A Method of Teaching English Speaking Learners to Produce Mandarin-Chinese Tones

Hui Shi
Evaluating the Impacts of Vaccination, Antiviral Treatment and School Closure on Influenza Epidemic

Xiaofei Shi

Thesis submitted to the Benjamin M. Statler College of Engineering and Mineral Resources at West Virginia University in partial fulfillment of the requirements for the degree of

Master of Science in Industrial Engineering

Feng Yang, Ph.D., Chair
Wafik Iskander, Ph.D.
Majid Jaridi, Ph.D.

Department of Industrial and Management Systems Engineering

Morgantown, West Virginia
2015

Keywords: Agent-based Simulation Model, Multi-object Optimization, Genetic Algorithm, Intervention Strategies

Copyright 2015 Xiaofei Shi
ABSTRACT

Evaluating the Impacts of Vaccination, Antiviral Treatment and School Closure on Influenza Epidemic

Xiaofei Shi

Multi-objective simulation optimization was performed to investigate the cost and benefits of the most commonly-used strategies for H1N1 epidemic mitigation: vaccination, antiviral treatment, and school closure. By simultaneously considering three intervention strategies, this study leads to findings that supplement those in the existing work, and provides additional insights regarding intervention decision making. Specifically, our investigation suggests that different vaccine prioritization strategies, the age-based versus ACIP recommendation, be implemented depending on vaccine availability; individual school closure policies are favored over their global counterparts, at least when both vaccination and antiviral treatment are implemented with relatively plentiful medicine supply. The trade-offs of cost and benefits of the intervention strategies were investigated, and can be used to support relevant decision making.
ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Feng Yang for her great guidance, for giving me the opportunity to work on this project, which could not have been done without her encouragement and support. I am also thankful to Dr. Majid Jaridi and Dr. Wafik Iskander for serving on my committee, and for their insightful ideas and assistance in preparing this thesis.
## Contents

List of Figures vi

List of Tables vii

List of Acronyms viii

1 Introduction 1

2 Methods 3

2.1 Agent-Based Simulation Model . . . . . . . . . . . . . . . . . . . . . . . . . 3

2.2 The FRED Framework . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

2.3 Genetic Algorithm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

2.4 Intervention Strategies and Decision Variables . . . . . . . . . . . . . . . . 6

2.4.1 Vaccines . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

2.4.2 Antiviral . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2.4.3 School Closure . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10

2.5 Performance Metrics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

2.6 Objective function . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

2.7 GA application . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14

2.7.1 Genetic Encoding . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

2.7.2 Fitness Calculation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

2.7.3 Selection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

2.7.4 Crossover and Mutation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
2.8 Determine the number of replications

3 Results and Discussions

4 Summary and Future Research

References
List of Figures

2.1 The Intervention Strategies and Decision Variables. ....................... 7
2.2 The Breakdown of Total Cost. ................................................. 12
2.3 The GA Workflow. ............................................................. 14
2.4 The Meaning of each digit. .................................................... 16
2.5 Genes Encoding ................................................................. 16
2.6 Tournament Selection ......................................................... 17
2.7 Crossover Operator ............................................................. 18

3.1 Solutions Resulting from the Simulation Optimization. .................... 21
3.2 Pareto Frontier and its Cost. ................................................. 21
List of Tables

2.1 Cost Parameters. ................................................................. 13

3.1 Three Selected Pareto Policies. ............................................. 22
List of Acronyms

**ABM**  Agent Based Model

**ACIP**  Advisory Committee on Immunization Practices

**GA**  Genetic Algorithm

**FRED**  Framework for Reconstructing Epidemiological Dynamics

**AR**  Attack Rate

**YLL**  Years of Life Lost

**TC**  Total Cost
Chapter 1

Introduction

Influenza is a serious public health problem that causes severe illness and death in high risk populations. To prevent an epidemic, a wide range of mitigation strategies have been investigated including vaccine allocation, use of antiviral, school closure policies, etc. In the current literature, agent-based simulation models (ABM) have been developed and served as an effective and popular method to investigate the mitigation effects of different intervention policies [1, 2, 3, 4]. However, the existing work is typically restricted to a small set of decision variables concerning a certain intervention aspect. For instance, Lee et al[3] compared the impacts of different vaccine prioritization policies when vaccine is in limited supply; Brown et al[4] evaluated various school closure policies specified by different closure lengths. These work focused on limited scenarios generated by varying one or two selected decision variables, evaluated those scenarios through ABM, and derived insights as to the main effects of the selected variables. Hence, they do not allow for the investigation of the potential interacting effects of mitigation policies (for example, the interacting effects of vaccine prioritization policies and the vaccine supply quantity), and neither are they adequate for the exploration of optimal policies concerning a range of mitigation strategies.

