EPIDEMIC MODELS FOR VULNERABLE POPULATIONS

A dissertation submitted to the University of Manchester
for the degree of Master of Science
in the Faculty of Science and Engineering

2022

Lei Wang
Department of Mathematics

Contents

List of Tables

List of Figures

Abstract

Epidemics have always affected different aspects of people's social lives for many years, and disease outbreaks are particularly prevalent in vulnerable environments. Limited medical resources, poor living conditions and also other factors lead to vulnerable populations facing greater hardship in the epidemic. In order to study the characteristics of epidemic transmission among vulnerable populations, epidemic models are fitted using the influenza data in the winter of 1920 as a study case. The parameters of the models are estimated and calibrated by Approximate Bayesian Computation algorithm to obtain posterior distributions and to simulate the effects of intervention measures. The result has shown that the basic reproduction number increases with the addition of model compartments, reaching 6.77 (95% CI: 5.579-7.344) in the SEIAR model. Latent state and asymptomatic infection state may cause vulnerable people to ignore their illness. Both vaccination and contact control are effective in reducing the number of infections and the size of the epidemic. This method has practical implications for epidemiological modelling studies and disease control of vulnerable populations. The models can be adjusted according to various situations and applied in more detailed studies in the future.

Declaration

No portion of the work referred to in the dissertation has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Intellectual Property Statement

i. The author of this dissertation (including any appendices and/or schedules to this dissertation) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this dissertation, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has entered into. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the dissertation, for example graphs and tables ("Reproductions"), which may be described in this dissertation, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this dissertation, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Dissertation restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's Guidance on Presentation of Dissertations.

Acknowledgements

Through the writing of the dissertation I have received a lot of support and assistance.

Firstly, I would like to thank my supervisor, Professor Hall, for your expertise in this field, which I greatly respect and admire. You guided me in every stage of my writing and gave me a lot of inspiration and encouragement. In addition, your kindness and patience eased my nerves from the first meeting. I really enjoyed the meetings with you because I always learned a lot from them, and your guidance helped me to overcome many difficulties. This project has been a valuable learning opportunity for me. I remember that when I applied for the Statistics programme, I wrote in my personal statement that statistics can contribute to public health safety by providing information for decision making about the epidemic like COVID-19. The completion of this dissertation might be considered as a response to my past statement. I would like to thank you sincerely for giving me this valuable opportunity.

I would also like to thank my friend Nuo. In July and August, we frequently went to the Engineering Building to write our own dissertations, always encouraging each other. Without you, I might not have been able to discipline myself so well. Companion is a special kind of power.

Finally, I would like to thank my parents for supporting me to study here and for always be there for me. This will be the end of this phase of my life, but the beginning of the next one. I will often miss what I have gained during this year and take it as motivation to keep going.

Chapter 1

Introduction

1.1 General background

In the course of human civilization, infectious diseases have exerted numerous impacts on human beings. In November 2002, Severe Acute Respiratory Syndrome (SARS) broke out in southern China. WHO issued a global alert on the disease the following year, then six SARS outbreaks followed in Southeast Asia, North America and Europe, leading to the first pandemic of the 21st century. In 2013, the Ebola virus, described by the UN Security Council as a "threat to International Peace and security", caused an outbreak of unprecedented scale, with 28646 cases reported and 11323 deaths reported in three years(Coltart et al., 2017). In addition, dengue is a major epidemic that has plagued people for many years, with up to 100 million cases of dengue occurring worldwide each year(Monath, 1994). Efforts to develop effective vaccines and drugs are still being pursued in many countries. In this regard, it is important to understand the characteristics of the epidemic and mitigate its impact on public health.

The current respiratory illness outbreak, COVID-19, is a new threat to global health, which rapidly spread across countries around the world in 2019 (Fauci et al., 2020). The pandemic of COVID-19 has made the topic of epidemic become the top priority in recent years. At present, the British government has effectively controlled and stabilized the epidemic in 2021 through vaccination, testing and quarantine measures. During this period, protecting the lives and livelihoods of citizens has been the main goal of the government's work (GOV.UK website). The prevalence of epidemic

seems to be equal to all populations, regardless of the regions, wealth, ages. However, in reality, the impact of the epidemic is more severe among vulnerable population, even widening the gulf between the vulnerable and other populations (Gray, 2021).

The specific needs of vulnerable groups are often overlooked in pandemic measures, such as patients with chronic conditions and the elderly who experience delays or interruptions in routine care, and nursing home residents who are also in a very vulnerable state with increased risk due to the care home's connection to the hospital and the staff's connection to the community. A report shows that there are about 400,000 elderly residents living in care homes, the majority of whom are over 80. The worst COVID-19 outcomes are connected with old age and fragility, with the oldest age groups experiencing the highest mortality rates(GOV.UK website). In addition, displaced refugees or people living in remote areas are facing limited access to health care for protecting against disease. According to the Internal Displacement Monitoring Centre, by the end of 2021, about 59.1 million people in the world are internally displaced on account of widespread violence, human rights or armed conflict (IDMC website). In particular, when information about the epidemic does not reach these groups, it leads to unreliable information or lack of information about the epidemic, which increases their difficulties in getting social attention and achieving treatment for the disease. Research has shown that at least 800 people may have died in 3 months in 2020 due to misinformation related to COVID-19 (WHO website). Therefore, focusing on vulnerable populations during the epidemic, helping them return from the margins of society to the reach of relevant agencies, and providing timely and targeted responses that not only meet the sustainable development goals, but also meet humanitarian needs.

Vulnerable populations are diverse due to different living environments, health conditions and other factors. But they have something in common. Higher exposure to infections, increased susceptibility to serious diseases, and limited access to medical care are all indicators of their vulnerability (Hutchins, 2009). Similarly, various kinds of epidemics have similarities. Currently, in the context of COVID-19, there are still many other epidemics spreading around the world. The UK government points out that studying other infectious diseases could provide more insight into this field, including influenza, which spreads in a similar way to COVID-19 and can cause serious

illness and complications. It is very important to maintain the monitoring and research of these infectious diseases (GOV.UK website). As a result, the extraction of similar parts of the epidemic to carry out the study of vulnerable populations in the epidemic is conducive to a more comprehensive reflection of their current situation and characteristics in the epidemic period.

In addition to the characteristics of epidemic transmission in vulnerable populations, how to make targeted epidemic prevention decisions is also a very critical step. In general, in response to an epidemic, commonly used measures to contain transmission include hospitalization and active treatment, tracking of close contacts, wearing masks, quarantine, and vaccination. In the case of COVID-19, during the severe period of the pandemic, Italy imposed lockdown measures, closed public gathering places and suspended all meetings and performances. The UK also went through a period of lockdown, by shutting down most places of business, canceling flights and asking people to stay at home. However, in the early stage of the epidemic, due to the lack of medical resources, people's insufficient understanding of the virus, and non-compliance with epidemic prevention measures, the epidemic often breaks out on a large scale and spreads to a wider area (Jafari and Gharaghani,2020). Measuring the effectiveness and necessity of epidemic prevention measures is very important for the control of the epidemic.

1.2 Research contents and objective

Previous research in the field of epidemiology has covered a very wide range of different epidemics, yet less attention has been paid to the epidemiology of vulnerable groups, and this paper will focus on vulnerable populations in the direction of epidemic modelling and Approximate Bayesian Computation. The Kelleys Island influenza outbreak in 1920 is chosen to be the study case in data analysis, to study epidemic transmission dynamics and characteristics, estimate the transmission parameters of the models and predict the epidemic scale. Social distancing and vaccination are introduced to simulate the impact of preventive measures on the spread of the epidemic.

The research objective of this paper is to provide applicable epidemic models for

vulnerable populations, which can not only offer rapid and reliable estimates in influenza or COVID-19 epidemics, but also facilitate the adjustment from characteristics of specific vulnerable population or other epidemics in future studies, which enables the model to respond to changing and complex situations.

1.3 Dissertation outlines

The rest of the dissertation is organized into following sections. The second chapter will review the current epidemic models of vulnerable populations and the application of Approximate Bayesian Computation in parameter calibration. The third chapter will introduce the epidemic models and the definition of their parameters, as well as the implementation process of ABC algorithm. Next the fourth chapter will be about data analysis and result discussion. The models will be fitted with epidemic data and numerical simulation is implemented to carry out the epidemic dynamics, and finally the conclusion is discussed in fifth chapter.

Chapter 2

Literature Review

This chapter reviews the the definition of vulnerability, key literatures on epidemic models for vulnerable population and the application of Approximate Bayesian Computing in epidemic models. Many Scholars have conducted extensive research on epidemiological models and applied them to different infectious diseases or different types of vulnerable populations, exploring key indicators such as model parameters, basic reproduction numbers, and prevalence rates. Approximate Bayesian Computing is a method with great advantages in the step of model parameter calibration, which is used by many scholars in the field of epidemiology. Due to the diversity of its algorithms and functions, some scholars have conducted comparative analysis and case analysis on its use. The review of literatures on Approximate Bayesian Computing is beneficial to the selection, adjustment and application of this method.

The first section defines the scope of vulnerable people. The second section is the sorting out of the epidemic models of vulnerable populations and the comparison of the relevant literature according to different populations. The third section is a summary of the application of Approximate Bayesian Computing calibration in epidemiology, showing the applicability of different Approximate Bayesian Computing algorithms and their advantages and disadvantages. Finally, the main content of literature review is summarized and research gap is proposed.

2.1 Vulnerability

In order to determine the scope of vulnerable populations, which is the research object of this paper, it is necessary to review the concept of vulnerability. Here, we focus on the epidemiological aspects of vulnerability.

In the social science and public health studies, vulnerable population are usually described as at-risk, special and vulnerable. Vulnerability is partly attributed to social inequalities. Social factors influence or shape the susceptibility of different groups to harm, and also govern their coping ability (Cutter et al., 2003).

Stanturf et al.(2015) studied the Ebola virus in West African countries, with a population census to construct a classification method of vulnerability. Potential indicators of poverty was set to be variables as a representative of vulnerability for cluster analysis. The study obtained that water quality, health services, food quality, whether displaced populations or disabled populations and lack of access to free health care can be used to measure the vulnerability of the regional population. Although the use of poverty to generalize vulnerability factors is inadequate, the study is still informative because the social, cultural and political dimensions are difficult to measure. Similarly, Shi and Stevens (2021) developed a comprehensive vulnerability model in which risk is divided into social and individual dimensions, with the individual dimension referring to health status and availability of health resources, and the social dimension including demographic location, socio-economic status and so on. They emphasized that vulnerability was not an individual deficiency, but rather an interaction of individual-level and community-level risks.

