A Study on the Effects of Containment Policies and Vaccination on the Spread of SARS-CoV-2

Vahid S. Bokharaie*

Abstract

This paper presents a method to predict the spread of the SARS-CoV-2 in a population with a known age-structure, and then, to quantify the effects of various containment policies, including those policies that affect each age-group differently. The model itself is a compartmental model in which each compartment is divided into a number of age-groups. The parameter of the model are estimated using an optimisation scheme and some known results from the theory of monotone systems such that the model output agrees with some collected data on the spread of SARS-CoV-2.

To highlight the strengths of this framework, a few case studies are presented in which different populations are subjected to different containment strategies. They include cases in which the containment policies switch between scenarios with different levels of severity. Then a case study on herd immunity due to vaccination is presented. And then it is shown how we can use this framework to optimality distribute a limited number of vaccine units in a given population to maximise their impact and lower the total number of infectious individuals.

Keywords— Mathematical Epidemiology, COVID-19, SARS-CoV-2, Compartmental Models, Parameter Estimation, Contact Rates, Global Optimisation, Monotone Systems, Nonlinear Dynamical Systems.

MSC subclass— 92C60, 92C50

1   Introduction

This manuscript presents a framework to model the spread of SARS-CoV-2 in a population with a known age-structure. The model itself is a compartmental model in which each compartment is divided into a number of age-groups. And a data-driven approach is presented to adapt the parameters of this model to the available data on the spread of SARS-CoV-2. The mathematical framework is adopted from [11] and [4], although they only consider the SIS case, while here I have used the SEIR model. The SEIR models are of course not new, but the main contribution of this work is to use an optimisation scheme to tune its parameters to the available data on the spread of SARS-CoV-2. The framework is general, as will be explained, although in this manuscript the data collected from early stages of the spread of the virus in Wuhan is used. With more detailed and comprehensive data, the outputs of the model can be more accurate. One advantage of the adopted framework is that it allows us to estimate the parameters based on data collected from one population in any other population with a known age-structure. Hence, using the data collected from a population in China will not limit us to model the spread of virus only in China.

There is also a publicly available Python library called MiTepid_sim [8] which was specifically developed as a part of this study. All the figures in this manuscript are generated using `MiTepid_sim`. It can be used to simulate the spread of the SARS-CoV-2 in each age group starting from any given initial condition and under any defined containment policy. Also, another library called `MiTepid_opt` [7] accompanies this work which is used to implement the optimisation scheme presented in this manuscript and is also publicly available.

The structure of the paper is as follows. In Section 2, I will explain the outline of the model and the optimisation schemes, without getting into the details. The mathematical details are explained in Appendices A.1 and A.2. Then various case studies are presented in Section 3. They include a study on how changing variables of the model affect the trajectory of the Infectious population. And then studying the effects of various suppression strategies, including those which might affect each age group differently. Then, long-term containment plans. These are the types of containment policies which we

---

*Vahid Samadi Bokharaie is with Max Planck Institute for Biological Cybernetics, Tübingen, Germany (email: vahid.bokharaie@tuebingen.mpg.de)

have already seen in many countries around the world. They usually include scenarios in which we switch between policies with different levels of severity. And finally, a study on the effects of vaccination on the population. Using the presented framework, we can see at what level of vaccination we can achieve herd immunity. It will also be shown in case we do not have enough vaccine units to reach herd immunity, how to use the limited number of vaccine units to have a maximum impact. That is another optimisation problem, details of which are explained in Appendix A.3. A discussion on the pros and cons of this approach and possible directions for improving this methodology is presented in Section 4.

2 The Method

What follows is based an earlier work [9], in which an SIR compartmental model was used as a basis for all simulations and optimisations. At the time of writing of [9] it was already shown that SARS-CoV-2 infers immunity, for Macaque monkeys in [6] and for humans in [5]. Further investigations by virologists and physicians seem to show that SARS-CoV-2 also infers an inhibition period, as explained in [14] and references therein. Hence an SEIR model is a more suitable choice than SIR. As a reminder for those who might not be familiar with the terminology, in SEIR models, the population is divided into compartments that capture different stages in the progress of that disease. The four compartments in SEIR model are *Susceptible*, S, which includes those who are healthy and can be infected; *Exposed*, E, those who are infected but not yet Infectious, *Infectious*, I, which includes those who are infected and can transmit the disease; and *Removed*, R, which includes those who were Infectious but no more. And the term SEIR shows the progress of the disease. If, for example, a virus does not have a latent period but infers immunity, we use SIR compartmental model. There are also some works which have used more detailed compartmental models, for example [13], in which Infectious compartment is further divided into Asymptotic and Symptomatic. In this work, such a differentiation is not necessary, for reasons that will be discussed when explaining the optimisation scheme.

In this manuscript, each compartment is divided into a certain number of groups. The choice of how to group the population is completely arbitrary from a theoretical perspective. But from a practical point of view, such a choice should be informed by the characteristics of the disease and also the availability of the relevant data. And in that regard, a very good choice for SARS-CoV-2 is to divide the population into age groups. More specifically, the population in each compartment S, E, I and R is divided into nine age groups 0-10, 10-20, ..., 70-80 and 80+. The subsequent model is a set of 4×9 nonlinear ordinary differential equations (ODEs) with some unknown parameters which we should estimate based on the known characteristics of the SARS-CoV-2 and the available data. As detailed in Appendix A.1, we can assume the population is constant, which is a reasonable assumptions in the time-scales which are of interest to us. Hence, we end up with a set of 3×9 ODEs. This set of ODEs have a number of unknown parameters which include *transfer rates*, *inhibition rates*, and *contact rates*. Transfer rates and inhibition rates, represented with γ_i and σ_i respectively for each of the age-groups $i = 1, \cdots, 9$, can be easily calculated if we know on average, how long is the inhibition period and for how long an individual is Infectious. I have assumed them to be 5 and 4.6 days, respectively, as reported in [14]. And assuming one day to be the unit of time, and assuming this time-period is on average the same for all age groups, then $\gamma_i = 1/5$ and $\sigma_i = 1/4.6$ for $i = 1, \cdots, 9$. Or we can consider different γ_i and σ_i for each age group if we have such data in each age-group available.

But the real challenge is in estimating the contact rates, represented by β_{ij} for $i, j = 1, \cdots, 9$ which denote the rate at which Susceptibles in age-group i are infected by Infectious individuals in age-group j. Contact rates are very difficult to estimate because they capture various characteristics of the population and the dynamics of the virus. Their value can depend on the average number of direct contacts between the members of different groups, which needs a comprehensive and detailed analysis of the behaviour and mobility of the individuals in the population. Contact rates can also depend on the differences in susceptibility of each age-group to the virus and also on the mechanisms of its transmission. Hence, to directly calculate the values of contact rates is a monumental task that even if possible, might be extremely difficult, time-consuming and expensive.

The main contribution of this work is to show how to overcome this challenge and how to estimate the contact rates. In order to do so, an optimisation scheme is used which is based on two distinct but important sets of data on the spread of COVID-19. One is the estimate for basic reproduction number, R_0, for COVID-19 in an uncontained population. There are various estimates for that parameter, but the one reported in [19], which is $R_0 = 2.95$ is used in this manuscript. A few other groups have also reported values very close to that [2]. The other piece of information is coming from [1], which shows the relative distribution of confirmed cases of COVID-19 in each age-group in Wuhan, China, as of February

11th, as shown in Table 5 in Appendix A.2. Up until two weeks before this date, Chinese authorities did not impose any meaningful containment strategies and the virus was spreading in an uncontained population. Hence we can assume these numbers are the results of the uncontained spread of the virus. Using these two sets of information, and using the optimisation scheme as detailed in Appendix A.2, we can estimate for the contact rates. One detail of the optimisation scheme which deserves to be highlighted is that it relies on the ratio of the number of reported cases in each age-group, not the absolute values. By doing so, we can avoid two potential issues that might arise in using such data. One is the fact that the number of confirmed cases in each age-group is an unknown fraction of actual Infectious numbers at each time. Another issue is the deliberate or non-deliberate errors that creep in while reporting these numbers. Relying on the ratios of the reported Infected numbers in the optimisation scheme makes it less sensitive to both of these issues, although it does not completely eliminate them. Using more data points as inputs to the optimisation scheme is one possible way to overcome these issues even more.

All the mathematical details of both the model and the optimisation schemes are explained in detail in Appendices A.1 and A.2. The important thing to keep in mind is that although the optimisation scheme uses data collected from China in a certain time-period and in an uncontained population, but the values we obtain can be adapted to any population with a known age-structure and under any containment policy, if we know how that policy affects contact between various age-groups, or how it will change the *Basic Reproduction Number* (defined in Section A.1.2). The model has some other advantages and also some disadvantages which are discussed in Section 4. Even Considering the disadvantages, it can serve as a good first step to quantify the effects of various containment policies in different countries. The actual values of the contact rates are accompanied with `MiTepid_sim` [8], a python library which can be used to simulate the model used in this paper or any other stratified compartmental model of interest.

3   Results

Using the mathematical framework explained in Section 2 and detailed in Appendices A.1 and A.2, we can answer various questions about how SARS-CoV-2 spreads in any population with a known age-structure. We can quantify the effects of various containment policies on the spread of the virus even if those policies affect age-groups differently and if the policies vary in time. It can be used to study the effects of partially vaccinating the population, how wide-spread the vaccination should be to reach herd immunity and how to optimally distribute limited vaccine resources in a population. In this section we will see how we can find answers to all these questions.

