Compartmental SIR-models for infectious diseases

Juuso Hapulahti

School of Science

Bachelor's thesis
Espoo 03.12.2021

Supervisor

Prof. Ahti Salo

Advisor

Dr. Tuomas Lahtinen

Aalto University
School of Science

Aalto University
School of Science

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi
Tekniikan kandidaatintyön tiivistelmä

Tekijä Juuso Hapulahti

Työn nimi Tartuntatautien SIR-mallinnus

Koulutusohjelma Teknillinen fysiikka ja matematiikka

Pääaine Matematiikka ja systeemitieteet **Pääaineen koodi** SCI3029

Vastuuopettaja Prof. Ahti Salo

Työn ohjaaja Dr. Tuomas Lahtinen

Päivämäärä 03.12.2021 **Sivumäärä** 21 **Kieli** Englanti

Tiivistelmä

COVID-19 pandemia on aiheuttanut haasteita tutkijoille ympäri maailmaa. Maailmalla on iso kysyntä matemaattisille malleille, jotka auttavat ymmärtämään, miten tartuntataudit leviävät väestössä ja auttavat yhteiskuntia valmistautumaan epidemioihin. Yksi tälläisistä malleista on 1900-luvun alkupuolelta peräisin oleva susceptible-infected-recovered (SIR) -malli.

Tässä työssä tarkastellaan, miten tutkijat ovat hyödyntäneet SIR-mallia. Työssä esitetään miten SIR-mallia ja sen muunnelmia on käytetty COVID-19-epidemian ja muiden tartuntatautien mallintamiseen. Sitten työssä tarkastellaan, miten SIR-mallin muunnelmia on käytetty ennustamiseen ja miten SIR-mallilla voidaan mallintaa COVID-19-rokotuskampanjoita ympäri maailmaa.

Käsiteltyjen tutkimusten valossa SIR-mallinnus soveltuu ennustamistarkoitukseen lyhyellä aikavälillä, mutta pitkän aikavälin ennustaminen on hankalaa. Käsitellyt rokotemallinnukset osoittavat, että rokotteiden priorisoinnilla on iso merkitys epidemian etenemiseen.

Avainsanat SIR-model, COVID-19, Vaccine

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the bachelor's thesis

Author Juuso Hapulahti

Title Compartmental SIR-models for infectious diseases

Degree programme Engineering Physics and Mathematics

Major Mathematics and Systems Sciences **Code of major** SCI3029

Teacher in charge Prof. Ahti Salo

Advisor Dr. Tuomas Lahtinen

Date 03.12.2021 **Number of pages** 21 **Language** English

Abstract

The COVID-19 pandemic has presented a challenge for researchers worldwide. There is a massive demand for models that help to understand how these kinds of diseases spread in a population and help governments to be well prepared. One of these models is the susceptible-infected-recovered (SIR) model, originating from the early 20th century.

In this thesis, we consider how researchers have used the SIR-model for different purposes. Initially, we look at how SIR-model and its many variations have been used to model COVID-19 and other infectious diseases. Afterwards, we consider how these SIR-model variations have been used for different kinds of predictions and then we investigate how SIR-model has been used to model the ongoing COVID-19 vaccination efforts around the globe.

We find that SIR-modeling works excellently when predicting over a short period of time but is less suited for predicting the long-term behaviour of a epidemic. We also find that vaccine prioritization to certain population groups with higher risk of exposure to the virus or a higher risk of death can be of great importance.

Keywords SIR-model, COVID-19, Vaccine

Contents

1 Introduction

The SARS-CoV-2 virus has caused a public health crisis worldwide. As of 30 July 2021, it has caused 197 million positive cases and 4,2 million deaths [1]. To efficiently control the ongoing pandemic, it is important that public health services are able to make informed decisions. This is where modeling becomes important, because it can help the policy makers understand the possible consequences of their actions. Existing models, among other things, have been used to study the spread of the infection and how to properly implement efficient intervention strategies such as face mask mandates, stay-at-home orders and vaccination. Many intervention strategies are designed to 'flatten the curve' which means lowering the maximum number of infected at a specific point of time so that hospitals and intensive care units would not get overloaded with patients which is likely to increase the number of deaths caused by the disease.

The most widely used modeling approach for infectious diseases is the 'basic' SIR-model, popularized by Kermack and McKendrick [2]. The basic SIR-model consists of three compartments denoted by the letters S (susceptible), I (infected) and R (recovered). A susceptible individual can become infected by being in contact with an infectious person and then will transition from susceptible population to the infected population. The rate at which this happens is called the transmission rate. An infected individual can transition to the recovered compartment either by recovering from the disease or by dying. The rate at which this happens is called the recovery rate. A very important constant in a SIR-model is the so-called reproduction number (denoted R_0), which is the expected number of infected people that a single infected can cause in a population where everybody is susceptible. Most models begin with most of the population being susceptible to an infectious disease without natural immunities or vaccination. The SIR-model can be expressed by the following ordinary differential equations.

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I,$$

where β is the transmission rate constant, γ is the recovery rate constant, S is the susceptible population, I is the infected population and R is the recovered/dead population. The total number of the population N can be expressed by the equation $N = S + I + R$.

One reason for why a SIR-model is common is that it can be easily tailored to specific needs. A common modification is the addition or removal of compartments from the model. A very popular modification is the addition of a compartent E (exposed) where an individual is carrying the disease but is not yet infectious. This variation of the model is called the SEIR-model. Other variations are formed similarly, popular added compartments are V (vaccinated) and D (dead).