In addition to the generalization of characteristics and conceptual framework, some scholars directly listed the types of vulnerable groups after concept identification. Sokat and Altay (2021) the vulnerable populations were classified into age, occupation, living conditions, people on the move, social status, mental health, physical health, and gender/sexual orientation. Among them the refugees, internally displaced people, healthcare workers, homosexual men received more attention in this research field. The Department of Health and Human Services in the United States defines high-risk individuals during an pandemic as those with special needs in a public health emergency, including people living in institutions, people from different cultures, people

with limited or non-English speaking skills, people with poor transportation, or people with drug dependence (Hutchins et al., 2009). It can be found that there are many categories of vulnerable groups. Though their risks come from different aspects, they face similar vulnerabilities and susceptibility in the face of epidemics.

2.2 Epidemic model for vulnerable populations

The spread of infectious diseases has always been the healthcare issue of great concern around the world, especially the prevalence of COVID-19 in 2020, which makes the research on infectious diseases more urgent. Epidemiological modelling is an important field for the study of the dynamics of infectious diseases, which can obtain key information of diseases based on the related data, so as to guide health institutions and governments to make decisions and propose timely solutions (Kypraios et al., 2017). Currently, disease outbreaks in vulnerable settings have been prominent during the pandemic (Hall et al., 2021). However, at present, the epidemiological modeling studies on vulnerable populations are still limited (Aylett-Bullock et al., 2022). The following part will mainly focus on epidemic studies in care home residents, schools and refugees. We will then review the epidemic models of several representative and well-studied vulnerable groups.

Care home is a place where vulnerable people live, with a stable living and working population. Due to the close relationship with community and its semi-closed nature, they are facing a high risk of illness and death during infectious disease outbreaks (SCWG, 2020). Hall et al. (2021) used care homes as infection units to study the outbreak of COVID-19 using the SIS model, and the growth rate was fitted with a Generalized Additive Model to estimate the disease prevalence in care homes without intervention. It is concluded that the worst-case prevalence was expected to be as high as 73%. Similarly, Morciano et al. (2021) also paid attention to the situation of care homes in COVID-19, demonstrated an application of local authority fixed-effect Poisson regression to explore the relationship between the excess mortality and the characteristics of care homes such as client type, size, ownership status. The study found that nursing homes had higher excess mortality rates than residential homes, with COVID-19 accounting for 64.7% of excess mortality. Overton et al. (2022)

focused on case fatality risk of care home residents, estimated the time trend of instantaneous case fatality risk, as well as measured case fatality risk of different ages and care needs. The results suggested that the case fatality risk could reach 30% at its peak and age factors on case fatality risk was more significant than care needs, with 85+ age group the highest, which was conducive to the surveillance of the epidemic and the proposal of measures. Based on these literature reviews, it can be found that care home residents have a high risk of death due to their age and health status in the epidemic.

Similar in structure to care homes, schools are also prone to infectious disease outbreaks. As young people tend to have longer physical contact with each other, they pose a greater risk of infection and may introduce infection into their respective households (House et al. al., 2011). For schools, closure is often referred to a non-pharmacological intervention to achieve epidemic prevention and control (Jackson et al., 2013). Therefore, many epidemic models that use schools as research objects aim to measure the impact of school closures on the epidemic. Stage et al. (2021) took the pandemics in Denmark, Norway, Sweden, and German as research cases to explore the relationship between transmission of COVID-19 and school status: closure and reopening. SEIR model fitting by Approximate Bayesian Computing was proposed to estimate the impact of school closures in the early stage of the epidemic, and Generalized Additive Model was used to estimate the instantaneous growth rate at different time points to assess the impact of reopening. Effectively reduction of the growth rate was founded after seven days of school closures. Di Domenico et al. (2020) presented a random age-structured SEIIR epidemiological model, integrating data on ages and social contacts to assess the effects of school closures. In this model, two infection states, pre-symptomatic infectious and symptomatic infectious, were considered. According to the fitting and simulation of the model, it was proposed that closing schools for 8 weeks and 25% of adults working remotely can delay the peak of the epidemic and reduce the incidence by about 40%, contributing to reduce the burden of healthcare system at the corresponding period.

Different from semi-closed environment, some vulnerable people have a more unique geographical distribution all over the world: refugee and internally displaced people. A lack of adequate sanitation and poor housing conditions are highly likely to lead

to disease outbreaks (Charnley et al., 2021), and limited health care increases their vulnerability, which makes it more difficult to resist the epidemic of infectious diseases. In response to these problems, epidemiological modeling has also been used by some scholars for this type of population. Truelove et al. (2020) constructed a stochastic SEIR model to explore the spread of COVID-19 in the Kutupalong-Balukhali Expansion Site refugee camp, with age as the main determinant of infection severity, and three infection scenarios: low, medium, and high, were hypothesized to estimate projected hospitalizations, deaths, and medical needs. The study addressed that a single infected person was likely to cause a large-scale outbreak, and the infection rate will exceed 70%-98% in the first year, which informed ongoing preparedness and response activities. On the one hand, the society can provide timely help for refugees and internally displaced people to resist the pandemic attack according to the model results. On the other hand, corresponding protective measures can also be implemented within such groups. Pascual- García et al. (2020) focused on Covid-19 outbreaks in Northwest Syria, simulating at internally displaced persons camps with a stochastic SEIR model. Social distancing, self-isolation, the establishment of safe zones and other interventions were included in the study as protections for vulnerable populations. He claimed that the mortality rate could be significantly reduced by providing one isolation tent per 200 persons as well as reducing contacts by 50%. However, these strategies, although feasible, are difficult to implement smoothly in camps plagued by epidemics. Fouad et al. (2021) analyzed that high overcrowding and insufficient level of community awareness within Syrian refugee families and informal tent settlements were the direct effects of transmission. The authors found that the absence of interventions, even a small increase in transmission among Syrian refugees could lead to a large increase in morbidity and total cumulative infection. Therefore, they need more help from the society, such as the publicity of epidemic prevention awareness, the establishment of isolation facilities, the expansion of hospital capacity, the provision of medical assistance and other practical actions.

Vulnerable groups are an important part of the current epidemic environment that cannot be ignored. However, the research on infectious disease modeling in this population are facing many difficulties and challenges, including obtaining high-quality data, and constructing models that apply to variable situations (Aylett-Bullock et

al., 2022). In addition, the use of epidemiological models needs to consider its implementation environment, that is, the social context. Incorporating sociology and anthropology into the construction of the model can enable the model to cope with a variety of environments and situations, otherwise it will be limited by the mathematical model and cause mistakes and blind spots(Rhodes et al., 2020). Taking the current COVID-19 as an example, the establishment of the model should not only base on the traditional infectious disease model itself, but should also consider the combination with the current social structure, population characteristics, individual behaviors and corresponding intervention measures. However, in different time and space, these factors are changeable, so it is difficult to capture stable features, and the models need to be continuously adjusted to improve their applicability. These factors explain why there are still many gaps in this field, which is where we need to explore and expand. Only by keeping advancing in this field can we be able to curb the spread of infectious diseases in vulnerable settings in time and reduce the impact of diseases. Such efforts are not only in line with the humanitarian concept, but also guarantee the stability and health of the entire society.

2.3 Approximate Bayesian Computation

In the study of epidemiological models, the model often needs to be fitted with observed data to obtain epidemiological parameters, which plays an important role in proposing timely and effective measures of public health events (Fraser et al., 2004). But in reality, the likelihood functions of model parameters are often untractable, in which case Approximate Bayesian Computing is a likelihood-free method with accuracy and efficiency (Beaumont et al., 2002). The estimation of parameter posterior distributions was achieved by performing numerical simulations and comparing the distance between the simulated data and the summary statistics of the target data (Saulnier et al., 2017). The following section will focus on the use of various Approximate Bayesian Computing (ABC) to estimate and calibrate parameters in epidemiological models.

Minter and Retkute (2019) presented three cases to demonstrate the application of ABC in infectious disease models, namely deterministic epidemic model, age-structure

stochastic model and spatial dynamics individual-based model. Studies have shown that ABC can be used in a variety of different structures. Among them, the effectiveness of ABC-SMC was more prominent than that of ABC-rejection. The ABC-SMC also used in another study for the stochastic SEIR model of H5N8, in which the proposal of within-flock transmission provided ideas for model building of virus transmission in closed spaces or communities and parameter estimation combined with the ABC algorithm (Vergne et al., 2021). Regression-ABC is also a commonly used algorithm when studying epidemiological parameters. In a comparative study, regression-based ABC was proved to be less computationally intensive than ABC-MCMC or ABC-SMC, and less sensitive to the dimension of summary statistics (Saulnier et al., 2017). Therefore, the choice of using the ABC algorithm can be weighed by multiple comparisons. In addition, scholars continue to extend ABC to make it applicable to more hypotheses and scenarios. Danesh et al. (2021) considered two types of hosts when studying the sexual transmission of HCV, providing more research ideas on host population heterogeneity in phylodynamics, which is still a very limited research area, and the ABC approach is enough to deal with this more precise and detailed model. Moreover, the algorithm determines its high applicability to the stochastic model. The model parameters are allowed to be assumed to be random, so the parameter variation is well accommodated in the modeling process, which means that it can better simulate the fluctuation and uncertainty cause by the randomness of the external environment (Sanchez et al., 2019). As a result, ABC is a very powerful and effective algorithm that allows for various epidemiological studies.

However, there are some matters need attention when applying the ABC algorithm. Different parameter settings will affect the efficiency and accuracy of the algorithm. The choices of tolerance range, particle number and summary statistics affect algorithm speed and accuracy (Minter et al., 2019). Therefore, after setting the parameter space, simulating multiple scenarios with different parameter values can verify the rationality and validity of the algorithm (Chong et al., 2018).

2.4 Summary

To sum up, many scholars have put forward models for vulnerable populations in the field of epidemiology, and proposed measures conducive to alleviating the spread of epidemics according to their current situation and environment. Different vulnerable groups will also face different dilemmas. For example, residents of care homes are generally older, have lower immunity, and have nursing needs. Therefore, the model focuses more on the analysis of age structure and mortality. The outbreak in schools will bring infectious diseases from individuals to families, accelerating the spread of epidemics in communities. Since the status of schools can be controlled, more attention of scholars is paid to closing schools to curb the spread. The outbreak in refugee camps is due to limited sanitation and weak awareness of epidemic prevention, so the speed of transmission is a major concern. In addition, the application of ABC in epidemiological models can also provide more accurate and reliable results with the extension of the algorithm. The applicable scenarios include parameter estimation, model selection, and the possibility of more compartments design for the model, such as population heterogeneity. Therefore, when studying the vulnerable population, the algorithm can make the processing of the model more convenient.

