

## **FEATURE-MODIFIED SEIR MODEL FOR PANDEMIC SIMULATION AND EVALUATION OF INTERVENTION APPROACHES**

Yingze Hou  
Hoda Bidkhor

University of Pittsburgh  
Department of Industrial Engineering  
Benedum Hall 3700 O'Hara St  
Pittsburgh, PA 15261, USA

### **ABSTRACT**

SEIR (susceptible-exposed-infected-recovered) model has been widely used to study infectious disease dynamics. For instance, there have been many applications of SEIR analyzing the spread of COVID to provide suggestions on pandemic/epidemic interventions. Nonetheless, existing models simplify the population, regardless of different demographic features and activities related to the spread of the disease. This paper provides a comprehensive SEIR model to enhance the prediction quality and effectiveness of intervention strategies. The new SEIR model estimates the exposed population via a new approach involving health conditions (sensitivity to disease) and social activity level (contact rate). To validate our model, we compare the estimated infection cases via our model with actual confirmed cases from CDC and the classic SEIR model. We also consider various protocols and strategies to utilize our modified SEIR model on many simulations and evaluate their effectiveness.

### **1 INTRODUCTION**

SIR and SEIR models have been used to study the dynamics of the epidemic by analyzing the change in population and utilizing parameters representing the nature of disease and population characteristics, as well as the effect of policies to control the spread. The classical SIR model classifies people into four states: susceptible (vulnerable to disease but not carrying virus), infected (symptomatic patient and can spread the virus to others), recovered (recovered from the disease), and dead. The SEIR (susceptible-exposed-infected-recovered) model is an extension of the SIR model that plays a role when there is a non-trivial incubation period; exposed refers to people being exposed to the virus but currently asymptomatic.

Since the burst of COVID pandemic, many studies using the SIR or SEIR models to predict the pandemic/epidemic and provide suggestions for controlling policies (Ellison 2020), (Moein, Nickaeen, Roointan, Borhani, Heidary, Javanmard, Ghaisari, and Gheisari 2021). Many papers studied the time evolution of populations to investigate the effect of parameters in SIR and forecast the spread (Alvarez, Argente, and Lippi 2021), (Cooper, Mondal, and Antonopoulos 2020), (López and Rodo 2021), (Efimov and Ushirobira 2021). These models utilize the population change in different states to predict the future spread of disease. Some evaluated and modified existing models for better accuracy and comprehensiveness (Djidjou-Demasse, Michalakis, Choisy, Sofonea, and Alizon 2020), (Engbert, Rabe, Kliegl, and Reich 2021). But due to the difficulty in parameter estimation, they addressed further improvements for a more effective mechanism.

Existing models suffer from the following issues: homogeneous assumption and difficulty in parameter estimation. Most existing works assume a homogeneous population and do not consider different demographic features that can affect the epidemic. Some papers that extended SEIR to consider different infection

parameters or interactive social levels, but the systematic frameworks and more comprehensive models require more effort for better prediction (Grimm, Mengel, and Schmidt 2021), (Ellison 2020), (Ram and Schaposnik 2021). Moreover, the existing numerical models are subjected to uncertainties in representing the nature of the disease and population features (Ellison 2020), (Ghostine, Gharamti, Hassrouny, and Hoteit 2021), (Rahimi, Gandomi, Asteris, and Chen 2021). One parameter they use requires a large dataset to estimate, and hence its reliability affects the prediction performance. Therefore, a more comprehensive SEIR model should be proposed to consider various population characteristics. Furthermore, to have a reliable estimation of the pandemic, easier-to-estimate parameters shall be considered.

**Our contribution** To overcome these challenges brought by homogeneity in population and parameter estimation, we established a feature-modified SEIR model. Our model uses a new approach to estimate infection and allows heterogeneity. Combining different contact rates and sensitivities to disease, we propose two versions of the feature-modified SEIR model, using contact rate and sensitivity to differentiate people and allow for more reliable parameter estimations. We provide a comparison with actual confirmed cases from CDC and the classic SEIR model to validate our model. A comparison with previous SEIR and real-world data validates our model. Next, using a wide range of information in simulation, our model evaluates epidemic control through different regulations. To illustrate the qualitative and quantitative impact of changing contact rate, sensitivity, and proportions, our numerical experiments assess various protocols for decreasing contact rate and sensitivity on alleviating the severeness of pandemic. acemoglu2020testingIt is worth mentioning that our model can also apply to other infectious diseases.

## 2 PRELIMINARY

The SEIR model is an extension of SIR when there is a non-trivial incubation period. It is used to describe the dynamics of infectious diseases by dividing the population into the following different states, Susceptible, Exposed, Infected, Recovered, Cured, and Death:

- Susceptible ( $S$ ), uninfected but vulnerable individuals who never encounter or do not carry the virus.
- Exposed ( $E$ ), infected but asymptomatic people who carry the virus and can infect others.
- Infected ( $I$ ), symptomatic patients from state  $E$ .
- Recovered ( $R$ ), people fully recovered from state  $E$ .
- Cured ( $C$ ), people previously infected but recover from the disease.
- Dead ( $D$ ), infected people who die due to the disease.

The total population ( $N$ ) is the sum of the population in all states except for  $D$ . Figure 1 shows the dynamics of the SEIR model, with different rates of change from one state to another.

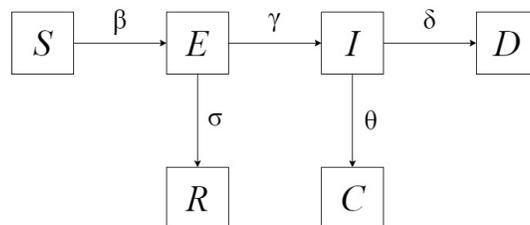


Figure 1: The SEIR model flow diagram.

This single direction flow assumes non-reinfection and non-infectiousness of  $R$  and  $C$ . Each rate of change has practical meaning and estimation:

- Effective contact rate ( $\beta$ ) counts the average number of new infections caused by effective contact, where virus transmission happens between one infectious individual ( $E$  or  $I$ ) and one susceptible.

