

# Merits of an Age-Compartmentalized SIR Model In Predicting Healthcare Burden From COVID-19 In Suffolk County, MA

Kenneth Cox

May 2020

Submitted in fulfillment of class requirements for 1.063 (Fluids and Diseases), taught by  
Prof. Lydia Bourouiba at MIT, Spring 2020

# 1 Introduction

On 31 December 2019, China reported to the WHO a cluster of pneumonia cases with no known origin or cause in Wuhan, China. The cause was identified on 7 January 2020 as a new coronavirus, and by 30 January, with 82 confirmed cases outside of China, WHO declared the epidemic a “public health emergency of international concern” [1]. Within months, the virus caused cases in every inhabited continent, and many governments declared shutdowns in order to slow its spread.

The combined social, psychological, and economic impacts of these shutdowns is unprecedented in the modern history of disease containment, so their effectiveness must be worth their risk. Preliminary data indicate the measures have reduced the spread of the new disease (named COVID-19), since the epidemic curves in many areas display an increasingly negative slope. However, the consequences of these shutdowns have prompted many officials to lift them, despite a high risk of a second wave of cases. Given the difficulty the healthcare system faced during the first wave, concerns abound about the ability of hospitals and other critical care facilities to take care of another influx of COVID-19 patients.

Therefore, careful implementation of shutdown measures are crucial to maintaining a strong, economically viable society, while at the same time ensuring that there are enough resources available for those who do get sick. The current plan, advocated by many public health experts, is a “phased reopening” that proceeds in a number of stages, where individuals can initially return to work with modifications (including social distancing and the use of personal protective equipment), and eventually builds up to a complete reopening of society [2]. However, this approach is novel, and leaves as many questions as answers. Foremost among these concerns is whether a phased reopening will produce a large volume of cases, and send society back into another complete lockdown.

To simulate the effect of a phased reopening on the number of COVID-19 cases in Suffolk County, Massachusetts, an age-compartmentalized SIR (Susceptible-Infected-Recovered) model was developed and implemented. The age compartmentalization (into a young class and old class) provides better control over the disease dynamics, which seem to depend heavily on patient age [3]. However, this extra flexibility means there are more parameters to be tuned, which introduce additional uncertainty into the model. Sensitivity analyses of some parameters demonstrate some shortcomings of this model, and highlight the need for models that are not so sensitive to initial conditions.

Interventions are modeled as taking immediate effect on the day they are put in place. For Massachusetts, the statewide stay-at-home order went into effect on 24 March 2020 [4], and it is the only intervention considered in the model to have an effect, because its magnitude is large in comparison to other intervention measures (e.g. increased handwashing, mask wearing). Furthermore, it encompasses other intervention measures. For example, schools were closed before the stay-at-home order went into effect, but the stay-at-home order includes school closures as well as work closures.

## 2 Background

Building mathematical models of diseases is important for prediction of the total number of cases (especially critical when cases can be severe, and therefore place a large burden on healthcare systems). A common model used for such purposes is the SIR model (abbreviation of Susceptible-Infected-Recovered), and related variants, including the SEIR model (where the E stands for “exposed”) [5]. The idea behind these models is to divide the population into several mutually exclusive compartments and derive equations for the movement of individuals between compartments. When the model is being used to predict eventual case numbers, the most important variable is the number of people in the infected compartment at any time  $t$ .

Unfortunately, any model so simple is bound to have its faults, and the SIR model is no exception. A major failing of the SIR model is its assumption of population homogeneity. In other words, all individuals in the population are assumed to be the same, and therefore are all at equal risk of contracting infection if they are susceptible, or are all equally likely to recover if infected. In reality, people are different, and some people are infected more readily or recover more quickly than others. So far in the COVID-19 epidemic, it appears that underlying conditions such as diabetes, lung diseases, and heart diseases, among others, influence an individual’s chance of getting sick and recovering or dying [3]. It also appears that older adults are at far higher risk of severe outcomes than younger adults. Clearly, then, a more accurate model takes into account underlying conditions, or at least age, when making predictions about the number of infections.

Such models have been developed. The current study is partly inspired by the “heterogenous transmission thesis” argued in [6]. The authors show that under a COVID-19 regime, where a part of the population is at a significantly higher risk of contracting fatal infection (in the case of COVID-19, mortality rates for adults  $\geq 60$  years of age are up to 50 times higher than the mortality rates of their young peers), maintaining viral spread at normal levels for the less vulnerable population while placing strict restrictions on the more vulnerable population can actually reduce the number of mortalities, despite increasing the number of cases. In theory, then, allowing regular activities for individuals under 60, while enacting stay-at-home and other mitigation practices for people over 60 will cause more infections but less deaths. The practical implications of such a containment strategy are enormous, since much of the economic burden resulting from universal stay-at-home orders can be undone. The same authors recently released another paper that generalized the model to an arbitrary number of age groups, and arrived at similar conclusions [7]. An even more sophisticated model with the same premise analyzed COVID-19 dynamics in Lombardy, Italy and Hubei, China [8]. Along with age structure, the authors include the presence of co-morbidities and household structure in the model, improving the model’s performance in predicting the number of severe cases and deaths.