When considering the above literature, there are some research gaps within the existing research. In this field of research, it is basically about the status and preventions in epidemic of specific vulnerable groups. However, due to the broad definition of vulnerable populations, there are still many vulnerable groups that have not been fully studied and proposed specific countermeasures. In addition, there are very few studies that treat vulnerable populations as a whole. If the common vulnerabilities and potential risks of different vulnerable groups are extracted, corresponding epidemiological models which are applicable to most groups are established, the transmission characteristics and medical needs of such groups can be better demonstrated. When an epidemic breaks out, the model can be quickly and effectively adjusted to adapt to a specific population according to their characteristic, so as to propose targeted conclusions and strategies.

Chapter 3

Methodology

In recent years, more and more dynamic models have been applied to the study of epidemiology to analyze various problems in infectious diseases. Most of the models add compartments or conditions related to the actual situation in the basic model, which makes the epidemic models more consistent with natural laws. In this section, we start with the simplest epidemic models, introduce the SIR model, SEIR model and SEIAR model respectively, and present the mathematical equations describing these models as well as the explanation of various model parameters. In terms of model parameter estimation, the general algorithm steps of Approximate Bayesian Estimation(ABC) are presented. In particular, the sequential Monte Carlo algorithm used in this paper is introduced in detail.

3.1 Epidemic models

In the construction of the model, we will focus on the acute infections, assuming that the pathogen causes a period of illness followed by immunity. In epidemiology, the infectious process of this epidemic can be described by the SIR model, which was first thoroughly studied by Kermack and McKendrick(1927). The model divides the population into three states, which are Susceptible(S), Infectious(I), and Recovered(R).

The demographic births, deaths and migrations are ignored. Under the assumption of a closed population, the model is divided into the transition from S state to I state and the transition from I state to R state(Figure 3.1). First, the process from S to I represents the spread of the disease, which is determined by the following factors: the

number of potential contacts per unit, the transmission probability under a certain number of contacts, and the prevalence of infection. When the susceptible and the infected contact with each other at a certain probability, the susceptible may become infected and transfer to the second state. The transfer from infectious to recovered depends on the length of the infection period. In general, clinical data can estimate the infection time of various epidemics, and the transfer probability from I to R is the inverse of the infection period, which can also call the recovery rate.

Figure 3.1: Transmissions of the SIR model

Based on the flow chart of state transitions, we assume uniform mixing of the population and constant epidemiological probability, leading to the equations:

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{3.1}$$

Where, β can be defined as $\beta = -\kappa \log(1-c)$. κ is the average number of contacts per unit of time produced by a susceptible individual. c is the probability of successful disease transmission after one contact. Parameter γ represents the recovery rate, whose inverse is the average duration of infection, T $= \frac{1}{\gamma}$. The initial states of the model is $S(0) > 0, I(0) > 0$ and $R(0) = 0$, $N = S + I + R$.

The key parameter that determines whether an epidemic begins or ends is the basic reproduction number R_0, which is one of the most important quantities in epidemiology and reflects the number of subsequent cases arising from an infection in a fully susceptible population (Dietz, 1993). In this model, $R_0 = \frac{\beta}{\gamma}$. Only when R_0 is greater than 1, that is, the infection rate is greater than the recovery rate, the disease can be prevalent in the population.

SIR model is the most classical model in epidemiology, with some limitations, such as not considering the subdivision of the population and the impact of various factors

in the epidemic.

Then we considers the incubation period on the basis of the SIR model. When transmission occurs, the pathogen enters an susceptible individual, and after a period of time, the pathogen multiplies rapidly in the host, undisturbed by the immune system. At this stage, pathogen abundance is too low for transmission to other susceptible individuals, but the pathogen is still present. Therefore, hosts cannot be classified as susceptible, infectious, or recovered (Keeling and Rohani, 2008). We introduce a new category for these individuals, who are infected but not yet infectious. The name of this state is exposure, denoted by E (Figure 3.2).

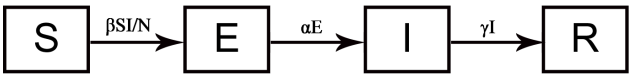

Figure 3.2: Transmissions of the SEIR model

After adding the exposure state, the equations of SEIR model are:

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dE}{dt} &= \frac{\beta SI}{N} - \alpha E \\ \frac{dI}{dt} &= \alpha E - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{3.2}$$

The initial state is $\mathrm{S}(0) = \mathrm{S}_0, \mathrm{I}(0) = \mathrm{I}_0, \mathrm{E}(0) = \mathrm{R}(0) = 0$ and the relationship between compartments is $N = S + E + I + R$. The incubation period is $\frac{1}{\alpha}$. The expression for the basic reproduction number is not affected and therefore remains to be $R_0 = \frac{\beta}{\gamma}$. It has been demonstrated that the incubation period slows an outbreak's initial growth rate and delays the start of an individual's infection period, leading to later onset of secondary infections(Nishiura, 2009).

In addition to the incubation period, it is also important to consider that a significant proportion of infected individuals never develop symptoms and go through an asymptomatic period. During this period they become infectious before entering the recovered compartment. Therefore, based on SEIR, the SEIAR model is proposed,

and it contains Susceptible(S), Exposed(E), Infectious(I), Asymptomatic(A) and Recovered(R). See Table 3.1 for a list of parameters used in the model.

Table 3.1: Description of model parameters

Parameters	Description
β	Rate of transmission
p	Proportion of symptomatic cases
κ	Rate of transition from exposed state to infectious state
α	Rate of recovery from infectious state
η	Rate of recovery from asymptomatic state
δ	Modification of transmission for asymptomatics

he specific transition processes between each compartment are as follows:

(i) In the initial stage, a small proportion of infected people I_0 in the total population N.

(ii) Susceptible individuals(S) enter the exposure compartment(E) with probability β after contact with asymptomatic or symptomatic infected individuals.

(iii) A fraction p of exposed members move to the infectious compartment(I) at the rate of κ, while the remaining members proceed to the asymptomatic compartment (A) at the same rate.

(iv) The infectious leave the compartment for the recovery compartment(R) at the rate of α.

(v) Asymptomatics are reduced in infectivity by a factor δ and go to the recovered compartment at the rate of η.

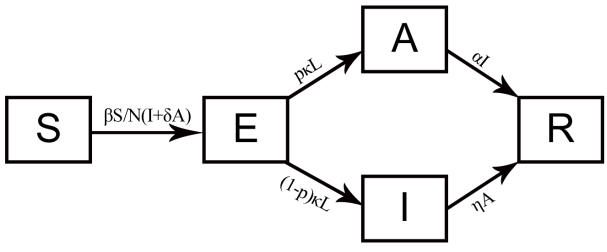

Figure 3.3: Transmissions of the SEIAR model

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta S}{N}(\mathrm{I}+\delta \mathrm{A}) \\
\frac{dE}{dt} &= \frac{\beta S}{N}(\mathrm{I}+\delta \mathrm{A}) - \kappa \mathrm{E} \\
\frac{dI}{dt} &= p\kappa \mathrm{E} - \alpha I \\
\frac{dA}{dt} &= (1-p)\kappa E - \eta A \\
\frac{dR}{dt} &= \alpha I + \eta A
\end{aligned} \tag{3.3}$$

The initial state of SEIAR model is $S(0) = S_0, I(0) = I_0, E(0) = A(0) = R(0) = 0$. The expression of the basic reproduction number is: $R_0 = \frac{\beta S_0}{N}[\frac{p}{\alpha} + \frac{\delta(1-p)}{\eta}]$. It can be interpreted as that an exposed member will become infectious with probability p when there is S_0 susceptible individuals, in which case he may cause $\frac{\beta S_0}{\alpha N}$ infections during an infection period of lenghth $\frac{1}{\alpha}$, or with a probability of 1-p, he/she become asymptomatic and cause $\frac{\delta\beta S_0}{\eta N}$ cases of infection in the asymptomatic period of length $\frac{1}{\eta}$.

3.2 Herd immunity

Vaccination is an effective policy to achieve epidemic control, but measurement of vaccination requires consideration of community-level issues such as the proportion of vaccination that effectively stops transmission and the impact of vaccination on infection and morbidity (Anderson, 1992). The threshold of herd immunity is related to the basic regeneration number in the epidemic model. In a homogenous and mixed population, the immune ratio to suppress epidemic infection is set as p. When the basic regeneration number is less than 1, the epidemic stops spreading, that is $R_0(1-p) < 1$, then $p > 1 - \frac{1}{R_0}$ can be obtained, and the value of p is the threshold value of herd immunity. If a susceptible individual is in contact with an immune host, he/she is indirectly protected from infection even if he/she is not immune. In this paper, vaccination is used as a measure to interfere with the spread of an epidemic, and the herd immunity threshold will be employed to assess the influence of vaccination on an epidemic in subsequent model fitting.

3.3 Parameter estimation

3.3.1 Bayesian estimation

There are usually two ways to predict the probability of events in real life. One is classical statistics, which uses only observational data for inference, the other is Bayesian estimation, which treats the inferred parameters as random variables and uses three kinds of information to infer the distribution of the variables. They are (i) prior probability: the probability of occurrence of an event θ judged according to experience. (ii) Sample information: y^*,obtained by sampling from the overall distribution. The more samples there are, the more accurate statistical inference is. (iii) Overall information: the overall distribution of samples.

$$\mathrm{P}\left(\theta \mid \mathrm{y}^*\right) = \frac{\pi(\theta)\mathrm{P}\left(y^* \mid \theta\right)}{\int_\theta \pi(\theta)\mathrm{P}\left(y^* \mid \theta\right) d\theta} \tag{3.4}$$

In the formula above, $\mathrm{P}\left(y^* \mid \theta\right)$ is the joint density function contains the population and sample information, namely the likelihood function. $\pi(\theta)$ is based on empirical information about the parameters θ. $\mathrm{P}\left(\theta \mid \mathrm{y}^*\right)$ is the posterior probability, which involves the information about the population, sample and the prior of θ. But in practice, the likelihood function is difficult to calculate, so Approximate Bayesian Computing replaces the likelihood function with distance comparison.

3.3.2 Approximate Bayesian Computation Calibration

For model parameter estimation in this paper, we apply an Approximate Bayesian Computation method(ABC). The advantage of the ABC method is that we do not need to compute the specific likelihood function. The general steps are:

A1 Generate the candidate parameter set θ^* from the selected prior distribution.

