) which is uniquely given by $\psi(t) = e^{-Vt}\psi(0)$. Furthermore, by lemma 1, the eigenvalues of V have positive real parts.

To obtain the number of new infections caused by the initially introduced infected individuals we can integrate $F\psi(t)$ with t from zero to infinity, which has as solution $FV^{-1}\psi(0)$. FV^{-1} can be interpreted as follows. If we have an infected individual entering at compartment k at the situation where the population only exists of susceptibles (so a disease free situation) then entry (j, k) of V^{-1} will be the average time this individual spends in compartment j, so it is the expected time that an individual currently in state k will be in state j in its future epidemiological life [6]. The (i, j)-th entry of F represents the rate at which new infections are caused in compartment i, by an infected individual in compartment j. In summary: the (i, k)-th entry of FV^{-1} represents the number of infected cases produced by an individual that was originally introduced into compartment k [21].

$FV^{-1} = \mathbf{M}$ is defined as the next generation matrix by [5] and we R_0 is defined as the dominant eigenvalue FV^{-1}, or more precisely the spectral radius ρ of FV^{-1} [6]:

$$R_0 = \rho(FV^{-1}) \tag{5.6}$$

By Theorem 2 of [21], the spectral radius of FV^{-1} (and thus R_0) os smaller than 1 if and only if the real part of the eigenvalues of the *infected subsystem* part of the next generation matrix have a negative real part. This implies dying out of the disease in the population. On the other hand, we have a $R_0 > 1$ if and only if (at least) one of eigenvalues the next generation matrix has a non-negative real part[21][6]. Again, we would like to stress that we treated this proof in a very general way and more details can be found in the articles cited above.

5.2 Transmission rate r

Let us conclude this chapter by returning to the article of [16] and the goal of obtaining the baseline transmission rate r from a *calculated* R_0 as well as a *calibrated* $\overline{R_0}$.

Note that for M as presented in equation 17 in [16], we have $F = -\Sigma$ for the transition part of the next generation matrix. We can divide all elements of $-\Sigma$ by the scalar r and denote the new next-generation matrix as $\mathbf{M}$. Note that we have $r\mathbf{M} = M$. Since we want to obtain the transmission rate r for which $\overline{R_0} = \rho(M)$, we have:

$$\overline{R_0} = \rho(M) \tag{5.7}$$
$$\overline{R_0} = \rho(r(\mathbf{M})) \tag{5.8}$$
$$\overline{R_0} = r\rho(\mathbf{M}) \tag{5.9}$$
$$r = \frac{\overline{R_0}}{\rho(\mathbf{M})} \tag{5.10}$$

As stated before, $\overline{R_0}$, is calibrated using data and ranges from, 1.5 to 2.5. However, an additional R_0 can be calculated epidemiological dynamics, the contact matrix, and the average time that people spend in sickness stage and age compartments which are all used in the next-generation matrix. Using the methods described earlier in this section, it is possible to extract r from the calculated R_0.

An important thing to note here, is that $\overline{R_0} \neq \rho(\mathbf{M})$. Of course, if they were equal the transmission rate would just be one and we don't want this to be the case, but other than that, $\rho(\mathbf{M})$ can be seen as the dominant eigenvalue of the next generation matrix *divided* by the scalar r. This is already implied by our assumption that $r\mathbf{M} = M$, but to validate our method of extracting from $\rho(\mathbf{M})$, recall the following: if we have matrix A with eigenvalue λ, and multiply (all elements of) A by scalar r, its eigenvalue is multiplied by r as well. This validates the steps made in equations (5.8) and (5.9).

We will close this chapter by providing more information about the calibration of variables and sensitivity analyses that is performed for certain scenarios, before moving on to some results.

5.3 Calibration of variables and sensitivity analysis

Now that we have explained a bit more about the next-generation matrix and R_0, further elaboration on the calibration of the variables should be given. For the calibration, data about the confirmed positive cases in Ontario, Canada, is used as well as seroprevalence data to make sure that under-reporting of cases did not influence the estimated transmission [7] [18] [14].