The above models also include discussion of the role of interventions. In these models, the infection parameter is reduced by some amount, corresponding mechanistically to a reduction in the number of contacts (due to stay-at-home orders) and/or a reduction in the

probability of infection given a contact (due to better sanitary practices and mask wearing). The authors of the first two papers presented above use a gradual reduction, while the authors of the latter paper use an immediate reduction. Because the latter paper is able to account for individual-level parameters, the authors are also able to simulate social distancing measures separate from sheltering measures. Another paper making use of gradual reductions is [9]; the authors decrease the infection parameter exponentially, and this is the approach chosen for the present study.

Despite the modeling improvements the above studies make, none truly replicate the most likely eventual intervention — a combination of several different measures, including social distancing, stay-at-home measures, and contact tracing/quarantine of suspected cases. A comprehensive study examining all the different types of measures is generally only feasible on the local level, and that is what the authors in [10] attempt. Their study, crucially, includes quarantining of suspected cases. Their results are different from those expected from a traditional SIR model (interestingly, an  $R_0$  less than one is still capable of producing exponential growth of cases). Although no model can pin down the exact dynamics of COVID-19, the SIR compartmental model serves as a decent approximation.

There are few studies that analyze the impact of COVID-19 on Suffolk county. The present study seeks to examine the effectiveness of an SIR model, compartmentalized further by age, in predicting the total case counts, particularly the severe case counts (since these are the ones that place a burden on the healthcare system).

### 3 The Model

The full model is an SIR-like series of differential equations, shown below.

$$\begin{aligned}
\dot{S}_{young} &= -\beta_{yy}S_{young}I_{mild,young}/N - \beta_{oy}S_{young}I_{mild,old}/N \\
\dot{S}_{old} &= -\beta_{oo}S_{old}I_{mild,old}/N - \beta_{yo}S_{old}I_{mild,young}/N \\
\dot{I}_{mild,young} &= \beta_{yy}S_{young}I_{mild,young}/N + \beta_{oy}S_{young}I_{mild,old}/N - \gamma_{mild,young}I_{mild,young} - a_{young}I_{mild,young} \\
\dot{I}_{severe,young} &= a_{young}I_{mild,young} - \gamma_{severe,young}I_{severe,young} - \mu_{young}I_{severe,young} \\
\dot{I}_{mild,old} &= \beta_{oo}S_{old}I_{mild,old}/N + \beta_{yo}S_{old}I_{mild,young}/N - \gamma_{mild,old}I_{mild,old} - a_{old}I_{mild,old} \\
\dot{I}_{severe,old} &= a_{old}I_{mild,old} - \gamma_{severe,old}I_{severe,old} - \mu_{old}I_{severe,old} \\
\dot{R}_{young} &= \gamma_{mild,young}I_{mild,young} + \gamma_{severe,young}I_{severe,young} \\
\dot{M}_{young} &= \mu_y I_{severe,young} \\
\dot{R}_{old} &= \gamma_{mild,old}I_{mild,old} + \gamma_{severe,old}I_{severe,old} \\
\dot{M}_{old} &= \mu_{old}I_{severe,old}
\end{aligned}$$

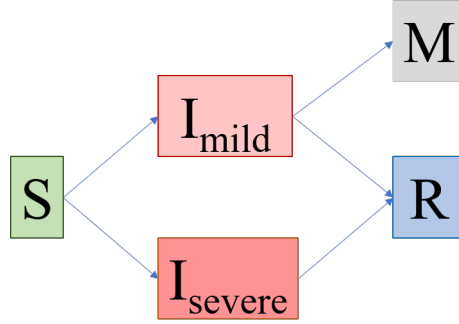


Figure 1: Model compartments.

### 3.1 Model Parameters

The compartments of the model are shown below.

It is a traditional SIR model, with a few additional modifications. One, it is further compartmentalized by age (where young refers to  $<60$  and old refers to  $\geq 60$ ), so both the young and old age compartments have the given model structure. Two, there is a mortality compartment (M). Finally, the infectious compartment is subdivided into mild and severe. Severe infection is defined as infection that requires hospitalization, so it serves as an approximation to the number of COVID-19 hospitalizations. It is important to note that severe infectives do not infected others in the model. The values and interpretations of the different model parameters are given on the next page.

## Parameters

Name	Represents	Value
N	Total population of Suffolk county	807252
$\beta_{yy}$	Transmissibility parameter between young individuals	0.304
$\beta_{oy}$	Transmissibility parameter of old individuals infecting young individuals	0.276
$\beta_{oo}$	Transmissibility parameter between old individuals	0.262
$\beta_{yo}$	Transmissibility parameter of young individuals infecting old individuals	0.276
$\gamma_{mild,young}$	Recovery parameter for mild young infectious	1/7
$a_{young}$	Controls rate that young severe infections occur	0.0065
$\gamma_{severe,young}$	Recovery parameter for severe young infectious	1/14.5
$\mu_{young}$	Mortality parameter for young individuals	1/400
$\gamma_{mild,old}$	Recovery parameter for mild old infectious	1/8
$a_{old}$	Controls rate that old severe infections occur	0.09
$\gamma_{severe,old}$	Recovery parameter for severe old infectious	1/18
$\mu_{old}$	Mortality parameter for old individuals	7/200

These parameters were obtained through a combination of existing data (particularly [11]; Suffolk county data obtained from the United States Census Bureau). For example, the recovery and death parameters were obtained by adding the length of hospital stays to the average recovery length for infections. The parameters  $a_{young}$  and  $a_{old}$  had to be estimated from the data. Different values were plotted against the number of severe infections in Suffolk county over the course of the pandemic, and the values that best fit the data are the ones used here.