A2 Generate simulation data set x^* based on conditional probability distribution $f\left(x \mid \theta^*\right)$.

A3 Use distance function d and tolerance ϵ to compare the simulated data with the observed data; If $d(x_0, x^*) \leq \epsilon$, accept x^*.

The set of sample parameters generated from $\pi\left(\theta \mid \mathrm{d}\left(\mathrm{x}_0, x^*\right)\right) \leq \epsilon$ is the output of the ABC method, whose distribution is a fair approximation of the posterior distribution $\pi\left(\theta \mid x_0\right)$ when ϵ is sufficiently small. In general, the distance function will

replace x^* and x_0 with summary statistics of simulated and observed data.

The existing ABC methods include rejection sampling, Markov chain Monte Carlo (MCMC) and Sequantial Monte Carlo (SMC). Among them, ABC-SMC method can avoid the disadvantages that exist in the first two methods, which is first developed by Sisson et al. (2007). It has the following advantages: (i) The sampler will never get stuck in the low probability region. (ii) It can avoid the mismatch between the initial sampling and the target distribution caused by the inefficiency in the rejection sampling. (iii) The complex posteriors can be explored more effectively. (iv) Samples are drawn from several distributions with various tolerances, which allows checking the robustness of the posterior or dynamic. (v) Unlike MCMC, the particles in SMC are uncorrelated so there is no need to determine the aging period or evaluate convergence (Toni et al., 2009).

The steps of ABC-SMC algorithm are as follows:

S1 Initialized $\epsilon_1, ..., \epsilon_T$, set t=0.

S2 Set the particle indicator i=1.

S3 Generate a particle(parameter vector) $\theta_t^{(i)}$:

If t=1, sample θ^{**} from prior of parameters $\pi(\theta)$.

If t¿1, sample θ^* from previous population $\theta_{t-1}^{(i)}$ with weights $\omega_{t-1}^{(i)}$ and use the perturbation kernel K_t for perturbing the particles to get $\theta^{**} \sim K_t\left(\theta \mid \theta^*\right)$.

S4 If $\pi(\theta^{**}) = 0$, go back to S3.

Else, simulate a data set $x^* \sim f\left(x \mid \theta^{**}\right)$.

If $d\left(S\left(x^*\right), S\left(x_0\right)\right) \geq \epsilon_t$ go back to S3.

S5 Set $\theta_t^{(i)} = \theta^{**}$ and adjust the weight for $\theta_t^{(i)}$,

$$\omega_t^{(\mathrm{i})} = \begin{cases} 1, \text{ if t} = 0 \\ \frac{\pi\left(\theta_{\mathrm{t}}^{(\mathrm{i})}\right)}{\sum_{\mathrm{j}=1}^{\mathrm{N}} \omega_{\mathrm{t}-1}^{(\mathrm{i})} \mathrm{K_t}\left(\theta_{\mathrm{t}-1}^{(\mathrm{j})}, \theta_{\mathrm{t}}^{(\mathrm{j})}\right)}, \text{ if t } > 0 \end{cases}$$

If $i < N$, set i=i+1, go back to S3.

S6 Normalized the weights:

If $t < T$, set t=t+1, go back to S2.

For model fitting in the following sections of this paper, ABC-SMC is used through Pygom package in Python (Tye et al., 2018). The system provides functions for solving, estimating and simulating odinary diffenrential equation system. When running the

algorithm, we set 100 samples in each generation, and generate for 10 times. The quantile used to specify the tolerance for future generations is 0.5.

3.4 Data

In our models, we used the case incidence of influenza on Kelleys Island in the winter of 1920 as the observation data. This is a classic case data about vulnerable population. Although it has been a long time, it is enough to show the characteristics of the epidemic spreading among vulnerable people, which can be used as a reference for modern epidemics.

The epidemic period was from 24th January to 22nd February in 1920. The dynamics of the infections of Kelleys Island can provide a case study for us to explore model parameters. Data were obtained from the epidemiological study conducted by Armstrong and Hopkins (1921). Kelleys Island, a political subdivision in Erie County, Ohio, is a small, isolated island that fits the features of vulnerable population. There was a serious influenza outbreak in January and February 1920, when there were 689 people on the island. Because it was winter, communication with the mainland was limited, and crossing the border was difficult and dangerous. The population can therefore be considered fixed during the epidemic.

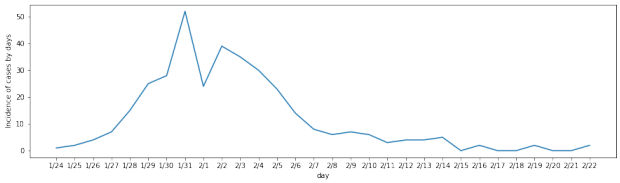

Figure 3.4: Incidence of cases by days in Kelleys Island

According to records, the 1920 epidemic began to become severe around January 24, peaked on January 31 with 52 new infections, and then the incidence of new cases declined until February 16, when there were almost no new cases (see Figure 3.4). The total number of infections in the epidemic reached 369, with an attack rate of 53.56%.

3.5 Summary

One of the purposes of establishing epidemic models is to study the key factors and trends affecting the spread of diseases and to provide theoretical guidance for government epidemic prevention. At the same time, the model also provides a theoretical tool for verifying our conjecture. Since it is impossible to conduct disease experiments on humans, the control of infectious diseases and the measures taken in the real world need theoretical verification and computer simulation first, and then implemented according to the actual situation.

There are deterministic model and stochastic model. Deterministic models often ignore individual differences and explore state transmission in the average situation, while stochastic models can simulate individual changes. However, when the model is closer to the actual situation, it will become more complex, the theoretical analysis is more difficult, and the parameter values are more difficult to obtain. Relatively speaking, simple models have fewer parameters and are easier to analyze and simulate.

In addition to model selection, the choice and estimation of parameters are very important. The parameter values are usually associated with important information about the outbreak, such as infectivity and recovery rate. The ABC algorithm introduced above can realize effective and quick parameter estimation, which is a very practical tool in epidemic models. It should be noted that due to the gap between the model and the real world, the results obtained by the algorithm are only an approximation to the real world.

When we take up the research problem, we will be faced with the selection of the model. In order to present the results more closely related to the actual situation, we will consider starting from the simplest model, gradually increasing the compartments. Fitting the observed data, calculating the model parameters by ABC method, and presenting the stochastic simulation results of some models to reflect the model changes under random conditions are the tasks we will complete in the following section.

Chapter 4

Data analysis and Results

4.1 Parameter estimation and numerical simulation

In this section, the parameters of the SIR, SEIR and SEIAR models are estimated respectively using the ABC algorithm, and the transmission dynamics of the epidemic are presented through numerical simulations.

4.1.1 SIR model fitting and simulation

The ABC algorithm was applied to 1920 influenza data in Kelleys Island and estimates of β,γ and R_0 were obtained. The prior distributions are summarised in the Table 4.1 with the results of the parameter estimation. The prior distributions are chosen to be uniformly distributed, so that the range is as large as possible and contains less information. The posterior distributions of the parameters β and γ generated by the ABC algorithm are shown in Figure 4.1. Their distributions all converge in a very small interval and it can be considered that the estimates accepted in the algorithm are very concentrated. The upper and lower 95% confidence intervals are very close and therefore not presented in the table. The median estimate for β is 1.868, γ is 1.256 and R_0 is 1.487. It can be inferred from this result that the average infectious period is only about 0.8 days. In fact, this is a very short period of time and we believe that the compartments of the SIR model are relatively simple that can not take into account the remaining possible states during this epidemic, so the results

may be slightly biased.

Table 4.1: Transmission parameters of SIR model obtained by ABC

Parameters	Prior distribution	Posterior median
β	Uniform(0,3)	1.868
γ	Uniform(0,3)	1.256
R_0		1.487

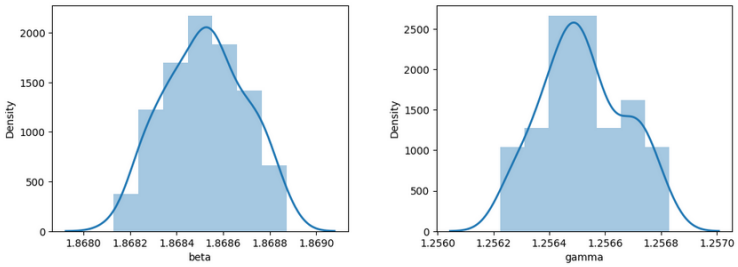

Figure 4.1: Posterior distributions of fitting parameters: beta(β) and gamma(γ)

The influenza dynamics in Kelleys Island were simulated by SIR model with the estimated parameters, presenting in the Figure 4.2. The number of infections in the simulation is relatively consistent with the observed data, reflecting the ability of the model to adequately simulate the evolution of the epidemic. Given the importance of contingency, we applied the parameters to the numerical simulation of the stochastic model, using a probabilistic jump process driving changes between states, and generating 10 realizations. The visualization produced by PYGOM is shown in Figure 4.3.

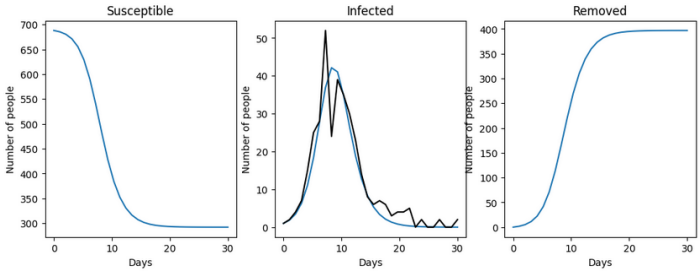

Figure 4.2: Observed epidemic curve(black line) along with deterministic simulation curves(blue lines)

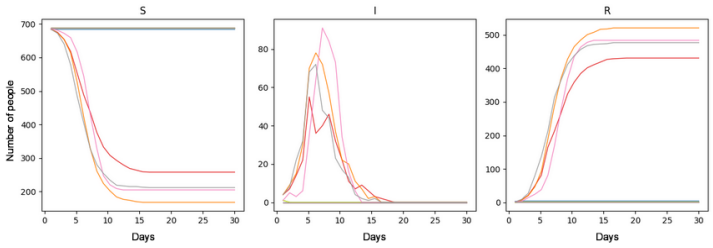

Figure 4.3: 10 realizations of SIR stochastic model

It can be noticed that the curves are relatively smooth in both the S and R states, and only the fluctuation in I state is obvious. The dynamics of infectious diseases are affected by stochasticity and the number of infections may even peak at 70. This is more realistic for the simulation of epidemics than the deterministic model. In addition, the jump process model is able to produce results that the disease is completely eliminate from the model before it breaks out. The horizontal lines at the top of the susceptible compartment and the bottom of the recovered compartment indicate that all members of compartment 'I' move to compartment 'R' before more individuals are infected.

