SENSITIVITY ANALYSIS AND UNCERTAINTY ANALYSIS IN A LARGE-SCALE AGENT-BASED SIMULATION MODEL OF INFECTIOUS DISEASES

by

Xiaozhi Zhou

B.S., Peking University, China, 2005

M.S., Texas Tech University, 2008

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

GRADUTE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Xiaozhi Zhou

It was defended on

December 4, 2013

and approved by

Douglas P. Landsittel, PhD, Professor
Department of Medicine
School of Medicine, University of Pittsburgh

Margaret Potter, J.D., Professor
Department of Health Policy and Management
Graduate School of Public Health, University of Pittsburgh

Hasan Guclu, PhD, Assistant Professor
Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

Jagpreet Chhatwal, PhD, Assistant Professor
Department of Health Services Research
MD Anderson Cancer Center, Houston, TX

Dissertation Advisor: John J. Grefenstette, PhD, Professor
Department of Biostatistics
Graduate School of Public Health, University of Pittsburgh

SENSITIVITY ANALYSIS AND UNCERTAINTY ANALYSIS IN A LARGE-SCALE AGENT-BASED SIMULATION MODEL OF INFECTIOUS DISEASES

Xiaozhi Zhou, PhD

University of Pittsburgh, 2013

ABSTRACT

The purpose of this study is to develop appropriate statistical methods and procedures for dealing with parameter uncertainty and for improving the computational efficiency of sensitivity analysis in a large-scale agent-based model of infectious disease. An agent-based model is a rule-based computational simulation model that can keep track of the dynamical activities of all agents and their interactions within an environment and analyze the course of a disease through the population and evaluate interventions. Sensitivity analysis is a method for quantifying uncertainty in a complex model by systematically changing inputs (parameters and initial conditions) of the model and quantifying the consequences for the output of the model. Sensitivity analysis and uncertainty analysis are used for agent-based model to analyze the uncertainty in the model.

The specific aims of the study are to (1) develop specific procedures and criteria to determine important input parameters in the FRED agent-based influenza model; (2) develop specific procedures and criteria to determine high sensitivity parameters in the FRED agent-based influenza model via local sensitivity analysis; (3) improve the computational efficiency of sensitivity analysis by comparing two sampling procedures for probabilistic sensitivity analysis

in agent-based models: simple random sampling and Latin Hypercube sampling; and (4) apply uncertainty analysis procedures to evaluate the cost-effectiveness for different school closure intervention strategies as well as the reliability of the uncertainty analysis in the FRED agent-based influenza model.

This study emphasizes the important role of sensitivity analysis, uncertainty analysis and statistical analysis in making better use of simulation results for decision-making in the control of infectious disease. In this study, the FRED (Framework for Replicating Epidemic Dynamics) influenza model is used to produce all the simulation results from sensitivity analysis. The methods and procedures that are developed in this study can be generalized to all kinds of disease models under the FRED framework.

In public health practice, this study will help to provide timely responses for decision-making when there is a public health crisis. It also provides important information for public health policy makers about how certainly the FRED framework can provide reliable intervention comparison results for decision-making.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

The work presented in this dissertation represents the technical side of obtaining a doctoral degree in biostatistics, for which I owe a great deal of thanks to several people. First, I would like to express my deepest gratitude to my advisor, Dr. John Grefenstette. I have learned so much from working with him over the past several years.

I would also like to thank Prof. Maggie Potter, who is my graduate student researcher advisor as well as a member of my dissertation committee. She provided solid financial support in my first several years of PhD study and also provided me with valuable insights in doing research in public health issues during my dissertation study.

Many thanks to members of my dissertation committee Dr. Douglas Landsittel, Dr. Hasan Guclu, and Dr. Jagpreet Chhatwal, who actively participated in my dissertation study and provided me with thoughtful suggestions. In particular, I'd like to thank Dr. Shawn Brown, who provided helpful suggestions at the beginning of my dissertation study.

Last, I would like to thank my family, to whom I cannot possibly fully express my gratitude for all that they have done for me. Mom, Dad, and Chuan, thank you for your unending support and love. Without that, none of this would have been possible.

CHAPTER 1 INTRODUCTION

Sensitivity analysis is a method for quantifying uncertainty in any type of complex model by systematically changing inputs (parameters and initial conditions) of a model and quantifying the resulting change to the outputs. This study addresses the important role of sensitivity analysis, uncertainty analysis and statistical analysis in making better use of agent-based simulation models for decision-making in the control of infectious disease. An agent-based model is a rule-based computational simulation model that can keep track of the dynamical activities of all agents and their interactions within an environment and analyze the course of a disease through the population and evaluate interventions. Since an agent-based model is a stochastic simulation model that usually involves a number of input parameters with estimated values, it requires sensitivity analysis and uncertainty analysis to analyze the uncertainty in the model.

This chapter introduces the background in modeling infectious disease (Section 1.1) and agent-based model (Section 1.2), reviews previous studies that used agent-based model in infectious diseases (Section 1.3), presents the background in the FRED framework (Section 1.4), introduces the background in uncertainty in the simulation model (Section 1.5), sensitivity analysis (Section 1.6) and sampling procedures (Section 1.7). The objectives of this research are stated in Section 1.8, and an outline of the remainder of the thesis is given in Section 1.9. A list of key definitions and acronyms used in this study is provided in Section 2.0.

1.1 MODELING INFECTIOUS DISEASES

Infectious diseases are passed between individuals by direct or indirect transmissions. The primary risk factor of infectious diseases is the presence of infectious cases in the local population. Due to the variety of transmission types and the dynamics of pathogen, infectious diseases with their complex epidemiological behaviors comprise the disease dynamics that require qualitative descriptions of disease dynamics from individual-level dynamics to population-scale epidemic as well as models that deal with variability in infection profiles, parameter values, and timescales (1).

The primary goal of studying infectious diseases is to improve control and reduce the infection in the population. In infectious diseases research, models provide a powerful tool for optimizing the use of limited resources, for planning the use of intervention measures more efficiently and therefore limiting the disease spread. In epidemiology, models not only provide translational information from a known set of conditions or from various scales of behaviors, but also allow us to predict the population-level epidemic dynamics from individual-level dynamics and elucidate how an infectious disease spreads in the real world. However, models also have limitations in their accuracy due to uncertainty of parameters(1).