The  $\beta$  values were estimated from  $R_0$ . The expression for  $R_0$  is derived in the usual way:

$$\dot{I} = \dot{I}_{mild,old} + \dot{I}_{mild,young} + \dot{I}_{severe,old} + \dot{I}_{severe,young} = 0$$

At  $t = 0$ , we have  $S_{young}/N \approx 0.831$ , which is the fraction of Suffolk county under 60 years of age. Similarly,  $S_{old}/N \approx 0.169$ , which is the fraction of Suffolk county over 60 years of age. The number of infected individuals at  $t = 0$  is scaled to reflect the demography, i.e.  $I_{young}(0) = I_{mild,young}(0) = 0.831$  and  $I_{old}(0) = I_{mild,old}(0) = 0.169$ . We will refer to the fraction of young individuals as  $f_y$  and the fraction of old individuals as  $f_o$ . We obtain

$$\frac{\beta_{yy}f_y^2 + \beta_{oy}f_of_y + \beta_{yo}f_of_y + \beta_{oo}f_o^2}{\gamma_{mild,old}f_o + \gamma_{mild,young}f_y} = R_0$$

Note that we recover  $R_0$  of the traditional SIR model if we let  $f_y \rightarrow 1, f_o \rightarrow 0, \beta_{yy} \rightarrow \beta, \gamma_{mild,young} \rightarrow \gamma$ . If we specify a value of  $R_0$  for each of the four scenarios (i.e. young infecting old, old infecting young, young infecting young, old infecting old), we can recover the  $\beta$  value for that scenario. For example, if we have a value for  $R_{0,yy}$ , the  $R_0$  for young infecting young, then, following the analysis in [6], we can obtain  $\beta_{yy}$  using

$$\frac{\beta_{yy}}{\gamma_{mild,old}f_o + \gamma_{mild,young}f_y} = R_{0,yy}$$

## 4 Modeling Interventions

The interventions were modeled using the technique presented in [9]. Stay-at-home orders are considered the primary intervention, and they are modeled by exponentially decreasing the value of  $\beta$ . In other words, for  $t > T$ , where  $T$  is the time of the stay-at-home intervention, we have

$$\beta = \beta_0 e^{-k(t-T)}, \quad k > 0$$

The same value of  $k$  is used to transform each  $\beta$  parameter in the present model.

The lifting of interventions is modeled in the same fashion, except as an increasing exponential. In other words,

$$\beta = \beta_0 e^{h(t-T)}, \quad k > h > 0$$

As we will see during the sensitivity analysis, modeling the interventions in this way makes them very sensitive to the parameters  $h$  and  $k$  (which are set at 0.01 and 0.03, respectively).

## 5 Results

The timescale of these results ranges from 24 January 2020, one week before the first infection in Suffolk county was reported<sup>1</sup>, to 11 August 2020 (200 days after 24 January). The cut-off point of 11 August is somewhat arbitrary, and of course results ranging that far into the future are to be viewed with suspicion accordingly.

### 5.1 Scenario 1: Full Reopening

First we consider the impact of a full reopening, i.e. a complete reversal of the previous stay-at-home order where all individuals are permitted to go about their daily lives with no interruption or precaution. Such a scenario is modeled by reverting back to the original  $\beta$  values as soon as the stay-at-home order is lifted. The stay-at-home order for Massachusetts is expected to be relaxed on 18 May 2020 at the time of writing [12]. The blue dotted lines represent the time of interventions: the first one represents the intervention going into place, and the second one represents the intervention being relaxed.

We are primarily interested in the healthcare burden on Suffolk county in each scenario. There are 6,800 hospital beds (the figure includes both regular and ICU beds) in Suffolk county (the horizontal black line is drawn accordingly, to indicate that 6,800 is the nominal county healthcare capacity) [13]. Note that this number is variable, and the number of hospital beds can be lower or higher than this number depending on local circumstances. Finally, proper care of COVID-19 patients includes isolation of patients, so a typical hospital set-up is insufficient; the 6,800 number should be viewed as a worst-case upper bound.

---

<sup>1</sup>This is just an estimate for the actual epidemic start time; it is nearly impossible to tell with certainty when the first person infected with COVID-19 came to Suffolk county.

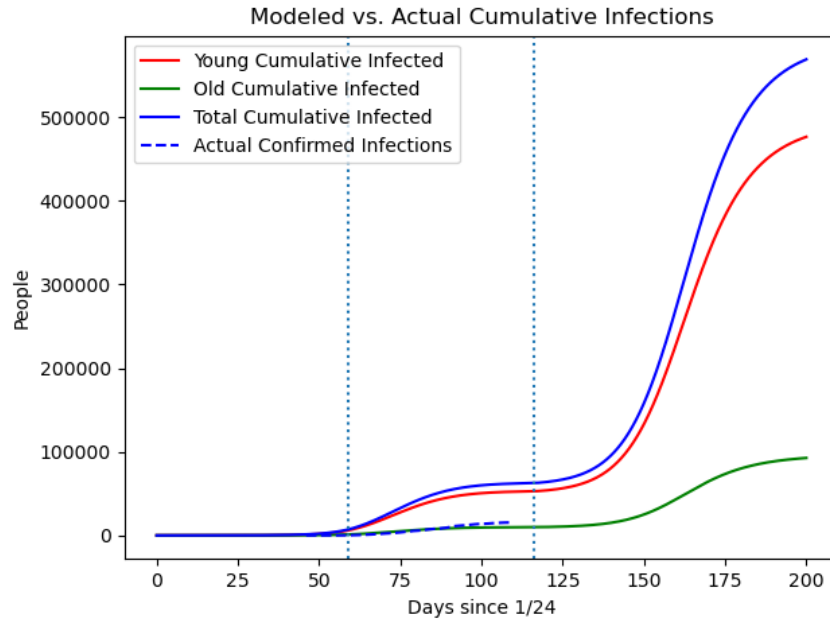


Figure 2: Modeled cumulative case counts during a full reopening.