3.1   Uncontained Spread of the Virus

Let's consider the case in which COVID-19 is spread uncontained, i.e. when people in the population interact with each other as in normal times, with no external or self-imposed restrictions in the interactions.

Figures 1.1 and 1.2 show the Infectious and Removed population in each age-group in Germany. As mentioned in the previous section, the output of the model is the evolution of the trajectories of E_i, I_i and R_i for $i = 1, \cdots, n$. We can then add up these values to obtain the aggregate trajectories of E, I and R compartments. That is how the Figures 1.3 and 1.4 are generated. For each country, the trajectory for

Country	Eventual Ratio of Removed Population (%)	Maximum Instantaneous Infectious Ratio (%)
Germany	91.77%	14.35%
Iran	82.15%	10.33%
Italy	91.23%	14.21%
Spain	90.35%	13.95%
China	87.06%	12.06%
USA	90.56%	13.63%
UK	91.99%	14.18%
France	91.95%	14.32%
South Africa	80.01%	9.72%

Table 1: Maximum Instantaneous Infectious ratio and eventual Removed ratio in different countries in the uncontained scenario for SEIR model. The Removed compartment includes those individuals who have been Infectious and then were cured or died.
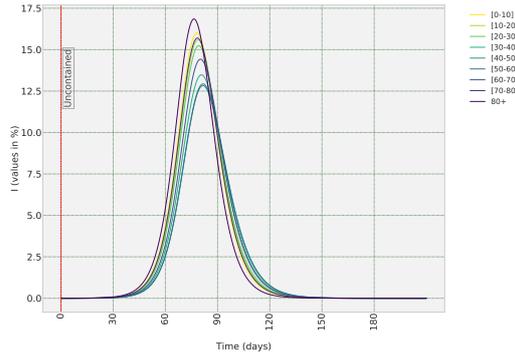
Figure 1.1: The Infectious ratio in the uncontained scenario in each age-group in Germany.
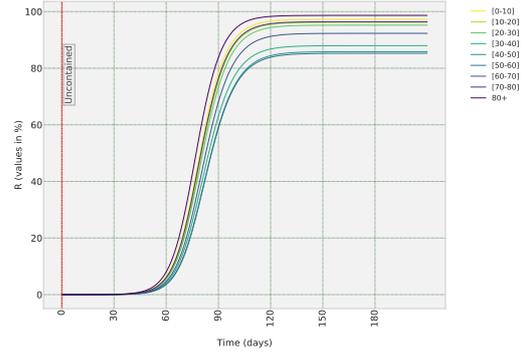
Figure 1.2: The Removed ratio (ratio of those who were infected and then were cured or lost their lives) in the uncontained scenario in each age-group in Germany.

each age-group is calculated, and then the aggregate is calculated based on the population distribution for that country. As can be seen, the predicted peak in the number of Infectious people and the eventual ratio of the Removed population varies considerably among different countries. That can be explained based on differences in the population distribution in these countries. For example, in Iran, 66.8% of the population are under 40 years old while that ratio is 39.8% in Italy [3]. The model predicts that countries with an older population would be affected worse if they let the virus spread uncontained, in agreement with for example [12] when comparing mortality rates in the USA and UK.

It should be noted that the model is not directly concerned with the mortality rate or the number of those who might need respiratory or intensive care in hospitals. Such information can be inferred from the number of Infectious and Removed compartments in each country, based on the available data or estimates on mortality rate and also the ratio of the Infectious and that need intensive care. Dealing with that is out of the scope of this manuscript, but interested reader can look, for examples, at [17, 15, 16, 20].

To solve any system of Ordinary Differential Equations (ODEs), apart from the equations themselves, we should also define the initial conditions, which in our case means the initial ratios of Infectious, Inhibited and Removed populations in each age group. In all the figures presented in this manuscript, I have assumed 1 in 100,000 in each age group is Infectious at time $t = 0$, and the Removed population is 0. We should be cautious in using a system of continuous ODEs for numbers lower than that. It should be noted that the peak values in Figures 1.3 and 1.4 are barely affected by the choice of initial conditions. But that is not true for the time it takes to reach the peak values. Hence, in order to predict the day in which the number of Infectious reaches the peak value, we should have a reasonable estimate of the initial conditions. But given that the number of confirmed cases are reported on a daily basis in many countries, and that there are suggestion on how to infer the total number of infected based on the reported number in each country, finding the right initial conditions for each country is a problem that can be solved.

We can also run simulations assuming different values of Basic Reproduction Number for uncontained populations, not just $R_0 = 2.95$ to see how that might change the trajectories. To clarify, I have assumed R_0 for uncontained population can be any value in $[2.05, 3.95]$ range with steps of 0.1. We can then estimate contact rates with each assume value for R_0 and then calculate the trajectories. As can be seen in Figures 1.5 and 1.6, higher R_0 leads to higher peaks in both I and R.

Also, I have run the same procedure varying T_I, the average time each individual remains Infectious. Figures 1.7 and 1.8 show what happens if the actual value of T_I changes in [5, 14] days range. Increasing T_I increases the peak value in I, but delays the time it takes for R to reach its maximum value.
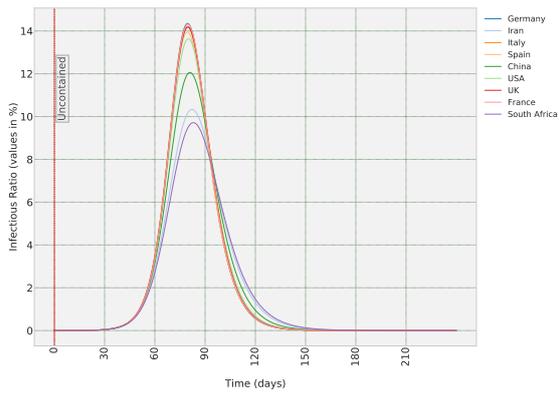
Figure 1.3: Aggregate ratio of Infectious in the uncontained scenario in different countries with SEIR model.
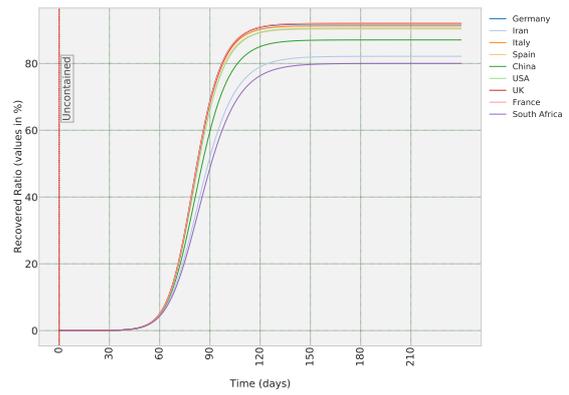
Figure 1.4: Aggregate ratio of Removed compartment in the uncontained scenario in different countries with SEIR model.
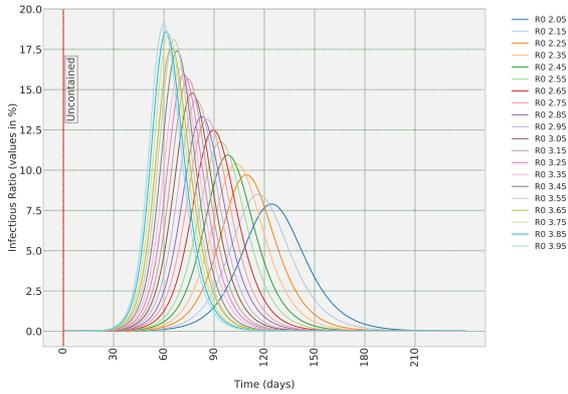
Figure 1.5: Aggregate ratio of Infectious in the uncontained scenario in Germany when R_0 in uncontained population is assumed to be different values.
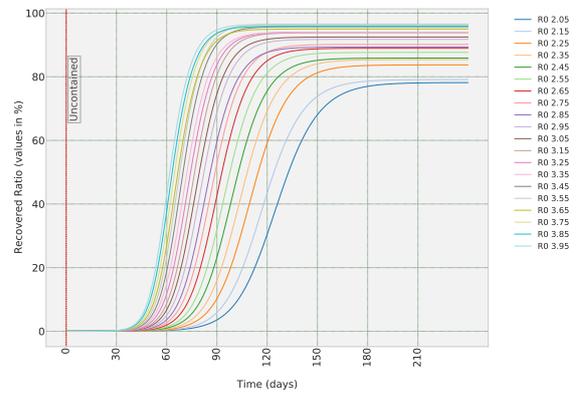
Figure 1.6: Aggregate ratio of Recovered in the uncontained scenario in Germany when R_0 in uncontained population is assumed to be different values.
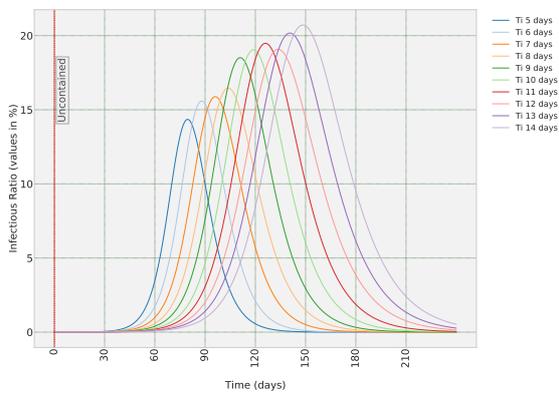
Figure 1.7: Aggregate ratio of Infectious in the uncontained scenario in Germany when assumed value of T_i (the average time-period hosts remain infectious) in uncontained population is varied.
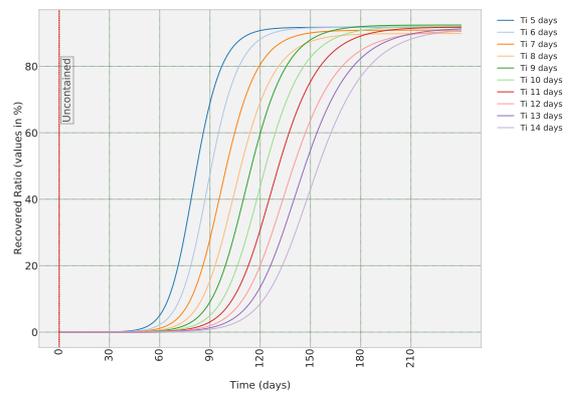
Figure 1.8: Aggregate ratio of Recovered in the uncontained scenario in Germany when assumed value of T_i (the average time-period hosts remain infectious) in uncontained population is varied.