4.1.2 SEIR model fitting and simulation

In this section, an exposed state was added to the SIR model, indicating individuals who have been exposed to an infected person but are not yet infectious. Similarly, we obtained estimates of the three parameters through the ABC algorithm, and the dynamic process of the SEIR model was simulated (Figure 4.4). It is observed that the number of exposed individuals peaked earlier than that of infected individuals. The peak of infection was delayed from the eighth day to approximately the eleventh day due to the addition of the exposure state, suggesting that the incubation period have slowed the course of the epidemic.

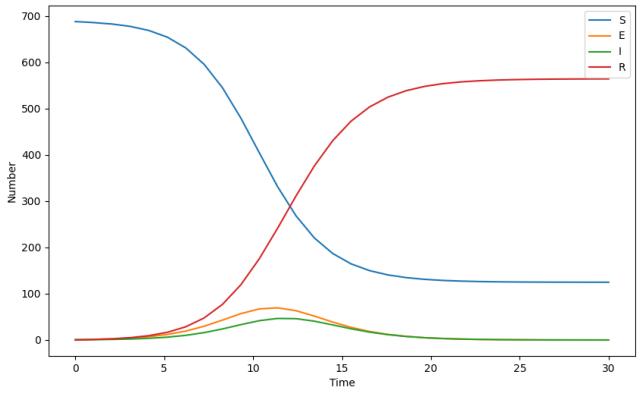

Figure 4.4: SEIR model simulation curves

Table 4.2 lists the estimation results and the settings of the prior distributions. According to the epidemiological study of Kelleys Island, the highest frequency of incubation period is 1-4 days(Armstrong and Hopkins, 1921), so the range of the prior distribution of α is limited to 0.1 to 1. According to the result, when a susceptible individual comes into contact with an infected individual, they will experience an incubation period of about one day before they become infected.

Table 4.2: Transmission parameters of SEIR model obtained by ABC

Parameters	Prior distribution	Posterior median
β	Uniform(0,3)	2.994
α	Uniform(0,1)	0.996
γ	Uniform(0,3)	1.436
R_0		2.085

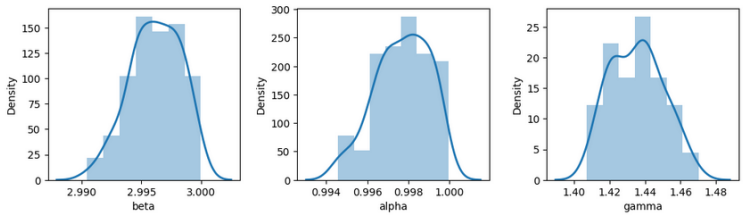

Figure 4.5: Posterior distributions of fitting parameters: beta(β), alpha(α) and gamma(γ)

As shown in Figure 4.5, in fact, the posterior distributions of parameter β, α and γ are affected by the range of the prior distribution, resulting in a rightward bias of the

distributions of two parameters, which are close to their maximum values. Therefore, we then consider further modifications to the model to explore more states of the epidemic.

4.1.3 SEIAR model fitting and simulation

Combining the characteristics of influenza, the A(asymptomatic) compartment compartment is added to SEIR to denote asymptomatic infected individuals. Since exposed members with probability p and probability 1-p will be symptomatic and asymptomatic infected respectively, we have set the upper limit of prior distribution for this parameter to be 1. δ represents the decreasing infectivity of an asymptomatic individual, so it is similarly restricted by Uniform(0,1). The posterior medians estimated by our ABC method has shown in the Table 4.3.

Table 4.3: Transmission parameters of SEIAR model obtained by ABC

Parameters	Prior distribution	Posterior median	95% Confidence interval
β	Uniform(0,5)	4.701	[4.667,4.736]
p	Uniform(0,1)	0.120	[0.119,0.122]
κ	Uniform(0,1)	0.831	[0.811,0.851]
α	Uniform(0,2)	0.240	[0.237,0.242]
δ	Uniform(0,1)	0.884	[0.870,0.898]
η	Uniform(0,2)	0.827	[0.755,0.900]
R_0		6.772	[5.579,7.344]

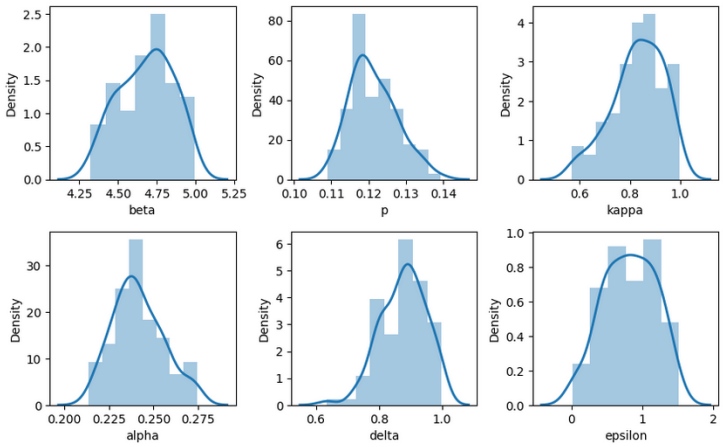

Figure 4.6: Posterior distributions of fitting parameters: beta(β), p, kappa(κ), alpha(α), delta(δ) and epsilon(ϵ)

Figure 4.6 shows the distribution curves. Based on the estimation results, we can obtain that the average latent period of influenza is 1.16 days, and the average infective period of symptomatic infection is 4.17 days. The recovery time of asymptomatic individuals is faster, and they will enter the recovery state after 1.21 days. At the same time, the infectivity of the asymptomatic is 0.116 times less than that of symptomatic infected persons, which means that a susceptible individual who contact with the asymptomatic are less likely to be infected. In this model, the introduction of an exposed member in 689 people on the island will result in approximately 20 cases during the infection period, or become asymptomatic and cause about 5 infections in the asymptomatic period of 1.21 days.

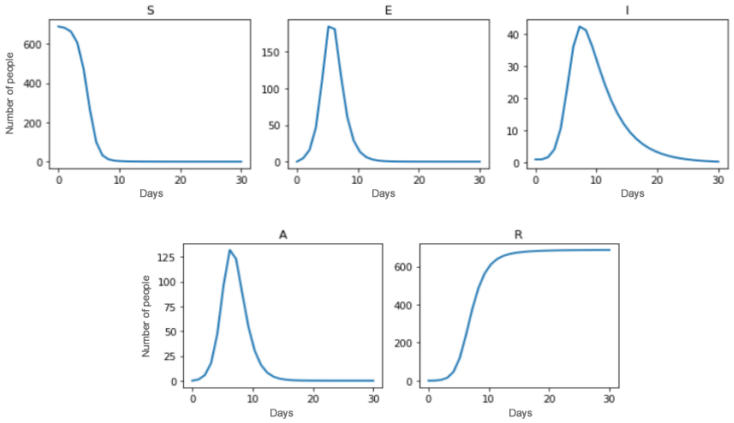

Figure 4.7: SEIAR deterministic model simulation curves

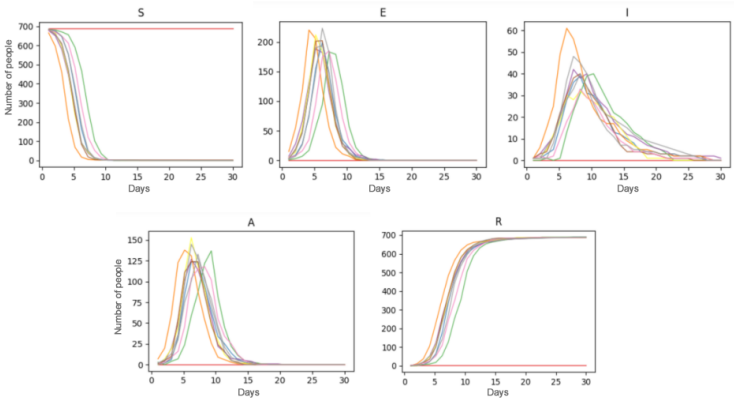

Figure 4.8: SEIAR stochastic model simulation curves(10 realizations)

The dynamic behaviors of the deterministic and stochastic models are presented in Figure 4.7 and Figure 4.8 respectively. It is straightforward to see that the number

of people in exposed the state peaks before the number in the infectious state, and the number of asymptomatic individuals almost reached 0 by about day 13. By contrast, the symptomatic have a longer period of infection, lasting until the 30th day before they are fully removed. Under the stochastic scenario, the peak of influenza infection may occur from day 6 to day 10, similar to the stochastic realisation of the SIR, demonstrating the case where influenza is suppressed before it starts, which is described the horizontal lines in the figure.

4.2 The impact of intervention on epidemic

4.2.1 Vaccination and immunity

Interventions for epidemics are introduced based on the SEIAR model to assess the effectiveness of different measures for epidemic control. Vaccination is one of the main methods of preventing and controlling epidemics. It provides direct protection to vaccinated individuals against viral attacks and is a healthy way to limit the spread of pandemics by gaining population immunity (Gostin and Salmon, 2020).

For the effect of the vaccine, we assume that the vaccine will directly immunize some susceptible individuals into the recovery compartment, and on this basis, we set the number of initially susceptible individuals to be $S(0) = (1-\theta)N$ .The initial value of recovery is $R(0) = \theta N$. Parameter θ indicates the proportion of people immunized after vaccination.

Figure 4.9 and 4.10 present the change in the number of infections when θ is 0, 0.1, 0.3 and 0.5. With immunization, fewer people will be infected. As immunization rates increase, the peak of infection in an outbreak decreases, and the peak occurs later. It is worth noting that even though symptomatic infections peak at different times, the trend of decline in Figure 4.9 after 15 days is very similar across the four conditions, with the curves almost overlapping, suggesting a more significant impact of vaccination in the early stages of the epidemic. The graph confirms that when the effectiveness of the vaccine allows people to become truly immune, it will cause the decreases in the number of the susceptible and the infectious.