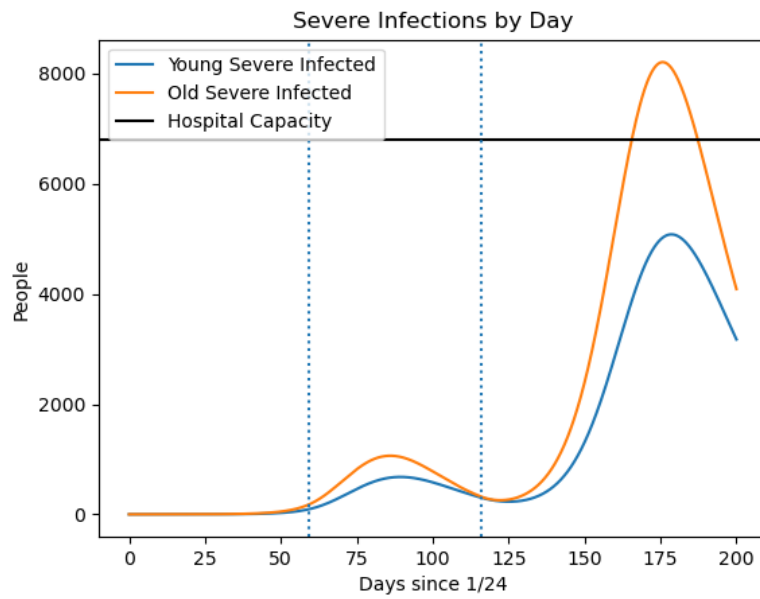


Figure 3: Modeled cumulative case counts during a full reopening.



## 5.2 Scenario 2: Phased Reopening

Another relaxation scenario, and in fact the one that Massachusetts is planning for [12], is a true “relaxation”, where individuals return to normal life gradually, over the course of months. As above, the dotted lines represent the date of the intervention changes, and the black horizontal line in the second plot represents the nominal county healthcare capacity.

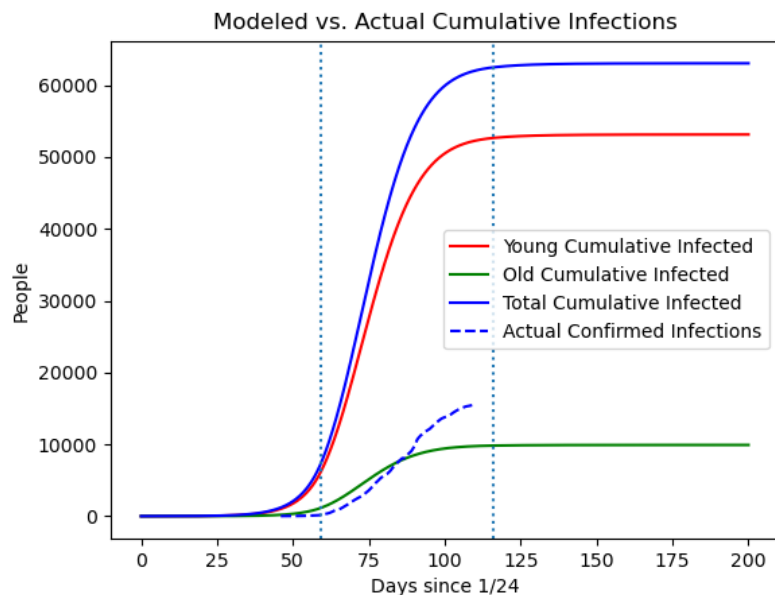


Figure 4: Modeled cumulative case counts during a full reopening.

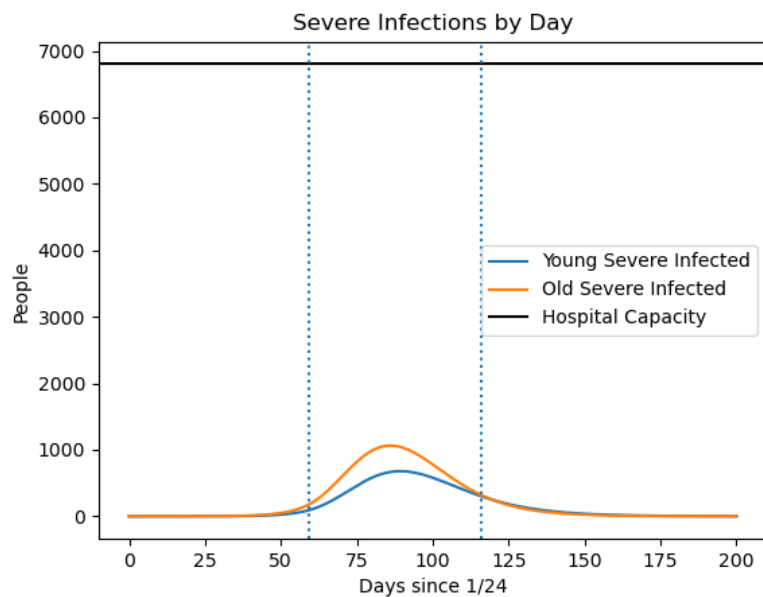


Figure 5: Modeled cumulative case counts during a full reopening.