Other than the calibration of certain variables, sensitivity analyses are performed that: "explore the effects of vaccine efficacy assumptions and a higher R_0, ascertainment rate, or social learning rate, and that incorporate dynamics of vaccinating behaviour [16]." This is a method to evaluate how much the uncertainty of certain individual aspects of a model contribute to the uncertainty as a whole. All though we will not discuss this topic in further detail, the outcomes of the sensitivity analyses on the earlier mentioned elements are presented in the appendix of [16].

6 Results

6.1 Initial conditions

Before the results are presented, we need to specify the initial conditions that are used in this study. Note that we used this data directly from [16]. As stated in 1, the simulation is situated in Ontario, Canada. $t = 0$ is chosen to be March 10th 2020, which is the day at which the number of COVID cases surpassed 50. This number is taken to rule out any extreme stochastic effects that are present at a lower 'start' number [10].

For this initial infected group, $I_a(t) + I_s(t)$, we can define an age distribution. Let ω_i be the number of initial cases in age group i. To determine the *true* amount of cases at time $t = 0$ we first need the ascertainment rate α_i that was introduced in 2. Recall that it denotes the fraction of the total cases in age group i that are reported. Dividing ω_i by α_i would hence account for the cases that remain unnoticed. Since we do not know the actual active number at $t = 0$, the constant I_0 is introduced. I_0 is treated as a fitted model variable and multiplied by $\frac{\omega_i}{\alpha_i}$ to determine the true number of cases.

The numbers of pre-symptomatic and exposed individuals are treated in a similar way and we end up with the following starting conditions:

$$I_{s_i}(0) = \eta I_0 \frac{\omega_i}{\alpha_i}, \quad I_{a_i}(0) = (1-\eta) I_0 \frac{\omega_i}{\alpha_i}, \quad P(0) = P_0 \frac{\omega_i}{\alpha_i}, \quad E(0) = E_0 \frac{\omega_i}{\alpha_i} \tag{6.1}$$

Where E_0 and P_0 are fitted accordingly and η denotes the fraction of symptomatic infections, which was set to 0.15 [17].

Furthermore, it is assumed that the total number of susceptible, unvaccinated individuals is equal to the whole population minus the initial infected group: $S_i^1(0) = N_i - (I_a(0) + I_s(0) + E(0) + P(0))$ and $S_i^2 = 0$ and $S_i^2(0) = 0$ and $R_i(0) = 0$ for all i. The fraction that adheres to NPIs at $t = 0$ is set to 1%, so $x(0) = 0.01$ and the closure level is set zero, $D(0) = 0$.

6.2 Vaccination strategies

When administering a vaccine, it is important to look at efficacy. For example, an efficacy against disease of 80% means that people who are vaccinated have an 80% lower chance of getting ill than people who are not. This does not mean that 80% will not get ill, it just lowers the risk by 80% and it is important to note this difference. Additionally, it is important to note that there is a difference between efficacy solely against

transmissibility and efficacy against *transmissibility and disease*. The first only refers to the reduction in change to transmit the virus to a new individual, whereas the latter also includes the reduction in getting ill. For this study, a vaccination efficacy of 75% [22] against transmissibility is taken and denoted by v_{T_i}.

The total number of people that are vaccinated is taken to be $\sum_{i=1}^{16} \phi \frac{S_i(t)}{N_i}$ per day and the total number of people that are completely vaccinated against transmission is $\sum_{i=1}^{16} v_{T_i} \frac{S_i(t)}{N_i - V_i}$ where $N_i - V_i$ represents the total number of individuals minus the number of vaccinated individuals in age group i. If there is a number of vaccinations left after full vaccination of a certain age group, these will be evenly distributed amongst the non-vaccinated age groups. This is the case if the number of the remaining individuals in age group S_i^1 is lower than ϕ_i.