There are some widely used epidemiological models in the studies of infectious diseases. The simple epidemic models are built based on the fundamental infection classifications: susceptible, exposed, infectious, and recovered (SEIR). SEIR and its reduced model SIR (ignoring exposed class) describe host-pathogen systems with lifelong protection after infection. There are some other reduced models from SEIR: SI (leading to death, usually in fatal infectious diseases), SIRS (immunity is not permanent), and SIS (susceptible after infection, usually in sexually transmitted infections). These models are usually deterministic compartmental

modelsthat assign individuals in the population to different subgroups or compartments, and each of compartments represents a specific stage of the epidemic (2). Since individuals in deterministic models are studied as groups, this kind of model may be less realistic than stochastic dynamic model, which depends on the chance variations in risk of exposure, disease and other illness dynamics. However, stochastic dynamic model may only be good for small population or rare diseases (1).

1.2 AGENT-BASED MODELS

Studies of infectious diseases, such as flu transmission, usually require a large population and dynamical network among individuals. Previous studies have usually used one of two main approaches to understand the spread of infectious diseases: theory and experiment. However, these two approaches may have limitations when studying a large population and the dynamical network among this population, because this system would be too complex to be captured by analytical expressions or experiments, or the extensive computation time would be too expensive to conduct such experiments (3). In recent years, a computational approach for simulating the actions of agents that interact within an environment is becoming more commonly used. This approach is called an agent-based model (ABM). An agent-based model is a rule-based, discrete-time, discrete-event computational modeling methodology. Rule-based is defined as using rules to make decisions, choices, or movements; discretetime is defined as measuring variables of interest at each separate time step; discrete-event is defined as a discrete sequence of events, each event occurs at a particular time period and makes a change to the system. In other words, an agent-based model may provide explanatory insight into the actions and interactions of agents

obeying certain rules in discrete time periods with a discrete sequence of events. In this work, agents in our model represent individuals in a population living in a specific geographic region. Each agent has recorded demographic information, activities and behavior strategies during an infectious disease breakout. Simulations are run based on above recorded information and the dynamic activities of all agents in the model. Those simulation results can be used to analyze the course of an infection through the population and to evaluate possible interventions. Like other simulation models, agent-based model requires sensitivity analysis and other methods for quantifying the uncertainty in the model. Some advantages of agent-based model include features in stochasticity, spatiality and heterogeneity. Stochasticity means that an agent-based model allows us to estimate the effects of random variations in events over time; spatiality means that agents and their actions are referenced to particular locations; heterogeneity means that agents differ from each other, for example, in age, sex, race or income level(3). Agent-based model also have advantages including the ability to measure precisely things that could not be measured in the real world, such as who infects whom, the ability to systematically vary the probability distributions of input parameters representing the estimates of model, coherence (agents' activities follow the ordinary rules for probability calculations and consistent decisions can be obtained from these probabilities) and explicit representation of process that introduces agent actions and interactions directly into an environment(3, 4). If large-scale computer resources such a supercomputer are used, the simulation procedure of agent-based model require short time to complete and provide timely results in response to crisis (5).

1.3 CURRENT PRACTICES IN AGENT-BASED MODELS

In previous studies, agent-based simulation models were used to study infectious disease transmission, and the effects of vaccination strategies or school closure interventions on pandemics. Epstein (6) stated that agent-based computational models can capture complex social networks and the direct contacts between individuals, adapt their behaviors based on disease prevalence in confronting H1N1 influenza pandemics. In Epstein's summary of modeling to contain pandemics, agent-based models helped to shape H5N1 flu policy and design containment strategies for smallpox. Burke et al. (7) developed an large-scale agent-based computational model to evaluate eight response scenarios at two epidemic scales for smallpox. They found that mass reactive vaccination of either 40% or 80% of the total population provided some additional protection of the population, reduced the mean number of infected people, and shortened the mean epidemic duration. However, school closure interventions without mass vaccination appeared to have little additional protection. The results indicated the importance of contact tracing and vaccination in smallpox epidemic and revealed that the agent-based simulation model provided to be a valuable tool in crafting policy in response to outbreak. Similar results were found in the study by Longini et al. (8) based on a discrete-time, stochastic simulation model of smallpox spread within a structural population. Halloran et al. (9) developed three stochastic, spatially structured, agent-based discrete time simulation models to access the effectiveness of a set of potentially feasible intervention strategies, such as quarantine, isolation, school closure, community and workplace social distancing etc., in influenza pandemic. The results showed that timely initiation of measure and school closure played an important role, and timely implementation of a combination of targeted household antiviral prophylaxis and social distancing measures could substantially lower the illness attack rate. Two similar studies

conducted by Lee et al. (10, 11) developed a set of computer simulation models to evaluate the economic impact of vaccination for older adults (65 years and older). The results showed that vaccination is more cost-effective than no vaccination, and the timing of annual influenza vaccination makes a difference. Both studies also discussed that such agent-based simulation models were simplifications of reality and may not be able to account for every possible factor or interaction. However, those agent-based simulation model provided useful information to decision makers about possible scenarios and relationships.

Previous studies also used agent-based simulation models to evaluate vaccination strategies and school closure intervention strategies for the 2009 H1N1 influenza pandemic. Shi (12) developed an agent-based simulation model and used data from the state of Georgia to investigate the effects of viral mutation and seasonality on the course of an influenza pandemic. The results showed that the time an epidemic started and the time viral mutations introduced had different impacts on seasonality and the initial wave's peak prevalence, respectively. Two studies conducted by Lee et al. (13, 14) employed an agent-based simulation model of the greater Washington, DC, metropolitan region to compare several vaccination strategies and determine the effects of employee vaccination to mitigate the 2009 H1N1 influenza pandemic. The results indicated that children should receive highest priority for vaccination when vaccine supply was limited. The conclusions supported adherence to the ACIP (Advisory Committee on Immunization Practices) prioritization recommendation. Results also revealed that timely vaccination of at least 20% of the large company workforce was as effective in mitigating the epidemic as vaccinating all workplaces. Previous studies on school closure strategies showed that individual school closures were more effective than entire school system closures, and longer duration of school closure (e.g., 8 weeks) provided additional time to implement a second

most effective intervention such as vaccination (15). A study showed that longer duration of school closure was more cost-effective than short duration of school closure, which suggested that school closure strategy that is targeted to school-aged children should be combined with interventions that target other age populations to yield a more effective intervention strategy for mitigation (5).