In light of the discussions above, an effort was made in this thesis to simultaneously consider a large number of decision variables that specify three intervention aspects: vaccination, antiviral treatment, and school closure policies. The stochastic ABM developed through the Models of Infectious Disease Agent Study (University of Pittsburgh MIDAS National Center of Excellence 2009) was adopted as the scenario evaluation tool for H1N1.
The ABM has been extended in this work to track a range of performance metrics including attack/death rate of the disease, infections, years of life lost, productivity losses, and the total cost incurred. Clearly, there are trade-offs between the intervention benefits and costs.

To explore the solutions in the large decision space while balancing the benefits and costs of invention strategies, we adapted a genetic algorithm (GA)-based multi-objective simulation optimization method. The Pareto frontier of the non-dominated solutions obtained was further investigated, and used to derive interesting insights regarding the intervention policies.
Chapter 2

Methods

2.1 Agent-Based Simulation Model

A large population and dynamic individuals network are usually required for studies of infections diseases, such as influenza. In previous studies, there have been two main approaches used to understand the spread of infection diseases: theoretical approach and experimental approach. However, when dealing with a large population and the dynamic individuals network among this population, these two approaches may have limitations. The system would be too complex to be mapped by analytical expressions or experiments, or the extensive computation time would be too expensive to build such experiments[5]. In recent years, a computational approach, which is called an agent-based model (ABM), is becoming more commonly used. This model is a rule-based, discrete-time, discrete-event computational modelling methodology for simulating the actions of agents that interact within an environment and other agents. Rule-based means using rules to make decisions; discrete-time means measuring variables of interest at each separate time step; discrete-event means a discrete sequence of events, each event occurs at a certain time period and makes a change to the system[6]. An agent-based model would provide explanatory insight into the actions and interactions of agents under certain rules in discrete time periods with a discrete sequence of events.

In this work, the ABM simulates the H1N1 influenza spread within Allegheny County, Pennsylvania, with a population of 1,164,779 individuals, 524,584 households, 532 schools, and 48,703 work places. Based on the data from H1N1 2009 influenza epidemic in the U.S.
(2009-2010), the reproductive rate of H1N1 is set as 1.8\cite{7}, and the simulation length is set as 180 days to reflect the length of the epidemic season\cite{8}.

### 2.2 The FRED Framework

FRED, the stochastic ABM available at the website of Framework for Reconstructing Epidemiological Dynamics (University of Pittsburgh Public Health Dynamics Laboratory 2013), was used in this simulation study. This ABM includes explicit representations of the county’s population with each agent (virtual person) possessing a set of sociodemographic characteristics and daily behaviors; provides a range of parameters to characterize a disease (e.g., the basic reproductive rate); and has a range of built-in actions for epidemic mitigation.\cite{9}

FRED has more than 400 parameters. The definition of most parameters can be found in FRED User Guide\cite{9}. Most of the input parameters can be divided into three categories: natural history parameters, transmission parameters, and health related behaviors. There are also several main sets of intervention parameters: the school closure intervention strategies, vaccination intervention strategies, and antiviral treatment strategies. The model outputs include fatality rate, symptomatic incidence, attack rate, reproductive ratio, etc. The FRED simulation model can be used to estimate the effect, cost, and efficiency of each intervention strategies and compare them for an influenza pandemic.

In the beginning of the simulation, FRED sets up the initial environment and assigns to each agent a range of properties. Each agent represents an individual living within the area considered. These properties include

- demographic information (e.g., age, gender etc.);
- health information (e.g., current health status, immunization, infectiousness etc.);
- location of social activity (e.g., household, neighborhood, school or workplace etc.);

and
• health-related behaviors (e.g., contact rate at different locations, probability of staying home when sick, probability of getting a vaccine etc.)

After initializing the environment and population, FRED inserts some number of infected cases into the population randomly. As the simulation proceeds, each agent will perform his daily activities. For instance, students will go to school, adults to work, and patients to hospital. Through these activities, a healthy agent may get infected when contacting an infected agent. A patient may be cured after accepting an antiviral treatment. At the end of a day, the property of agents will be updated and get ready for the next-day simulation. This circle is repeated until the last day of simulation period.

2.3 Genetic Algorithm

Genetic Algorithm is a heuristic search algorithm that mimics the process of natural selection, a biological process in which stronger individuals are likely be the winners in a competing environment. The potential solution of a problem is assumed as an individual and can be represented by a set of parameters, which are regarded as the genes of a chromosome and can be structured by a string of values in binary form. The fitness value is used to reflect the degree of goodness of the chromosome for solving the problem, and this value is closely related to its objective value.