## 6 Sensitivity Analyses

There are several parameters in the model that are highly sensitive and produce very different results in changed, even if the deviation is small. Of particular concern are the parameters  $k$  and  $h$ , which control the severity of intervention measures; the value of  $R_0$ ; and the values of  $a_{young}$  and  $a_{old}$ , which control the rate of young and old severe infections respectively. Following are plots that display the sensitivity of the overall number of cases to these parameters.

**Sensitivity of the intervention parameters  $k$  and  $h$  to cumulative infections**

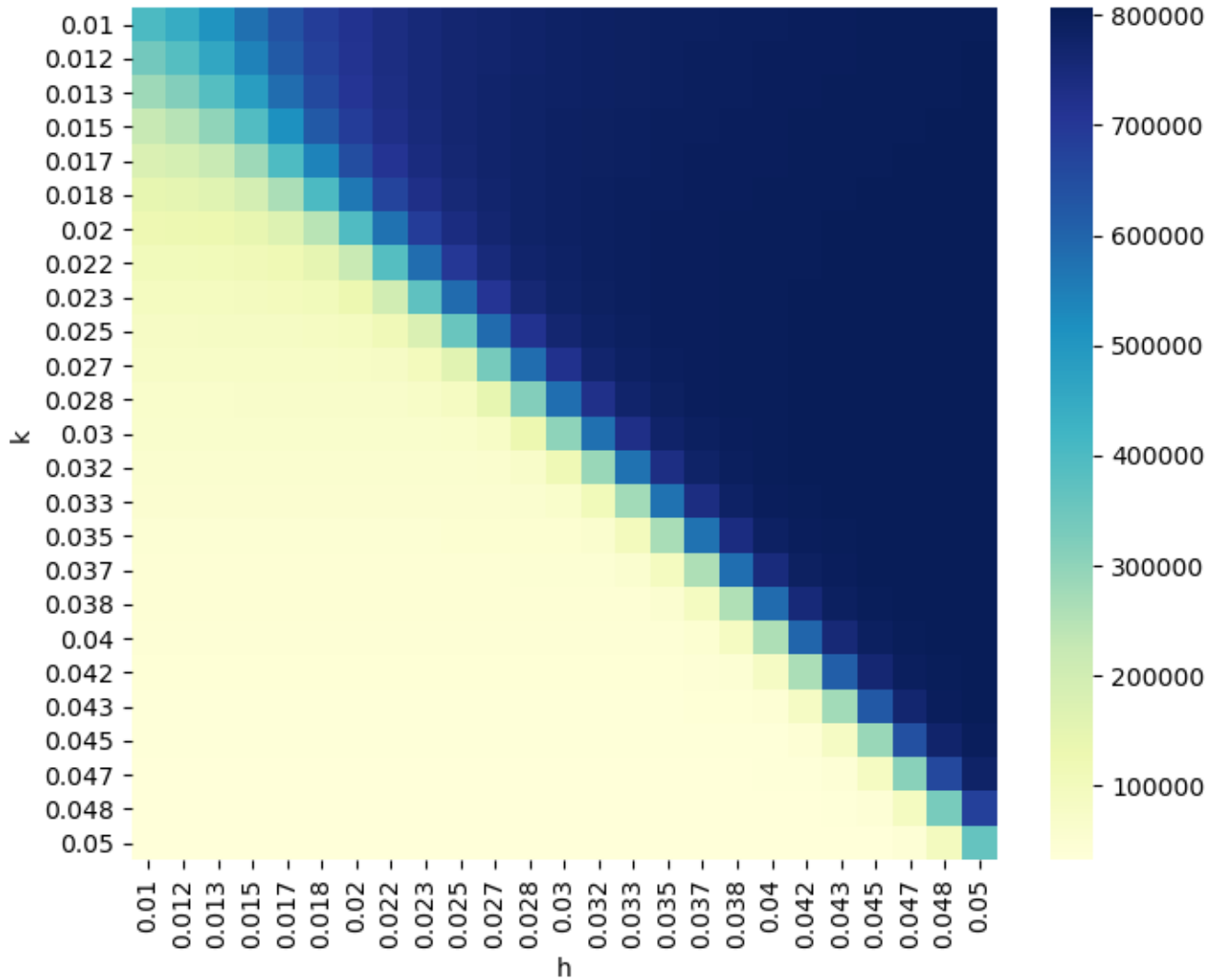


Figure 6: Sensitivity analysis for  $k$  and  $h$ .