1.4 THE FRED FRAMEWORK

In this study, a large-scale agent-based framework of infectious diseases developed by the University of Pittsburgh Public Health Dynamics Lab (PHDL) will be used. This framework is an open source, object-oriented system called FRED (Framework for Replicating Epidemic Dynamics)(16). In the FRED framework, each agent represents an individual person living within a population (e.g., Allegheny County, Pennsylvania). Each agent has its demographic information (e.g., age, gender etc.), health information (e.g., current health status, date of infection, level of symptoms, length of being symptomatic, infectiousness, susceptibility, etc.), location of social activity (e.g., household, neighborhood, school or workplace etc.), and health-related behaviors (e.g., probability of staying home when sick, probability of getting a vaccine, contact rate at different locations etc.). The FRED framework can be used to generate a large number of different disease models by setting specific sets of parameters.. In this study, we will focus on the FRED influenza model. However, all the methods and procedures developed in this study apply to any FRED model.

In the FRED influenza model, there are three important sets of input parameters: natural history parameters, transmission parameters and health related behaviors, and two main sets of

intervention parameters: the school closure intervention strategies and vaccination intervention strategies. The model outputs include attack rate, clinical attack rate, reproductive ratio, number of cases, etc. The FRED model can be used to compare different intervention strategies for an influenza pandemic. One advantage of using such simulation model is that we can include no-treatment intervention option in the analysis since it is difficult and unethical to obtain the no-treatment samples from real world health care trials.

Initially, the FRED model initializes an epidemic by inserting some number of infected cases into a population. For each run of the simulation, the model will consider the dynamical activities within the whole population and their consequent effects on the model outputs, and provide the details of model outputs for each day of the full period of epidemic outbreak. The time required to complete one run of the simulation depends on the size of population chosen in the model and the processing speed (CPU) of the computer. For example, it would take about one minute to run a simulation of Allegheny County, PA with a population of about 1.2 million people in a quad-core 2.8 GHz i7 computer, while it would take about 15 seconds to simulate a small county such as Washington County, PA with 200,000 population in the same computer. For each set of input parameters, repeated runs (e.g., 20 runs) of simulations will be performed to obtain a stable and reliable mean model outputs. The FRED model could computationally be very intensive when dealing with large sets of samples from input parameters with several intervention conditions. Long running time is a very important issue because longer time of simulation runs could be very expensive, especially in large-scale simulation models. We address the appropriate approaches for improving the computational efficiency of sensitivity analysis in this study.. Computational efficiency is defined as reduced sample size and simulation time with the same or better accuracy of the outputs.

1.5 UNCERTAINTY IN SIMULATION MODELS

Input parameters for most mathematical models have a limited degree of certainty because of natural variation, or a lack of current techniques to measure them. A commonly used algorithm for dealing with the uncertainty in computational models is Monte Carlo simulation, which includes the technique of statistical sampling to solve quantitative problems. A Monte Carlo simulation uses random numbers sampled from probability distributions to perform multiple model evaluations that can be used to both determine the uncertainty in model output and perform sensitivity analysis (17).

There are two main types of uncertainty: model uncertainty and parameter uncertainty. Model uncertainty refers to uncertainty about the basic assumptions made and abstractions used to build the model. Model uncertainty reflects that fact that the model can only express an approximation to reality. Parameter uncertainty is defined as the degree to which the exact values of parameters are unknown. While model uncertainty is perhaps more difficult to evaluate, parameter uncertainty can be addressed through a process called sensitivity analysis.

Sensitivity analysis is a method for quantifying uncertainty in any type of complex model by systematically changing inputs (parameters and initial conditions) of a model and quantifying the resulting change to the outputs. Previous studies have shown that sensitivity analysis not only plays an important role in model verification and validation throughout the course of model development and refinement(18), but also provides insight into the robustness of model results for decision making (19, 20).The general procedure of performing sensitivity analysis involves the following steps (21):First, identify key input parameters and define distributions to characterize the uncertain parameters. Second, use a sampling procedure (random sampling, importance sampling, and Latin Hypercube sampling etc.) to generate the samples of the

uncertain parameters. Third, propagate the samples to generate the responses from the model. Fourth, perform statistical methods such as regression analysis or analysis of variance to assess the impact of the uncertain parameters on the model output. More details of the procedures of performing sensitivity analysis will be discussed in Chapter 3 and 4.

1.6 SENSITIVITY ANALYSIS

Sensitivity analysis is a technique for systematically changing inputs (parameters and initial conditions) of a model and quantifying the changes in the outputs. It allows one to study the relationships between inputs and outputs, and identify the most significant factors or variables affecting the model outputs (22). Methods of sensitivity and uncertainty analyses are based either on deterministic or statistical concepts. The traditional approach used for sensitivity measure is computing numerical values for input parameters in multiple simulations. Other approaches such as probabilistic sensitivity analysis is encoding probability distributions rather than point estimates for key model parameters (17, 21).

There are two types of sensitivity analyses: local sensitivity analysis and global sensitivity analysis. Local sensitivity analysis studies how the model behaviors when varying each input parameter one at a time. Global sensitivity analysis varies all parameters over their full ranges and looks at several different output measures to understand the full range of model behavior.

Based on the techniques used in the analyses, there are three main classes of sensitivity analysis methods: (1) mathematical methods, (2) statistical methods, and (3) graphical methods (19). Mathematical methods assess the sensitivity of a model output by a range of variation of an

input. Mathematical methods are used for local sensitivity analysis in deterministic models. Those methods include nominal range sensitivity analysis, difference in log-odds ratio method etc(19). For example, nominal range sensitivity is used to evaluate the effects of how a model output affected by individually varying one of the model inputs across its entire possible range of values while controlling for the other inputs at base levels or average values. It is relatively simple and easily applied, and works well with linear models. The difference in log-odds ratio method is similar to nominal range sensitivity analysis but is more useful when the model output is a probability. Both methods have limitations to correlated inputs, nonlinear interactions and the combinational explosion of possible cases, but are efficient for verification and validation of linear models (19, 23).

The second class of sensitivity analysis methods consists of statistical method. For example, probabilistic sensitivity analysis, which is also called uncertainty analysis, involves running simulations based on samples of inputs from anassigned probability distributions and assessing the effect of variances from inputs on the outputs. This method is usually used in global sensitivity analysis for accessing the uncertainty of a set of input parameters. Statistical methods are used for probabilistic models, for example regression analysis and analysis of variance (ANOVA). Regression analysis is performed on an independent sample of data and fits the linear relationship of inputs and outputs. It can also be used for some nonlinear cases where the nonlinear models can be transformed into linear models in the case of generalized linear models (GLMs) such as logistic regression models and Poisson regression models (24). The advantage of using regression analysis is that it allows to evaluate sensitivity of individual model inputs while taking into account the simultaneous effects of other model inputs on the outputs (19, 23). However, regression analysis may not work well when its normality and independence

assumptions are violated. If the relationships between the input parameters and the responses are nonlinear, the regression model will perform poorly. The analysis of variance (ANOVA) also assumes that the output is normally distributed. It can address categorical inputs and groups of inputs. However, it may become computationally intensive if there are a large number of inputs. In such case, a nominal range sensitivity analysis can be used for screening out low sensitivity inputs and reducing the number of inputs before performing an analysis of variance (19, 25).