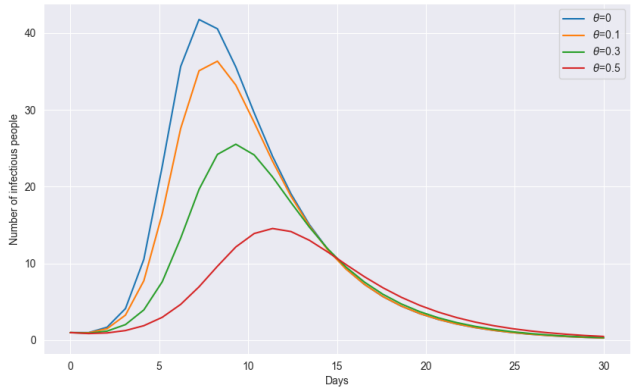

Figure 4.9: Number of symptomatic infectious people with different immunization rates

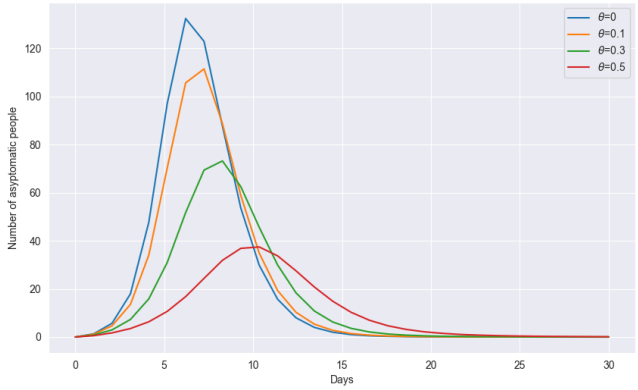

Figure 4.10: Number of asymptomatic infectious people with different immunization rates

Therefore, it is important to achieve vaccine coverage by estimating the threshold of herd immunity for the implementation of vaccination to mitigate the epidemic or prevent the next outbreak.

Figure 4.11 presents the threshold curves for herd immunity at different basic reproduction numbers. The proportion of population needing to be immunized becomes larger with the increase of basic regeneration number. The black dashed line indicates that in the case of Kelleys Island, the estimation of base reproduction number is 6.772, corresponding to a threshold for herd immunity of approximately $1 - \frac{1}{R_0} = 85.23\%$, which means that when 85.23% of the population is immune, the virus can not cause

an epidemic. It is important to note that this threshold only applies to new epidemics and that current epidemics will not stop when the immunisation rate reaches the threshold.

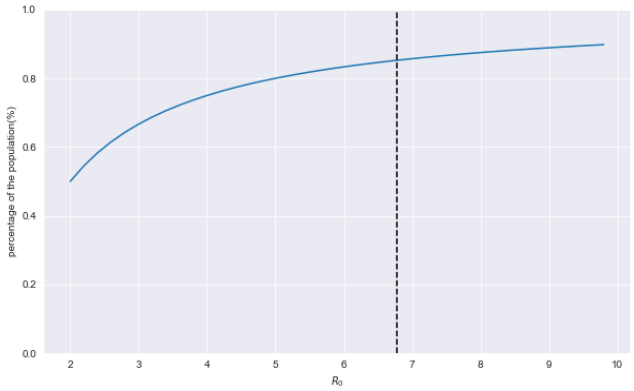

Figure 4.11: The threshold of herd immunity under different R_0 conditions

4.2.2 Contact rate reduction

Social distancing, isolation and wearing masks are the main human behaviors that affect contact rates during the pandemic. Mathematically, the contact rate correlate with the infection rate β. The equation is $\beta = \rho \times n$, where ρ represents the probability of being infected after contact with an infectious individual and average n contacts per unit of time (Keeling and Eames, 2005). Wearing masks reduces ρ , the probability of being infected, and keeping social distance and isolation actually reduces the number of people exposed to an infected individual n, so these behaviours all lead to lower infection rates β. We set the effect as parameter ϕ, indicating that the implementation of these measures will reduce the infection rate by 1-ϕ%.

It can be seen from Figure 4.12 that the decrease in contact rate leads to a slower decline in the number of susceptible individuals. On the contrary, when the contact rate is not controlled, all the susceptible will experience the epidemic on about the tenth day. When the contact rate is reduced by half, a small number of susceptible individuals will not be involved in the outbreak, making the final size of the epidemic smaller (Figure 4.13).

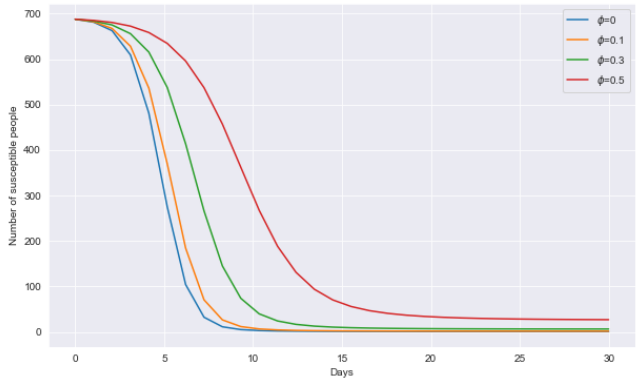

Figure 4.12: Number of susceptible people with different degree of reduced contact rates

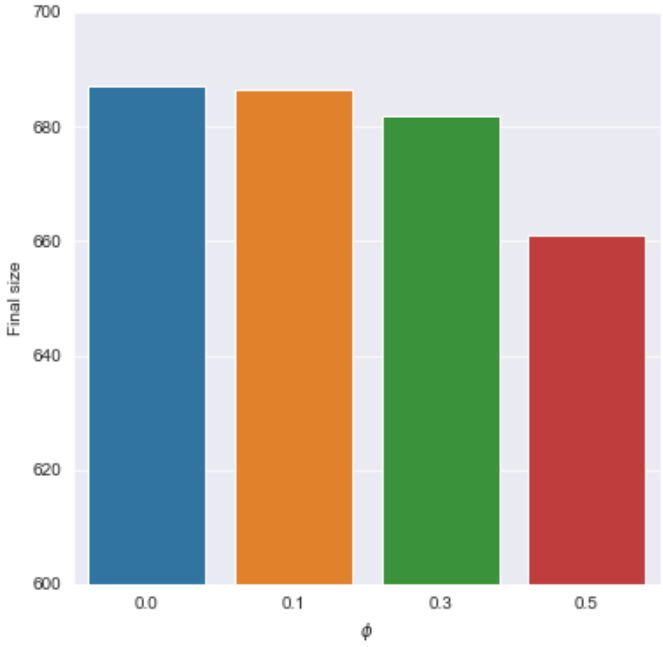

Figure 4.13: The final size with different degree of reduced contact rates

For the symptomatic infectious (Figure 4.14), the effect of reducing the contact rate is similar to that of vaccination, with the peak value decreasing as the rate decreases and taking a little longer to reach its peak. However, the difference is that the decrease of contact rate does not have a significant impact on the decline rate of the infection cases. When the contact rate is reduced by 0.5, the infected individuals are still

not cleared by the last day. In contrast, at different contact rates, the number of asymptomatic infected people (Figure 4.15) basically reaches a steady state within 30 days, with other characteristics similar to those of symptomatic infected state.

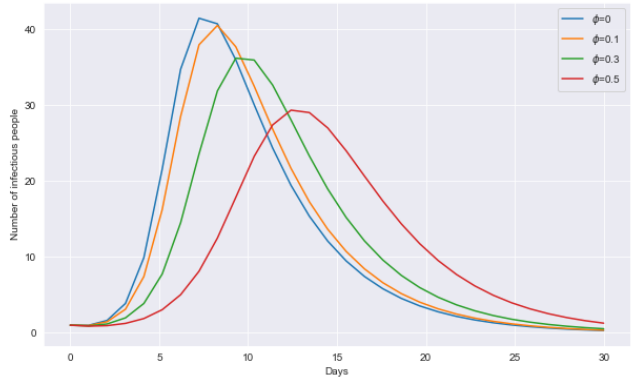

Figure 4.14: Number of symptomatic infection with different degree of reduced contact rates

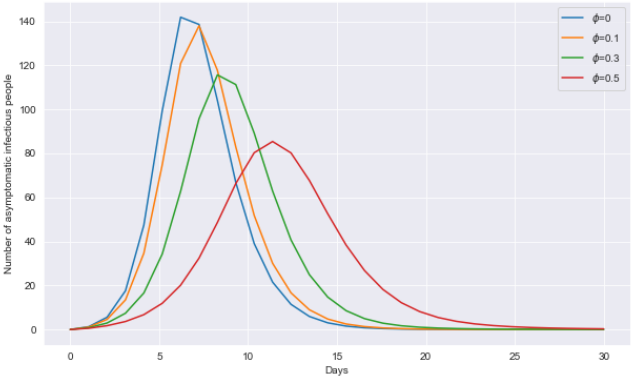

Figure 4.15: Number of asymptomatic infection with different degree of reduced contact rates

In summary, the results suggest that the decrease of contact rate is beneficial in reducing the size and transmission of the epidemic, and therefore encouraging social distancing and mask wearing can contribute to controlling the spread of the epidemic among vulnerable populations.

Chapter 5

Conclusion

5.1 Key findings

During the COVID-19 pandemic, epidemiology has once again received a great deal of attention and this field of research is expanding. Due to the individual heterogeneity in living, health and medical conditions, there are vulnerable population in society who are more difficult to survive in the epidemic than the general population. Understanding the characteristics of vulnerable populations and their dilemma in the epidemic can help to develop targeted disease mitigation measures, so as to not leave behind every individual in society in the process of resisting or co-existing with the disease.

On the basis of epidemiological models, this dissertation aims to study the infection situation and coping strategies of vulnerable population in epidemic. SIR, SEIR and SEIAR are respectively used to fit influenza data and estimate corresponding model parameters in the data analysis section. It can be observed that due to the complexity of the reality, simple models are limited in the generalization of the dynamic behaviors of the epidemic, and the more compartments, the more situations can be concerned. After adding the exposure compartment and asymptomatic infection compartment, the parameters estimated by ABC are used to simulate the dynamic behaviors of the epidemic, which have shown the change of the number of people in different compartments over time. The peak of the epidemic is delayed, and the weak infectivity of asymptomatic individuals lead to a slight transmission rate in the overall population.

In addition, we have measured the change in the number and the peak of infections

after the addition of intervention in the model. Vaccination and control of the contact rate are easier measures to implement in vulnerable population. We have confirmed that vaccination can delay the peak of an epidemic, meaning it can give governments and agencies more time to respond and take action. Scheduling vaccinations according to the threshold of herd immunity and providing more ways for free vaccination can effectively ensure that epidemics like influenza, which are likely to break out every winter, can be prevented in advance and controlled in the early stage. In terms of the intervention of the social contact, social distancing can effectively reduce the size of the epidemic and infections, so compulsory isolation when necessary is suggested to be considered. Wearing a mask is another measure in the aspect of social distancing, which can also effectively alleviate the spread of the epidemic. Whether in nursing homes, refugee camps or remote islands, raising awareness of epidemic prevention can encourage voluntary compliance.