However, understanding the limitations of SIR-modeling is as important as understanding the model itself. Natural births and deaths are often omitted from the models because they complicate the model without significant gain. To name a few examples, acquiring realistic parameters for transmission and recovery rates is often difficult and existing data is very often used to estimate the parameters but this has obvious downsides like becoming very inaccurate after some time has passed [3]. Another issue is the large scope of SIR-modeling, as it portrays a part of the population in a compartment while ignoring important variations among individuals. Some of these issues can be solved by adding compartments to the model but this can cause other issues and the models are usually kept simple with only a few additions if the phenomenon that is being studied can be studied with a simple model. Despite advanced modeling, modeling and forecasting the spread of infectious diseases has remained a challenge.

In this paper, we investigate how researchers have used SIR-modeling as a tool to study different phenomena with focus on vaccination and the future of the COVID-19 epidemic. The research papers we study are in Table 1 with a description of the model they have used and the main problem the research is trying to tackle. The papers were chosen to give a good overview of SIR-modeling from multiple points of view. Discussion of mathematical methods used for modeling optimization is not covered and the primary focus is on the purpose and the results of the research and the implementation of the SIR-models.

Research	Model	Main problem
Giordano et al. (2020)	SIDARTHE	Modeling the COVID-19 epidemic and implementation of population-wide interventions
Moein et al. (2021)	SIR	Problems with forecasting the COVID-19 epidemic with SIR-modeling
Ghostine et al. (2021)	SEIQRDV	Forecasting the COVID-19 epidemic with an extended SEIR-model
Berkane et al. (2021)	SIRV	Predicting the evolution of COVID-19 epidemic in Canada
Bubar et al. (2020)	SEIRV	Quantifying the impact of COVID-19 vaccine prioritization strategies
Yu et al. (2016)	SEIRV	Efficient intervention strategies to control an epidemic
Agarwal et al. (2021)	SEIR	Trade-off between vaccination prioritization and vaccination speed
Karaivanov et al. (2020)	SIRD	Estimating the impact of public policy measures on the spread of COVID-19

Table 1: Main models used in the research papers listed and the main problems the research is trying to tackle [3] - [10]

Before discussing about vaccination and the forecasting of the ongoing COVID-19 epidemic, it is pertinent to investigate how recent research has used SIR-modeling to model infectious diseases and the COVID-19 epidemic in particular. One of the main challenges of SIR-modeling is the parametrization of the model. Parameters are often obtained from earlier similar studies or using reliable government sources. A drawback of this type of parametrization is that parameters change over time and data may become obsolete. We next discuss examples how the studies in Table 1 have acquired their parameters.

Moein et al. [3] demonstrate the inefficiency of SIR-modeling for COVID-19 and assume a mortality rate of 2,5% for Wuhan and Italy that is provided by China CDC. When validating the data by running the built SIR-model, it provided realistic results for Italy, but using a similar model and assuming a similar virus in both countries resulted in a simulation curve that would have resulted in more deaths than reported in Wuhan. Chinese authorities later declared inaccuracies in reported deaths in Wuhan [11].

Bubar et al. [7] evaluate the impact of vaccine prioritization strategies by age and use parameters provided by Davies et al. [12] study named "Age-dependent effects in the transmission and control of COVID-19 epidemics" that uses data from China CDC and Singapore Ministry of Health and a few other sources. The raw data used is validated with a dynamic model using a few assumptions like only 10% of clinical cases were reported Yu et al. [8] model different intervention strategies for the H1N1-epidemic and use data provided by the Hong Kong government and infection rates from a previous similar study [13].

To model inaccuracies in given data, modelers sometimes conduct sensitivity analysis to validate their results. Usually this means changing a single input variable (such as vector efficiency or vaccination release time) and measuring the resulting changes in the output variables. As an example, Bubar et al.[7] use sensitivity analysis to test their initial results while varying the previously assumed constant R_0 (the reproductive number).

As Table 1 indicates, there is a large range of different versions of the SIR-model. Some researchers prefer using a basic SIR-model and then add one compartment to the model to study the phenomena they deem important. Berkane et al. [6] use a SIRV-model where they have added a compartment V for vaccination to model the effect of vaccination on infectious diseases. Another common variation of the SIR-model is the SEIR-model in which E represents the exposed compartment where individuals are infected but not yet infectious. This type of approach is used by Agarwal et al. [9] in its simplest version but a similar vaccination modification to the SEIR-model is done by Bubar et al. [7] and Yu et al. [8] as Berkane et al. [6] used in a SIR-model.