In each cycle of genetic operation, the genes of the parents are to be mixed and recombined to produce the offspring in the next generation. It is expected that from this process of evolution, the chromosome with higher fitness will create a large number of offspring, and thus has a higher chance of surviving in the subsequent generation, such as the survival-of-the-fittest mechanism in nature. Throughout the evolving process, the fitter chromosome has the tendency to yield good-quality offspring, which means a better solution to the problem.[10]

A general GA process usually includes the following steps: initialization, selection, genetic operators and termination. For a given problem, the chromosomes are coded to
represent each possible solution. The initial population is generated randomly. Several parameters such as population size and mutation rate are set to process the evolving cycle.

To generate good offspring, a proficient parent selection mechanism is necessary. This is a process used to determine the number of trials for one particular individual used in reproduction.

The genetic operators include crossover operator and mutation operator. In this process, a new offspring is created which typically shares many of the characteristics of its parents using the above methods of crossover and mutation. These processes ultimately result in the next generation population of chromosomes that is different from the previous generation. Generally the average fitness will have increased by this procedure for the population.

The evolving cycle is repeated until a termination condition has been reached. For example: the fixed number of max generations is reached or a solution is found which satisfies minimum criteria.

### 2.4 Intervention Strategies and Decision Variables

Figure 2.1 provides an overview of the intervention strategies considered in this work. A total of 15 decision variables are used to define the vaccine, antiviral and school closure policies, and to specify a scenario of the combination strategies. As detailed in the rest of this subsection, a range of experimental values are selected for each variable to investigate the variable's quantitative impacts, which leads to 10,459,400 alternative policy combinations.

#### 2.4.1 Vaccines

Vaccination is the principal preventive strategy to prevent an outbreak of influenza epidemic and to reduce the disease burden of infectious diseases. Large quantities of vaccines are produced and distributed all over the world to prevent the outbreak of influenza such as H1N1 and H5N1. How do the vaccine availability and prioritization affect the effectiveness
of vaccination to mitigate an influenza epidemic? The vaccination plan and policies are specified by seven variables (V1)-(V7) in this work.

(V1) enable_vaccination: It is a Boolean variable, which can be either 0 or 1. A value of 0 corresponds to no implementation of vaccination strategies at all; and when (V1) is equal to 0, the variables (V2)-(V7) play no part in the simulation model. When (V1) takes a value of 1, the variables (V2)-(V7) come into play and their settings specify a vaccination plan to be implemented.

(V2) vaccine_prioritize_acip and (V3) vaccine_prioritize_by_age: The variable pair (V2, V3) is a two-digit Boolean vector, which could be either (1, 0) or (0, 1) corresponding to different types of vaccine prioritization policies detailed as follows.
If (V2, V3) is set as (1, 0), then the ACIP vaccine prioritization policy is adopted. ACIP refers to the vaccine prioritization recommended by the Advisory Committee on Immunization Practices (ACIP) on July 29, 2009. By adopting the ACIP, the following groups of people have higher priority to receive the influenza vaccine [3]: pregnant women, household contacts and caregivers for children younger than 6 months of age, healthcare and emergency medical services personnel, all people from 6 months through 24 years of age, and persons aged 25 through 64 years who have health conditions associated with higher risk of medical complications from influenza. The ACIP recommendation is primarily based on occupational risk of vulnerable populations.

If (V2, V3) is set as (0, 1), then the age-based prioritization policy is employed. The age-based policy requires dividing the population into several age groups. In this work, five age groups[11] are considered: (i) 0-5 years of age, (ii) 6-24 years of age, (iii) 25-44 years of age, (iv) 45-64 years of age, and (v) 65 years old and above. An age-based policy is defined by a selected age group with priority over other groups, and the selected group is specified by the variable pair (V4, V5) as below.

(V4) vaccine_priority_age_low and (V5) vaccine_priority_age_high: These two variables (V4, V5) specify the lower and upper bounds of the priority age group. Based on the five age groups (i)-(v) considered, (V4, V5) can take values (0, 5), (6, 24), (25, 44), (45, 64) and (65, 120) assuming that 120 is the highest possible age.

(V6) vaccine_total_avail: It is the total amount of vaccine that will be made available during the influenza season. Timely production and delivery of sufficient vaccine still remains a challenge, and variable (V6) constrains the maximum number of people that can be vaccinated. According to Fedson[12], the doses of influenza vaccine distributed in North America is about 26.5% of the population. Thus, in our exploration, (V6) is allowed to range from 5% to 20% of the population size. The population of Allegheny County, PA is about 1.2 millions, so the experimental value of (V6) is set as: 60,000, 80,000, 100,000, 150,000, 200,000, and 250,000.
(V7) vaccine\_additional\_per\_day: This variable represents the additional daily supply of vaccines. It is required that (V7) is less than or equal to (V6). Starting from the first day of the simulated influenza season, an additional amount of (V7) will be made available on each day until the total planned vaccine supply (V6) has been reached. Given (V6), the variable (V7) largely determines the vaccine availability over time. We allow (V7) to take the following values: 1,000, 2,000, 4,000, 6,000, 8,000, and 10,000 subject to the constraint (V7) ≤ (V6).