- Exposed-infected rate ( $\gamma$ ) is the percentage of exposed people developing symptoms, estimated by the incubation period.
- Recovery rate for exposed ( $\sigma$ ) is the percentage of exposed people recovering. It is estimated by the corresponding recovery time.
- Cured rate for infected ( $\theta$ ) is the percentage of infected people recovering. It is estimated by the corresponding recovery time.
- Death rate ( $\delta$ ) is the percentage of infected people who die due to the disease. It is estimated by case fatality rate, the proportion of deaths compared to the total number of people diagnosed with the disease for a particular period.

The following system of equations (1) to (7) summarizes the law of motion for the SEIR model with discrete time. Each term with a subscript  $t$  refers to a state's population at the beginning of the  $t$ -th period.

$$S_{t+1} = S_t - \beta(E_t + I_t)S_t, \tag{1}$$

$$E_{t+1} = E_t + \beta(E_t + I_t)S_t - \sigma E_t - \gamma E_t, \tag{2}$$

$$R_{t+1} = R_t + \sigma E_t, \tag{3}$$

$$I_{t+1} = I_t + \gamma E_t - \theta I_t - \delta I_t, \tag{4}$$

$$C_{t+1} = C_t + \theta I_t, \tag{5}$$

$$D_{t+1} = D_t + \delta I_t, \tag{6}$$

$$N_{t+1} = N_t - \delta I_t. \tag{7}$$

We assume that effective contact happens only between  $S$  and  $E$  in (1) in the rest of our discussion if infected individuals can be isolated.

### 3 FEATURE-MODIFIED SEIR MODEL

The existing SEIR model assumes homogeneous individuals. We generalize the model to consider different social activity levels (contact rate) and health conditions (sensitivity). Moreover, existing models estimate new infections using  $\beta$  in equation (1). This state-wide parameter is not accurate for a smaller population. We modify this estimation in an analogous but more precise way.

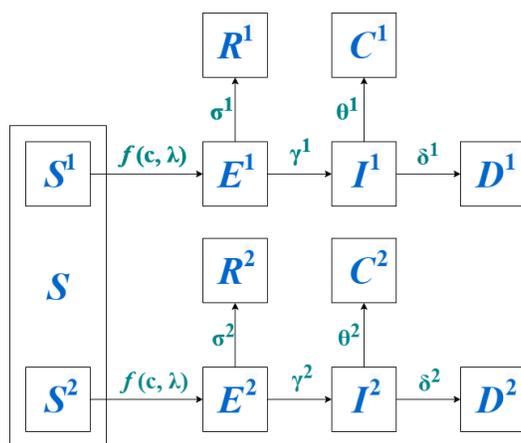


Figure 2: Feature-based (sensitivity) SEIR model with  $M = 2$ .

2020testing We extend the SEIR model by adapting different contact rates and sensitivity (rates of change). Graph 2 shows the dynamics of the sensitivity-modified SEIR model. We assume  $M$  number of

different feature groups. Within each group, their people are assumed to be identical.  $f(c, \lambda)$  represents the rate of change depending on both contact rate  $c$  (or  $c^k$ ) and infection probability  $\lambda$ . We assume uniform  $\lambda$ , since it only depends on the form of contact and irrelevant to sensitivity (Agrawal and Bhardwaj 2021), (Manski and Molinari 2021). Notations are explained in Table 1.

Table 1: Notations for feature-based SEIR model.

Symbols	Description (Population)	Symbols	Description (Feature)
$t$	Index for period starting at $t$	$T$	Terminal time
$S_t$	Susceptible population	$\beta$	average number of new infection per contact (virus-transmission happens)
$E_t$	Exposed population	$\lambda$	Infection probability from $S$ to $E$
$R_t$	Recovered population	$\gamma$	Exposed-infected rate
$I_t$	Infected population	$\sigma$	Recovery rate for exposed
$C_t$	Cured population	$\theta$	Recovery rate for infected
$D_t$	Dead population	$\delta$	Death rate for infected
$X$	General notation for population state $X \in \{S, E, I, R, C, D, N\}$	$s, s^i, s^m$	General notation for sensitivity $s \in \{\lambda, \gamma, \sigma, \theta, \delta\}$
$p^i$	Population proportion of $i$ -th feature group	$c, c^j, c^k$	Contact rate
$X_t^i$	Population of state $X$ with $i$ -th feature at the beginning of $t$ -th period	$i, j, k, m$	Feature index ( $i, m = 1, \dots, M$ ) ( $j, k = 1, \dots, N$ )

**Sensitivity as feature:** First of all, we assume people sharing the same contact rate and distinguish them based on their sensitivity to disease. The population change for susceptible in  $i$ -th sensitivity group regarding contact situations with exposed people of all sensitivities is:

$$S_{t+1}^i - S_t^i = -\lambda \frac{S_t^i}{N_t - I_t} c \sum_{m=1}^M E_t^m, \quad j = 1, \dots, M. \quad (8)$$

To reach (8), we consider the number of contacts between  $S_t^i$  and  $E_t^m$ . The number of contact happening is:

$$C(S_t^i, E_t^m) = E_t^m \cdot c \cdot P(E_t^m \text{ meet } S_t^i) = \frac{S_t^i}{N_t - I_t} \cdot c \cdot E_t^m. \quad (9)$$

$E_t^m$  is exposed population in  $m$ -th group, and  $c$  is the average number of people met by an exposed person for a given period of time. Its value can be estimated via social network simulation (Zhou, Sornette, Hill, and Dunbar 2005), (Del Valle, Hyman, Hethcote, and Eubank 2007), (Van de Kasstele, van Eijkeren, and Wallinga 2017). A  $c$  value of 15 means that on average, one person contact 15 other people closely.  $cE_t^m$  counts all people contacted by  $E_t^m$ . Among them, approximately  $P(E_t^m \text{ meet } S_t^i)$  is  $S_t^i$ . This probability is estimated by the proportion of  $S_t^i$  among all people active in social contact ( $N_t - I_t$ ). Same idea is used in (Ram and Schaposnik 2021).