In addition to the model that is used, an option to include vaccination refusal is presented in the article, which is very similar to the dynamics that describe NPI adherence. Another extension is made where a difference between vaccinating against transmissibility, and transmissibility and disease, by reducing the mortality rate. However, we will not include these features in this thesis.

6.3 Results

Before presenting some results, let us reconsider the different vaccination strategies that were used to determine whether or not practicing certain age groups for vaccination would result in fewer casualties from COVID-19:

- Vaccinating people who are of ages 60 and older (oldest first strategy)
- Vaccinating people aged 20 and younger (youngest first strategy)
- Vaccinating uniformly by age (uniform strategy)
- Contact-based strategy

For the contact-based strategy we look at the contact matrix C_{ij}, which is discussed in chapter 4, to see which age groups are responsible for the most transmission. This strategy tends to practice vaccination for individuals in the age class 15-19 first, followed by individuals from age classes 20-59, as can be seen in figure 6.1.

A lot of figures are presented in [16], for example the effect of a different vaccination rate, or variation in individual cost to practice NPIs (low, moderate, high). As said in 5, additional sensitivity analysis is performed for certain variables to examine how much other variables are affected by their uncertainty. We decided to only include a few of these figures which were deemed most relevant for this thesis.

Firstly it is important to validate the model by using additional data. This was done for a total number of cases per age class. As we can see in figure 6.2, the actual number of cases does not lie within the modelled interval for some of the youngest and oldest age groups. This might have to do with the fact that only the parameters regarding the number of individuals an age group N_i, the COVID-19 case fatality rate μ_i (not

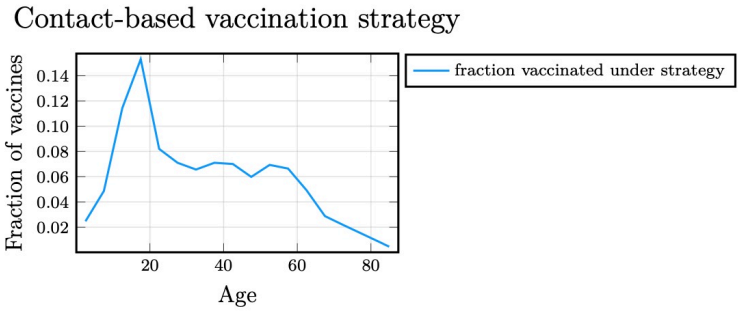

FIGURE 6.1: Figure and caption from ([16], appendix figure S6) **Age distribution of vaccination under the contact-based strategy.** This strategy vaccinates proportionally to the leading eigenvector of the full contact matrix, $C(0)$, to vaccinate people who will, approximately, produce the most secondary infections in a linearized regime.