2.4.2 Antiviral

Antiviral drugs are a class of medication used specifically for treating viral infections, and hence antiviral policies are for treatment as opposed to prevention of the disease. In an influenza epidemic, the availability of antiviral drugs is a major concern, besides the effectiveness of the drugs[13]. The following four variables are used to specify an antiviral policy.

(A1) enable\_antivirals: It is a Boolean variable, which can be either 0 or 1. A value of 0 corresponds to no implementation of antiviral strategies at all; and when (A1) is equal to 0, the variables (A2)-(A4) play no part in the simulation model. When (A1) takes a value of 1, the variables (A2)-(A4) come into play and their settings specify an antiviral strategy to be implemented.

(A2) av\_total\_avail: It is the total amount of antiviral doses that will be made available in the epidemic season. A range of values have been selected for (A2): 10,000, 50,000, 100,000, 200,000, 300,000 and 400,000.

(A3) av\_initial\_stock: It is the initial stock of antivirals with healthcare agencies and hospitals at the beginning of epidemic season. The experimental values of (A3) are set as: 1,000, 5,000, 10,000, 20,000, 30,000, and 40,000, with (A3) ≤ (A2).

(A4) av\_additional\_per\_day: It is the daily antiviral doses added to supply, from the first day of the influenza season until the total planned amount (A2) has been reached. In
this work, the variable (A4) is allowed to be 1,000, 5,000, 10,000, 20,000, 30,000 and 40,000, subjecting to the constraint: (A3)+(A4) ≤(A2).

2.4.3 School Closure

As an indirect intervention strategy, school closure has been debated and questioned [14, 15, 16]. It has been noted that school closure could be very costly and perhaps not worthwhile, whereas at the same time, transmission among school children is also considered to be a primary mode of disease propagation[8, 1]). In this work, four decision variables are associated with school closure policies.

(S1) school_closure_policy: This variable is used to set the type of school closure policy, and it can be any of the three categories: none, global, and individual. When (S1) is set as “none”, school closure policy is disabled, and the variables (S2)-(S4) play no part in the simulation model. When (S1) is set as “global”, the global school closure policy is adopted and specified by the variables (S2) and (S4) below. When (S1) is set as “individual”, the individual school closure policy is employed and specified by the variables (S2), (S3) and (S4), as will be seen.

(S2) school_closure_threshold: When the global school closure policy is employed, (S2) is the threshold attack rate (defined later in Section2.5), above which all schools are to be closed. In Lee et al[2], a threshold of 1% is used, which means closure of all schools once the attack rate of the population being considered exceeds 1%. Herein, we allow (S2) to vary over the range of [0.5%, 1.5%], and take the following values: 0.5%, 0.7%, 0.9%, 1.1%, 1.3%, and 1.5%.

(S3) school_closure_cases: Under the individual school closure policy, a school will be closed if the number of infection cases at that individual school reaches the threshold (S3) or the global attack rate reaches threshold (S2). As in Brown et al[4], (S3) is set to range from 1 to 30, with experimental values: 1, 5, 10, 15, 20, 25, and 30.
(S4) Weeks: It is the duration of school closure once a closure has been triggered. Following Brown et al[4], (S4) ranges within [1, 8] weeks, and takes the following values in our work: 1, 2, 4, 6, and 8 weeks.

2.5 Performance Metrics

We have extended FRED to evaluate the five most commonly-used performance metrics [17, 11, 18, 19] for each scenario specified by a combination of the intervention policies (i.e., a setting of the decision variables). These performance metrics are described as follows.

1. ATTACK RATE (AR): An attack rate is the proportion of a well-defined population that develops illness over a specific period of time[20], where the numerator is the number of new cases that occurred during that period and the denominator is the size of the population at risk (i.e., the individuals with a probability of acquiring the disease in the area considered). Attack rate is an important indicator used to aid in marshalling resources for delivery of medical care as well as production of vaccines and/or anti-viral and anti-bacterial medicines[21].

2. DEATHS: Deaths is the outcome indicator that counts the number of deaths during an epidemic. Since deaths news tends to have severe mental and psychological effects on people, which may lead to psychological panic and even social chaos, the number of deaths is always one of primary indicators for epidemic severity[11]. A disease-associated death results in substantial social and economic losses.