On the contrary to previous simple approaches, two of the papers use a seemingly complicated version of the SIR-model. Ghostine et al. [5] use a so-called SEIQRDV-model in which the dead compartment is separated from the recovered compartment and Q represents a quarantined compartment where all confirmed infected will end up. A major problem with accurate modeling of infectious diseases, especially if parameters are derived from observed data, is counting the actual number of infected. Some people will get infected and will stay outside the range of public health services to not spread the disease further but this causes inaccuracies in the actual number of infected people. Ghostine et al. [5] address this problem with the quarantined compartment. Giordano et al. [4] also use a seemingly complicated model aswell which they call the SIDARTHE-model. On top of the normal S and I compartments, we have D for diagnosed, A for ailing, R for recognized, T for threatened, H for healed and E for extinct. In this model, the researchers separate some compartments based on whether the individuals in them are symptomatic, acutely symptomatic or

asymptomatic. For example, in their research, the compartment I represents infected that are asymptomatic, infected and still undetected and A for ailing represents symptomatic infected that are undetected. Similarly D for diagnosed represents individuals that are asymptomatic, infected and undetected and R for recognized represents symptomatic, infected and detected. T for threatened represents the compartment where infected individuals have life-threatening symptoms. Separating some of the compartments based on whether they are symptomatic or not makes sense when modeling COVID-19. Existing research suggests that people can become infectious a few days before symptoms start appearing [14]. All in all, the model by Giordano et al. [4] provides a good foundation to further build upon for a model that is complex but can reasonably accurately depict different phenomena. The downside of simpler models often leave out bits and pieces that can be very important.

2 Summary of earlier research

This section describes how the papers in Table 1 have used their version of the SIR-model in tackling the given problem. The chosen research covers three main topics. First, SIR-modeling as a tool to model COVID-19 which is one thing all of them have in common. Second, forecasting how the COVID-19 epidemic will continue or how accurately models have predicted the future of the epidemic. Third, how SIR-modeling is used to model the ongoing vaccination programs around the globe.

Giordano et al. [4] use a SIDARTHE-model which in their words "discriminates between detected and undetected cases, between different severity of illness, non-life-threatening and life-threatening cases". They attempt to measure how the timing and stringency of population-wide interventions affected the spread of COVID-19 in Italy in 2020 using data from February 2020 to April 2020. Model parameters were updated over time to reflect the changes in population-wide restrictions. Their study period was 350 days and their model predicted that 0.61% of the population will contact the virus and 0.06% of the population of Italy will die from COVID-19 which would mean around 36000 deaths. As of February 2021, there had been 90000 deaths caused by COVID-19 in Italy [15]. According to the authors, the adopted lockdown measures in Italy were vital to contain the epidemic and could have been even more restrictive.

Moein et al. [3] focus on how accurately SIR-modeling can predict the number of infections and the patterns of the disease in the long term. They start with a rough estimation on how a SIR-epidemic behaves using reported data from Wuhan and Italy. As stated before, in this stage the model was very inaccurate in predicting the number of deaths in Wuhan. After this, their simple SIR-model was run with three different scenarios with different parameters (namely R_0) to reflect that scenario. These scenarios were named Good, Feasible and Bad. Good scenario reflected strict interventions to control the spread of the disease, feasible scenario represented closing down schools, limiting social gatherings and non-critical businesses and bad scenario assumed a normal pre-outbreak way of life. When the model was run using these

three scenarios, the generated values were far from actual data in later time intervals but the model was able to forecast the disease patterns in the short term. The model predicted that the disease would be over in six months when using data from February 2020 to April 2020 from Isfahan, a city in Iran. According to the authors, similar inconsistencies can be found in other studies based on SIR-models. Does this research give us a reason to doubt the accuracy of simple SIR-modeling or is there something SIR-modeling ignores about the behaviour of COVID-19 that is vital to its spreading mechanism? Developing new and more complicated models like Giordano et al. [4] and Ghostine et al. [5] have done might be an easy answer to this problem.

Ghostine et al. [5] use an extended SEIR-model called SEIQRDV to counteract the inaccuracy of predictive SIR-modeling. To model the accuracy of their predictions, they ignore the impact of vaccinations and assumed no vaccinations yet. They use the model to make short-term two-week predictions using data from Saudi Arabia. The model performs very well and the calculated Relative Mean Absolute Error (RMAE) was 3-5% for most of the predictions which demonstrates excellent accuracy when it comes to short-term predictive modeling. The authors also modeled the impact of vaccination on the total number of confirmed cases and deaths. Using data from January 2021 when there were 300 000 vaccinated in Saudi Arabia and assuming a 95% vaccine efficacy resulted in less than 10% reduction in the number of total cases and deaths. The model also showed that intensifying the vaccination efforts resulted in a significant reduction in the number of total cases and deaths. Doubling the vaccination rate (α) resulted in a 16% and 13% reduction in the total number of cases and deaths, respectively. Multiplying the vaccination rate by 8 resulted in a 39% and 34% reduction in the total number of cases and deaths, respectively. Thus the correlation is not completely linear as one could expect, but increasing the number of given vaccines can have a massive impact on the outcome of a pandemic.

Ghostine et al. [5] assume a 95% efficacy for the vaccine and used a so-called "all-or-nothing" vaccine meaning you either get a full immunity to the disease with a 95% probability or no immunity with a 5% probability. The common "two-dose" vaccination was omitted from their model. Modeling was carried out by assuming a single dose would either give you an immunity or not. The model was based upon the vaccine preventing the spread of infections but you could also model a vaccine that prevents the symptoms of the disease. How do you track/model whether a vaccinated person who is asymptomatic but is still carrying the disease spreads it further? This poses a very difficult challenge for public health services and modelers alike.