The third class of sensitivity analysis methods is graphical methods, which provide the sensitivity in the form of graphs, charts, or surfaces. Graphical methods are used for probabilistic models(19). Since graphical methods provide the means to qualitatively visualize sensitivity, they are used as either a screening method at the beginning of model building or a complement result for mathematical and statistical methods. A commonly used graphical method is scatter plot, which visually assesses the effects of inputs on outputs and identifies potentially complex dependencies between inputs and outputs. It is frequently used as a first step before performing a regression analysis. However, it may not work well when there are a lot of inputs and outputs (19).

More details about local sensitivity analysis and probabilistic sensitivity analysis will be discussed in Chapter 3 and 4, respectively.

1.7 SAMPLING PROCEDURES

In order to perform sensitivity analysis more efficiently, a reasonable and appropriate sampling procedure should be considered carefully. There are several sampling procedures that can be used to sample the values of input parameters. For example, simple random sampling procedure

is commonly used to randomly draw samples from the sample space without any restrictions of the samples. This method is the simplest way for simulation models but may less efficient in obtaining reliable results for complex simulation models because it may require large sample size. Another commonly used sampling procedure is the stratified random sampling method, which will weight the samples from different strata and provide more precise estimates for population means and variances. An alternative method to the stratified random sampling is unequal-probability sampling procedure, which will draw stratified samples based on different probability in each stratum and reduce the sampling variance from strata. For sampling the values of input parameters, two methods are commonly used, systematic sampling and Latin Hypercube sampling. A systematic sample consists of equally spaced units, and will perform analysis from all spaced units. This method would be good for input parameters with small ranges. The Latin Hypercube sampling method splits parameter space into equal probability intervals and samples from each interval without replacement. This method is recommended for parameters with wider ranges to reduce the sample sizes of input values and improve the computational efficiency of the sensitivity analysis process(26). More details about sampling procedures will be discussed in Chapter 4.

1.8 RESEARCH OBJECTIVES

The purpose of this study is to develop appropriate statistical methods and procedures for dealing with parameter uncertainty and improving the computational efficiency of sensitivity analysis in a large-scale agent-based model of infectious disease. In particular, this study uses local sensitivity analysis, probabilistic sensitivity analysis (uncertainty analysis), statistical

sampling procedures, and statistical methods to develop specific procedures and criteria to (1) determine important input parameters in the FRED influenza model, (2) improve the computational efficiency of sensitivity analysis by comparing two sampling procedures: simple random sampling and Latin Hypercube sampling, and (3) apply the developed procedures to evaluate the effects and the cost-effectiveness for different influenza intervention strategies.

This study emphasizes the important role of sensitivity analysis, uncertainty analysis and statistical analysis in computer simulation models and the use of those analyses in such models to improve the computational efficiency of sensitivity analysis. The results of this study will promote the improved use of simulation models for decision-making by making clear the level of uncertainty in the reported results. This study also considers the public health aspects of using agent-based models to prevent and manage diseases, evaluate public health interventions and provide more accurate information for public decision-makers.

1.9 THESIS OUTLINE

Chapter 2 presents the results of a literature review on previous studies to determine the reference values and distributions for potential important parameters in the FRED influenza model. Chapter 3 focuses on the local sensitivity analysis approach for identifying high-priority input parameters for future research in FRED model. Chapter 4 compares simple random sampling and Latin hypercube sampling procedures in probabilistic sensitivity analysis (uncertainty analysis) to improve the computational efficiency of sensitivity analysis. Chapter 5 applies the procedures developed in chapter 4 to evaluate several different intervention strategies of school closure. Chapter 6 contains general conclusions and discussions of this study.

1.10 LIST OF KEY DEFINITIONS AND ACRONYMS

Table 1 provides a list of key definitions and acronyms used in this study.

Table 1. List of key definitions and acronyms

Term	Definition	Acronym	Section
Agent-based model	A rule-based, discrete-time, discrete-event computational modeling methodology	ABM	1.2
FRED framework	An open source, object-oriented agent-based modeling system for different diseases. FRED stands for Framework for Replicating Epidemic Dynamics	FRED	1.4
Uncertainty analysis	The process of obtaining the probability associated with a policy decision based on a model, given the probability distributions of the input parameters of the model		1.5, 5.1
Sensitivity analysis	A method for quantifying uncertainty in any type of complex model by systematically changing inputs (parameters and initial conditions) of a model and quantifying the resulting change to the outputs		1.5, 1.6
Local sensitivity analysis	Studies how the model behaviors when varying each input parameter one at a time	LSA	1.6, 3.1
Global sensitivity analysis	Varies all parameters over their full ranges and looks at several different output measures to understand the full range of model behavior		1.6, 4.1
Probabilistic sensitivity analysis	The process of obtaining the probability distribution of the responses of a model given the probability distributions of the input parameters of the model	PSA	4.2
Random sampling	A subset of individuals (a sample) chosen from a larger set (a population), each individual is chosen randomly and entirely by chance		1.7, 4.3

Table 1 continued

Term	Definition	Acronym	Section
Latin Hypercube sampling	A sampling procedure that splits parameter space into equal probability intervals and samples from each interval without replacement	LHS	1.7, 4.3
Attack rate	Cumulative incidence of infection in a population observed over a period of time during an epidemic		Chapter 3-5
Cumulative distribution function	A function describes the probability that the random variable X takes on a value less than or equal to a value x	CDF	3.3
Probability density function	A function that describes the relative likelihood for the random variable X to take on a given value x	PDF	3.3

CHAPTER 2 KEY PARAMETERS IN FRED INFLUENZA MODEL

This chapter presents the results of literature review for natural history parameters (Section 2.1.1), transmission parameters (Section 2.1.2) and health related behavior parameters (Section 2.1.3) in FRED influenza model. A summary of reference parameter values is present in Section 2.2, followed by discussions and conclusions in Section 2.3. This chapter provides literature review for reference parameter values that will be used in Chapter 3-5.