# Sensitivity of the infection severity parameters $a_{young}$ and $a_{old}$ to cumulative infections

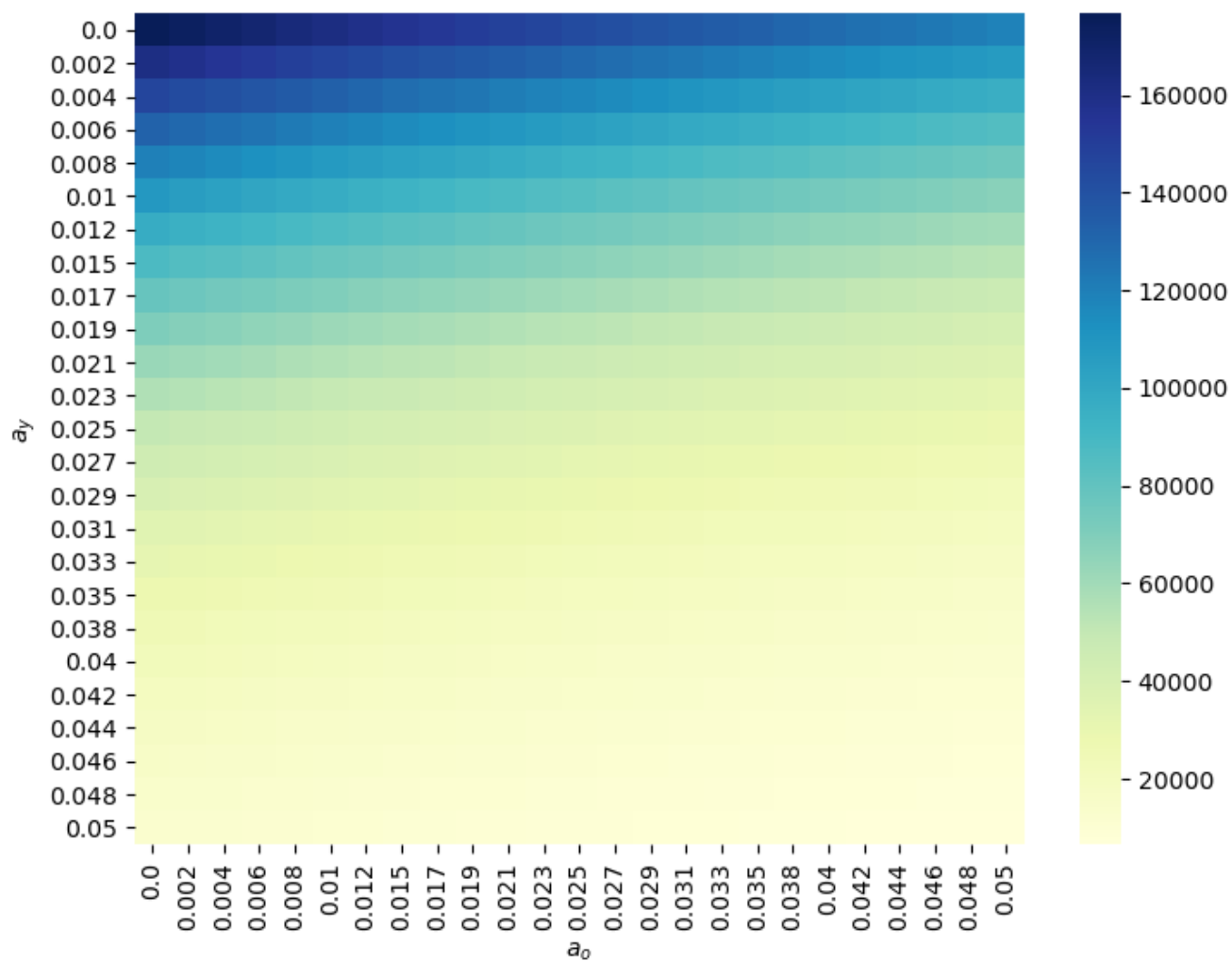


Figure 7: Sensitivity analysis for  $a_{young}$  and  $a_{old}$  with respect to cumulative infections.