Berkane et al. [6] used a SIRV-model to model the evolution of COVID-19 pandemic in Canada with four different vaccination scenarios and two vaccine efficacy scenarios. These four different vaccination scenarios were no vaccination ($\alpha = 0$), low vaccination rate ($\alpha = 0.0005$), moderate vaccination rate ($\alpha = 0.001$) and high vaccination rate ($\alpha = 0.002$) and the two vaccine efficacies were 60% (pessimistic) and 95% (optimistic). Currently major vaccine manufacturers Pfizer and Moderna advertise a 95% vaccine efficacy [16]. In comparison, Ghostine et al. [5] used vaccination rates of $\alpha = 0.0003$, $\alpha = 0.0006$ and $\alpha = 0.0024$ for their three different scenarios which are considerably larger. Berkane et al. [6] also did an assessment of

the situation for different provinces of Canada and a nation-wide assessment.

Table 2 show the approximate number of predicted lives lost in their model by the end of 2021 in their nation-wide assessment. Even with a pessimistic 60% efficacy in a high vaccination rate scenario, their model predicts a 64% reduction in the number of lives lost. With a more effective vaccine with 95% efficacy, that number rises to 76%. Keeping in mind the previous depictions about the inaccuracies of these long-term predictions, these numbers might seem huge. These kinds of models can serve as a helpful tool for public health officials to plan appropriate responses for outbreaks.

	$\alpha = 0$	$\alpha = 0.0005$	$\alpha = 0.001$	$\alpha = 0.002$
60% efficacy	219000	167000	127000	78000
95% efficacy	219000	142000	94000	53000

Table 2: Approximate numbers of predicted lives lost by the end of 2021 in 8 different vaccination rate-efficacy combinations [6]

We previously discussed different ways to model vaccines like "all-or-nothing" vaccines and having different vaccine efficacies. Bubar et al.[7] use an age-stratified SEIRV-model to study the impact of vaccine prioritization strategies but they also study multiple different types of vaccines, mainly infection- and transmission-blocking vaccines with different vaccine efficacies and use this kind of approach to study the optimal vaccine prioritization strategy when people are separated into different agegroups and later combining agegroups with serostatus. An individual being seropositive means that he is already carrying the necessary antibodies in his immune system to properly thwart an infection, similarly seronegative individual is not carrying them. The researchers divided the population into 10-year intervals and used a country dependent contact matrix to model the spread of the infection across agegroups and then picked 5 strategies to capture a realistic picture of the possible outcomes. These strategies were 1) prioritize under 20-year-olds, 2) adults between ages 20 and 49, 3) adults over 20, 4) adults over 60 and 5) no prioritization. Among these five strategies with different combinations of vaccine supply between 1% and 50% of the total population and different vaccine efficacies, strategy number 4 (adults over 60) reduced the total number of deaths substantially more than alternative strategies. Results indicate that the ranking of the vaccine prioritization strategies remains the same whether the vaccine is infection- or transmission-blocking and similarly for "all-or-nothing" vaccines and leaky vaccines. Using these five strategies, the authors also implemented vaccination paired with serological tests to their model. Seropositive individuals would not be vaccinated and therefore allowed for a higher total vaccine supply.

Yu et al. [8] discuss different ways to manage the epidemic such as vaccination and quarantines. Their major contribution is the SEIRV-model for prioritizing vaccination strategies in a similar manner as Bubar et al. [7]. Their research also studies the effects of vaccine coverage and vaccine release times and later combining all three of these to model the number of accumulated infected with a certain strategy, certain vaccine coverage and certain vaccine release time. The authors used data from the

2009 H1N1 swine flu epidemic and used contact data and population sizes for different districts of Hong Kong.

For their model, the authors picked four different vaccination strategies, S1 (based on population size), S2 (based on contact pattern matrix), S3 (based on the infection rate), S4 (based on the infection risk). Their simulations had the population divided to 5 different age groups, A1 (5-14 years old), A2 (15-24 years old), A3 (25-44 years old), A4 (45-64 years old), A5 (65+ years old).

Yu et al. [8] carried out three simulations. The first simulation which assesses how the infectious disease spreads without vaccination. For their first simulation, they started with 1 infected individual in age group A3 and observed that age groups A1 and A3 accumulated the most number of infected during the course of the simulation. The second simulation considered the effects of vaccine coverage. Vaccine coverage varied from 2% of the total population to 10% of the total population with an increment of 2%. The results of the simulation indicate that more vaccines means less infected without a significant breakpoint in their vaccine coverage range. The third simulation studied the effect of vaccine release times. They picked 5 vaccine release times to consider: Day 1, Day 50, Day 100, Day 150 and Day 200. According to USA CDC, H1N1 vaccine was released 171 days after the new H1N1 virus was detected during the 2009 pandemic [17]. Results of this vaccine release time simulation indicate that it does not matter when the vaccine is released if its at a later time in the epidemic but generally, earlier vaccine release time means smaller numbers of infected individuals. After these findings, they combined the second and third simulations with the aforementioned vaccination strategies S1, S2, S3 and S4. The authors list two interesting observations from this simulation. More vaccine coverage and earlier vaccine release time might prevent the outbreak very early and more realistic vaccine prioritization strategies like S4 require fewer doses of vaccine to achieve herd immunity. The authors also simulated the vaccination strategies on their own and the results indicate that S4 is the best vaccine prioritization strategy and S1 is clearly the worst one when it comes to the number of total of infected individuals. Therefore according to their results, infection risk is the best variable to consider when allocating limited numbers of vaccine doses.