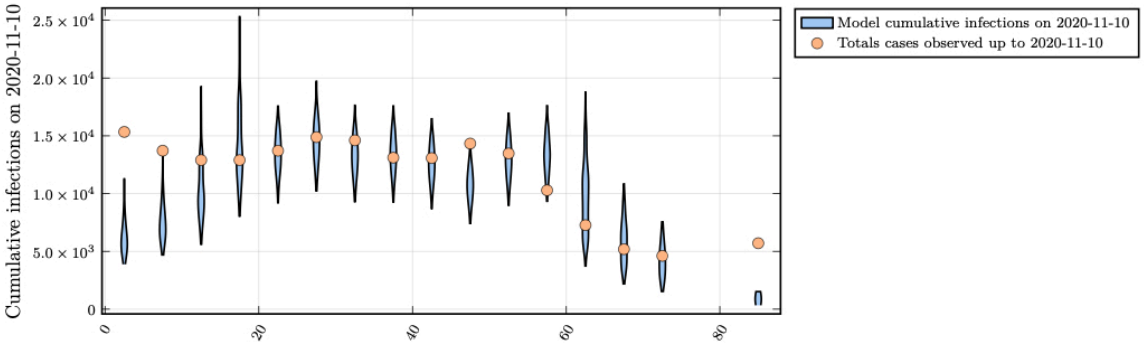

FIGURE 6.2: Figure and caption from ([16], appendix figure S4). **Fig. S4. Empirical data of cumulative infections due to COVID-19 by age and model posterior predictions.** The age-specific total cases at the end of the fitting window, were used to calibrate the model, in an age dependent way. We used only three parameters to capture age-specific effects and therefore trade-off some accuracy in the youngest and oldest age groups.

treated separately but obtained from [11] and interpolated to the 16 age classes), and the susceptibility ρ_i, are age-specific. Since three is a rather insufficient number of parameters to include all age-specific effects in this model, the accuracy of the number of total cases in the youngest and oldest age classes is somewhat lacking.

Three different regimes for the model dynamics were identified: an early regime, an intermediate regime, and a late regime. The regimes were tested under all four vaccination strategies for infection incidence over time, percentage reduction in mortality and the total number of COVID-casualties that were prevented. We described these below and also included modelled results for infection incidence in figure 6.3 and percentage reduction in mortality in figure 6.4. Furthermore, a few claims are made regarding the best vaccination strategies in certain regimes for which the exact numbers can be found in chapter 1.

For the first (early) regime, we have that vaccination starts at January 1st, 2021 with a vaccination rate that is $\geq 1.0\%$ of the population per week. In this regime a possible

third wave of infections is avoided, and the oldest first strategy saves most lives.

The second regime is an 'intermediate' scenario; either vaccination starts on January 1st 2021 and the rate at which vaccination is performed is low, $\leq$ 0.5%, or the vaccination starts September 1st 2021 and the vaccination rate is high ($\geq$ 1.5%). In this regime, a third wave of infections is not avoided, since not enough individuals will be vaccinated to stop it. As a consequence, the oldest-first strategy performs worse than the uniform and contact-based strategies if we look at the incidence of infection and mortality rate as also can be seen in 6.3 6.4. However, if we increase the vaccination rate enough, the effectiveness of all vaccination strategies will converge as it will not take too long before the population is vaccinated completely [16].

In the third (late) regime, the vaccination starts September 1st, 2021 and the vaccination rate is low, $\leq$ 1.0% of the population per week. Even though there is a relatively high group immunity present in the population, the slowness of vaccination will not result in protection against a third infection wave. Consequently, the oldest first strategy will be the most effective for saving lives.

In this chapter, we have presented the most important results, based on the modelled evolution of the COVID-19 virus in a landscape that changes based on epidemiological, as well as social components. In view of these results, conclusions and recommendations about vaccination strategies can be made and will be presented and discussed in the next chapter.