3.2 Suppression Strategies

The advantage of having a stratified model is that we can quantitatively evaluate the effects of various containment strategies that affect different age-groups differently. Table 2 has listed some of the more common policies and how they are defined. The column under *Policy Description* defines to what extent the contacts of age-groups are assumed to change under each policy. These values are chosen intuitively, but any other definitions and any other policies can be easily defined in mitepid_sim package that is developed as a part of this study [8].

As can be seen in Figures 2.1 and 2.2, all the containment policies decrease both the peak instantaneous Infectious ratio and eventual Removed ratio, which is to be expected. The most effective policy among the ones listed in the Table 2 is *Lock-down*. It is worth noting that although in Lock-down policy the contact are limited to half of Social Distancing policy, the decrease in the peak Infectious value is much less noticeable.

Table 2 also shows the basic reproduction number, R_0, for each policies. As can be seen in Figure 2.1, and as it is well-known in epidemiology, when $R_0 < 1$, the disease starts to disappear from the population. But as importantly, bringing the ratio of the Infectious population down to a small enough ratio of the total population can take months, even if we impose a total lock-down strategy (the exact time depends on the ratio of Infectious when we start the policy). And maintaining such a strategy might not be feasible in many countries with a more fragile economy. In the next section, we will see what happens if we switch between different strategies with different degrees of severity as a long-term plan.

Policy Label	Policy Name	Policy Description	Eventual Removed Ratio (%)	Maximum Instantaneous Infectious Ratio (%)	R_0
UN	Uncontained	All interactions as in normal circumstances	91.77%	14.35%	2.95
KI	Only Schools and Universities Closed	interactions of [0-20] age groups decreases to 20%	81.85%	11.02%	2.48
EL	Only Elderly Social Distancing	interactions of 70+ age groups decreases to 25%	84.45%	11.21%	2.37
KIEL	Schools Closed Elderly Social Distancing	combination of KI and EL	72.73%	7.98%	1.88
KIOF	Schools, Offices and Companies Closed	Combination of KI and interactions of [20-70] age range decreases to 50%	58.59%	5.81%	1.67
ADEL	Adults and Elderly Self-isolate	Interactions of Elderly down to 25% Adults to 20%	34.19%	4.29%	1.05
SD	Social Distancing	combination of KI and EL and interactions of people in [20-70] age range reduce to 20%	22.27%	3.94%	0.63
LD	Lock-down	interactions of all individuals reduced to 10%	16.41%	3.84%	0.30

Table 2: Effects of different policies in Germany. Please note that Eventual Removed Ratio is the total ratio of the population which have been infected at any time during the spread of the virus.
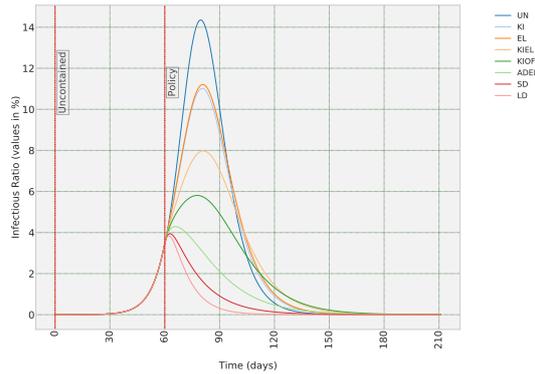
Figure 2.1: How Infectious ratios change under different containment policies in Germany or any country with similar age-structure. Policies are defined in Table 2
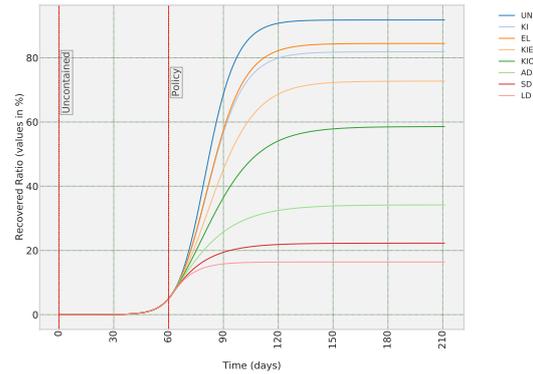
Figure 2.2: Removed ratios under different policies. Removed population refers to those who were Infectious, and then recovered or died.

3.3 Mitigation Strategies and Long-term Plans

So far, we have seen what happens if we leave the population uncontained and let the virus spread freely. It can lead to, for example in the case of Germany, around 92% of the population being eventually infected, which can lead to hundreds of thousands of death. Then we saw that some suppression strategies can be hugely effective in containing the spread of the virus and stop such humanitarian disasters. But a complete and long-lasting lock-down strategy, even if possible, can have a huge economic cost, apart from its psychobiological toll on the members of the society. A reasonable compromise is to switch between containment strategies of various degrees of strictness, depending on the observed trend in the number of confirmed cases. What results is what is known in control theory as a *switched system*. Implementing that using this framework is quite easy. We start with the first set of parameters, contact rates, at time $t = 0$, run the simulations until time $t = t_1$, and use the final state of the previous ODE as the initial conditions for the new one. All of this can be easily implemented in `MiTepid_sim` software package [8]. What follows in this section is simply a few examples of how such strategies can or cannot be useful. This by no means is meant to be considered a comprehensive list of effective mitigation strategies but merely examples to highlight the capabilities of this framework. All the example discussed in the following is based on the population distribution of Germany.

Let's start with a case in which we switch between uncontained case and Social-Distancing, as defined in Table 2. It means the case in which contacts between children and adults is brought down to 20% of uncontained case, and for elderly to 25%. When for uncontained $R_0 = 2.95$, in Social Distancing it is decreased to $R_0 = 0.63$. We assume uncontained case for the first 30 days, starting from the case in which 1 in 100,000 is Infectious. As can be seen in Figures 3.1 and 3.2 the policy manages to initially contain the virus, but as the number of Infectious during uncontained phases increase, the exponential growth causes the number of Infectious to reach its peak value at 150 days and then start to decrease. Although this policy is obviously not able to contain the number of Infectious in a manageable level, we can still see some benefits in it. The peak value of Infectious is now less than what it was in uncontained case, 9.6% vs 14.3%. Also, the time to reach the peak value has increased from around 80 days to 150 days. But more importantly, the eventual recovered ratio, i.e. the total number of individuals who were infected at some stage, is decreased from around 92% to around 62%. That means although this policy will eventually lead to, most probably, overload of the health system resources, but it can limit the mortality rate significantly.

Now let's repeat the same scenario, but now we, after the first 30 days of uncontained policy, we replace uncontained policy with the case in which $R_0 = 1.5$. Given $R_0 = 2.95$ for uncontained population, this means contacts between members of the population is assumed to be cut to half. Figures 3.3 and 3.4 show how Infectious and Recovered populations changes. This time, the policy is successful in containing the spread of the virus. Obviously, it does not make sense to go back to uncontained at 330 days, but I have left it there so we can see the potential the population has for a new outbreak if we relax the resuscitations.