Thus, we have the new exposed people caused by such effective contact, which is the number of contact happening multiplied by infection probability  $\lambda$ :

$$S_{t+1}^i - S_t^i = - \sum_{m=1}^M \lambda \cdot C(S_t^i, E_t^m). \quad (10)$$

A negative sign is due to people leaving state  $S$ .  $\lambda$  is the infection probability measuring the possibility of virus-transmission when  $E$  and  $S$  meet each other (Agrawal and Bhardwaj 2021). (10) estimates new exposed in an analogous way to (1). In (1),  $E_t S_t$  estimates all possible contacts ( $I_t$  is ignored by isolation assumption).  $\beta \cdot E_t S_t$  gives the new exposed people who get virus-transmitted from effective contact.

Similarly, (10) estimates the contact number by  $C(S_t^i, E_t^m)$ .  $\lambda$  plays a similar role to  $\beta$ , since they both count the average number of effective contact (virus transmitted) happening. The summation over sensitivity index  $m$  counts the change in  $S_t^i$  caused by contacts with exposed people having different sensitivities. The relation between  $E_t^m$  and  $E_t$  is:

$$X_t = \sum_{m=1}^M X_t^m, \quad X \in \{S, E, I, R, C, D, N\}.$$

Corresponding changes are made to exposed state, while other changes remain the same as existing SEIR model, except for replacing  $X_t$  by  $X_t^i$  and sensitivity  $s$  by  $s^i$ :

$$\begin{aligned} E_{t+1}^i &= E_t^i + \lambda \frac{S_t^i}{N_t - I_t} \sum_{m=1}^M c E_t^m - \sigma^i E_t^i - \gamma^i E_t^i, \\ R_{t+1}^i &= R_t^i + \sigma^i E_t^i, \\ I_{t+1}^i &= I_t^i + \gamma^i E_t^i - \theta^i I_t^i - \delta^i I_t^i, \\ C_{t+1}^i &= C_t^i + \theta^i I_t^i, \\ D_{t+1}^i &= D_t^i + \delta^i I_t^i, \\ N_{t+1}^i &= N_t^i - \delta^i I_t^i. \end{aligned}$$

**Contact rate as feature:** Now, we consider people with the same sensitivity but having different contact rates. The population change for susceptible in  $j$ -th contact group regarding contact situations between exposed people of all contact rates is:

$$S_{t+1}^j - S_t^j = -\lambda \frac{S_t^j}{N_t - I_t} \sum_{k=1}^N c^k E_t^k, \quad j = 1, \dots, N.$$

Analogous to (9), first, we consider the number of contacts between susceptible  $S_t^j$  and exposed  $E_t^k$ :

$$C(S_t^j, E_t^k) = E_t^k \cdot c^k \cdot P(E_t^k \text{ meet } S_t^j) = \frac{S_t^j}{N_t - I_t} \cdot c^k \cdot E_t^k. \quad (11)$$

Since we allow different contact rates,  $c^k$  is the average number of the person met by an exposed person (with  $k$ -th contact rate) for a given period of time. Similar to (10), we multiply (11) by  $\lambda$  and sum over all  $N$  contact rates to consider contacts between  $S_t^j$  and exposed people of all contact rates:

$$S_{t+1}^j - S_t^j = \sum_{k=1}^N \lambda \cdot C(S_t^j, E_t^k).$$

Corresponding changes are made to exposed state, while other changes remain the same as the existing SEIR model), except for replacing  $X_t$  by  $X_t^j$ :

$$\begin{aligned} E_{t+1}^j &= E_t^j + \lambda \frac{S_t^j}{N_t - I_t} \sum_{k=1}^N c^k E_t^k - \sigma E_t^j - \gamma E_t^j, \\ R_{t+1}^j &= R_t^j + \sigma E_t^j, \\ I_{t+1}^j &= I_t^j + \gamma E_t^j - \theta I_t^j - \delta I_t^j, \\ C_{t+1}^j &= C_t^j + \theta I_t^j, \\ D_{t+1}^j &= D_t^j + \delta I_t^j, \\ N_{t+1}^j &= N_t^j - \delta I_t^j. \end{aligned}$$

## 4 EXPERIMENTS

We utilize the feature-modified SEIR model to conduct several numerical experiments evaluating regulations on contact rate, and provide suggestions on medical approaches to reduce sensitivity to disease. Many of the parameters of our experiments are chosen based on the current COVID-19 pandemic. The effectiveness is measured by the highest infection proportion and death proportion. The highest infection proportion is the infection proportion among the current population, and we take the highest value over time to measure the severity. Death proportion is the cumulative death divided by the total initial population, calculating the proportion of death due to disease.

First of all, we introduce the numerical choice for sensitivity, contact rate, and population. Then, four sets of numerical results are presented: In Section 4.1, we compare the estimated infection cases via our sensitivity-based SEIR with actual confirmed cases from CDC and classic SEIR model to validate our model; in Sections 4.2 and 4.3, we discuss the effect of changing contact rate and sensitivity; in Section 4.4, we discuss the effect of the partial change to contact rate where only the contact rate of high-sensitivity people are adjusted. Results show that while decreasing both contact rate and sensitivity alleviate the spread of disease, changing contact rate brings more improvement than sensitivity.

**Setup:** We divide sensitivity into 2 groups. For instance, exposed-infected rate  $\gamma = (1/5, 1/7)$ , where the first one being more vulnerable. The value of sensitivity parameters is chosen based on the actual length of duration for a particular state (Alvarez, Argente, and Lippi 2021). Table 2 gives the range of choice.

Table 2: Sensitivity parameters and values in SEIR model.