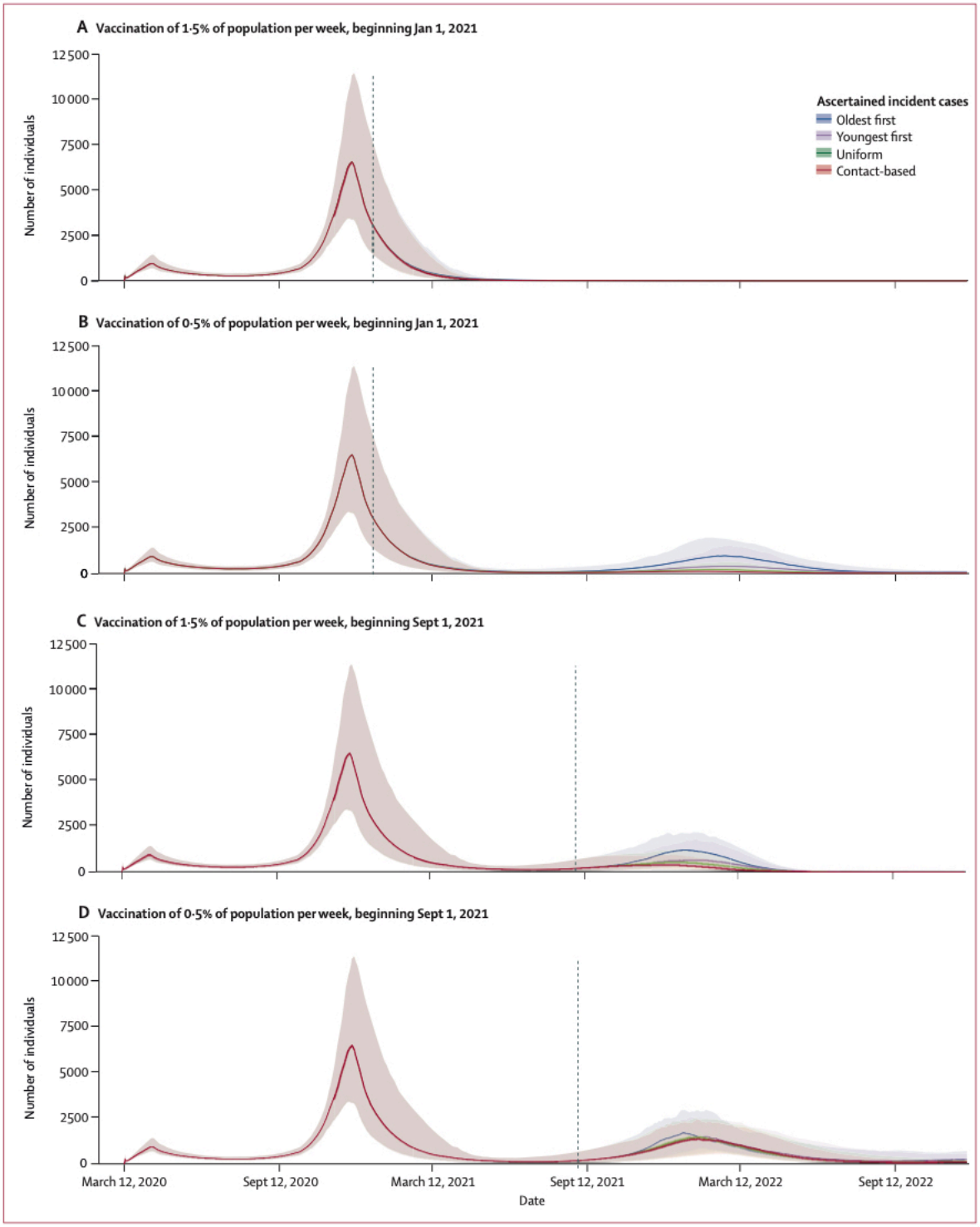

FIGURE 6.3: Figure and caption from ([16], figure 3). **Incident cases by vaccination strategy across three model regimes.** Projections of ascertained incident COVID-19 cases if vaccination begins in January (A, B) or September (C, D), and if the rate of vaccination is 1.5% (A, C) or 0.5% (B, D) of the population per week. These scenarios represent three main model regimes: timely vaccination (A), partial vaccination and indirect protection (B, C), and slow and late vaccination (D). Projections are based on the Ontario population size of 14.6 million and shutdown occurring at 200% of peak cases in the first wave. (Other parameter values are provided in the appendix of [16] (pp 1–11).