But reaching the same level of restrictions as we have defined under the Social-distancing policy has proved very difficult (at least in Germany and many other countries.). So, lets look at a more realistic scenario, in which after the first 30 days of uncontained population, we switch between cases in which $R_0 = 1.50$ and $R_0 = 0.9$, which is not far from what has already happened in Germany. Figures 3.5

7
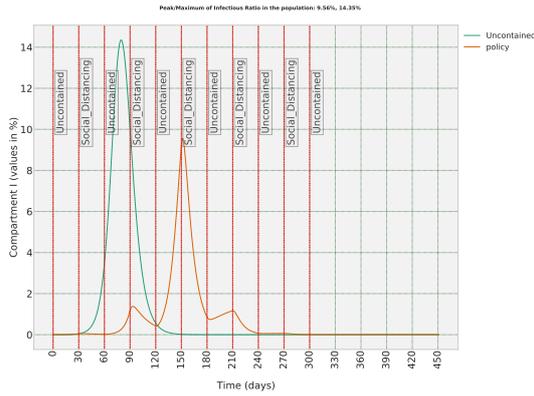
Figure 3.1: Switching between uncontained and social-distancing scenarios. The Infectious ratio.
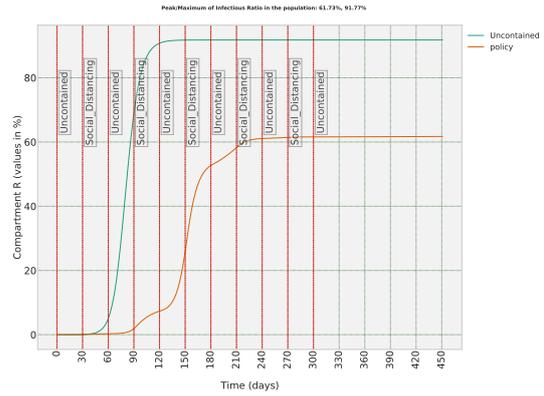
Figure 3.2: Switching between uncontained and social-distancing scenarios. The Removed ratio.
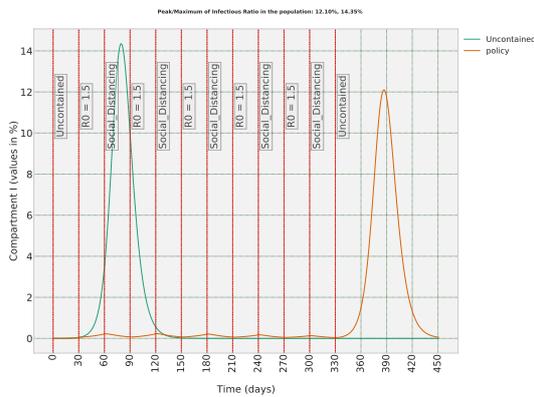
Figure 3.3: A less strict long-term strategy: switching between social distancing and $R_0 = 1.5$ scenarios. The Infectious population.
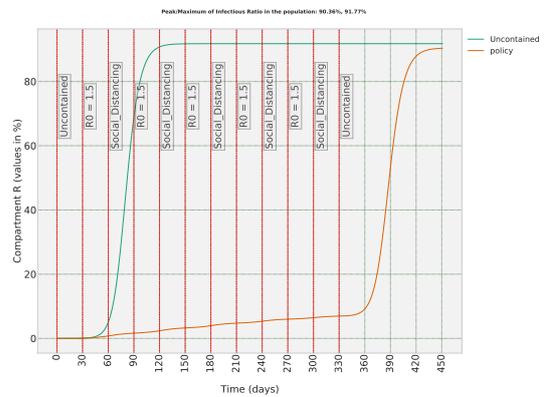
Figure 3.4: Switching between social distancing and $R_0 = 1.5$ scenarios. The Removed population.

and 3.6 show the outcome. It is no surprise to see that the number of Infectious is now more than the previous case but still the maximum number of Infectious is far from that of the uncontained scenario. But what happens if we start imposing this switching policy after 60 days of uncontained population, not 30 days. Figures 3.7 and 3.8 shows the price we have to pay for an extra 30 days of inaction in containing the spread of the virus. Number of Infectious starts to get out of hand, and although the peak is much less than uncontained case 5.58% vs 14.35%, it can still be overwhelming for public health resources of any country.

There are various other possibilities that can be studied using this framework, all of which can be easily implemented using the `MiTepid_sim` python package [8].
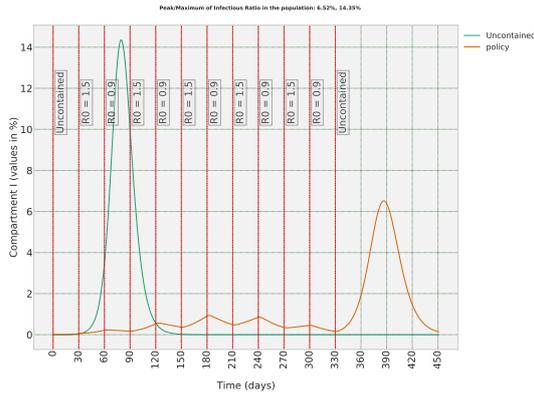
Figure 3.5: A less stringent long-term strategy when we switch between situations where $R_0 = 0.9$ and $R_0 = 1.5$. Containment policy can contain the spread of the virus when imposed 30 days after 1 in 100,000 in population is Infectious.
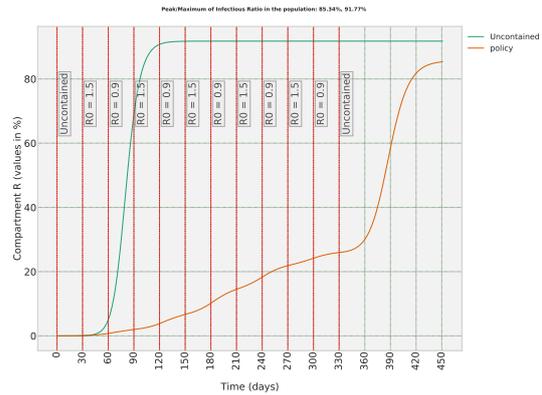
Figure 3.6: Removed population when we switch between $R_0 = 0.9$ and $R_0 = 1.5$ scenarios.
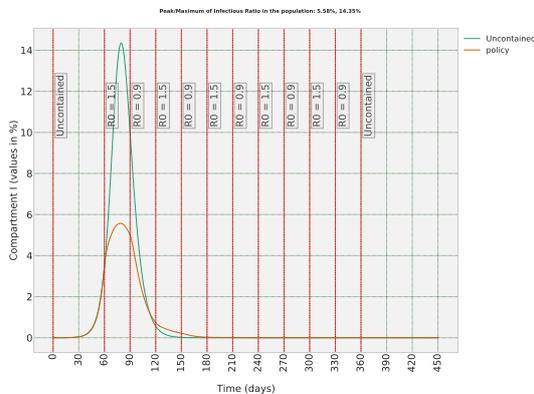
Figure 3.7: Same as in Figure 3.5, only difference is that policy is imposed 60 days after Infectious population reaches 1 in 100,000. Only extra 30 days of inaction causes the number of Infectious to get out of hand.
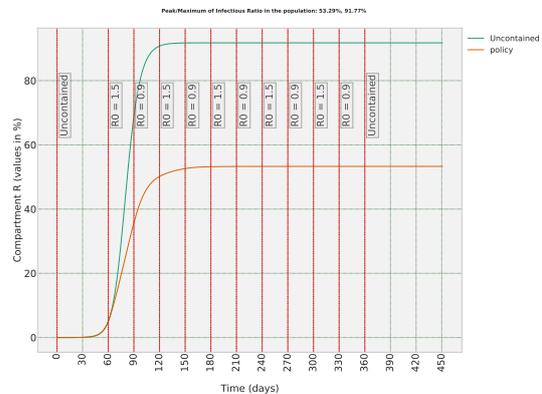
Figure 3.8: Recovered ratio for scenario described in Figure 3.7.

3.4 Herd Immunity due to Vaccination

In Section 3.1 we saw that even in the uncontained case, not all of the individuals in the population need to be infected for the spread of the virus to be stopped. As can be seen in Table 1 and Figure 1.4, in an uncontained population in which SARS-CoV-2 is spread freely, when the total number of recovered reaches a certain value, the disease dies out, a phenomenon which is called *Natural Herd Immunity*. The exact value depends on the dynamics of the virus and also age-structure of the population in case of SARS-CoV-2. But the price we pay for natural herd immunity is catastrophic. For example, in Germany, natural herd immunity is reached when around 92% of the population is eventually infected, a parameter that I call *Herd Immunity Threshold (HIT)*. And with a mortality rate of 0.6% that amounts to around 460 thousand deaths.

We have already seen in the previous section that imposing any kind of containment policy, can reduce the HIT significantly. Now let's assume a vaccine for SARS-CoV-2 is produced and is approved by health authorities. Let's see how does that changes HIT. In particular, we want to know what ratio of the population should be vaccinated to reach herd immunity? Answering that question is very easy using the framework we have in place. There is no need to define a new compartment for the vaccinated people. We can assume the vaccinated population as the initial conditions for the Recovered compartment and then see how the trajectories of Infectious and Recovered evolve.