Parameters	Symbols	Value	Description
Infection probability	$\lambda$	0.01~0.2	Probability of infection (only related to form of contact)
Exposed-infected rate	$\gamma$	1/14~1/5	5 to 14 days incubation period
Recovery rate for exposed	$\sigma$	1/14	14 days quarantine period
Recovery rate for infected	$\theta$	1/20~1/10	10 to 20 days to recovery for $I$
Death rate	$\delta$	2%~2.6%	Case fatality rate

Note that the more sensitive a person, the higher risk that person will be affected by the disease, and the higher  $\lambda$  and  $\delta$  are, but the lower  $\sigma$  and  $\theta$  are. The values of sensitivity can also be decided by medial research on infection risk (Manski and Molinari 2021). We set two contact rate groups. Namely,  $c = (c^h, c^l)$ , where high contact rate  $c^h \in \{25, 20, 15, 10\}$  and low contact rate  $c^l \in \{15, 10, 5\}$  (contact rate of first group is always higher than the second group). The numbers are chosen based on the simulation in (Zhou, Sornette, Hill, and Dunbar 2005), (Del Valle, Hyman, Hethcote, and Eubank 2007), (Van de Kastele, van Eijkeren, and Wallinga 2017). Their values can also be decided through regulations on social activity. Proportion of different feature (e.g. high contact and low contact) varies from  $p = (p^h, p^l) = (0.9, 0.1)$  to  $(0.1, 0.9)$  with increment no bigger than 0.1. We use  $p_s$  for sensitivity and  $p_c$  for contact rate. We set the initial exposed population taking up a small proportion of the susceptible population. Other states are zero at the beginning.

### 4.1 Comparison with Actual Confirmed Cases

To show the validity of our model, we compare the new simulated infection cases (using the classical and sensitivity-based SEIR) with the actual confirmed cases in Allegheny county from CDC's dataset (CDC 2022). We select the time window from late April 2020 to June 2021. The starting time is when the population of virus carriers (exposed and infected people) has accumulated to a non-trivial amount (confirmed cases exceeded 1000). The time ends before vaccination is widely distributed.

We set three stages for the spread. Different number of population is assigned to each stage, based on the total confirmed cases and related regulations. The first stage begins in late April 2020, when the spread of COVID was about to start again (new daily confirmed cases started to rise after the time of declining). The second stage begins in the middle of August when the new confirmed cases become stable. The last

stage begins in late October 2020, when the new confirmed cases started to rise again. The beginning time of these stages is referred historical data, but can also be decided based on medical predictions on the next coming wave.

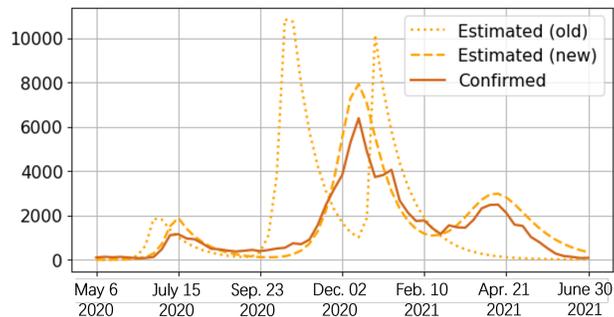


Figure 3: Weekly confirmed and estimated cases of new infection.

The new infection cases are estimated by the product of the exposed population and their corresponding sensitivity ( $\gamma$ ). Compared to the old SEIR estimation (orange), the new model (blue) provides a better estimate for actual confirmed cases (red). The higher infection population in blue is because the confirmed cases only include the reported ones, and underestimated the actual number. The number of total infections can be 3 to 20 times higher than the number of total confirmed infections (Wu, Mertens, Crider, Nguyen, Pokpongkiat, Djajadi, Seth, Hsiang, Colford, Reingold, et al. 2020). The sudden surge in the old SEIR model is caused by Equation (1), which gives a rough estimate of the new exposed population. Thus, its performance is largely affected by the uncertainty in  $\beta$ , resulting in a sudden surge at the very beginning. Even though there might be some over-fitting for the three stages, with given input and predicted starting time, our multi-feature SEIR can predict the spread much more accurately.

#### 4.2 Effects of Changing Contact Rate on Pandemic

As we discussed before, we consider two performance metrics, highest infection proportion and death proportion, and we study the effect of changing contact rate and its proportion on these metrics. We conduct simulations on changing high and low contact rates for  $T = 50$  periods, respectively. Figure 4 and Figure 5 compare the effect of changing low contact rate and high contact rate, respectively. Both figures also consider the change in contact rate proportion. For these experiments, we fix sensitivity as  $\lambda = 0.05$ ,  $\gamma = 1/10$ ,  $\sigma = 1/14$ ,  $\theta = 1/14$ ,  $\delta = 0.025$ , but results are similar with other values. We set the initial populations as  $S_0 = 100000$  and  $E_0 = 50$ .

In Figure 4, horizontal axis is  $p_c^h$ , proportion of high contact rate group (first group with  $c^h$ ), changing from 0.1 to 0.9 with an increment of 0.1. The vertical axis in Figure 4a to Figure 4d denotes the proportion of infection among the current population, and its highest value over time is exhibited. The vertical axis in Figure 4e to Figure 4h denotes the proportion of cumulative death among the total population. Within each subfigure, we hold a constant high contact rate ( $c^h$ ) to see the influence of changing low contact rate ( $c^l$ ).

Figure 4 demonstrates that the lower the contact rate and the lower the proportion of the high-contact-rate group, the lower highest infection and cumulative death are. Moreover, within each figure, each curve declines more for smaller  $p_c^h$ . When low-contact-rate people take the majority, the effect of decreasing their contact rate is more observable. Additionally, in each subfigure, gap between each curve increases when low contact rate is decreased. For instance, in Figure 4a, gap between the curves (25, 10) and (25, 5) is much larger than the gap between (25, 15) and (25, 10). As a result, if possible, it is beneficial further to decrease the low contact rate through social distancing protocols. Results in terms of cumulative death are analogous.

Utilizing the same data, Figure 5 focuses on the effect of changing the high contact rate and its proportion. Horizontal axis is  $p_c^h$ , the proportion of high-contact-rate group. Vertical axis in Figure 5a to

Figure 5c is the highest value over the time of infection proportion among the current population. Vertical axis in Figure 5d to Figure 5f is the proportion of cumulative death among the total population. Sensitivities are the same as Figure 4. Within each subfigure, we only change the high contact rate ( $c^h$ ) to see its influence. Note that values on the vertical axis are not the same, due to the small variation in percentage.