It is worth noting that although asymptomatic individuals recover quickly and have a lower infection rate, it is difficult to monitor them because they have no obvious symptoms. Moreover, according to the model results, the number of the asymptomatic is not small. So positive self-detection awareness and necessary social distance control should be advocated. For vulnerable populations, limited medical conditions constrain their access to timely treatment, and trajectory tracking is difficult for people living in remote or displaced areas. Therefore, there is a limit to what can be achieved for them. Providing free vaccinations, distributing masks, sterilization supplies, and building isolation cabins are useful attempts on quarantine measures for vulnerable populations.

5.2 Limitations

Theoretical and methodological limitations have constrained current research in several ways: First, the data selected in this paper are influenza data occurring in a remote island, where the infected population is in line with the characteristics of vulnerable population. However, age, health status and other factors are not considered in the study. These factors may influence the spread of the epidemic. For example, age structure plays a key role in immunity, treatment-seeking behavior, and disease

transmission (Ram and Schaposnik, 2021). Without including age structure, we have to assume that individual ages are evenly mixed and that immunity and infection rates are consistent across ages: all of these assumptions are not accurate enough. Secondly, the models selected in this paper are still classical models, which is convenient to fit the observed data, but it is still insufficient in simulating the real situation. In addition, we have assumed that is a fixed proportion of immunity, actually vaccination will be affected by the age and the effectiveness of the vaccine, and this requires specific epidemic data of the corresponding vaccine. The purpose of this paper is to study the epidemic spread of the vulnerable population, and therefore, specific vaccine-related information is not considered.

5.3 Future research

In future studies, the heterogeneity of vulnerable populations is worth taking into account in epidemic models. In addition to achieving human immunity and controlling contact rates, other targeted measures can also be evaluated in the model, such as hospitalization to improve recovery rates. What's more, different categories of vulnerable populations lead to their dispersed geographical locations. The combination of spatio-temporal statistical model and epidemiological model is also a direction for further research. The epidemic model applied in this paper can be adjusted to the real situation for the study of a specific epidemic or a certain type of vulnerable population, and the ABC method can still be used for model parameter estimation in future in-depth research.

Finally, the characteristics of vulnerable populations determine that they will encounter more difficulties in the outbreak of an epidemic, and it is imperative to implement quick and accurate measures to mitigate the impact of them in the epidemic. Therefore, the vulnerable population need more attention from all sectors of society and the government. In the field of epidemiology, the study of vulnerable population is of great significance and worthy of continuous promotion.

Bibliography

[1] Anderson R. M., (1992), "The concept of herd immunity and the design of community-based immunization programmes". Vaccine, 10(13), 928–935.

[2] Armstrong, C., Hopkins, R., (1921), "An epidemiological study of the 1920 epidemic of influenza in an isolated rural community". Public Health Reports (1896-1970): 1671-1702.

[3] Aylett-Bullock, J., Gilman, R. T., Hall, I., et al. (2022), "Epidemiological modelling in refugee and internally displaced people settlements: challenges and ways forward". BMJ global health, 7(3), e007822.

[4] Beaumont, M.A., Zhang, W., Balding, D.J., (2002), "Approximate Bayesian computation in population genetics". Genetics, 162 (4), 2025–2035.

[5] Charnley, G.E.C., Kelman, I., Gaythorpe, K.A.M. et al.(2021), "Traits and risk factors of post-disaster infectious disease outbreaks: a systematic review". Scientific Reports, 11, 5616.

[6] Chong, K. C., Zee, B., Wang, M. H. (2018), "Approximate Bayesian algorithm to estimate the basic reproduction number in an influenza pandemic using arrival times of imported cases". Travel medicine and infectious disease, 23, 80–86.

[7] Coltart, C. E., Lindsey, B., Ghinai, I., Johnson, A. M., Heymann, D. L. (2017), "The Ebola outbreak, 2013–2016: old lessons for new epidemics". Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1721): 20160297.

[8] Cutter S.L., Boruff B.J., Shirley W.L., (2003), "Social vulnerability to environmental hazards". Social Science Quarterly, 84 (2) , pp. 242-261.

[9] Danesh, G., Virlogeux, V., Ramière, C., Charre, C., Cotte, L., Alizon, S. (2021), "Quantifying transmission dynamics of acute hepatitis C virus infections in a heterogeneous population using sequence data". PLoS pathogens, 17(9), e1009916.

[10] Di Domenico, L., Pullano, G., Pullano, G., Hens, N., Colizza, V., (2020), "Expected impact of school closure and telework to mitigate COVID-19 epidemic in France". EPIcx Lab, 15

[11] Dietz K., (1993),"The estimation of the basic reproduction number for infectious diseases". Statistical methods in medical research, 2(1): 23-41.

[12] European Commission, 2021. How vulnerable groups were left behind in pandemic response. https://ec.europa.eu/research-and-innovation/en/horizon-magazine/how-vulnerable-groups-were-left-behind-pandemic-response

[13] Fauci, A. S., Lane, H. C., Redfield, R. R. (2020), "Covid-19—navigating the uncharted". New England Journal of Medicine, 382(13): 1268-1269.

[14] Fraser, C., Riley, S., Anderson, R. M., Ferguson, N. M. (2004), "Factors that make an infectious disease outbreak controllable". Proceedings of the National Academy of Sciences, 101(16), 6146–6151.

[15] Gostin, L. O., Salmon, D. A., (2020), "The dual epidemics of COVID-19 and influenza: vaccine acceptance, coverage, and mandates". Jama, 324(4): 335-336.

[16] GOV.UK, 2020. SCWG: Care homes analysis-12 May 2020. https://www.gov.uk/government/publications/care-homes-analysis-12-may-2020

[17] GOV.UK, 2022. COVID-19 Response: Living with COVID-19. https://www.gov.uk/government/publications/covid-19-response-living-with-covid-19/covid-19-response-living-with-covid-19

[18] Hall I, Lewkowicz H, Webb L, et al. (2021), "Outbreaks in care homes may lead to substantial disease burden if not mitigated". Philosophical Transactions of the Royal Society B, 376(1829), 20200269.

[19] House T, Baguelin M, Van Hoek A J, et al. (2011), "Modelling the impact of local reactive school closures on critical care provision during an influenza pandemic". Proceedings of the Royal Society B: Biological Sciences, 278(1719), 2753-2760.

[20] Hutchins, S. S., Truman, B. I., Merlin, T. L., Redd, S. C., (2009), "Protecting vulnerable populations from pandemic influenza in the United States: a strategic imperative". American journal of public health, 99(S2): S243-S248.

[21] Jackson C, Vynnycky E, Hawker J, et al. (2013), "School closures and influenza: systematic review of epidemiological studies". BMJ open,3(2), e002149.

[22] Jafari, H., Gharaghani, M. A.(2020), "Cultural challenges: the most important challenge of COVID-19 control policies in Iran". Prehospital and Disaster Medicine, 35(4): 470-471.

[23] Keeling, M. J., Eames, K. T., (2005), "Networks and epidemic models". Journal of the royal society interface, 2(4): 295-307.

[24] Keeling, M.J., Rohani, P. (2008), Modelling Infectious Diseases in Humans and Animals, (1st edition), Princeton University Press, Princeton.

[25] Kermack, W. O., McKendrick, A. G. (1927). "A contribution to the mathematical theory of epidemics". Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, 115(772), 700-721.

[26] Khan, R., Ahmed, A., Zeitounie, R., Khandekar, R., (2021), "Impact of influenza vaccine in reduction of incidence and severity of influenza-like illness". Eastern Mediterranean Health Journal= La Revue de Sante de la Mediterranee Orientale= Al-majallah Al-sihhiyah Li-sharq Al-mutawassit, 27(9): 884-891.

[27] Kypraios, T., Neal, P., Prangle, D. (2017), "A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation". Mathematical biosciences, 287, 42-53.

[28] Minter, A., Retkute, R. (2019), "Approximate Bayesian Computation for infectious disease modelling". Epidemics, 29, 100368.

[29] Monath, T. P. (1994), "Dengue: the risk to developed and developing countries". Proceedings of the National Academy of Sciences, 91(7): 2395-2400.

[30] Morciano M, Stokes J, Kontopantelis E, et al.(2021), "Excess mortality for care home residents during the first 23 weeks of the COVID-19 pandemic in England: a national cohort study". BMC medicine, 19(1): 1-11.

[31] Nishiura H.,(2009), Mathematical and statistical estimation approaches in epidemiology. Dordrecht:: Springer Netherlands.

[32] Overton C E, Webb L, Datta U, et al.(2022), "Novel methods for estimating the instantaneous and overall COVID-19 case fatality risk among care home residents in England". arXiv preprint arXiv, 2202, 07325.

[33] Pascual-García A, Klein J, Villers J, et al.(2020), "Empowering the crowd: feasible strategies to minimize the spread of covid-19 in high-density informal settlements". medRxiv.

[34] Ram, V., Schaposnik, L. P.,(2021), "A modified age-structured SIR model for COVID-19 type viruses". Scientific reports, 11(1): 1-15.

[35] Rhodes T, Lancaster K, Lees S, et al. (2020), "Modelling the pandemic: attuning models to their contexts". BMJ global health, 5(6): e002914.

[36] Sanchez, F., Barboza, L. A., Vásquez, P. (2019), "Parameter estimates of the 2016-2017 Zika outbreak in Costa Rica: An Approximate Bayesian Computation (ABC) approach". Mathematical biosciences and engineering: MBE, 16(4), 2738–2755.

[37] Saulnier, E., Gascuel, O., Alizon, S. (2017), "Inferring epidemiological parameters from phylogenies using regression-ABC: A comparative study". PLoS computational biology, 13(3), e1005416.

[38] Shi, L., Stevens, G. D. (2021), Vulnerable populations in the United States, (3rd edition), John Wiley Sons.

[39] Shivayogi P. (2013), "Vulnerable population and methods for their safeguard". Perspectives in clinical research, 4(1): 53.

[40] Sisson, S. A., Fan, Y., Tanaka, M. M., (2007), "Sequential monte carlo without likelihoods". Proceedings of the National Academy of Sciences, 104(6): 1760-1765.