Agarwal et al. [9] consider two things, vaccine prioritization and vaccination speed and how many vaccine doses one would need to achieve the same impact with increasing your vaccination speed than allocating your vaccines efficiently like in the previous paper by Yu et al. [8]. We first look at the vaccine prioritization strategies used in their model. They considered three vaccine prioritization strategies, No Prioritization, Optimal Prioritization and NASEM's guidelines. Optimal prioritization is a strategy that minimizes the total number of deaths and NASEM is short for National Academies of Sciences, Engineering, and Medicine who developed a framework for vaccine allocation to assist policymakers in allocating vaccine doses. NASEM's guidelines has 4 phases, phase 1 allocates vaccines to high-risk health workers and people with conditions that put them at higher risk, phase 2 allocates vaccines to essential workers, teachers and staff, phase 3 consists of young adults and children and phase 4 is for everybody who did not receive a vaccine in the previous phases.

Agarwal et al. [9] simulate NASEM's vaccination strategy using US population data and their simulation were done by using a constant 30 million vaccine doses per month. Most of their results compare the differences in the total number of deaths and infections when using the two mitigation strategies, Optimal prioritization and NASEM's guidelines. According to their results, its optimal to vaccinate the elderly in both of these mitigation scenarios. In the optimal vaccine prioritization strategy, individuals over 60 are vaccinated first, after that comes individuals with the highest Infection Fatality Rate (IFR) and then individuals aged between 30-59. Total number of infected is the highest when using the optimal prioritization strategy but by definition, it saves the most lives. In all three strategies, the most number of deaths is among the elderly because of their high IFR. Mortality among the elderly is 10-20% higher when following NASEM's guidelines or using no optimization strategy when compared to Optimal prioritization.

Agarwal et al. [9] consider vaccination speeds ranging from 15 million vaccinations per month to 40 million vaccinations per month and compared the total number of deaths to a scenario where an epidemic is not controlled with vaccinations at all. As expected, increasing the vaccination speed decreases the total number of deaths in all considered scenarios. The total number of prevented deaths is the highest when using the Optimal prioritization strategy followed by NASEM strategy which is followed by No prioritization strategy. After this observation, the authors considered the tradeoff between vaccine prioritization and vaccination speed. This was done using the same range of vaccination speed as previously from 15 million vaccinations per month to 40 million vaccinations per month and then measured how big of a increase in vaccination speed was needed to acquire the same effect as changing prioritization policy from No Prioritization strategy to NASEM strategy. The resulting graph is almost a linear curve from an increase of 8 million vaccinations per month to 20 million vaccinations per month for the whole range which means that nearly a 50% increase in vaccination speed is required to acquire the same effect as using a vaccine prioritization strategy when it comes to the number of deaths. These results assume a vaccine with 80% efficacy. The results indicate that vaccine prioritization seems very important and attempting to replace vaccine prioritization with increased vaccination speed does not seem too realistic. Understanding this relationship between vaccine prioritization and vaccination speed can be very important when designing new and effective vaccination strategies and public policy measures to mitigate the impact of infectious diseases.

We have considered many aspects of SIR-modeling from transmission dynamics to predicting the future of an outbreak to mitigating an outbreak with vaccination. Vaccination is only one of many ways to mitigate an outbreak. An straightforward way to mitigate an outbreak is by enforcing strict temporary laws that are designed to hinder the spread of infectious diseases. Karaivanov et al. [10] estimate the impact of face mask mandates and other non-pharmaceutical interventions (NPI) in Canada. Previously we covered the predicted number of deaths in Canada by Berkane et al. [6] but Karaivanov et al. [10] use the total number of cases as their results instead which makes comparing these two research papers difficult.

Data used by Karaivanov et al. [10] is from March 2020 to August 2020 and during

this time period, Canada enforced mask mandates in all ten Canadian provinces. The implementation dates of these mask mandates and other policy measures were acquired from official trusted government sources. These mask mandates were introduced in Canada when other policy measures were relaxed which in turn makes it hard to quantify the impact of mask mandates specifically. The authors use a causal model where adopted policies affect health outcome of individuals directly or indirectly by altering their behaviour. These policies are assigned a dummy variable from 0 to 1, where 0 indicated no restrictions and 1 indicates heavy restrictions and numbers between 0 and 1 indicate partial restrictions. For a mask mandate, there are only 2 possible 'states', 1 means mask mandate is in effect and 0 means no mask mandate. This kind of dummy variable was also assigned to other policy measures like school closures, non-essential travel regulations and business regulations to study the effect of other non-pharmaceutical interventions. Behavioral data used is provided by Google, where mobility data measures the changes in people's location data from January 2020 to February 2020 to the study period.

Karaivanov et al. [10] indicate a 25 to 46% reduction in the number of weekly cases during the first weeks of enforced mask mandates and while their sample data does not support further prediction, it can be concluded that mask mandates can be a powerful tool to temporarily hinder the spread of COVID-19 in Canada. The study follows the approach of Chernozhukov et al. [18] who perform a similar evaluation of the impact of public policy measures for USA and their results indicate a 19% to 47% reduction in the number of deaths nationally in May 2020 if mask mandates were enforced early on in the epidemic. Both of these results have a huge margin for error which might be caused by the used causal model. Karaivanov et al. [10] also implemented a 14-day or 28-day lag in their data to replicate the laggy changes in human behaviour, this caused some of their calculations to only have 9 days of data to run their model, this limitation of data makes these results quite unreliable.