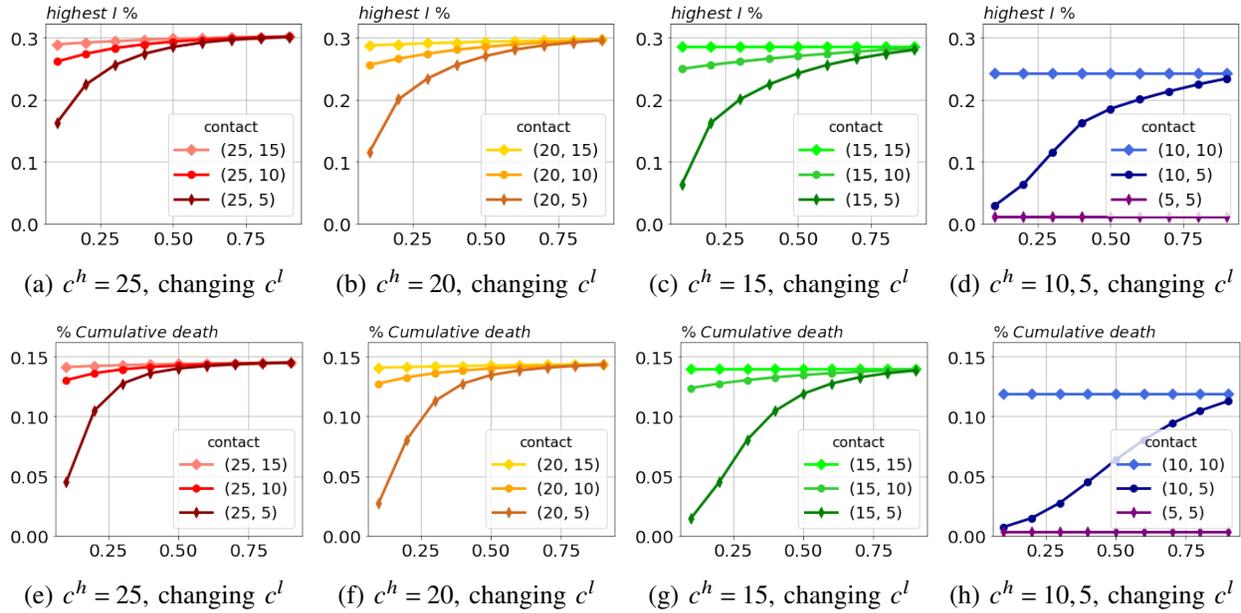


Figure 4: Highest infection proportion and death proportion with decreasing low contact rate ( $c^l$ ).

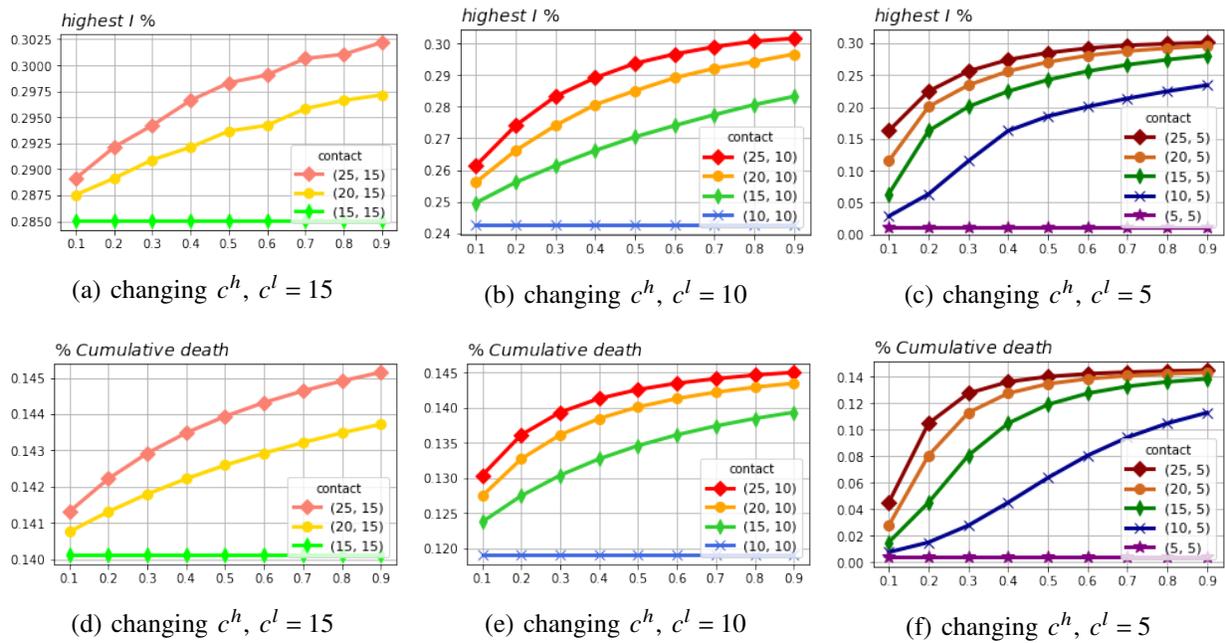


Figure 5: Highest infection proportion and death proportion with decreasing high contact rate ( $c^h$ ).

Comparing each curve in subfigures of Figure 5, for given contact-rate proportion, we observe that decrease in the highest infection proportion is more when the high contact rate is further decreased. For example, in Figure 5b, gap between the curves labeled as (25, 10) and (20, 10) is smaller than the gap between (20, 10) and (15, 10). Additionally, in Figure 5a and 5d, where population have a higher contact rate than other situations, gap between each curve increases when  $p_c^h$  increases. So, the effect of decreasing the high contact rate is obvious when these people take the majority. However, it is not always the case when high contact rate is much higher than low contact rate. Gap between (25, 10) and (20, 10) in Figure 5b, as well as the gap between (25, 5), (20, 5), and (15, 5) in Figure 5c, is not always increasing when  $p_c^h$  increases. Thus, a small decrease in the high contact rate does not bring a significant improvement in these situations. Results are similar using the cumulative death performance metric in (5d) to (5f).