[41] Social care working group, 2020. Care Homes Analysis. Scientific Advisory Group for Emergencies, https://www.gov.uk/government/publications/care-homes-analysis-12-may-2020

[42] Sokat, K. Y., Altay, N. (2021), "Serving vulnerable populations under the threat of epidemics and pandemics". Journal of Humanitarian Logistics and Supply Chain Management, 11(2), 176-197

[43] Stage, H. B., Shingleton, J., Ghosh, S., Scarabel, F., Pellis, L., Finnie, T., (2021), "Shut and re-open: the role of schools in the spread of COVID-19 in Europe". Philosophical Transactions of the Royal Society B, 376(1829), 20200277.

[44] Stanturf, J. A., Goodrick, S. L., Warren Jr, M. L., Charnley, S., Stegall, C. M., (2015), "Social vulnerability and Ebola virus disease in rural Liberia". PLoS One, 10(9): e0137208.

[45] Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M. P.,(2009), "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems". Journal of the Royal Society Interface, 6(31): 187-202.

[46] Truelove, S., Abrahim, O., Altare, C., Lauer, S. A., Grantz, K. H., Azman, A. S., Spiegel, P.,(2020), "The potential impact of COVID-19 in refugee camps in Bangladesh and beyond: A modeling study". PLoS medicine, 17(6), e1003144.

[47] Tye, E., Finnie, T., Hall, I., Leach, S., (2018), "PyGOM-A Python Package for Simplifying Modelling with Systems of Ordinary Differential Equations". arXiv preprint arXiv:1803.06934.

[48] Vergne, T., Gubbins, S., Guinat, C., Bauzile, B., Delpont, M., Chakraborty, D., Gruson, H., Roche, B., Andraud, M., Paul, M., Guérin, J. L. (2021), "Inferring within-flock transmission dynamics of highly pathogenic avian influenza H5N8 virus in France, 2020". Transboundary and emerging diseases, 68(6), 3151–3155.

[49] Wang, X., Gao, D., Wang, J., (2015), "Influence of human behavior on cholera dynamics". Mathematical biosciences, 267, 41–52.

[50] World Health Organization, 2021. Fighting misinformation in the time of COVID-19, one click at a time. https://www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time

[51] Zhong, N. S., Zheng, B. J., Li, Y. M., et al. (2003), "Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February". The Lancet, 2003, 362(9393): 1353-1358.

Appendix A

Python code

```
#import packages
import pygom
from pygom import DeterministicOde, Transition, SimulateOde, TransitionType
import numpy as np
from pygom import SquareLoss
from pygom.model import common_models
import pygom.approximate_bayesian_computation as pgabc
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import scipy.integrate
import scipy.stats

#print the figure of infections
data=pd.read_excel('kelley.xlsx',index_col=None)
x = data.loc[:, 'day']
y = data.loc[:, 'infected']
fig=plt.figure(figsize=[15,4])
plt.plot(x,y)
plt.ylabel('Incidence of cases by days')
plt.xlabel('day')
plt.show()

# ABC for SIR model
N = 689.0
ode = common_models.SIR_N({'beta':0.5, 'gamma':1.0/3.0, 'N':689.0})
x0 = [N-y[0], y[0], 0]#the initial state, normalized to one
t = np.linspace(0, 30, 30)#set the time sequence
ode.initial_values = (x0, t[0])
# setting the parameters in the inference
parameters = [pgabc.Parameter('beta', 'unif', 0, 3, logscale=False),
              pgabc.Parameter('gamma', 'unif', 0, 3, logscale=False)]
# creating the loss and abc objects
sir_obj = pgabc.create_loss(SquareLoss, parameters, ode, x0, t[0],
t[1::], y[1:] , ['I'])
sir_abc = pgabc.ABC(sir_obj, parameters)
# getting the posterior sample
sir_abc.get_posterior_sample(N=100, tol=np.inf, G=10, q=0.5)
sir_abc.continue_posterior_sample(N=100, tol=sir_abc.next_tol, G=10, q=0.5)
med_est = np.median(sir_abc.res, axis=0)
```

```
#plot posterior distributions
beta_sir=sir_abc.res[:,0]
gamma_sir=sir_abc.res[:,1]
fig,ax=plt.subplots(1,2,figsize=[12,4])
sns.distplot(beta_sir,ax=ax[0])
ax1=ax[0]
ax1.set_xlabel('beta')
sns.distplot(gamma_sir,ax=ax[1])
ax2=ax[1]
ax2.set_xlabel('gamma')
plt.subplots_adjust(wspace=0.3)
plt.show()

# SIR model simulation(deterministic)
ode_fit=common_models.SIR_N({'beta':med_est[0],'gamma':med_est[1],'N':689.0})
x0_fit=[N-y[0], y[0], 0]
t_fit=numpy.linspace(0,30,30)  #np.linspace(start stop number)
ode_fit.initial_values = (x0_fit, t_fit[0])
sol_fit = ode_fit.integrate(t_fit[1::])
plt.rcParams['figure.figsize'] = [12, 4]
ode_fit.plot()

# SIR model simulation(stochastic)
states = ['S', 'I', 'R']
params = ['beta', 'gamma', 'N']
transitions = [Transition(origin='S', destination='I',
               equation='beta*S*I/N',
               transition_type=TransitionType.T),
               Transition(origin='I', destination='R', equation='gamma*I',
               transition_type=TransitionType.T)]
# initial conditions
N = 689.0
in_inf = 1  #round(0.0000001*N)
init_state = [N - in_inf, in_inf, 0.0]
# time
t = np.linspace (0 , 30 , 30)
# deterministic parameter values
param_evals = [('beta', med_est[0]), ('gamma', med_est[1]), ('N', 689.0)]
# construct model
model_j = SimulateOde(states, params, transition=transitions)
model_j.parameters = param_evals
model_j.initial_values = (init_state, t[0])
# run 10 simulations
simX, simT = model_j.simulate_jump(t[1::], iteration=10, full_output=True)
plt.rcParams.update(plt.rcParamsDefault)
plt.rcParams['figure.figsize'] = [12, 4]
model_j.plot(simX,simT)

# Define SEIR model
def SEIR_N(param=None):
    state = ['S', 'E', 'I', 'R']
    param_list = ['beta', 'alpha', 'gamma','N']
    transition = [
        Transition(origin='S', destination='E', equation='beta*S*I/N',
                   transition_type=TransitionType.T),
        Transition(origin='E', destination='I', equation='alpha*E',
                   transition_type=TransitionType.T),
```

```
        Transition(origin='I', destination='R', equation='gamma*I',
                   transition_type=TransitionType.T)
        ]
    ode_obj = DeterministicOde(state, param_list, transition=transition)
    if param is None:
        return ode_obj
    else:
        ode_obj.parameters = param
        return ode_obj

# ABC for SEIR model
ode = SEIR_N({'beta':0.4, 'alpha':1.0/3.0 ,'gamma':1.0/3.0, 'N':689.0 })
x0 = [689-y[0], 0, y[0], 0]
t = np.linspace(0, 30, 30)
ode.initial_values = (x0, t[0])
parameters = [pgabc.Parameter('beta', 'unif',1,3,
              logscale=False),
              pgabc.Parameter('alpha', 'unif',0.1,1,
              logscale=False),
              pgabc.Parameter('gamma', 'unif', 0.1, 3,
              logscale=False)]
# creating the loss and abc objects
seir_obj = pgabc.create_loss(SquareLoss, parameters, ode, x0, t[0],
                          t[1::], y[1:] , ['I'])
seir_abc = pgabc.ABC(seir_obj, parameters)
# getting the posterior sample
seir_abc.get_posterior_sample(N=100, tol=np.inf, G=10, q=0.5)
seir_abc.continue_posterior_sample(N=100, tol=seir_abc.next_tol, G=10, q=0.5)
med_est_SEIR = np.median(seir_abc.res, axis=0)

# Define SEIAR model
def Influenza_SEIAR_N(param=None):
    state = ['S', 'E', 'I', 'A', 'R']
    param_list = ['beta', 'p', 'kappa', 'alpha', 'delta', 'epsilon','N']
    ode = [
        Transition(origin='S', equation='-beta*S*(I + delta*A)/N',
                   transition_type=TransitionType.ODE),
        Transition(origin='E', equation='beta*S*(I + delta*A)/N - kappa*E',
                   transition_type=TransitionType.ODE),
        Transition(origin='I', equation='p*kappa*E - alpha*I',
                   transition_type=TransitionType.ODE),
        Transition(origin='A', equation='(1-p)*kappa*E - epsilon*A',
                   transition_type=TransitionType.ODE),
        Transition(origin='R', equation='alpha*I + epsilon*A',
                   transition_type=TransitionType.ODE),
    ]
    # initialize the model
    ode_obj = DeterministicOde(state, param_list, ode=ode)
    if param is None:
        return ode_obj
    else:
        ode_obj.parameters = param
        return ode_obj

# ABC for SEIAR model
ode = Influenza_SEIAR_N({'beta':0.8, 'p':0.8 ,'kappa':0.5, 'alpha':0.5,
'delta':0.5,'epsilon':0.5, 'N':689 })
x0 = [N-y[0], 0, y[0],0,  0]
```

```
t = np.linspace(0, 30, 30)
ode.initial_values = (x0, t[0])
# setting the parameters in the inference
parameters = [pgabc.Parameter('beta', 'unif', 0, 5, logscale=False),
              pgabc.Parameter('p', 'unif', 0, 1,
              logscale=False),
              pgabc.Parameter('kappa', 'unif', 0, 1, logscale=False),
              pgabc.Parameter('alpha', 'unif', 0, 2, logscale=False),
              pgabc.Parameter('delta', 'unif',0, 1,
              logscale=False),
              pgabc.Parameter('epsilon', 'unif', 0, 2, logscale=False)]

# creating the loss and abc objects
seiar_obj = pgabc.create_loss(SquareLoss, parameters, ode, x0, t[0],
                              t[1::], y[1:] , ['I'])
seiar_abc = pgabc.ABC(seiar_obj, parameters)

# getting the posterior sample
seiar_abc.get_posterior_sample(N=100, tol=np.inf, G=10, q=0.5)
seiar_abc.continue_posterior_sample(N=100, tol=seiar_abc.next_tol,
                                    G=10, q=0.5)
med_est_SEIAR = np.median(seiar_abc.res, axis=0)

# Confidence interval
def median_confidence_interval(data,confidence=0.95):
    a = 1.0 * np.array(data)
    n = len(a)
    m,se = np.median(a), scipy.stats.sem(a)
    h = se * scipy.stats.t.ppf((1 + confidence)/2.,n-1)
    return m, m-h, m+h
```