3. INFECTIONS: The total number of infections in an epidemic is the major factor that determines the medical cost and productivity loss[11].

4. YEARS OF LIFE LOST (YLL): YLL measures the total number of life years lost due to premature death[11]. The definition of premature death varies in studies, and in this work, YLL are calculated following the 2008 US Life Expectancy data[22]. YLL measures health loss from early death, taking into account the age that death occurred.
5. TOTAL COST (TC): The total cost seeks to quantify the economic and social losses incurred in an epidemic. Figure 2.2 illustrates the breakdown of the total cost, which can be divided into two parts: the intervention cost, and the infection cost. (i) The intervention cost includes the cost to implement the intervention policies: the manufacturing and delivery cost of vaccines/antivirals, the cost accounting for the productivity loss of teachers, and school students missed cost (including student daily cost and parents productivity loss due to missing work), etc. (ii) The infection cost includes the cost triggered by disease infections, for which we distinguish two types of infected populations: the asymptomatic population that does not incur any infection cost yet (while being at increased risk for becoming symptomatic), and the symptomatic population. The cost incurred by symptomatically infected individuals is divided into two categories: death cost, and non-death cost caused by inpatients, outpatients and those who only need over-the-counter drugs. In each category, a patient incurs a cost of treatment depending on his/her age, vaccine condition, and a cost of productivity loss due to absenteeism. The cost parameters are given in Table 2.1.
Table 2.1: Cost Parameters.

<table>
<thead>
<tr>
<th>Cost</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vaccination cost (per dose)</td>
<td>$37.26</td>
</tr>
<tr>
<td>Antiviral cost (per course)</td>
<td>$134.30</td>
</tr>
<tr>
<td>Medical cost (per death)</td>
<td></td>
</tr>
<tr>
<td>age 0-19</td>
<td>$3435</td>
</tr>
<tr>
<td>age 20-64</td>
<td>$7605</td>
</tr>
<tr>
<td>age 65+</td>
<td>$8309</td>
</tr>
<tr>
<td>PV earning lost (per death)</td>
<td></td>
</tr>
<tr>
<td>age 0-19</td>
<td>$1016101</td>
</tr>
<tr>
<td>age 20-64</td>
<td>$1037673</td>
</tr>
<tr>
<td>age 65+</td>
<td>$65837</td>
</tr>
<tr>
<td>Non-death cost (per infection) (vaccinated)</td>
<td></td>
</tr>
<tr>
<td>age 0-19</td>
<td>$309.24</td>
</tr>
<tr>
<td>age 20-64</td>
<td>$377.42</td>
</tr>
<tr>
<td>age 65+</td>
<td>$517.46</td>
</tr>
<tr>
<td>Non-death cost (per infection) (unvaccinated)</td>
<td></td>
</tr>
<tr>
<td>age 0-19</td>
<td>$330.28</td>
</tr>
<tr>
<td>age 20-64</td>
<td>$418.02</td>
</tr>
<tr>
<td>age 65+</td>
<td>$557.97</td>
</tr>
<tr>
<td>Teacher productivity loss (per teacher per day) (2009 US Dollars)</td>
<td>$208</td>
</tr>
<tr>
<td>Student school missed cost (per student per day) (2009 US Dollars)</td>
<td>$19.22</td>
</tr>
</tbody>
</table>

2.6 Objective function

The ultimate goal is to find the intervention policies that lead to the best performance while minimizing the total cost involved. The objective function is formulated as: “Minimize both AR and TC with respect to the intervention policies specified by the decision variables (V1)-(V7), (A1)-(A4), (S1)-(S4).”

Since the first four performance metrics: AR, DEATHS, INFECTIONS and YLL are correlated with each other while emphasizing different aspects of epidemic severity, and since AR is most-widely used in the existing literature, we consider AR as the performance objective of the optimization. The cost objective is of course to minimize the total cost incurred for the epidemic. Hence, the objective function is generated.

With the decision variables (along with their possible settings) presented in Section 2.4, there are totally 10,459,400 scenarios of different intervention policies. It takes about 3 minutes to complete a single simulation run by a 16-core Dell workstation. In light of the
computational burden, a GA algorithm is adapted to search for the Pareto frontier of the multi-objective optimization problem.