Table 3 shows the eventual recovered and maximum Infectious ratios under different initial ratio of vaccinated population, assuming that after vaccination, a containment strategy is in place such that $R0 = 1.50$. Remember that $R0 = 2.95$ in an uncontained population, hence $R = 1.50$ means interactions in the populations are reduced to around half of that in an uncontained case. As can be seen in Table 3,

Ratio of Vaccinated	Eventual Recovered Ratio	Total Infected	Maximum Infectious Ratio
0%	55.63%	55.63%	3.10%
5%	52.99%	47.99%	2.32%
10%	50.13%	40.13%	1.63%
15%	46.97%	31.97%	1.05%
20%	42.45%	22.45%	0.58%
25%	32.47%	7.47%	0.24%
26%	30.85%	4.85%	0.17%
27%	29.95%	2.95%	0.11%
28%	29.72%	1.72%	0.06%
29%	29.99%	0.99%	0.03%
30%	30.57%	0.57%	0.01%
31%	31.33%	0.33%	0.01%
32%	32.20%	0.20%	0.00%
33%	33.13%	0.13%	0.00%
34%	34.09%	0.09%	0.00%

Table 3: *How vaccinating different ratios of the population changes the total and maximum Infectious Ratios. It is assumed the population is under a containment policy that keep $R_0 = 1.50$.*
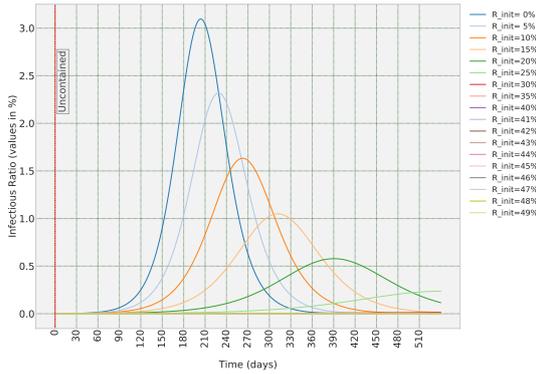
Figure 4.1: *The effect of vaccinating different ratio of the population of Germany on the evolution of the Infectious trajectory. It is assumed population is contained to $R_0 = 1.5$.*
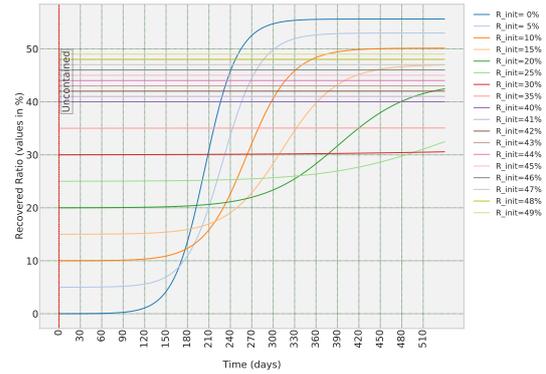
Figure 4.2: *The effect of vaccinating different ratio of the population of Germany on the evolution of the Recovered trajectory. It is assumed population is contained to $R_0 = 1.5$.*

when we inoculate around 30% of the population, we have already reached the herd immunity level. If we leave the population uncontained, we need to vaccinate around 65% of the population to reach HIT. Please note that these values are for a population with the same age-structure as Germany, for countries with a younger population, HIT is lower.

Figures 4.1 and 4.2 shows the evolution of Infectious and Recovered trajectories when we start from different ratios of vaccinated population when we keep $R_0 = 1.5$ in our population and Figures 4.3 and 4.4 show evolution of trajectories for an uncontained population. But in these figures, we have assumed the same ratio in each age-group is vaccinated. In other words we have ignored an important capability of the model, the ability to impose different conditions on different age-groups. So, let's say we have enough vaccine units to only inoculate 15% of the population of Germany. We can turn this problem into an optimisation problem, in which we keep overall number of vaccinated population at 15% while minimising the eventual Recovered population if we vaccinate different ratios of each age-group. Table 4 show the difference that non-uniform vaccination in age-groups can have. And Figures 4.5 and 4.6 show the evolution of trajectories corresponding to uniform and optimised cases. Obviously, it is not really realistic that we vaccinate only 15% of the population and then none. This simple example is merely used to highlight this feature of this framework. The code used to run this optimisation approach is also available in `mitepid_opt` [7] package.
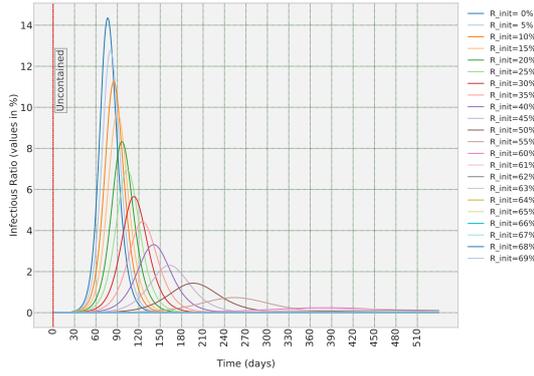
Figure 4.3: Same as Figure 4.1 but it is assumed the population is uncontained.
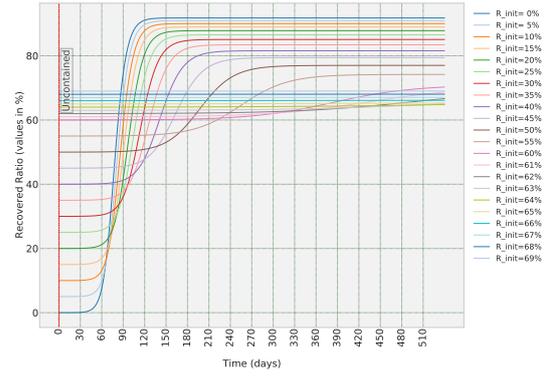
Figure 4.4: Same as Figure 4.2 but it is assumed the population is uncontained.

case	Eventual Recovered	Total Infected	Maximum Infectious
uniform	46.97%	31.97%	1.05%
optimised	44.28%	29.28%	0.87%

Table 4: Difference between the case in which 15% of each age-group is vaccinated and the optimised case, in which different fractions of each age-group are vaccinated, but still 15% of the population in total. Population is assumed to be under a containment policy that keeps $R_0 = 1.50$.
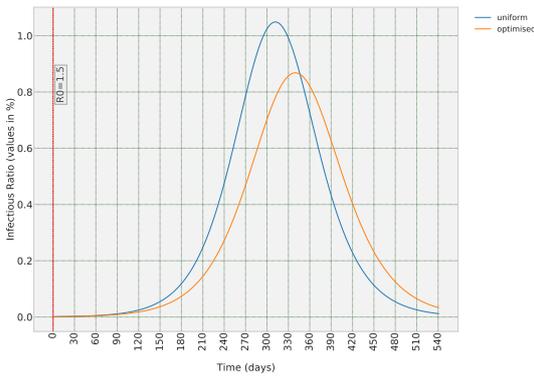
Figure 4.5: Difference between uniform vaccinations and optimally distributed vaccination in Infectious ratio.
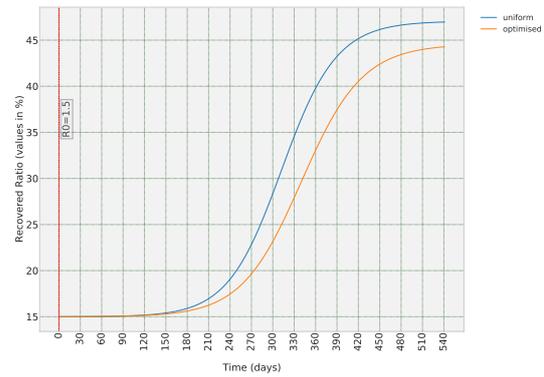
Figure 4.6: Difference between uniform vaccinations and optimally distributed vaccination in Recovered ratio.

4   Discussions

The method presented in this manuscript can be used to predict the spread of SARS-CoV-2 in each age-group of a population with a known age-structure. Not just the aggregate values, but also the evolution of the number of Infectious and Recovered individuals in each age-group. It was shown how we can use various containment strategies and how to quantify their effects. Strategies which might be time-varying and might affect each age group independently. The main advantage of this framework is that it allows us to estimate the contact rates of the model without a detailed knowledge of how and to what extent different age-groups in a population interact with each other, or how the virus affects each age-group. The contact rates are estimated based on the available data on the spread of COVID-19. The model itself is a set of nonlinear Ordinary Differential Equations (ODEs), hence simulating various containment policies in different time frames does need any special computational power.

To better estimate the contact rates, any other available data can be used as the input to the optimisation scheme. We need only two points in time with a known number of reported infected cases in each age groups while the Basic Reproduction Number, R_0, remains constant and known in between the two time-points. Given the ever-changing containment polices imposed on various populations in the

past few months, finding such population with a constant and known R_0 is difficult. That is the reasons that I have relied on the data collected from Wuhan, China, in the early weeks of the spread of the virus when no containment policy was imposed on the population. As already mentioned, if we obtain such estimation from data collected in China, we can easily adapt them to any other population of any country/region/city with a known age-structure.

In spite of all its strength, the methodology has some shortcomings that should kept in mind while interpreting the results. The method assumes the virus affects individuals (of the same age) in different countries similarly. If it is the case that the differences in genetic backgrounds or vaccination histories in different countries can affect the infection rate, such differences would be lost while transforming the contact rates from one population to another. Also, it is an implicit assumption in the model that the general interactions between people in different countries and societies are in general similar to each other. To clarify the point, if, for example, people older than 70 years in one country live with or have more contacts with the next generations, and in another country, they live in isolation or in nursing homes, then differences in contact rates caused by such social and cultural differences would also be lost. But if we estimate the contact rates based on two separate set of data collected in two different countries, and then compare the resulting normalised contact rates, we can discover such minute differences in how different age groups in the two populations interact with each other which is a useful corollary of this method.

The method can be extended in various directions to make the results even more useful. We can, for example, divide each age-group into sub-groups based on their vulnerability to the virus, or based on the relative amount of interactions with other individuals in the population. An obvious choice is people who work from home or are unemployed and people who have to go to their offices. Or people in 60-69 age range who are retired or those who still work. Even a rough estimate of the relative numbers of these sub-groups in each age group can give us more insight into more effective ways to contain the spread of the virus with less social and economic impact.

Given the fact that nowadays the population structure in different cities/regions in almost every country is known, we can use the model to describe the spread of the virus for each region or city, and then, assuming in and out-flow traffic to each city is known, consider them as exogenous inputs to our switched system. That would allow us to consider time-lags that might exist in the spread of COVID-19 to different parts of a country. But even as it is, this model and its estimated parameters can be a useful tool for different policy-makers in different countries in particular those countries with limited computational resources or people with expertise in mathematical epidemiology.