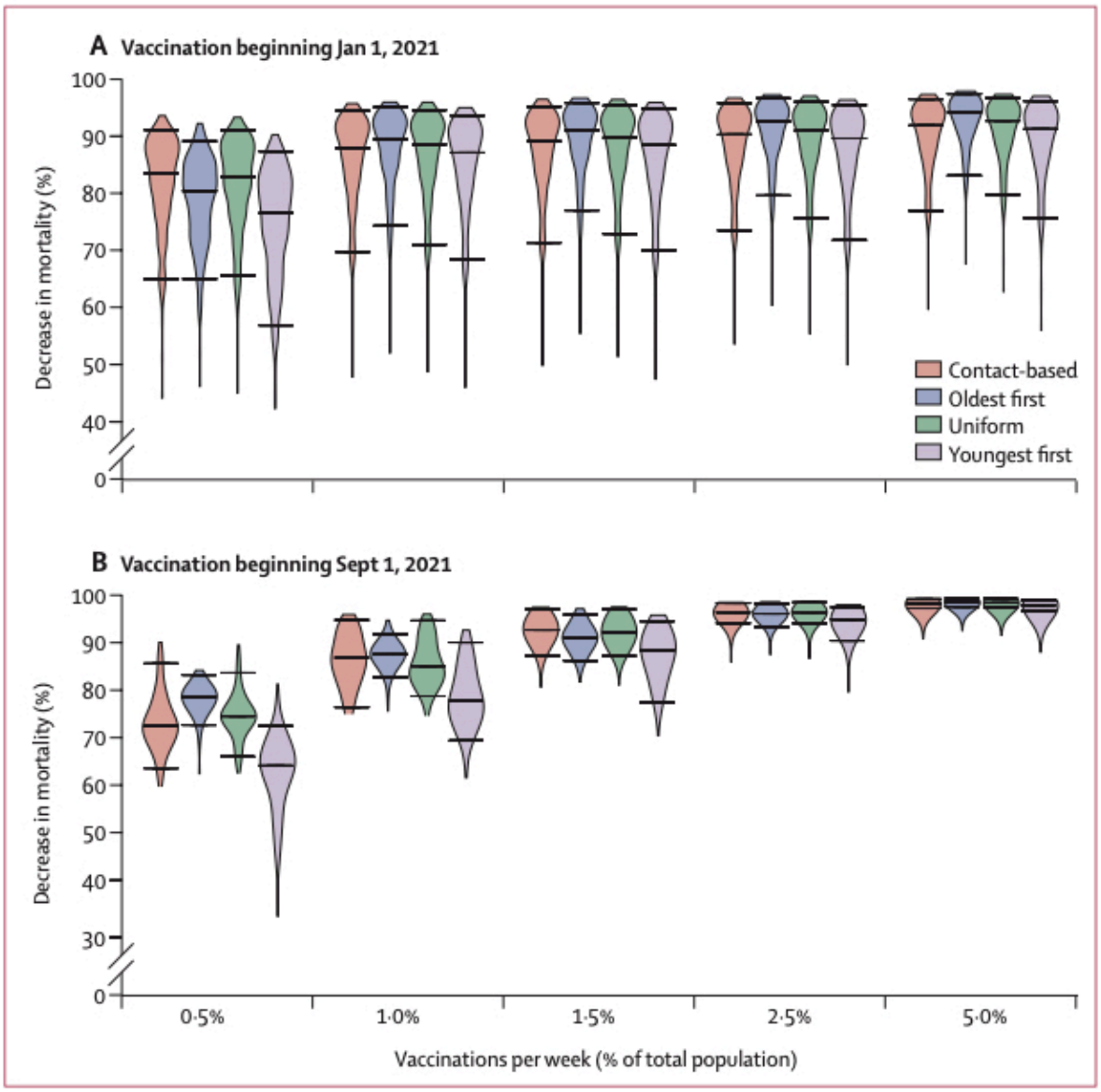

FIGURE 6.4: Figure and caption from ([16], figure 4). **Effects of vaccination strategy and start date on percentage reduction in mortality.**Violin plots of the percentage reduction in mortality under the four vaccine strategies, relative to no vaccination, as a function of the vaccination rate, for vaccination beginning on Jan 1, 2021 (A), and Sept 1, 2021 (B). Horizontal lines represent the median values and 95% credible intervals of posterior model projections. Projections are based on the Ontario population size of 14.6 million and shutdown occurring at 200% of peak cases in the first wave. Other parameter values are provided in the appendix (pp 1–11). The projected number of deaths in the absence of vaccination was 72 000 (95% credible interval 40 000–122 000) from Jan 1, 2021, to March 14, 2025, and 60 000 (31 000–108 000) from Sept 1, 2021, to March 14, 2025.