2.7 GA application

The GA framework is given in Figure 2.3. A brief overview of Genetic Algorithm is provided below. It starts with the random generation of an initial population of candidate policies. The population size is, say 60. For each candidate in the current population, simulation experiments are carried out to evaluate the performance and cost associated with that candidate; \( N_{C_i} \) \( (i = 1, 2, \ldots, 60) \) simulation replications are obtained for each candidate policy using different random streams; and the simulation length of each replication is 180 days. Here, both optimization objectives are to be minimized. Following the fitness evaluation, a tournament selection, crossover and mutation are carried out to produce a new population of candidates for further evaluation and reproduction.
2.7.1 Genetic Encoding

At the beginning of Genetic Algorithm, we need to encode a string of genes which can be considered as chromosomes in the organism. In general, there are three encoding methods: a fixed length binary values (0,1), Arabic numbers (1,2,3,...) and alphabetical characters (A,B,C,...). In this study, the genes are encoded by 13-digit Arabic numbers. The 1st to 5th, 6th to 9th, and 10th to 13th genes represent Vaccine strategy, Antiviral strategy, and School Closure strategy respectively.

In Figure 2.4, the first row represents the value of each digit. The first column denotes the location of the digit. The second column represents the corresponding decision variable. The blank in the table implies that for a certain digit, it cannot take that value. For example, the value “6,000” in row 6, column 6 means that if the value of 5th digit in a string is equal to 3, the value of variable ”V7” is equal to 6000. Figure 2.5 gives an example of a string of genes.

As shown in Figure 2.1, with the experimental settings in Section 2.4, there is a total of 216 candidate vaccination policies, 199 candidate antiviral policies, and 240 candidate school-closure policies. A policy combination of the three intervention strategies is encoded as a 13-digit genetic string in the GA.

2.7.2 Fitness Calculation

An initial population containing 60 individuals should be first generated. Each individual is a random 13-digit string corresponding to a strategy combination. Feeding each of the 60 policies into FRED leads to the corresponding performance metrics of these policies.

Fitness evaluation is performed by using the Pareto-ranking method proposed by Goldberg [23]. The Pareto-ranking method explicitly utilizes the Pareto dominance rule to rank candidates, and assigns a fitness value to each candidate based on its ranking in the current population. To determine the rank, individuals in the population are considered as points which have two dimensions: Attack Rate as x-axis and Total Cost as y-axis. The
<table>
<thead>
<tr>
<th>Value</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st digit</td>
<td>V1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd digit</td>
<td>V2, V3</td>
<td>(1, 0)</td>
<td>(0, 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd digit</td>
<td>V4, V5</td>
<td>(0, 5)</td>
<td>(6, 24)</td>
<td>(25, 44)</td>
<td>(45, 64)</td>
<td>(65, 120)</td>
<td></td>
</tr>
<tr>
<td>4th digit</td>
<td>V6</td>
<td>60,000</td>
<td>80,000</td>
<td>100,000</td>
<td>150,000</td>
<td>200,000</td>
<td>250,000</td>
</tr>
<tr>
<td>5th digit</td>
<td>V7</td>
<td>1,000</td>
<td>2,000</td>
<td>4,000</td>
<td>6,000</td>
<td>8,000</td>
<td>10,000</td>
</tr>
<tr>
<td>6th digit</td>
<td>A1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7th digit</td>
<td>A2</td>
<td>10,000</td>
<td>50,000</td>
<td>100,000</td>
<td>200,000</td>
<td>300,000</td>
<td>400,000</td>
</tr>
<tr>
<td>8th digit</td>
<td>A3</td>
<td>1,000</td>
<td>5,000</td>
<td>10,000</td>
<td>20,000</td>
<td>30,000</td>
<td>40,000</td>
</tr>
<tr>
<td>9th digit</td>
<td>A4</td>
<td>1,000</td>
<td>5,000</td>
<td>10,000</td>
<td>20,000</td>
<td>30,000</td>
<td>40,000</td>
</tr>
<tr>
<td>10th digit</td>
<td>S1</td>
<td>Individual</td>
<td>Global</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11th digit</td>
<td>S2</td>
<td>0.5</td>
<td>0.7</td>
<td>0.9</td>
<td>1.1</td>
<td>1.3</td>
<td>1.5</td>
</tr>
<tr>
<td>12th digit</td>
<td>S3</td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>13th digit</td>
<td>S4</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.4: The Meaning of each digit.

V1=1 (V2,V3)=(0,1) (V4,V5)=(25,44) V6=150,000 V7=10,000
A1=1 A2=100,000 A3=1,000 A4=5,000
S1=“Individual” S2=0.9 S3=20 S4=4

Figure 2.5: Genes Encoding

points (individuals) which are closest to x-axis and y-axis at the same time among the whole population have rank one. Remove these points, and then find the set of points with rank two following the same scheme. Repeat this Pareto-ranking method and obtain the ranks of all the individuals in the current population.