### 4.3 Effects of Changing Sensitivity on Pandemic

To compare the effect of changing on the pandemic, we conduct simulations on changing high and low sensitivity with  $T = 50$ , respectively. Figure 6 and Figure 7 compare the effect of changing low-risk sensitivity and high-risk sensitivity, respectively. Both figures also consider the change of sensitivity proportion. We fix contact as  $c = 15$ , but results are similar with other values. Populations are  $S_0 = 100000$ ,  $E_0 = 50$ . Although it is hard to change sensitivity, such comparison is still practical. When a virus has its variants or a vaccine is developed, sensitivity can be affected.

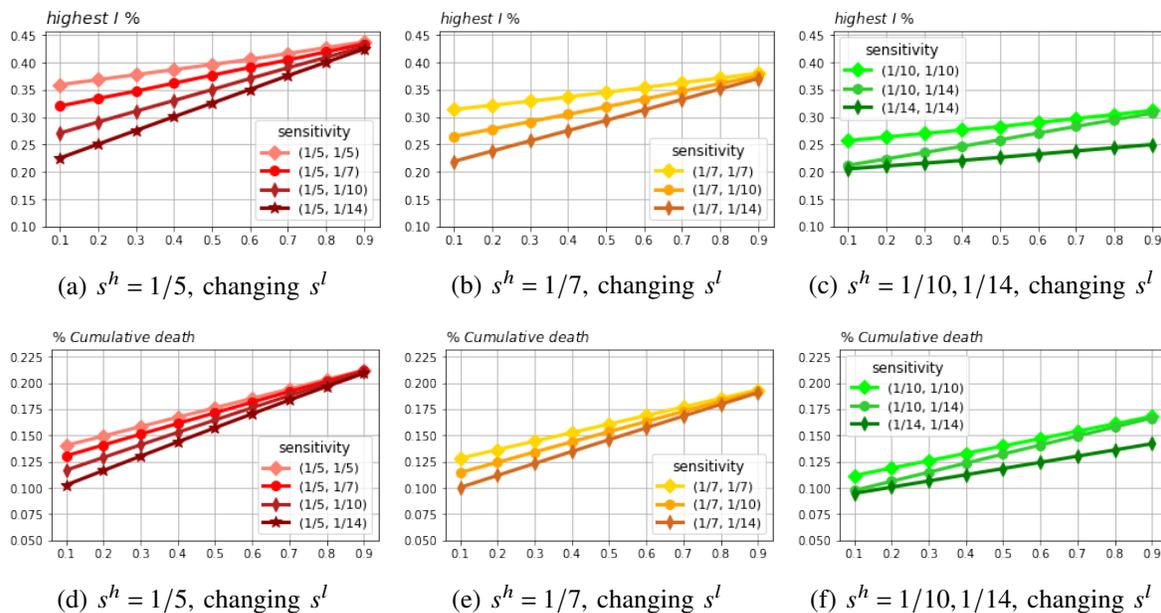


Figure 6: Highest infection proportion and death proportion with decreasing low-risk sensitivity ( $s^l$ ).

In Figure 6, horizontal axis is  $p_s^h$ , the proportion of high-risk sensitivity group. Vertical axis in Figure 6a to Figure 6c is the proportion of infection among the current population, and its highest value over time is exhibited. Vertical axis in Figure 6d to Figure 6f is the proportion of cumulative death among the total population. Within each subfigure, we only change the low-risk sensitivity to see its effect. Unlike contact rate, discussing the change of sensitivity is quite hard, for their being difficult to measure. In our simulation, we only change the exposed-infected rate, for this incubation period being easily observed. Other sensitivities for two groups are constant:  $\lambda = (0.05, 0.05)$ ,  $\sigma = (1/14, 1/14)$ ,  $\theta = (1/20, 1/10)$ ,  $\delta = (0.026, 0.023)$ .

Figure 6 gives observations that are quite different from changing low contact rate. When there are more low-sensitivity people (lower  $p_s^h$ ), pandemic loss decreases quite linearly, unlike the expedited decline

in Figure 4. But the gap between each curve increases when we decrease  $p_s^h$ . So, such decreasing effect is more observable when these people take up the majority. Additionally, for given sensitivity proportion, gaps between curves do not change significantly when we decrease the low-risk sensitivity. For instance, in Figure 6a, gap between (1/5, 1/5) and (1/5, 1/7) is quite the same as gap between (1/5, 1/7) and (1/5, 1/10). Hence, the influence of decreasing the low-risk sensitivity (holding high-risk unchanged) does not fluctuate much under given sensitivity proportion.

Utilizing the same data, Figure 7 focuses on the effect of changing high-risk sensitivity and its proportion. Horizontal axis is the proportion of high sensitivity group ( $p_s^h$ ), changing from 0.1 to 0.9. The vertical axis in Figure 7a to Figure 7c is the proportion of infection among the current population, and its highest value over time is exhibited. The vertical axis in Figure 7d to Figure 7f is the proportion of cumulative death among the total population. For the same reason discussed before Figure 6, we only change the exposed-infected rate, while other sensitivities and contact rate are the same as in Figure 6.