3 Observations on SIR-models

We had 3 main topics, (I) how SIR-modeling is used as a tool to model infectious diseases and specifically the COVID-19 epidemic, (II) what are the possibilities and limitations of SIR-modeling for prediction purposes and (III) how SIR-modeling has been used to model the ongoing vaccination around the globe. Now we could ask ourselves a question: Are the SIR-model and its extended versions good models to study these three topics or should we implement something better? Compartmental models provides a solid framework to build upon and more complicated versions of the model can better depict the actual biology of the studied infectious disease. An alternative that could be used is an individual-based model (IBM, also called agent-based-model ABM). Comparing compartmental models like SIR-model to IBMs, SIR-model is clearly a macroscale implementation and IBM is a microscale implementation of the same phenomena. IBMs provide an another way to verify your findings that you've found using a simpler compartmental model. To compare the main differences in IBMs and compartmental models, IBMs are more dependent on

data than compartmental models because you need a good representation of social patterns to make your model accurate and IBMs require more computing power because of its implementation.

As Moein et al. [3] highlighted in the previous section, SIR-modeling has limitations in predicting over a longer period of time. Basic SIR-model makes some strong assumptions that may not be that realistic, for example: the entire population is well-mixed and each individual has an equal probability to get infected. Also SIR-models assume that every individual is identical and there might be something about human biology that is relevant to the transmission of the studied infectious disease. SIR-model does not have an incubation period for the disease, when an individual gets infected, he will become infectious immediately. Taking into consideration all of these aforementioned assumptions, should we take any of predictions made with a basic SIR-models seriously? The main approach to combat these challenges is to develop more elaborate models but that is not a perfect solution either. More "realistic" models usually take one or more of these heavy assumptions and relax it. All in all, we should not completely trust very simple models unless they were confirmed with more realistic and elaborate models.

With vaccination, there are two main ways to protect the population that is the most vulnerable to the disease. You either vaccinate them directly or vaccinate the individuals that are spreading the disease which is called indirect vaccination.. According to Bubar et al. [7] results, indirect protection does not seem as effective as direct vaccination of vulnerable population.

There are multiple ways to model vaccination programs using a SIR-model as the backbone of your model. Some methods are described in the papers covered in section 2. Common approaches involve adding a vaccinated compartment to a SIR-model or a SEIR-model and then varying some of their starting parameters or the deployment of the vaccine to acquire interesting results

The rapid spread of COVID-19 has highlighted our vulnerability to infectious diseases and this pandemic should be used as a learning experience to prevent further spread of COVID-19 and future pandemics. This learning can be achieved through better modeling decisions or better preparation.

An interesting question to consider is: Why do we model infectious diseases or model anything at all? We might want to gain extra information about the phenomena that would be too costly to obtain by observation in real-life or maybe explain the core dynamics of the phenomena and that way offer options for future problems or maybe something way different. Epstein [19] lists 17 different reasons for modeling, including predicting, explaining, discovering new questions and educating the general public just to name a few. Epstein claims that the most important reason to model is to enforce a scientific habit of mind. In essence, this means having an open mind to challenge the views of others and eventually base your beliefs on scientific evidence. To start modeling, you require an open mind, and modeling works as a tool to gather that scientific evidence.

4 A numerical example

To highlight and verify some of the results in the covered research papers, I implemented a SIR-model in MATLAB. The goal of the model was to study COVID-19 vaccination and how vaccinating different agegroups and vaccine release times affected the outcome of the pandemic. The model used for these simulations was a SIRV-model. Natural births and deaths and deaths attributed to overcrowded health services were omitted from the model. A SEIRS-model was also considered where immunity from a previous infection would not last indefinitely but according to existing statistics, reinfection rates appear to be negligible [20].

The simulation starting parameters included 50 infected with vaccine coverage being one third of the entire population. The simulation had four different vaccination scenarios, vaccinating on day 1, vaccinating in four occasions, vaccinating every day and vaccinating as a response to a infection peak. These strategies were called Upfront, Periodical, Constant and Response respectively. The timeframe of the simulation was 300 days and the entire simulation was implemented using MATLAB.

The simulation yields three interesting output variables, the number of infected people at the end of the simulation, the peak number of infected people and the day the peak happened. According to the simulations, responding to an outbreak with a mass vaccination program is by far the worst option when considering the peak of infections as nearly 40% of the population is infected at one point while other vaccination strategies keep the percentage of infected people below 20%. The three other strategies, Upfront, Periodical and Constant are very similar in that regard. The main difference in the strategies is the day when the number of the infected people reaches its peak. Response strategy causes the peak to happen earlier than other strategies on day 93 of the simulation, Constant strategy peak occurs on day 108 of the simulation, Pulse strategy peak occurs on day 127 of the simulation and Upfront strategy peak occurs on day 140 of the simulation. So in essence, the earlier you vaccinate, the later the peak occurs.