Declarations

Bibliography

[1] Breakdown of 44,672 sample patients infected with novel coronavirus COVID-19 in china as of february 11, 2020, by age group. Accessed: 2020-04-01. [Online]. Available: https://www.statista.com/statistics/1095024/china-age-distribution-of-wuhan-coronavirus-covid-19-patients/

[2] "MIDAS 2019 novel coronavirus repository," https://github.com/midas-network/COVID-19, accessed: 2020-04-07.

[3] "Population pyramids of the world," https://www.populationpyramid.net/world/2019/, accessed: 2020-04-07.

[4] M. Ait Rami, V. S. Bokharaie, O. Mason, and F. Wirth, "Stability criteria for sis epidemiological models under switching policies," *Discrete and Continuous Dynamical Systems: Series B*, vol. 19, no. 9, pp. 2865–2887, 2014.

[5] F. Amanat, T. Nguyen, V. Chromikova, S. Strohmeier, D. Stadlbauer, A. Javier, K. Jiang, G. Asthagiri-Arunkumar, J. Polanco, M. Bermudez-Gonzalez, D. Caplivski, A. Cheng, K. Kedzierska, O. Vapalahti, J. Hepojoki, V. Simon, and F. Krammer, "A serological assay to detect sars-cov-2 seroconversion in humans," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/03/18/2020.03.17.20037713

[6] L. Bao, W. Deng, H. Gao, C. Xiao, J. Liu, J. Xue, Q. Lv, J. Liu, P. Yu, Y. Xu, F. Qi, Y. Qu, F. Li, Z. Xiang, H. Yu, S. Gong, M. Liu, G. Wang, S. Wang, Z. Song, W. Zhao, Y. Han, L. Zhao, X. Liu, Q. Wei, and C. Qin, "Reinfection could not occur in sars-cov-2 infected rhesus macaques," *bioRxiv*, 2020. [Online]. Available: https://www.biorxiv.org/content/early/2020/03/14/2020.03.13.990226

[7] V. S. Bokharaie, "MiTepid_opt: An optimisation package to estimate the contact rates of a stratifeid compartmental epidemiel-ogical model." https://github.com/vahid-sb/MiTepid_opt, accessed: 2020-09-20.

[8] ——, "MiTepid_sim: A python-based library to simulate the spread of COVID19 in a stratified population." https://github.com/vahid-sb/MiTepid_sim, accessed: 2020-09-20.

[9] ——, "A stratified model to quantify the effects of containment policies on the spread of covid-19," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/04/16/2020.04.10.20060681

[10] O. Diekmann, J. Heesterbeek, and J. Metz, "On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations," *Journal of mathematical biology*, vol. 28, no. 4, pp. 365–382, 1990.

[11] A. Fall, A. Iggidr, G. Sallet, and J. Tewa, "Epidemiological models and Lyapunov functions," *Math. Model. Nat. Phenom.*, vol. 2, no. 1, pp. 62–68, 2007.

[12] N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg *et al.*, "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and health-care demand," 2020.

[13] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri, "Modelling the covid-19 epidemic and implementation of population-wide interventions in italy," *Nature Medicine*, pp. 1–6, 2020.

[14] S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, and M. Lipsitch, "Projecting the transmission dynamics of sars-cov-2 through the postpandemic period," *Science*, vol. 368, no. 6493, pp. 860–868, 2020.

[15] M. O'Driscoll, G. R. Dos Santos, L. Wang, D. A. Cummings, A. S. Azman, J. Paireau, A. Fontanet, S. Cauchemez, and H. Salje, "Age-specific mortality and immunity patterns of sars-cov-2 infection in 45 countries," *medRxiv*, 2020.

[16] R. Pastor-Barriuso, B. Perez-Gomez, M. A. Hernan, M. Perez-Olmeda, R. Yotti, J. Oteo, J. L. Sanmartin, I. Leon-Gomez, A. Fernandez-Garcia, P. Fernandez-Navarro *et al.*, "Sars-cov-2 infection fatality risk in a nationwide seroepidemiological study," *medRxiv*, 2020.

[17] J. Perez-Saez, S. A. Lauer, L. Kaiser, S. Regard, E. Delaporte, I. Guessous, S. Stringhini, A. S. Azman, S.-P. S. Group *et al.*, "Serology-informed estimates of sars-cov-2 infection fatality risk in geneva, switzerland," *medRxiv*, 2020.

[18] P. van den Driessche and J. Watmough, "Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission," *Math. Biosci.*, pp. 29–48, 2002.

[19] Y. Wang, X. You, Y. Wang, L. Peng, Z. Du, S. Gilmour, D. Yoneoka, J. Gu, C. Hao, Y. Hao *et al.*, "Estimating the basic reproduction number of covid-19 in wuhan, china," *Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi*, vol. 41, no. 4, pp. 476–479, 2020.

[20] H. Ward, C. J. Atchison, M. Whitaker, K. E. Ainslie, J. Elliot, L. C. Okell, R. Redd, D. Ashby, C. A. Donnelly, W. Barclay *et al.*, "Antibody prevalence for sars-cov-2 in england following first peak of the pandemic: React2 study in 100,000 adults," *medRxiv*, 2020.

[21] Z. Wu and J. M. McGoogan, "Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention," *Jama*, 2020.

Appendices

A.1 The Mathematical Model

To understand the model discussed in this section, it is enough to know some basic concepts in Ordinary Differential Equations (ODEs) and linear algebra. Knowing the following notations and definitions can be helpful.

A.1.1 Notations and Some Basic Definitions

$\mathbb{R}$ is the field of real numbers and $\mathbb{R}_+$ is The set of non-negative real numbers. $\mathbb{R}^n$ is The space of column vectors of size n of real numbers and $\mathbb{R}^{n \times n}$ is The space of $n \times n$ matrices of real numbers. I use x_i to represent The ith entry of the vector x in $\mathbb{R}^n$, for $i \in \{1, \cdots, n\}$. Please note that x_0 is a vector in $\mathbb{R}^n$ that usually represents initial condition. Notation a_{ij} is used for (i, j) entry of the matrix A. $D = \text{diag}(x)$ is an $n \times n$ diagonal matrix in which $d_{ii} = x_i$ for all i. A^{-1} is The inverse of the matrix A. I is the identity matrix of proper dimensions and 0 is the zero matrix of proper dimensions. $\sigma(A)$ is the set of all eigenvalues (spectrum) of the matrix A. $\rho(A)$ is the spectral radius of the matrix A, i.e. the maximum of the absolute values of all eigenvalues. $\mu(A)$ is The spectral abscissa of the matrix A.

$A \gg B$ means $a_{ij} > b_{ij}$, for all $i, j \in \{1, \cdots, n\}$. It should not be mistaken with Positive Definite (PD) matrices. $A > B$ means $a_{ij} \geq b_{ij}$, for all $i, j \in \{1, \cdots, n\}$ and $A \neq B$ and $A \geq B$ means $a_{ij} \geq b_{ij}$, for all $i, j \in \{1, \cdots, n\}$. $\mathbb{R}^n_+$ is The positive orthant of $\mathbb{R}^n$, given by $\{x \in \mathbb{R}^n : x \geq 0\}$.

Knowing the following basic definitions can help in understanding the text better.

A matrix A is called *Hurwitz*, if $\mu(A) < 0$.

A real $n \times n$ matrix $A = (a_{ij})$ is *Metzler* if its off-diagonal entries are non-negative.

The matrix A is *irreducible* if and only if for every non-empty proper subset K of $N := \{1, \cdots, n\}$, there exists an $i \in K$, $j \in N \setminus K$ such that $a_{ij} \neq 0$. When A is not irreducible, it is *reducible*.

For any subset $\mathcal{U}$ of $\mathbb{R}^n$, a point x_0 is called an *interior point* of $\mathcal{U}$ if there is an open ball around x_0 which is wholly contained in $\mathcal{U}$. The set of all interior points of $\mathcal{U}$ is called the interior of $\mathcal{U}$ and is denoted by $\text{int}(\mathcal{U})$.

Consider a continuous-time nonlinear systems of the form:

$$\dot{x}(t) = f(x), \;\; x(0) = x_0 \tag{1}$$

where $f : \mathcal{D} \mapsto \mathbb{R}^n$ is a nonlinear vector field on a subset $\mathcal{D}$ of $\mathbb{R}^n$ and $x_0 \in \mathcal{D}$ is called the initial condition.

A.1.2 SIS Model

The SIS model, although not used in the main text, is presented here, both in the interest of completeness and to provide a theoretical basis for the SIR/SEIR discussions. The formulation presented in this section is adopted from [11] and [4]. In SIS model, the population of interest is first divided into two compartments S, Susceptibles, and I, Infectious. Each compartment is sub-divided into n groups. These groups can represent different age groups, different health conditions, professions, etc. In this manuscript, I have divided the population in each compartment into $n = 9$ age-groups defined as 0-10, 10-20, ..., 70-80 and 80+.

Let $I_i(t)$ and $S_i(t)$ be the number of Infectious and Susceptibles at time t in group i for $i = 1, \cdots, n$, respectively. Also, let $N_i(t) = S_i(t) + I_i(t)$ be the total population of group i. The total population of each group is assumed to be constant; formally, $N_i(t) = N_i$. This does not oversimplify the model, especially when the total population is significantly greater than the number of dead and newborn. But even if that assumption is not deemed realistic for a population, the formulation stated below can still be used as we will shortly see.