# Sensitivity of the infection severity parameters $a_{young}$ and $a_{old}$ to peak severe infections

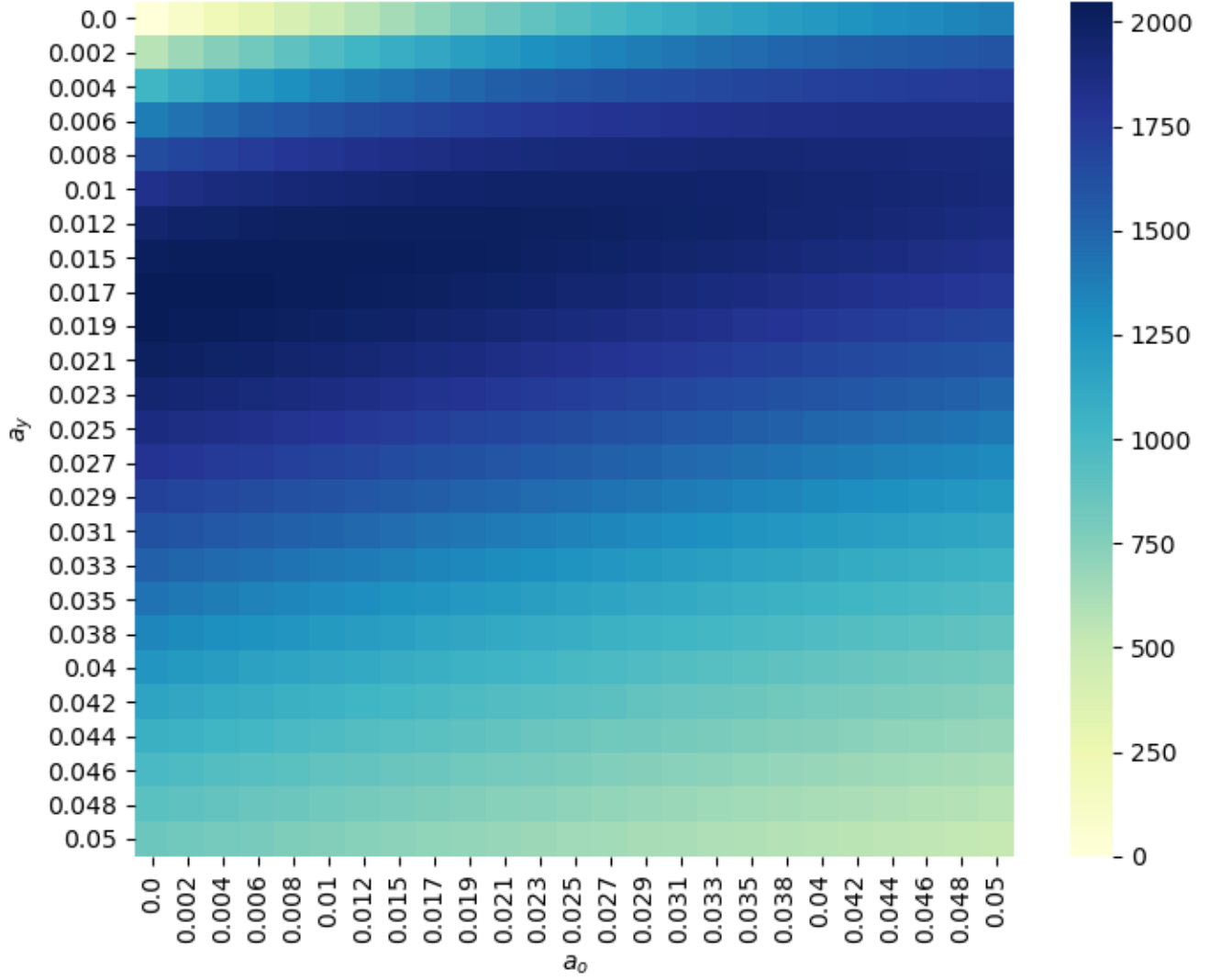


Figure 8: Sensitivity analysis for  $a_{young}$  and  $a_{old}$  with respect to peak severe infections.

## Sensitivity of $R_0$

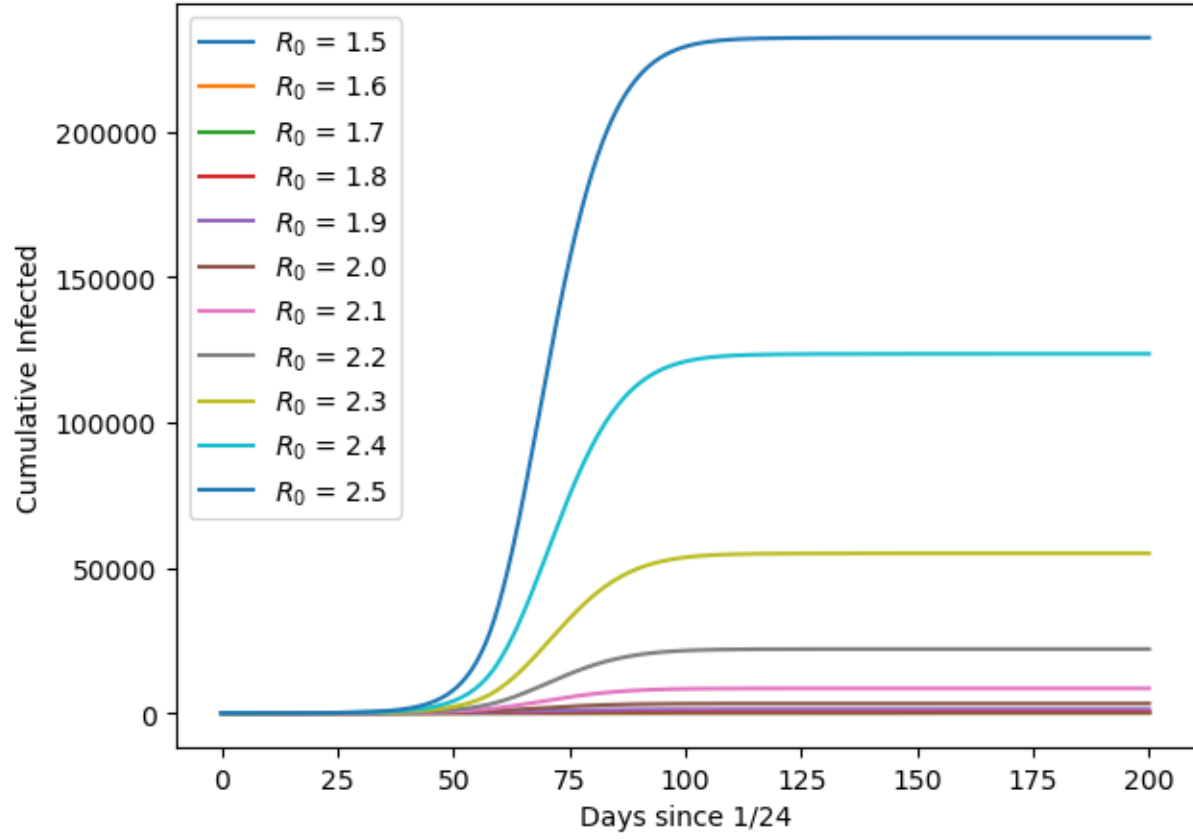


Figure 9: Sensitivity analysis for  $R_0$ . Note that this analysis uses the parameters from the phased reopening scenario.