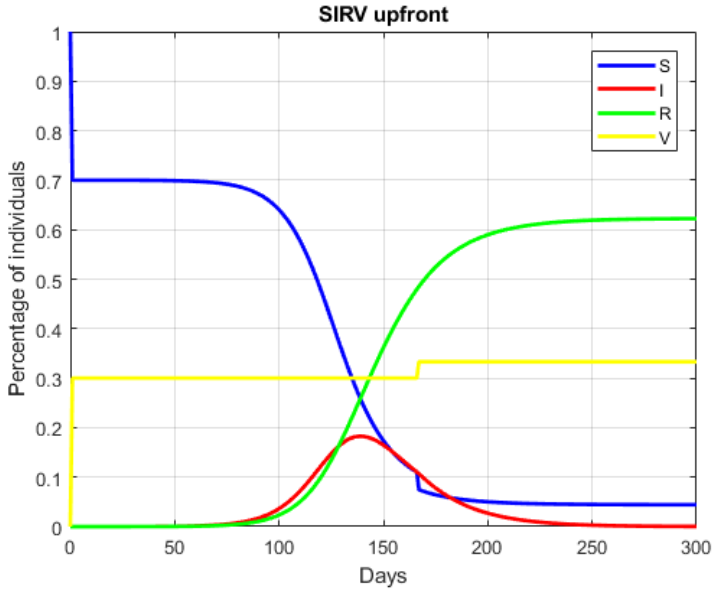

Figure 1: Result of the simulation when vaccinating on day 1.

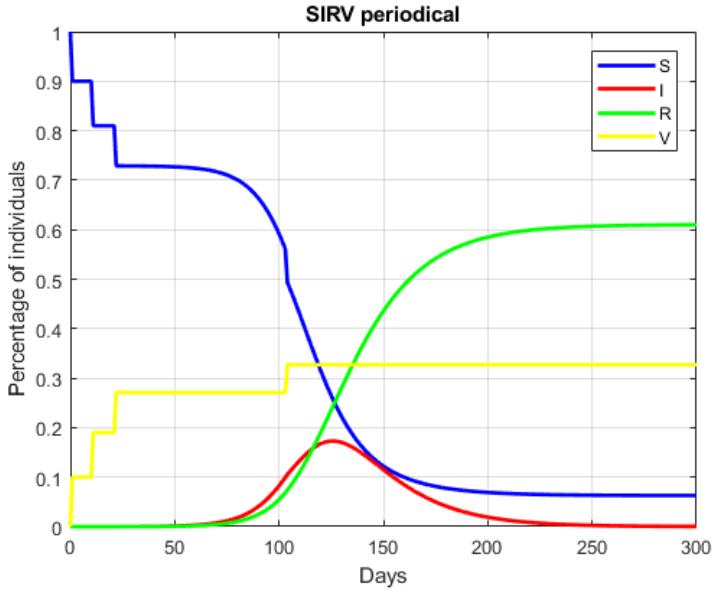

Figure 2: Result of the simulation when vaccinating on four occasions.

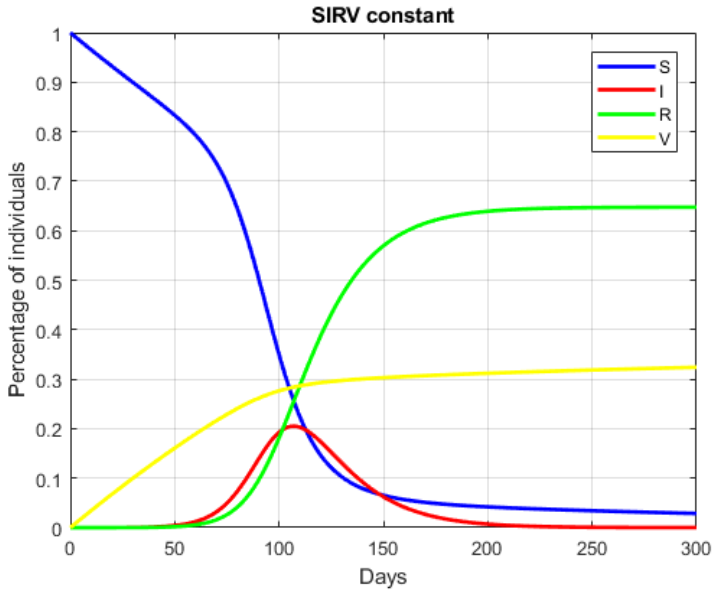

Figure 3: Result of the simulation when vaccinating every day.

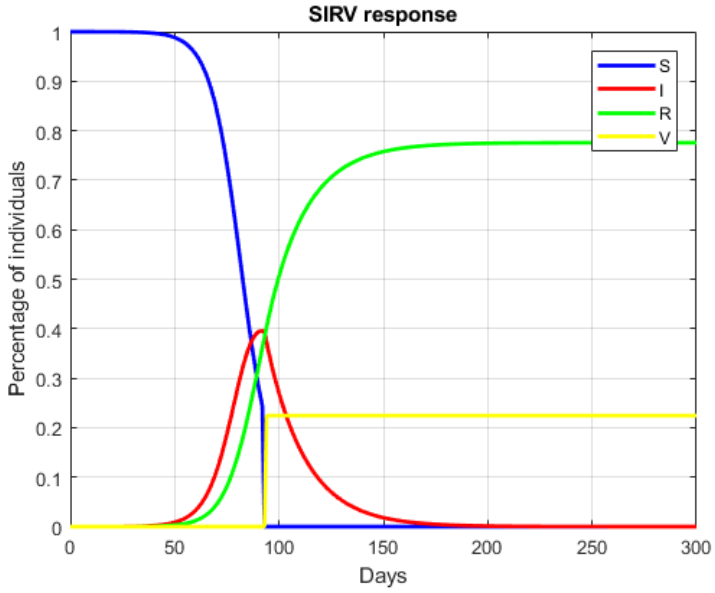

Figure 4: Result of the simulation when vaccinating as a response to a infection peak.