2.1 LITERATURE REVIEW FOR PARAMETERS IN FRED INFLUENZA MODEL

In the FRED framework, there are three categories of important epidemiological parameters that are used to generate models of particular diseases: natural history parameters (such as latent period, symptomatic period, symptomatic rate, mortality rate etc.), transmission parameters (such as the probability of transmission, asymptomatic infectivity etc.), and health related behavior parameters (such as the probability of staying home when sick, contact rate at different locations etc.). This Section describes the results of a literature review to identify the reference values and distributions for some important epidemiological parameters in FRED model. All the reviews for parameters in FRED model are mainly focused on influenza H1N1 pandemic.

2.1.1 Natural history parameters

In FRED model, the natural history parameters for influenza pandemic include latent period, asymptomatic period, symptomatic period, symptomatic rate, immunity loss rate, and mortality rate. Latent period is defined as the interval between the receipt of infection and the onset of the first symptoms of the illness. Mortality rate for seasonal flu is vary widely from year to year and is also differed by the type of influenza. Symptomatic period is defined as the period of being infectious and symptomatic. In contrast, asymptomatic period is defined as the period of being infectious but asymptomatic. Symptomatic rate in influenza pandemic is defined as the probability that infected persons will be symptomatic. Table 2 summarizes the definitions of natural history parameters in FRED model.

Table 2. Definitions of natural history parameters in FRED model

Parameter	Definition
Latent period	The interval between the receipt of infection and the onset of the first symptoms of the illness
Symptomatic period	The period of being infectious and symptomatic
Asymptomatic period	The period of being infectious but asymptomatic
Symptomatic rate	The probability of infected persons will be symptomatic
Immunity loss rate	Rate at which a person loses immunity after recovering from infection
Mortality rate	The probability of an infected person dying

In this Section, we will focus on identifying the reference values for latent period, symptomatic period, and symptomatic rate.

2.1.1.1 Latent period and symptomatic period

A literature review was performed to ascertain the reference assumptions on latent period and symptomatic period in influenza pandemic. The literature search was carried out using the PubMed database with the keywords “influenza” (all fields), “latent period” (all fields), or “symptomatic period” (all fields). A review to the bibliographies of published studies was also performed to find additional related articles. The search was limited to English-language articles. A total of 6 articles were found which were published between 1990 and 2013 (Figure 1). Those articles were filtered with selection criteria, that only the ones that clearly stated the assumptions for latent period and symptomatic period in influenza pandemic have been selected. The total number of articles reviewed was five.

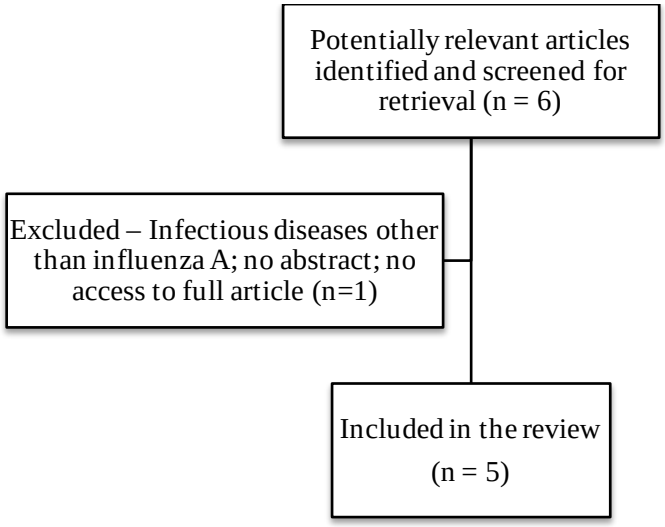

Figure 1. Identification of eligible articles for latent period and symptomatic period

The details of reviewed articles are described as follows:

Carrat et al. (27) performed systematic review on 56 published studies on human influenza from PubMed database (1965 to 2005) and reported an average of latent period at 1.1 days and the mean duration of symptoms at 4.4-5 days for H1N1 influenza, 3.7 days for H3N2

influenza, 4.6 days for H2N2 influenza and 4.1 days for influenza B. Cori et al. (28) reported an average of 1.63 days of latency period for influenza with a standard deviation 0.26 days using the same data from Carrat et al. (27). Tuite et al. (29) studied laboratory confirmed H1N1 cases in Ontario and estimated that the incubation periods followed a log-normal distribution with mean at 4.3 days and a 95% confidence interval at 2.6-6.6 days. They also estimated that the duration of symptoms followed a log-normal distribution with mean at 9.3 days and a 95% confidence interval at 2.6-24.2 days. Lessler et al. (30) performed systematic review of 38 publications with data from PubMed database (no time limits - 2009) and fitted a log-normal distribution for incubation period for influenza A with mean at 1.4 days and a 95% confidence interval at 1.3-1.5 days. Baguelin et al. (31) used a realistic generation time distribution and estimated that average latency period for H1N1 influenza was 1.40 days with a 95% confidence interval of [0.19–3.90].

Table 3 summarizes the data sources and reference values for latent period and symptomatic period in reviewed articles.

Table 3. Summary of data sources and reference values for latent and symptomatic periods

Articles	Data sources	Reference values
Carrat et al. (27)	Systematic reviewed papers from PubMed (1965-2005)	Symptoms: 4.4-5 (influenza A)
Cori et al. (28)	Data from Carrat et al. (27)	Latency: 1.63
Tuite et al. (29)	Laboratory confirmed H1N1 cases in Ontario	Latency: 4.3 (2.6, 6.6) Symptoms: 9.3 (2.6, 24.2)
Lessler et al. (30)	Systematic reviewed papers from PubMed (no time limits - 2009)	Latency: 1.4 (1.3, 1.5)
Baguelin et al. (31)	Epidemiological modeling (realistic generation time distribution)	Latency: 1.4 (0.19, 3.90)

2.1.1.2 Symptomatic rate

A literature review was performed to ascertain the reference assumptions on symptomatic rate in influenza pandemic. The literature search was carried out using the PubMed database with the keywords “influenza” (all fields), “symptomatic” (all fields), or “asymptomatic” (all fields). A review to the bibliographies of published studies was also performed to find additional related articles. The search was limited to English-language articles. A total of 739 articles were found which were published between 1990 and 2013 (Figure 2). Those articles were filtered with selection criteria, that only the ones that clearly stated the assumptions for symptomatic rate in influenza pandemic have been selected. The total number of articles reviewed was six.