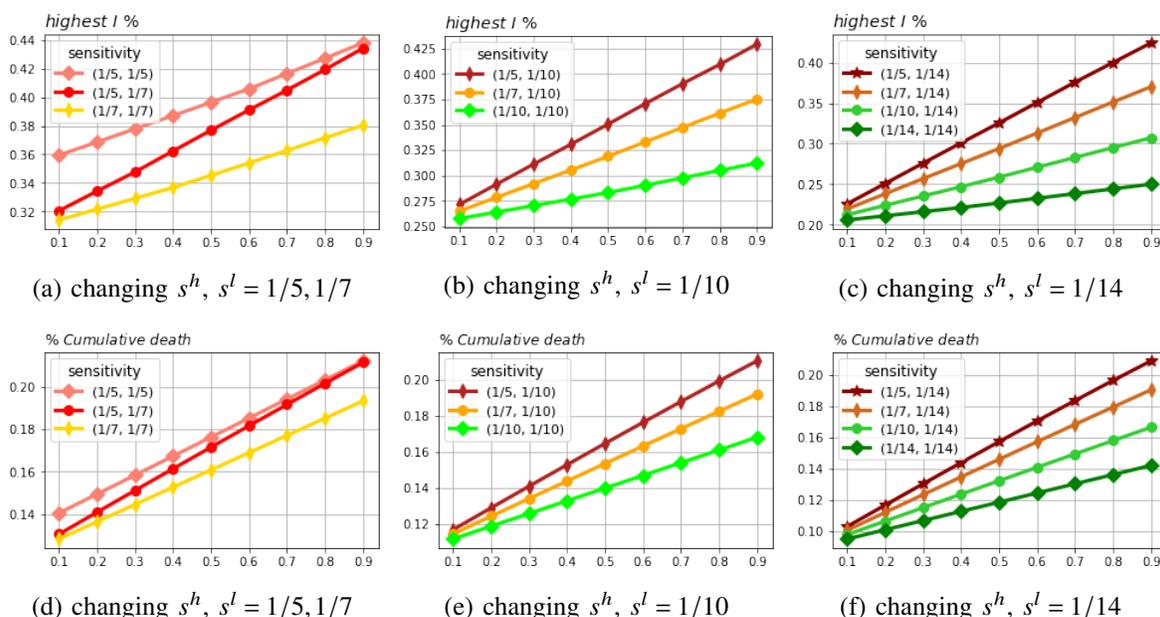


Figure 7: Highest infection proportion and death proportion with decreasing high-risk sensitivity ( $s^h$ ).

Figure 7 demonstrates the effects of changing high-risk sensitivity being similar to changing the low-risk one. When we increase  $p_s^h$ , pandemic loss increases quite linearly, and the gap between each colored curve is also amplified. Thus, decreasing high-risk sensitivity brings a more obvious effect when these people take up more proportion of the population. Besides, within each subfigure, all gaps between every two curves are mostly the same for given sensitivity proportion. For instance, in Figure 6f, gap between sensitivity situation (1/5, 1/14) and (1/7, 1/14) is almost the same as the gap between (1/7, 1/14) and (1/10, 1/14). So, the influence of such a decrease (holding low sensitivity unchanged) does not fluctuate a lot.

#### 4.4 Effect of Changing Contact Rate Based on Sensitivity on Pandemic

In previous sections, we discuss the effect of reducing pandemic loss with changing contact rates or sensitivity of all people. In this section, we analyze the effect of partial change on contact rate, where only the contact rate of high-risk sensitivity people is adjusted. We omit the discussion of a partial change in sensitivity due to the difficulty of estimating the sensitivity parameters, which may lead to biased predictions.

We use sensitivity as the feature in our modified SEIR, which makes all people having the same contact rate initially. A decreased contact rate for high-risk sensitivity people is implemented to see how the

situation is improved compared to the one without such change. In Figure 8, for different initial contact rates, we show the improvement, reduction in highest infection, and cumulative death between situations with and without the change under several sensitivity (exposed-infected rate) inputs.

In Figure 8, the difference in infection and death between situations with and without halving contact of high-risk sensitivity people is measured. Horizontal axis is  $p_s^h$ , proportion of high-risk sensitivity group, increasing from 0 to 1 with an increment of 0.025. Vertical axis is the uniform initial contact rate, from 1 to 25. For each situation, we use a different exposed-infected rate. Other sensitivities are  $\lambda = (0.05, 0.05)$ ,  $\sigma^E = (1/14, 1/7)$ ,  $\sigma^I = (1/20, 1/10)$ ,  $\delta = (0.026, 0.023)$ .  $S_0 = 100000$ ,  $E_0 = 50$ , and  $T = 50$ .

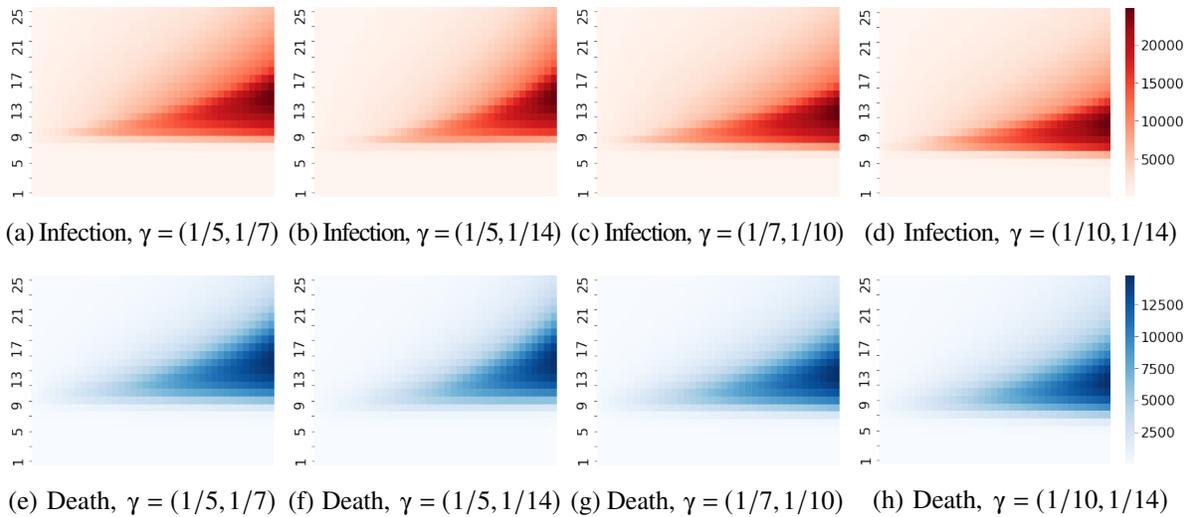


Figure 8: Reduced highest infection and cumulative death by halving contact rate of high-risk sensitivity.

In Figure 8, when  $p_s^h$  is increasing, the reduction in highest infection is increasing, meaning that more people are saved from getting infected. This shows that when a reduced contact rate is applied to high-risk sensitivity people, and if they take more proportion, improvement of such change is more significant. In addition, when general contact rate for all people is lower, such change is more beneficial. Nonetheless, when contact rate is initially low (e.g.  $c \leq 10$ ), pandemic is not severe, since virus does not spread quickly in these situations. As a result, significant improvement is not observed, mainly because of the low population of infected people. Besides, comparing Figure 8d to Figure 8a, the sensitivity turns from low-risk to high-risk, and a slightly larger dark region is observed, meaning that more people are saved. Results in terms of death are consistent with the results for highest infection.