Here, β_{ij}, the contact rate between groups i and j, denotes the rate at which Susceptibles in group i are infected by Infectious in group j for $i, j = 1, \cdots, n$. Further, γ_i, the transfer rate, is the rate at which an infectious individual in the group i leaves the Infectious compartment and join the Susceptible compartment. We also consider birth and death in the population, but as discussed, we set the birth

and death rates in each age-group to be the same value μ_i to keep the total population in each group constant. Using the mass-action law, the basic SIS model is then described as follows [11]:

$$\begin{cases} \dot{S}_i(t) = \mu_i N_i - \mu_i S_i(t) - \sum_{j=1}^{n} \beta_{i,j} \frac{S_i(t).I_j(t)}{N_i} + \gamma_i I_i(t) \\ \dot{I}_i(t) = \sum_{j=1}^{n} \beta_{i,j} \frac{S_i(t).I_j(t)}{N_i} - (\gamma_i + \mu_i) I_i(t) \end{cases} \qquad (2)$$

Since the population of each group is constant, it is sufficient to know $I_i(t)$. If we set $x_i(t) = I_i(t)/N_i$ and $\tilde{\beta}_{i,j} = \beta_{i,j} N_j / N_i$ and $\alpha_i = \gamma_i + \mu_i$, we obtain the following differential equation:

$$\dot{x}_i(t) = (1 - x_i(t)) \sum_{j=1}^{n} \tilde{\beta}_{i,j} x_j(t) - \alpha_i x_i(t), \qquad (3)$$

for all $i = 1, \cdots, n$. Based on the definition, $x \in B_n$ where $B_n := \{x \in \mathbb{R}_+^n : x \leq 1\}$. We can write the differential equation (3) in compact form as:

$$\dot{x} = [D + B - \mathrm{diag}\,(x)B]x \qquad (4)$$

where $D = -\mathrm{diag}\,(\alpha_i)$ and $B = (\tilde{\beta}_{ij}) > 0$. As a reminder, here $>$represents an element-wise inequality, as defined in Section A.1.1, and does not refer to Positive Definiteness.

The following properties of (4) are easy to check. Interested reader can look at [4] for proofs.

(i) $f(x) = [D + B - \mathrm{diag}\,(x)B]x$ with D and B defined as above is C^1 in $\mathbb{R}^n$, therefore, the solution for every initial condition in $\mathbb{R}^n$ exists and is unique for all $t \geq 0$.

(ii) The origin is an equilibrium point of (4). This equilibrium is referred to as the disease-free equilibrium (DFE) of the system (4).

(iii) System (4) may have an equilibrium in $\mathrm{int}\,(\mathbb{R}_+^n)$ (also referred to as an endemic equilibrium). Conditions for existence of endemic equilibrium for the system (4) depends on parameter R_0, explained below.

One important parameter in mathematical epidemiologically is the *basic reproduction number*, R_0. There are different definitions for the basic reproduction number. Probably the most common definition is as follows.

Definition A.1.1 (Basic reproduction number) *The basic reproduction number is the expected number of secondary cases produced, in a completely susceptible population, by a typical infective individual during its entire period of Infectiousness [10].*

For the SIS model (4), following [11], it can be proved that $R_0 = \rho(-D^{-1}B)$. The reproduction number can be used to characterise the existence and stability of the equilibria of (4). As shown in [11, Theorem 2.3], the disease-free equilibrium, i.e. the origin, is a globally asymptotically stable equilibrium of the system (4) if and only if $R_0 < 1$ (if matrix B is irreducible). And the endemic equilibrium, an equilibrium in $\mathrm{int}\,(\mathbb{R}_+^n)$, is globally asymptotically stable if and only if $R_0 > 1$. In other words, the necessary and sufficient condition to eradicate a disease for a population is to satisfy $R_0 < 1$.

A.1.3 SIR Model

The SIR model is quite similar to SIS, with a minor difference, namely, those who are cured, join the Removed, R, population, not S. Hence, the formulation for an SIR model is as follows:

$$\begin{cases} \dot{S}_i(t) = \mu_i N_i - \mu_i S_i(t) - \sum_{j=1}^{n} \beta_{i,j} \frac{S_i(t).I_j(t)}{N_i} \\ \dot{I}_i(t) = \sum_{j=1}^{n} \beta_{i,j} \frac{S_i(t).I_j(t)}{N_i} - (\gamma_i + \mu_i) I_i(t) \\ \dot{R}_i(t) = \gamma_i I_i(t) - \mu_i R_i(t) \end{cases} \qquad (5)$$

15

Again, assuming $N_i(t) = S_i(t) + I_i(t) + R_i(t)$ is constant, similar to what was done in the previous section, if we set $x_i(t) = I_i(t)/N_i$ and $y_i(t) = R_i(t)/N_i$ and $\tilde{\beta}_{i,j} = \beta_{i,j} N_j/N_i$ and $\alpha_i = \gamma_i + \mu_i$, $\mu_i = 0$, we obtain the following differential equation:

$$\begin{cases} \dot{x}_i(t) = (1 - x_i(t)) \sum_{j=1}^{n} \tilde{\beta}_{i,j} x_j(t) - \alpha_i x_i(t) \\ \dot{y}_i(t) = \gamma_i x_i(t) \end{cases} \tag{6}$$

$\forall i = 1, \cdots, n$. In compact from, (6) can be written as follows:

$$\begin{cases} \dot{x} = [D + B - \text{diag}\,(x)B]x \\ \dot{y} = \Gamma x \end{cases} \tag{7}$$

where $D = -\text{diag}\,(\alpha_i)$ and $B = (\tilde{\beta}_{ij}) > 0$ and $\Gamma = \text{diag}\,(\gamma_i)$ for $i = 1, \cdots, n$.

The system (7) has the following properties.

(i) $f(x) = [D + B - \text{diag}\,(x)B]x$ and $g(y) = \Gamma x$ with D, B and Γ defined as above are C^1 in $\mathbb{R}^n$, therefore, the solution for every initial condition in $\mathbb{R}^n$ exists and is unique for all $t \geq 0$.

(ii) To calculate the equilibria of the system we set $f(x) = 0$ and $g(x) = 0$. One equilibrium is the origin, Disease-Free equilibrium (DFE), and the other one, corresponds to the case in which the disease has swept through the population and a significant ratio of the population (and not necessarily all the population) belongs to Recovered compartment and I=0.

(iii) Basic Reproduction Number for (7), can be calculated using the same formula $R_0 = \rho(-D^{-1}B)$, when all R_i and I_i are close to 0.

Property (iii) follows from the discussion in [18, Section 3] and the fact that this equation is derived from the model linearised around the origin. This also means that as the more Infectious ratio increases, the effective R_0 becomes less and less than $\rho(-D^{-1}B)$.

A.1.4 SEIR Model

The SEIR model is an extension of SIR in which the susceptibles enter a latent, E, compartment, before becoming Infectious. The formulation of an SEIR model can be stated as follows:

$$\begin{cases} \dot{S}_i(t) = \mu_i N_i - \mu_i S_i(t) - \sum_{j=1}^{n} \beta_{i,j} \frac{S_i(t).I_j(t)}{N_i} \\ \dot{E}_i(t) = \frac{S_i(t)}{N_i} \sum_{j=1}^{n} \beta_{i,j} I_j - \mu_i E_i - \epsilon_i E_i \\ \dot{I}_i(t) = \epsilon_i E_i - \mu_i I_i - \gamma_i I_i \\ \dot{R}_i(t) = \gamma_i I_i(t) - \mu_i R_i(t) \end{cases} \tag{8}$$

Again, we assume $N_i(t) = S_i(t) + I_i(t) + R_i(t)$ is constant. Now we define $x_i(t) = I_i(t)/N_i$ and $y_i(t) = R_i(t)/N_i$ and $z_i = E_i/N_i$ and $\tilde{\beta}_{i,j} = \beta_{i,j} N_j/N_i$ and $\alpha_i = \gamma_i + \mu_i$, $\mu_i = 0$, we obtain the following differential equation:

$$\begin{cases} \dot{z}_i(t) = (1 - x_i - y_i - z_i) \sum_{j=1}^{n} \tilde{\beta}_{i,j} x_j(t) - (\mu_i + \epsilon_i)z_i \\ \dot{x}_i(t) = \epsilon_i z_i - (\mu_i + \gamma_I)x_i \\ \dot{y}_i(t) = \gamma_i x_i(t) - \mu_i y_i \end{cases} \tag{9}$$

$\forall i = 1, \cdots, n$. In compact from, (9) can be written as follows:

$$\begin{cases} \dot{z} = [B - \text{diag}\,(x)B]x - \text{diag}\,(y)Bx + [-\text{diag}\,(z)B + D]z \\ \dot{x} = \Xi z + Dx \\ \dot{y} = \Gamma x \end{cases} \tag{10}$$

where $D = -\text{diag}\,(\alpha_i)$ and $B = (\tilde{\beta}_{ij}) > 0$ and $\Gamma = \text{diag}\,(\gamma_i)$ and $\Xi = \text{diag}\,(\epsilon_i)$ for $i = 1, \cdots, n$.