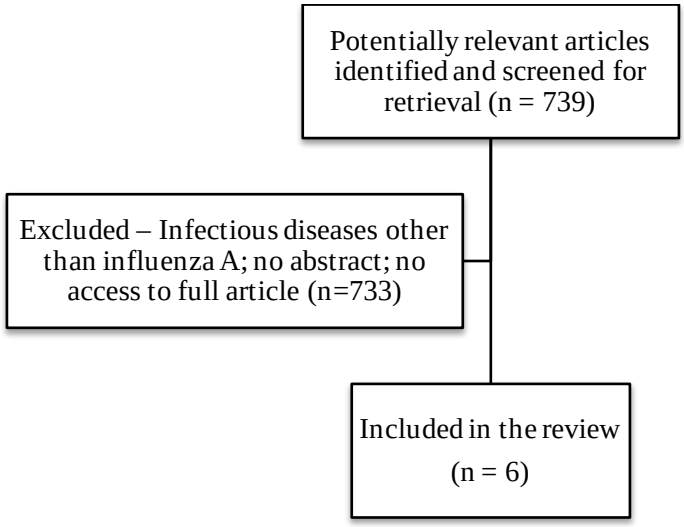

Figure 2. Identification of eligible articles for symptomatic rate

The details of reviewed articles are described as follows:

Longini et al. (32) assumed that the probability that an infected person will be symptomatic is 0.67 based on population-level influenza cohort studies in United States. Stilianakis et al. (33) assumed that an average of 50% of infected persons become sick with clinical symptoms. Papenburg et al. (34) identified the probability of infected persons will be

symptomatic is 0.67 by the secondary attack rates within household based on the data of the 2009 Pandemic A/H1N1 Influenza. Mathews et al. (35) used the data from 1918 H1N1 influenza pandemic to calculate an average of 0.38 symptomatic rate with 95% confidence interval between 0.28 and 0.60. Kuster et al. (36) conducted a systematic review and meta-analysis from OVID MEDLINE (1950 to 2010) and EMBASE (1947 to 2010), and estimated an average of 0.69 symptomatic rate in the annual incidence of influenza among healthy adults and healthcare workers. Carrat et al. (27) reviewed 56 published studies on human influenza from PubMed database (1965 to 2005) and reported an average of 0.669 symptomatic rate for all types of influenza and 0.708 for H1N1 influenza by GEE estimates.

Table 4 summarizes the data sources and reference values for symptomatic rate from reviewed articles.

Table 4. Summary of data sources and reference values for symptomatic rate

Articles	Data sources	Reference values
Mathews et al. (35)	Reported cases for H3N2 in Tristan da Cunha in 1971	0.38 (0.28, 0.60)
Longini et al. (32)	Experts' assumption based on natural history of influenza	0.67
Stilianakis et al. (33)	Experts' assumption	0.50
Papenburg et al. (34)	Prospective observational study for 2009 pH1N1 pandemic in Quebec City, Canada	0.67
Kuster et al. (36)	Systematic reviewed papers from OVID MEDLINE (1950-2010) and EMBASE (1947-2010)	0.689
Carrat et al. (27)	Systematic reviewed papers from PubMed (1965-2005)	0.669 (0.583, 0.745) 0.708 (0.504, 0.852)

2.1.2 Transmission parameters

In FRED model, transmission parameters include asymptomatic infectivity, transmissibility and transmission probability. Asymptomatic infectivity is defined as the proportion of an infected and asymptomatic person transmits an infection to an infected and symptomatic person transmits an infection. Transmissibility is defined to describe an infectious disease in its overall number of illnesses and deaths in the population during an epidemic. In FRED model, transmissibility is set by calibration, and will not be discussed in this chapter. Transmission probability is defined as the probability that a contact transmits an infection per contact per unit time (e.g., per day). In a given place type, the transmission probability generally depends on the age of the infectious person and the susceptible person. The definition of transmission probability in the references was very vague, and some references referred secondary attack rates to transmission probability. It is probably because this parameter may or may not be able to be measured in real life. In this Section, the term of transmission probability is reviewed with the referred unit time scale to make those reference values comparable. Table 5 summarizes the definitions of transmission parameters in FRED model.

Table 5. Definitions of transmission parameters in FRED model

Parameter	Definition
Asymptomatic infectivity	The proportion of an infected and asymptomatic person transmits an infection to an infected and symptomatic person transmits an infection
Transmission probability	The probability that a contact transmits an infection per contact per unit time
Transmissibility	Overall number of illnesses and deaths in the population during an infectious disease pandemic

In this Section, the results of a literature review will be used to identify the reference values of transmission probability.

2.1.2.1 Asymptomatic infectivity

In literature review, very few studies discussed asymptomatic infectivity. Some studies assumed that infectious but asymptomatic people usually have reduced infectivity compared with infectious and symptomatic people. For example, Longini et al. (32) and Yang et al. (37) assumed that an asymptomatic infection is only 50% as infectious as a symptomatic infection. Patrozou et al. (38) assumed that the proportion of transmission by asymptomatic individuals is one-third to one-half that of influenza-infected symptomatic individuals.

2.1.2.2 Transmission probability

A literature review was performed to ascertain the reference assumptions on transmission probability in influenza pandemic. A literature search was carried out using the PubMed database with the keywords “influenza” (all fields), “transmission probability” (all fields). A review to the bibliographies of published studies was also performed to find additional related articles. The search was limited to English-language articles. A total of 736 articles were found which were published between 1990 and 2013 (Figure 3). Those articles were filtered with selection criteria, that only the ones that clearly stated the assumptions for transmission probability in influenza pandemic have been selected. The total number of articles reviewed was eight.

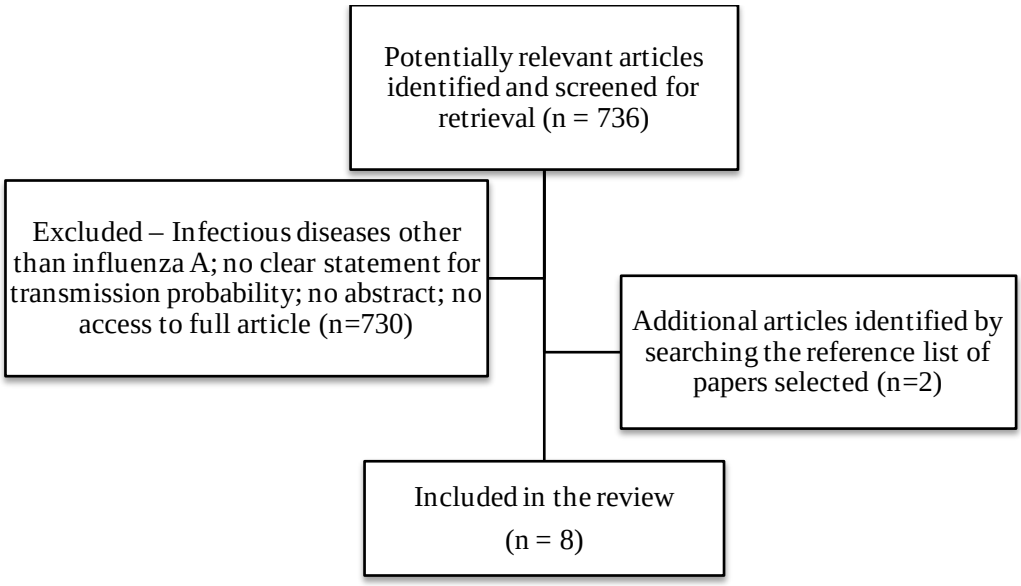

Figure 3. Identification of eligible articles for transmission probability

The details of reviewed articles are described as follows:

To estimate the household and community probabilities of transmission, Longini et al. (39, 40) have suggested that it is possible to estimate them from the distribution of the final number of cases in households at the end of the epidemic. The authors described a final-size distribution for transmission probability in household and community, and estimate the daily probability in household at 0.044 and the probability in community at 0.144 during the course of epidemic by the Asian influenza epidemic household data previously examined by Sugiyama (41). Reed et al. (42) reviewed data from past influenza seasons and pandemics, and characterized severity and transmissibility in both low-moderate and moderate-high levels as well as in a 7-scale measurement. For example, the authors estimated that the transmissibility rates in 2009 H1N1 influenza pandemic were 0.13 in household and 0.20 in the community. Piso et al. (43) reported a low transmission rate of H1N1 influenza at 0.019 during a long-distance public bus trip from Spain to Switzerland. Klick et al. (44) found the transmission probability was at 0.18 for H1N1 influenza among children in Hong Kong, and was at 0.06 among adults.

Merl et al. (45) used Monte Carlo simulation to explore the distributions of numbers of susceptible, infected, recovered individuals, and the total accrued cost, as functions of time. The estimates by Murray (46) were that the transmission rate 0.00218 and recovery rate 0.4. Then the authors estimated their negative binomial transmission function and implied that disease transmission occurs followed a Poisson process with a gamma distributed encounter rate.

Table 6 summarizes the data sources and reference values for transmission probability from reviewed articles.

Table 6. Summary of data sources and reference values for transmission probability

Articles	Data sources	Reference values	Unit
Longini et al. (39, 40)	Asian influenza epidemic household data in 1960	Household: 0.044	Daily
Sugiyama (41)		Community: 0.144	Epidemic
Reed et al. (42)	Published influenza pandemic data in 2009, 1968,1957,1918 and non-pandemic influenza seasons in 1978-79, 2006-07, 2007-08	Household: 0.13	Epidemic
		Community: 0.20	
Piso et al. (43)	Collected 2009 H1N1 influenza pandemic data in European public transportation systems	Community: 0.019	About a day
Klick et al. (44)	Recruited and collected household data for influenza pandemic in Hong Kong in 2008-09	Children: 0.18 (0.12, 0.25)	Epidemic
		Adults: 0.06 (0.03, 0.11)	
Merl et al. (45) Murray (46)	Previous studies	Transmission rate: 0.00218	unknown

2.1.3 Health related behavior parameters

One of important health-related behavior parameters in FRED model is the probability of staying home when sick. It is defined as the probability of withdrawal from work or school due to symptoms appearance. Another important health behavior parameter is contact rate. Contact rate is defined as the rate at which persons meet persons, and is measured as individuals per unit time, for example, number of contacts per day. Contact rate could take place in several locations: household, community, school and workplace. Table 7 summarizes the definitions of health related behavior parameters in FRED model.

Table 7. Definitions of health related behavior parameters in FRED model

Parameter	Definition
Probability of staying home when sick	The probability of withdrawal from work or school due to symptoms appearance
Contact rate	The rate at which persons meet persons, and is measured as individuals per unit time

In this Section, focus will be given to identify the reference values for the probability of staying home when sick and contact rate.

2.1.3.1 The probability of staying home when sick

A literature review was performed to ascertain the reference assumptions on the probability of staying home when sick in influenza pandemic. A literature search was carried out using the PubMed database with the keywords “influenza” (all fields), “absenteeism” (all fields) or “sick

leave (all fields)". A review to the bibliographies of published studies was also performed to find additional related articles. The search was limited to English-language articles. A total of 390 articles were found which were published between 1990 and 2013 (Figure 4). Those articles were filtered with selection criteria, that only the ones that clearly stated the assumptions for the probability of staying home when sick during influenza pandemic have been selected. The total number of articles reviewed was fourteen.

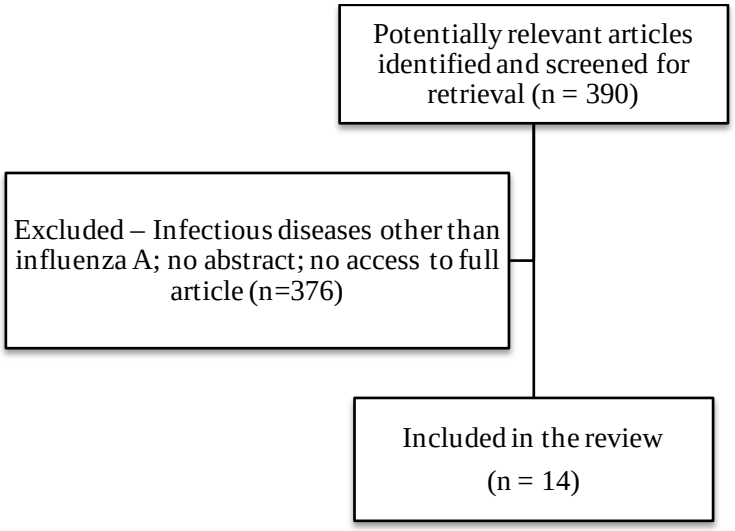

Figure 4. Identification of eligible articles for the probability of staying home when sick

The details of reviewed articles are described as follows:

Chan (47) studied the data from 2004-2005 seasonal influenza in Hong Kong and found that 55% of people with influenza-like illness took sick leave from work. The proportion would be lower if the person received vaccine treatment (30.3%). Schanzer et al. (48) conducted a Canadian labor force survey and confirmed that the estimates of absenteeism from workplace due to seasonal influenza typically ranged from 5% to 20% given the range of clinical attack rate from 15% to 35%. Mikolajczyk et al. (49) observed 49.3% absenteeism due to influenza-like illness for school children in Germany. Lau et al. (50) reported an approximate of 75% absenteeism for elementary school students due to influenza-like illness. Considine et al.