2.7.3 Selection

Based on the fitness list, 60 individuals are chosen as parents to generate offsprings by tournament selection. Tournament selection involves running several “tournaments” among
a few individuals chosen at random from the population. The winner of each tournament (the one with the best fitness) is selected as parent(Figure 2.6). One of the most important factors to affect the result of tournament selection is tournament size. If the tournament size is large, weak individuals will have a small chance to be selected, but it is easier to fall into local optimum. In this study, a range of tournament sizes have been tested and it is finally set as one tenth of the population size.

2.7.4 Crossover and Mutation

Crossover operator for the chromosome is based upon the two-point crossover scheme. Two breakpoints are generated randomly, then the genes between two breakpoints are switched to generate two children (Figure 2.7).

Crossover leads to a fixed number of genes to be mutated. The mutation rate is recommended to be five percent[24] of the total number of genes in the whole population. There are two commonly-used mutation methods. One is to add 1 or -1 to the selected gene. The other is to generate a random number and to replace the selected gene by that number. In my study, the second method is applied.
The GA algorithm was implemented in Python, and integrated with the simulation model in C++ programming language. For each candidate solution generated by GA in Python, the corresponding parameter settings are written in a text file, which serves as input to the simulation model in C++; and the simulation results are sent back to the GA in a text file as well.

2.8 Determine the number of replications

For a candidate solution in the optimization search, multiple simulation replications are carried out to estimate its performance metrics. The number of replications, \( N_{C_i} \), required for each candidate, \( \{C_i; i = 1, 2, ..., 60\} \), which corresponds to the \( i^{th} \) candidate in a generation, to obtain estimates of \( E[AR(C_i)] \) and \( E[TC(C_i)] \). Denote a performance metric as \( M(C_i) \) in general, where \( M(C_i) \) can be \( AR(C_i) \) or \( TC(C_i) \). The two-stage method[25, 26] is utilized to determine the number of replications \( N \) so that a desired precision can be achieved for \( \hat{E}[M(C_i)] \).

\[
\frac{\hat{Std}[M(C_i)]}{\hat{E}[M(C_i)]} < \gamma \%
\] (2.1)

where \( \hat{Std}[\hat{E}[M(C_i)]] \) is the estimated standard deviation of \( \hat{E}[M(C_i)] \), and \( \gamma \% \) the target precision level. At the first stage, an initial number of replications \( N_0 \) are generated. Denote \( \hat{E}[M(C_i)] \) as the estimated mean of \( M(C_i) \) and \( \hat{Std}_0[\hat{E}_0[M(C_i)]] \) as the estimated standard
deviation of $\hat{E}_0[M(C_i)]$ from the $N_0$ replications. Then the number of replications $N$ that is likely to achieve the target precision level $\gamma \%$ is estimated as

$$N_{C_i} = \left\lceil \frac{\hat{Std}_0[M(C_i)]}{\hat{E}_0[M(C_i)] \times \gamma \%} \right\rceil$$

At the second state, $N_{C_i} - N_0$ additional replications are generated and the data collected at both stages are used to estimate the performance metrics.
Chapter 3

Results and Discussions

Since the objective function has two dimensions: AR and Total Cost, we use Pareto frontier to find a good solution in a set of points. Pareto frontier is a set of points where it is impossible to make any one objective value better off without making at least one objective value worse. In other words, there is no other point that can be closer to X-axis and Y-axis at same time than the Pareto frontier.

In the GA-based simulation optimization, 1470 candidate policies were evaluated, and their corresponding objective values, AR and Total Cost, are represented as dots in Figure 3.1. The dotted curve depicts the Pareto frontier of solutions, and a total of 8 intervention policies lie on the frontier. For each Pareto solution/policy, its pair of AR and school closure cost is plotted as square in Figure 3.2. In addition, the (AR, vaccination cost) pair is depicted as a cross, the (AR, antiviral cost) as a time symbol, the (AR, non-death infection cost) as diamond, and (AR, death infection cost) as star; the vaccination and antiviral costs are very low compared to the school closure cost, and the squares and crosses both lie along the bottom of Figure 3.2. For example, the three Pareto solutions/policies (marked as Pu, Pm and Pd in Figure 3.1) are given in Table 3.1, and their five performance metrics are also provided. As can be seen from Table 3.1, AR is positively correlated with DEATHS, INFECTIONS, and YLL. The simulation optimization results support the following findings.

- Vaccination

This study confirms that vaccination has significant impact on mitigating a H1N1 epi-
demic. All the policies included in the Pareto frontier call for a maximum total quantity of available vaccines by setting (V6) vaccine_total_avail at its maximum 250,000.

When vaccine supply is relatively plentiful (with the available vaccine quantity being more than 20% of the population as in these Pareto solutions), the age-based vaccine prioritization (with the priority group being the one from 0 to 5 years of age) is favored.
Table 3.1: Three Selected Pareto Policies.