7 Conclusion & Discussion

To conclude whether or not prioritising certain age groups for vaccination will result in fewer deaths from COVID-19, we analysed a coupled social-epidemiological model presented in [16]. We tried to shed a light on the mathematics used, and explained them in an accessible way, using additional available literature. In this chapter we would like to restate some important results and discuss some elements of this research.

As seen in the Results chapter, 6, the order in which vaccination occurs *does* influence the amount of COVID-19 casualties. For clarity: with 'early' we mean starting on January 1st, 2021, and with late we mean starting on September 1st, 2021. In the early (first) regime with a relative fast vaccination rate, as well as the late (third) regime with a relative slow vaccination rate, indirect protection from the vaccine is not high enough to protect the population against a third wave of infections, making the oldest first strategy the most effective to protect against COVID-19 deaths. For the intermediate regime, with either vaccination starting early but administered at a slow rate, or starting late with a fast rate, a different scenario arises. Although not enough people are protected to entirely prevent a third wave, its severeness is significantly reduced due to the effects of vaccination, with the uniform and contact-based strategies being the most effective.

To improve this research, the introduction of more discrete classes could be introduced to include the influence of geography, gender, or socioeconomic status [16]. Furthermore, the inclusion of social dynamics in an epidemiological model, which is a relatively new way to look at the evolution of a disease, could be improved by making better use of mobility data. Since this data is provided through new technological ways, e.g. through location-based networks, it gives rise to the question of whether it is not crossing the privacy boundary. It will be interesting to look if more accurate mobility data can have a large enough effect on stopping virus transmission to out rule any emerging privacy arguments.

If we look at the number of COVID-19 deaths, the results found in this study indicate that the rate at which vaccination occurs and the presence of pre-existing immunity strongly influence the outcome. This might seem like a rather insignificant conclusion since vaccination is nearly completed in a lot of countries. These only refer to western countries however, because while 32% of the entire world population has received at least one vaccination dose, and 24% is fully vaccinated, only 1.3% of the population in low-income countries has received at least one dose [9] on the 28th of August, 2021. For these countries, the contact-based or uniform vaccination would probably be more effective to reduce the amount of casualties, depending on the vaccination rate. Therefore, the conclusions made are still highly relevant in the combat

against COVID-19 and for future epidemiological outbreaks.

Bibliography

[1] Chris T. Bauch and Samit Bhattacharyya. "Evolutionary Game Theory and Social Learning Can Determine How Vaccine Scares Unfold". In: *PLOS Computational Biology* 8.4 (Apr. 2012), pp. 1–12. DOI: 10.1371/journal.pcbi.1002452. URL: https://doi.org/10.1371/journal.pcbi.1002452.

[2] Abraham Berman and Robert J Plemmons. *Nonnegative matrices in the mathematical sciences*. SIAM, 1994.

[3] J Bjoernerstedt et al. *The rational foundations of economic behaviour*. 1996.

[4] JosÃ© M. Carcione et al. "A Simulation of a COVID-19 Epidemic Based on a Deterministic SEIR Model". In: *Frontiers in Public Health* 8 (2020), p. 230. ISSN: 2296-2565. DOI: 10.3389/fpubh.2020.00230. URL: https://www.frontiersin.org/article/10.3389/fpubh.2020.00230.

[5] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. "On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations". In: *Journal of Mathematical Biology* 28.4 (1990), pp. 365–382.