## 5 CONCLUSIONS & FUTURE RESEARCH

We establish a feature-modified SEIR model to study the time evolution of infectious diseases for heterogeneous populations. Our model allows heterogeneity among people and improves the estimation of the increase in virus-transmitted people in the original SEIR model. The comparison with the actual confirmed cases validates the effectiveness of our proposed model.

Our numerical studies simulate the COVID-19 pandemic under different situations to examine the trend of each population change and evaluate the effectiveness of contact regulation and medical approaches to decreasing sensitivity to disease. Results show the effectiveness of lower social activity level (contact rate) and infection risk (sensitivity to disease) in alleviating the pandemic. Additionally, decreasing contact rate reduces more infection and death than sensitivity. The model can inform regulators how and how much contact rate and sensitivity should be reduced to control the pandemic to the desired level. In summary, utilizing our model, we can easily evaluate the impact of certain changes, including reducing social contacts,

and the variant of virus affecting infection risk. This provides insights on the interaction of social activity level and sensitivity to disease in controlling the spread.

In the future, we plan to develop more comprehensive SEIR models. Meanwhile, other intervention approaches, such as vaccination, should be considered and equipped inside the SEIR model. [acemoglu2020testing](#)

## REFERENCES

- Agrawal, A., and R. Bhardwaj. 2021. “Probability of COVID-19 infection by cough of a normal person and a super-spreader”. *Physics of Fluids* 33(3):031704.
- Alvarez, F., D. Argente, and F. Lippi. 2021. “A simple planning problem for COVID-19 lock-down, testing, and tracing”. *American Economic Review: Insights* 3(3):367–82.
- CDC 2022. “COVID Data Tracker. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention”. <https://covid.cdc.gov/covid-data-tracker>, accessed on June 10<sup>th</sup> 2022.
- Cooper, I., A. Mondal, and C. G. Antonopoulos. 2020. “A SIR model assumption for the spread of COVID-19 in different communities”. *Chaos, Solitons & Fractals* 139:110057.
- Del Valle, S. Y., J. M. Hyman, H. W. Hethcote, and S. G. Eubank. 2007. “Mixing patterns between age groups in social networks”. *Social Networks* 29(4):539–554.
- Djidjou-Demasse, R., Y. Michalakakis, M. Choisy, M. T. Sofonea, and S. Alizon. 2020. “Optimal COVID-19 epidemic control until vaccine deployment”. *Cold Spring Harbor Laboratory Press*.
- Efimov, D., and R. Ushirobira. 2021. “On an interval prediction of COVID-19 development based on a SEIR epidemic model”. *Annual reviews in control* 51:477–487.
- Ellison, G. 2020. “Implications of heterogeneous SIR models for analyses of COVID-19”. *National Bureau of Economic Research*.
- Engbert, R., M. M. Rabe, R. Kliegl, and S. Reich. 2021. “Sequential data assimilation of the stochastic SEIR epidemic model for regional COVID-19 dynamics”. *Bulletin of mathematical biology* 83(1):1–16.
- Ghostine, R., M. Gharamti, S. Hassrouny, and I. Hoteit. 2021. “An extended SEIR model with vaccination for forecasting the COVID-19 pandemic in Saudi Arabia using an ensemble Kalman filter”. *Mathematics* 9(6):636.
- Grimm, V., F. Mengel, and M. Schmidt. 2021. “Extensions of the SEIR model for the analysis of tailored social distancing and tracing approaches to cope with COVID-19”. *Scientific Reports* 11(1):1–16.
- López, L., and X. Rodó. 2021. “A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics”. *Results in Physics* 21:103746.
- Manski, C. F., and F. Molinari. 2021. “Estimating the COVID-19 infection rate: Anatomy of an inference problem”. *Journal of Econometrics* 220(1):181 – 192. Themed Issue: Pandemic Econometrics / Covid Pandemics.
- Moein, S., N. Nickaeen, A. Roointan, N. Borhani, Z. Heidary, S. H. Javanmard, J. Ghaisari, and Y. Gheisari. 2021. “Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan”. *Scientific Reports* 11(1):1–9.
- Rahimi, I., A. H. Gandomi, P. G. Asteris, and F. Chen. 2021. “Analysis and prediction of covid-19 using SIR, SEIQR and machine learning models: Australia, Italy and UK cases”. *Information* 12(3):109.
- Ram, V., and L. P. Schaposnik. 2021. “A modified age-structured SIR model for COVID-19 type viruses”. *Scientific Reports* 11(1):1–15.
- Van de Kastelee, J., J. van Eijkeren, and J. Wallinga. 2017. “Efficient estimation of age-specific social contact rates between men and women”. *The Annals of Applied Statistics* 11(1):320–339.
- Wu, S. L., A. N. Mertens, Y. S. Crider, A. Nguyen, N. N. Pokpongkiat, S. Djajadi, A. Seth, M. S. Hsiang, J. M. Colford, A. Reingold et al. 2020. “Substantial underestimation of SARS-CoV-2 infection in the United States”. *Nature communications* 11(1):1–10.
- Zhou, W.-X., D. Sornette, R. A. Hill, and R. I. Dunbar. 2005. “Discrete hierarchical organization of social group sizes”. *Proceedings of the Royal Society B: Biological Sciences* 272(1561):439–444.

## AUTHOR BIOGRAPHIES

**YINGZE HOU** is a Ph.D. candidate at the Department of Industrial Engineering at the University of Pittsburgh. He earned his Master’s degree in Financial Mathematics from the Johns Hopkins University. His research interests include machine learning, optimization, and artificial intelligence. His email address is [YIH81@pitt.edu](mailto:YIH81@pitt.edu).

**HODA BIDKHORI** is an assistant professor at the Department of Industrial Engineering at the University of Pittsburgh. She earned her Ph.D. in Applied Mathematics from the Massachusetts Institute of Technology (MIT), where she subsequently spent several years as a postdoctoral researcher and lecturer in Operations Research and Statistics. Her current research focuses on the theory and applications of data-driven decision-making. Her email address is [bidkhori@pitt.edu](mailto:bidkhori@pitt.edu).