5 Conclusions

SIR-models are simple tools to help understand how a infectious disease spreads in a population. Because of their simplicity, they have a few limitations. Some of them can be accounted for with modified SIR-models such as a SEIR-model. The major shortcoming of the basic SIR-model is the assumed homogeneous mixing of the population, which means assuming all individuals in a population have an equal probability of infecting one another. In this paper, we have covered a few different papers where SIR-models were used for different purposes, highlighting the flexibility of SIR-modeling. From these papers, we arrive at a few conclusions: SIR-models do not accurately predict the number of infected over a longer period of time, and different vaccination strategies and vaccine release times can have a huge impact on the evolution of an epidemic even if the number of vaccines stays the same.

References

[1] Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D. and Roser, M., 2021. Coronavirus Pandemic (COVID-19). [online] Our World in Data. Available at: <https://ourworldindata.org/coronavirus> [Accessed 29 November 2021].

[2] Kermack, W. and McKendrick, A., 1927. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A Available at: <http://doi.org/10.1098/rspa.1927.0118> [Accessed 29 November 2021].

[3] Moein, S., Nickaeen, N., Roointan, A., Borhani, N., Heidary, Z., Javanmard, S., Ghaisari, J. and Gheisari, Y., 2021. Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan. Scientific Reports, 11(1), p.1-9.

[4] Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A. and Colaneri, M., 2020. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Medicine, 26(6), p.855-860.

[5] Ghostine, R., Gharamti, M., Hassrouny, S. and Hoteit, I., 2021. An Extended SEIR Model with Vaccination for Forecasting the COVID-19 Pandemic in Saudi Arabia Using an Ensemble Kalman Filter. Mathematics, 9(6), 636 (article number: 636)

[6] Berkane, S., Harizi, I. and Tayebi, A., 2021. Modeling the Effect of Population-Wide Vaccination on the Evolution of COVID-19 Epidemic in Canada.

[7] Bubar, K. M., Reinholt, K., Kissler, S. M., Lipsitch, M., Cobey, S., Grad,Y. H., Larremore, D. B., 2021. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. Science. 371(6532), p.916-921.

[8] Yu, Z., Liu, J., Wang, X., Zhu, X., Wang, D. and Han, G., 2016. Efficient Vaccine Distribution Based on a Hybrid Compartmental Model. PLOS ONE, 11(5), p.e0155416.

[9] Agarwal, N., Komo, A., Patel, C., Pathak, P. and Ünver, M., 2021. The Trade-off Between Prioritization and Vaccination Speed Depends on Mitigation Measures. National Bureau of Economic Research, Inc, NBER Working Papers 28519.

[10] Karaivanov, A., Lu, S., Shigeoka, H., Chen, C. and Pamplona, S., 2021. Face masks, public policies and slowing the spread of COVID-19: Evidence from Canada. Journal of Health Economics, 78, p.102475.

[11] Griffiths, J. and Jiang, S., 2021. Wuhan officials have revised the city's coronavirus death toll up by 50%. [online] CNN. Available at: <https://edition.cnn.com/2020/04/17/asia/china-wuhan-coronavirus-death-toll-intl-hnk/index.html> [Accessed 29 November 2021].

[12] Davies, N., Klepac, P., Liu, Y., Prem, K., Jit, M. and Eggo, R., 2020. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nature Medicine, 26(8), p.1205-1211.

[13] Liu, J. and Xia, S., 2011. Toward Effective Vaccine Deployment: A Systematic Study. Journal of Medical Systems, 35(5), p.1153-1164.

[14] Pollock, A. and Lancaster, J., 2020. Asymptomatic transmission of covid-19. BMJ, p.m4851.

[15] Worldometer. 2021. Worldometer - real time world statistics. [online] Available at: <https://www.worldometers.info> [Accessed 31 July 2021].

[16] Polack, F., Thomas, S., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J., Pérez Marc, G., Moreira, E., Zerbini, C., Bailey, R., Swanson, K., Roychoudhury, S., Koury, K., Li, P., Kalina, W., Cooper, D., Frenck, R., Hammitt, L., Türeci, Ö., Nell, H., Schaefer, A., Ünal, S., Tresnan, D., Mather, S., Dormitzer, P., Şahin, U., Jansen, K. and Gruber, W., 2020. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. New England Journal of Medicine, 383(27), p.2603-2615.

[17] Centers for Disease Control and Prevention. 2021. 2009 H1N1 Flu Pandemic Timeline. [online] Available at: <https://www.cdc.gov/flu/pandemic-resources/2009-pandemic-timeline.html> [Accessed 31 July 2021].

[18] Chernozhukov, V., Kasahara, H. and Schrimpf, P., 2021. Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. Journal of Econometrics, 220(1), p.23-62.

[19] Epstein, M. J., 2008. Why Model?. [online] Jasss.org. Available at: <https://www.jasss.org/11/4/12.html> [Accessed 31 July 2021].

[20] Qureshi, A., Baskett, W., Huang, W., Lobanova, I., Hasan Naqvi, S. and Shyu, C., 2022. Reinfection With Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Patients Undergoing Serial Laboratory Testing. Clinical Infectious Diseases, 74(2), p.294-300.