[6] O Diekmann, J A P Heesterbeek, and M G Roberts. "The construction of next-generation matrices for compartmental epidemic models." In: *J R Soc Interface* 7.47 (2010), pp. 873–885.

[7] Inc. Google. *COVID-19 community mobility reports (2020)*. https://www.google.com/{COVID}19/mobility/. [Online; accessed 09-August-2021].

[8] Günter Haag, Ulrich Mueller, and Klaus G Troitzsch. *Economic evolution and demographic change: formal models in social sciences*. Vol. 395. Springer Science & Business Media, 2012.

[9] Lucas Rodés-Guirao Cameron Appel Charlie Giattino Esteban Ortiz-Ospina Joe Hasell Bobbie Macdonald Diana Beltekian Hannah Ritchie Edouard Mathieu and Max Roser. "Coronavirus Pandemic (COVID-19)". In: *Our World in Data* (2020). https://ourworldindata.org/coronavirus.

[10] Dr. Alan Hastings and Dr. Louis Gross. *Encyclopedia of Theoretical Ecology*. no. 4. Berkeley: University of California Press, 2012, pp. 706–712.

[11] Ontario Agency for Health Protection and Promotion (Public Health Ontario). *COVID-19 case fatality, case identification, and attack rates in Ontario.* Toronto, ON: Queen's Printer for Ontario; [Online; accessed 12-August-2021].

[12] Herbert Hethcote. "The Mathematics of Infectious Diseases". In: 42 (Jan. 2000), pp. 599–653.

[13] Herbert W. Hethcote. "Qualitative analyses of communicable disease models". In: *Mathematical Biosciences* 28.3 (1976), pp. 335–356.

[14] Joe Hilton and Matt J. Keeling. "Estimation of country-level basic reproductive ratios for novel Coronavirus (SARS-CoV-2/COVID-19) using synthetic contact matrices". In: *PLOS Computational Biology* 16.7 (July 2020), pp. 1–10.

[15] Josef Hofbauer, Karl Sigmund, et al. *Evolutionary games and population dynamics*. Cambridge university press, 1998.

[16] Peter C Jentsch, Madhur Anand, and Chris T Bauch. "Prioritising COVID-19 vaccination in changing social and epidemiological landscapes: a mathematical modelling study". In: *The Lancet Infectious Diseases* (2021). ISSN: 1473-3099. DOI: https://doi.org/10.1016/S1473-3099(21)00057-8. URL: https://www.sciencedirect.com/science/article/pii/S1473309921000578.

[17] Kenji Mizumoto et al. "Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020". In: *Eurosurveillance* 25.10 (2020).

[18] Treasury Board Secretariat of Ontario. *Confirmed positive cases of COVID-19 in Ontario*. https://data.ontario.ca/dataset/confirmed-positive-cases-of-{COVID}-19-in-ontario/resource/455fd63b-603d-4608-8216-7d8647f43350. [Online; accessed 09-August-2021].

[19] Corey M Peak et al. "Individual quarantine versus active monitoring of contacts for the mitigation of COVID-19: a modelling study". In: *The Lancet Infectious Diseases* 20.9 (2020), pp. 1025–1033.

[20] Christina Savvides and Robert Siegel. "Asymptomatic and presymptomatic transmission of SARS-CoV-2: A systematic review". In: *medRxiv : the preprint server for health sciences* (June 2020), p. 2020.06.11.20129072.

[21] P. van den Driessche and James Watmough. "Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission". In: *Mathematical Biosciences* 180.1 (2002), pp. 29–48.

[22] WHO. *WHO Target Product Profiles for COVID-19 Vaccines*. https://www.who.int/publications/m/item/who-target-product-profiles-for-covid-19-vaccines. [Online; accessed 09-August-2021].

[23] Emilio Zagheni et al. "Using Time-Use Data to Parameterize Models for the Spread of Close-Contact Infectious Diseases". In: *American Journal of Epidemiology* 168.9 (Sept. 2008), pp. 1082–1090.