<table>
<thead>
<tr>
<th>Pareto Policies</th>
<th>Decision Variables</th>
<th>Outcomes</th>
<th></th>
<th>YLL</th>
<th>Total Cost($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pu</td>
<td>$V_1=1, V_2=0, V_3=1, V_4=0, V_5=5, V_6=250000, V_7=10000, A_1=1, A_2=400000, A_3=10000, A_4=30000, S_1=\text{individual}, S_2=0.5, S_3=25, S_4=8$</td>
<td>15.738</td>
<td>38</td>
<td>191,809</td>
<td>1,598</td>
</tr>
<tr>
<td>Pm</td>
<td>$V_1=1, V_2=0, V_3=1, V_4=25, V_5=44, V_6=250000, V_7=10000, A_1=1, A_2=400000, A_3=40000, A_4=40000, S_1=\text{individual}, S_2=1.5, S_3=10, S_4=1$</td>
<td>24.534</td>
<td>61</td>
<td>298,991</td>
<td>2,412</td>
</tr>
<tr>
<td>Pd</td>
<td>$V_1=1, V_2=0, V_3=1, V_4=65, V_5=120, V_6=250000, V_7=10000, A_1=1, A_2=400000, A_3=10000, A_4=10000, S_1=\text{individual}, S_2=1.5, S_3=1, S_4=2$</td>
<td>30.993</td>
<td>74</td>
<td>377,725</td>
<td>3,073</td>
</tr>
</tbody>
</table>

over the ACIP recommendation. This complements the findings of Lee et al[3]: the ACIP prioritization should be adhered to when vaccine is in limited supply. By examining the GA candidate solutions with the minimum vaccine supply (with available vaccine quantity being 5% of the population), we also confirmed the superiority of ACIP in term of vaccine scarcity.

- School closure

The cost and benefits of school closure are reflected in Figure 3.1. Compared to vaccination and antiviral strategies, the cost of school closure is very high, which is a cause for debate whether school closure is worthwhile[4].
All the Pareto solutions adopt the individual school closure policy, which is consistent with the CDC recommendation[27] that school closures should be implemented at the discretion of local authorities based on local considerations. Lee et al. (2010)[2] was not able to show any significant differences between individual and global closure policies, at least partly because that in their work, the impacts of different school closure policies were evaluated with all the other intervention policies fixed at certain levels, levels different from those in our Pareto solutions.

The individual school closure policies involved in these Pareto solutions vary in terms of their thresholds for closure (i.e., the values for the variable (S3) school_closure_cases) and their closure periods (i.e., the values for the variable (S4) Week). As illustrated in Table 3.1, the Pareto solutions are similar in terms of the vaccination and antiviral treatment policies. Hence, the times symbols in Figure 3.1 for school closure cost vs. AR indicates: The individual closure parameters, (S3) school_closure_cases and (S4) Week, affect both the school closure cost and the AR performance. This finding is again not given by Lee et al[3], who claim that AR is not sensitive to such variables. The trade-offs between cost and benefits are displayed in Figure 3.1, and can be employed for decision making regarding school closures.

- Antiviral Policies

For each of the Pareto solutions, a large quantity of antiviral drugs is utilized with the variable (A2) av_total_avail being in its highest allowed value. As vaccination, antiviral strategies are relatively cost-effective, and hence tend to be pushed to its upper limit in the optimization iteration.
Chapter 4

Summary and Future Research

A GA-based simulation optimization method was employed to explore the large decision space spanned by the decision variables for three types of intervention strategies: vaccination, antiviral treatment, and school closure. Our investigation suggests that different vaccine prioritization strategies be implemented depending on the vaccine availability: Age-based priority (with the group from 6 to 24 years of age being the priority group) is favored when vaccine supply is plentiful; and the ACIP is preferred in time of vaccine shortage. As far as school closure strategies are concerned, individual school closure policies appear to work better than their global counterparts, at least when both vaccination and antiviral treatment are both implemented with relatively plentiful medicine supply. In addition, the threshold and duration variables used to specify an individual closure policy tend to affect its cost and resulting benefits, the trade-offs of which have been graphically illustrated and can be used to support decision making.

Simulation-based optimization could be extremely time-consuming, and hence does not support real-time decision making. In the future work, metamodeling approach will be investigated to build a metamodel relating the performance metrics to the various decision variables. A metamodel is expected to have the high-fidelity of simulation as well as to allow for real-time “what if” analysis. Metamodel-based optimization will be substantially faster than its simulation-based counterparts, and thus has to potential to assist in real-time decision making.
References