## 7 Discussion and Limitations

The results from Section 4.1 show quantitatively that a premature complete stay-at-home reversal would simply lead to the large wave of cases we would experience if we had no lockdown in the first place, and place an unrecoverable burden on the healthcare system. But it is slightly more nuanced than that. Importantly, the model estimates the *total* number of cases, not the confirmed number, which is almost certainly an underestimate. Therefore, the model shows that the small fraction of the population infected so far is completely insufficient to guarantee even a little herd immunity. The percentage of the population necessary to sustain herd immunity is a goal of future research.

A phased reopening, as modeled here, has great potential to relieve pressure on the healthcare system. As long as it proceeds slowly enough (corresponding to lower values of the parameter  $h$ ), the model predicts that another spike in cases is unlikely. Unfortunately, the model does not reveal *how* a phased reopening should proceed — i.e. how many businesses should remain open at any time, and with what capacity; when schools should reopen; when social distancing should end; and so forth.

Of course, the model contains a number of limitations, some of them so severe that they prevent the model from making determinations accurate enough to inform policy. The less potent limitations are the ones that all SIR models share: the weakness of the population homogeneity assumption (despite having two age groups, each age group suffers from the consequences of this assumption), the assumed distribution of infectivity (it is not clear what distribution COVID-19 infectivity satisfies), a non-mechanistic description of disease progression (from mild infection to severe infection to mortality), and several other issues. However, the greatest limitation to these results is the highly sensitive nature of the parameters, especially  $h$  and  $k$ . It is not surprising that this should be the case, since those two numbers represent the whole of the model’s representation of COVID-19 response policy. Even a small perturbation in their values produces a massive shift in the total number of cases, and consequently, the predicted healthcare burden.

The severe specificity that these parameters require is hampered by the fact that there is no clear way to obtain them from data. In fact, several model parameters share this defect, even some that would appear to be easy to obtain. Hospitalization and death data exists, but the way they are counted has changed during the course of the pandemic (for example, in April, Massachusetts changed the way hospitalizations are counted [13]). Moreover, specific data is not available for some areas (specific length of hospital stays could not be located for Suffolk county, so data from Georgia had to be used).

While there is little that can be done about the data issues, an important goal of COVID-19 research should be an accurate modeling of interventions. Such modeling should be robust under a wide variety of parameter choices, and preferably have some mechanistic underpinning. The current study demonstrates that a highly sensitive intervention modeling paradigm can disrupt an otherwise functional model.

## References

- [1] Zunyou Wu and Jennifer M. McGoogan. Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in china. *JAMA*, 323(13):1239–1242, 2020.
- [2] Caitlin Rivers et al. Public health principles for a phased reopening during covid-19: Guidance for governors. Technical report, 2020.
- [3] CDC COVID-19 Response Team. Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019 — united states, february 12–march 28, 2020. *MMWR*, 69(13):382–386, 2020.
- [4] Governor’s Press Office. Governor charlie baker orders all non-essential businesses to cease in person operation, directs the department of public health to issue stay at home advisory for two weeks. Available at <https://www.mass.gov/news/governor-charlie-baker-orders-all-non-essential-businesses-to-cess-in-person-operation> (2020/05/14).
- [5] Simon Mongey David W. Berger, Kyle F. Herkenhoff. An seir infectious disease model with testing and conditional quarantine. *NBER WORKING PAPER SERIES*, 2020.
- [6] Maria Chikina and Wesley Pegden. Fighting covid-19: the heterogeneous transmission thesis. Unpublished Manuscript, 2020.
- [7] Maria Chikina and Wesley Pegden. Modeling strict age-targeted mitigation strategies for covid-19. Unpublished Manuscript, 2020.
- [8] Charpignon et al Wilder. Modeling the impact of demography on covid-19 dynamics in hubei and lombardy. Unpublished Manuscript, 2020.
- [9] Glenn F Webb Zhihua Liu, Pierre Magal. Predicting the number of reported and unreported cases for the covid-19 epidemics in china, south korea, italy, france, germany and united kingdom. Unpublished Manuscript, 2020.
- [10] Tom Kalvari Ofir Reich, Guy Shalev. Modeling covid-19 on a network: super-spreaders, testing and containment. Unpublished Manuscript, 2020.
- [11] Szablewski CM et al Gold JA, Wong KK. Characteristics and clinical outcomes of adult patients hospitalized with covid-19 — georgia, march 2020. *MMWR*, 69(18):545–550, 2020.
- [12] Staff and wire reports. Massachusetts’ 4-phase reopening plan, explained. Available at <https://www.nbcboston.com/news/local/massachusetts-4-phase-reopening-plan-explained/2122810/> (2020/05/14).
- [13] Massachusetts Department of Public Health. Massachusetts department of public health covid-19 dashboard. Technical report, 2020.