The same as in 7, the origin is also an equilibrium for 10. And using the same arguments as presented in the previous Section, we can use $R_0 = \rho(-D^{-1}B)$ as a reasonable estimate for Basic Reproduction Number, R_0 when the trajectory is around the origin. When the trajectory moves away from the origin, $\rho(-D^{-1}B)$ over-estimates the actual value of R_0.

Age Group	Population Ratio in % (C1)	Confirmed COVID-19 Ratio in % (C2)	C2/C1	C2/C1 normalised to first age-group
0-10	11.9	0.9	0.075	1.0
10-20	11.6	1.2	0.107	1.4
20-30	13.5	8.1	0.600	8.0
30-40	15.6	17.0	1.089	14.5
40-50	15.6	19.2	1.230	16.4
50-60	15.0	22.4	1.493	19.9
60-70	10.4	19.2	1.846	24.6
70-80	4.7	8.8	1.872	25.0
80+	1.7	3.2	1.882	25.1

Table 5: Population distribution in China and the distribution of Confirmed COVID-19 cases in China as of Feb. 11th. To normalise the distribution of confirmed cases, we can divide the values over the population ratio of that age group. The resulting values are then used to estimates parameters of the epidemiological models.

A.2   An Optimisation Scheme to Estimate the Parameters of the Epidemiological Model

In order to solve ordinary differential equations (ODEs) (4) or (7) or or (10), we need to have a reliable estimate of $\tilde{\beta}_{i,j}$ and γ_i for all i, j. Parameters and σ_i are easy to estimate. If for example the average duration that individuals in a group i are Infectious is 5 days, then $\gamma_i = 1/5 = 0.20$, given that we have chosen one day to be the unit of time. Estimating $\tilde{\beta}_{i,j}$, on the other hand, is very difficult, and this section explains how we can estimate contact rates based on real-world data on the spread of COVID-19.

Column C2 in Table 5 shows the distribution of confirmed cases of COVID-19 in different age groups in China as of Feb. 11th [1]. We normalise these numbers to the relative distribution of each age group in the general population. By doing so, we can compare the differences in how different age groups are affected by the virus. We can do so by dividing values in Column C2 to those of Column C1. We can further divide the resulting numbers to the smallest of them, which happens to be the first row. Then we obtain the last column of Table 5, which shows the normalised relative distribution of infective people in the Chinese population as of Feb. 11th.

The optimisation scheme aims to find the contact rates such that at a given time t_g, the ratio of the values of the states in Systems (7) or (4) matches the values reported in the last column of Table 5. At the same time, we should keep the basic reproduction number R_0 equal to its estimated value for the spread of COVID-19 in an unconstrained population. There are various estimates for R_0 for COVID-19, some are listed in [2]. Most estimates fall in $[1.5, 3.5]$ range. I have chosen $R_0 = 2.95$ as reported in [19].

Hence, the optimisation problem we need to solve is as follows: finding matrix B such that the following two conditions are satisfied:

(i) For a given dialogical matrix D and scalar $R_0 = 2.95$, $R_0 = \rho(-D^{-1}B)$

(ii) The relative values of the states of system (4) or (7) at a given time t_g and initial condition x_0 satisfy the values of the last column of Table 5.

But how to choose t_g and x_0? For that, I have relied on reports that the spread of the virus has probably started in Wuhan city, in late November [21]. Hence I have set $t_g = 75$ days (meaning the spread of the virus has initiated 75 days before Feb 11th). I have also assumed 75 days before our target date of Feb. 11th, 2020, 1 in $100,000$ of the population are infected. That seems like a reasonable assumption, although we cannot be really sure. To overcome this uncertainty, adding more data points to the optimisation scheme, rather than only two data points as used here, can help.

Now that all the required parameters are set, we can solve the optimisation problem to find a B_{opt}. In order to solve the optimisation problem, I have used `sqp` algorithm in `globalsearch` function in `Global Optimisation Toolbox` in Matlab©. Optimisation is done in two steps, in the first step, initial values for matrix B are chosen randomly from a uniform distribution. When the optimisation algorithm converges to a solution, the optimisation procedure is repeated, this time with the optimum value obtained in the first step as the initial values.

The objective function in the optimisation scheme is the weighted sum of two terms. One is the 2-norm of the difference between the ratio of trajectories of the set of ODEs at time t_g and the desired ratios extracted from Table 5. The second term is the difference between $\rho(-D^{-1}B)$ and the desired

basic reproduction number of $R_0 = 2.95$. The weights are tuned manually such that no term is obscured by the other one during the optimisation steps.

To implements the optimisation scheme, I have used an in-house package called `MiTepid_opt` which is publicly available [7]. The optimisation algorithm runs in 15-20 minutes on 40 hyper-threaded CPUs of type `Intel(R) Xeon(R) CPU E5-2687W v4 @ 3.00GHz`.

Note A.2.1 *It should be noted that the values obtained from the optimisation schemes are $\tilde{\beta}_{i,j}$ which are usable only for population distribution in China. But using the relationship $\beta_{i,j} = \tilde{\beta}_{i,j} N_i / N_j$, as defined in Section A.1.4, we can obtain normalised values that can be used for all population densities, where N_i, N_j are the population ratios in age groups i and j. For each other target population we can use the equation $\tilde{\beta}_{i,j} = \beta_{i,j} N_j / N_i$ to calculate the $\tilde{\beta}_{i,j}$ in (10) and then solve the ODEs to obtain the evolution of the trajectories.*

Note A.2.2 *This methodology can be applied to any uncontained or contained population if the effective basic reproduction number, R_0 is known. But given the difficulty in estimating R_0 in a contained population, and the fact that most countries in the world now have a form of containment policy in place, the data collected in early stages of the spread of the virus in China seems to be the best available option for the optimisation scheme.*

A.3   Optimising the distribution of limited vaccine units

This Section briefly explains the optimisation scheme used in Section 3.4 to optimise the distribution of a limited number of vaccine units in the population. The optimisation algorithm used for this problem is the exact same global optimisation method used in Section A.2 and I will not repeat its details. But the objective function is obviously different. In this case, the objective function we aim to minimise is a weighted sum of two terms. The main term, is the final aggregate value of the Recovered compartment. As a reminder, the value of compartment R at each time represents the total number of those who were infected at any time in the past and are not Infectious any more. The second term is devised to make sure the total number of vaccine units remains constant. Simply, turning a constrained optimisation problem into an unconstrained optimisation problem to save computational time.

A.4   SIR Case

As already mentioned in Section 1, the SEIR model seems to be a suitable choice for the spread of SARS-CoV-2 in a population, compared to SIR [14]. But as a thought experiment, in this section we can have a look at how the virus spreads in a population if instead of 4.6 days in Exposed period and 5 days in Infectious period, there is no latent period and 9.6 days of Infectious period. Table 6 show the eventual Recovered and maximum Infectious ratio of the same countries as in Table 1. Comparing the two tables we can see that the eventual Recovered Ratio (which is the total number of people who are infected at some stage) is less in SIR case, but the peak of instantaneous Infectious has decreased.

Figures 0.1 and 0.2 shows the evolution of Infectious and Recovered ratios in these countries.

Country	Eventual Ratio of Removed Population (%)	Maximum Instantaneous Infectious Ratio (%)
Germany	81.80%	18.40%
Iran	71.31%	13.61%
Italy	81.84%	18.38%
Spain	82.01%	18.40%
China	77.39%	15.71%
USA	81.20%	17.87%
UK	82.07%	18.34%
France	82.81%	18.76%
South Africa	69.20%	12.80%

Table 6: Maximum Instantaneous Infectious ratio and eventual Removed ratio in different countries in the uncontained scenario for SIR model.
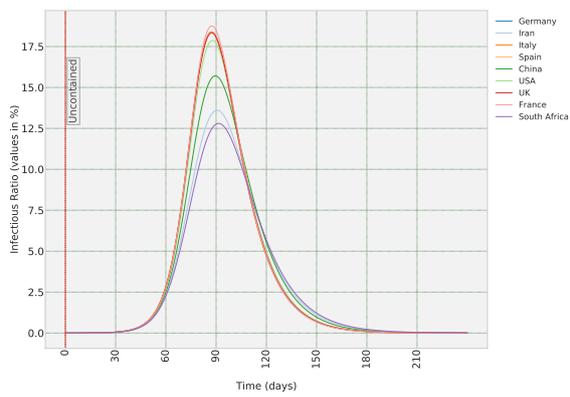
Figure 0.1: Aggregate ratio of Infectious in the uncontained scenario in different countries with SEIR model.
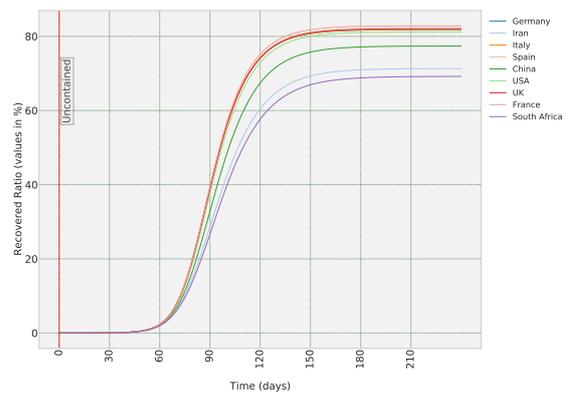
Figure 0.2: Aggregate ratio of Removed compartment in the uncontained scenario in different countries with SEIR model.