(51)reported that 57% of health care workers with influenza-like illness took sick leave during the 2009 H1N1 influenza pandemic in Australia. Campbell et al. (52) conducted a prospective, non-randomized, non-placebo control trial in six North Carolina textile plants and found that 31 out of 64 unvaccinated people who had influenza-like illness missed work (48.4%), compared to 15 out of 26 vaccinated people missed work due to influenza-like illness (57.7%). Neuzil et al. (53) observed an approximate of 63% absenteeism for school children during influenza season. Rousculp et al. (54) reported that 29.1% of employees with influenza-like illness took sick leave in 2007-2008 season. Turnberg et al. (55) examined annual influenza vaccination and sick leave practices and perceptions among health care workers and found only 31% of respondents routinely took sick leave when they had influenza-like symptoms. Study conducted by Tora et al. (56) for comparing absence due to sickness for 2009 H1N1 influenza indicated that women had a higher proportion of sickness absence due to influenza-like illness (52.2% in Catalonia and 49.7% in Andalusia). Nichol et al. (57) conducted a double-blind, placebo-controlled trial in healthy working adults in 1994-1995 season to investigate the effectiveness of vaccination against influenza, and found that 40.9% of placebo recipients and 35.6% of vaccine recipients took sick leave. Similarly, Bridges et al. (58) conduct randomized controlled trial to evaluate the effectiveness and cost-benefit of influenza vaccine among healthy working adults in 1997-1998 season and 1998-1999 season. They found that 39% of unvaccinated persons took sick leave from work due to influenza-like illness compared with vaccinated persons at 28% in 1997-1998 season. In 1998-1999 season, 54% of unvaccinated persons had workday loss due to influenza-like illness compared with vaccinated persons at 55%. Seale et al. (59) conducted cross-Sectional survey for health care workers in Sydney and reported that 81.2% of respondents would not present to work if they had influenza-like illness, and the proportion would be lower at 58.6%

during a severe staff shortage. Blendon et al. (60) reported that in the nation-wide survey for investigating the population behaviors in a pandemic, surprisingly 94% of the population stated that they would be able to stay home if they had influenza-like illness.

Table 8 summarizes the data sources and reference values for the probability of staying home when sick from reviewed articles.

Table 8. Summary of data sources and reference values for the probability of staying home when sick

Articles	Data sources	Reference values
Chan (47)	2004-05 seasonal influenza data in Hong Kong	0.50
Schanzer et al. (48)	Canadian labor force survey for seasonal and pandemic influenza between 1998 and 2009	(0.33-0.57)
Mikolajczyk et al. (49)	Collected data from 24 primary schools in Germany in 2004 in form of questionnaire	0.493
Lau et al. (50)	Cohort study from 2 Chicago public elementary schools in 2009-10	0.75
Considine et al. (51)	Electronic survey for emergency nursing and medicine in Australia in 2009	0.57
Campbell et al. (52)	Prospective control trial in six North Carolina textile plants in 1997	0.484
Neuzil et al. (53)	Prospective survey study in an elementary school in Seattle in 2000-01	0.63
Rousculp et al. (54)	Prospective study of employees from three US employers with access to FSLPs	0.291

Table 8 continued

Articles	Data sources	Reference values
Turnberg et al. (55)	Self-report questionnaire in Washington for health care workers in 2005	0.31
Tora et al. (56)	Surveillance data reported in Andalusia and Catalonia in the period 2007-2009	0.522 (Catalonia) 0.497 (Andalusia)
Nichol et al. (57)	Subjects were recruited from Minneapolis–St. Paul area in 1994	0.409
Bridges et al. (58)	Data collected from full-time employees of Ford Motor Co, Michigan during 1997-98 and 1998-99	0.39 (1997-98) 0.54 (1998-99)
Seale et al. (59)	Cross-Sectional investigation of health care workers in 2007 in Sydney, Australia	0.812
Blendon et al. (60)	Questionnaire survey conducted by Harvard School of Public Health in 2006	0.94

2.1.3.2 Contact rates

A literature review was performed to ascertain the reference assumptions on contact rate in influenza pandemic. A literature search was carried out using the PubMed database with the keywords “influenza” (all fields), “household contact” (all fields), or “social contact” (all fields). A review to the bibliographies of published studies was also performed to find additional, related articles. The search was limited to English-language articles. A total of 198 articles were found which were published between 1990 and 2013 (Figure 5). Those articles were filtered with

selection criteria, that only the ones that clearly stated the assumptions for contact rate in influenza pandemic have been selected. The total number of articles reviewed was seven.

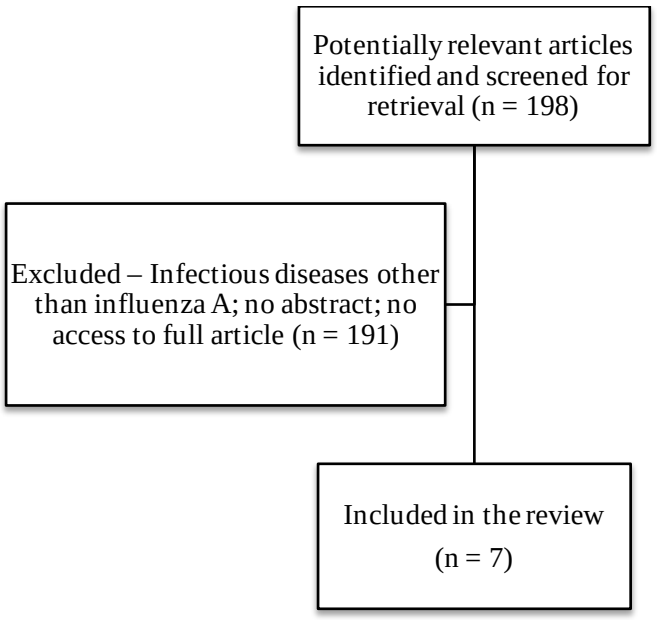

Figure 5. Identification of eligible articles for contact rate

The details of reviewed articles are described as follows:

Mikolajczyk et al. (49) conducted a survey about the social contacts of school children in Germany and found a total mean number of contacts per day at 32.7 with 95% confidence interval from 1 to 78, while the mean number with children was 25.1 and with adults was 7.5. They also suggested the negative binomial distribution to estimate the number of contacts at school with a shifted mean of 12.0 and a variance of 77.3. Chen et al. (61) found the mean number of contacts for elementary school children was ranging from 9.44 to 11.18 per day in Taiwan, and was lower for middle school children and high school children at 5.66 with a standardize deviation of 6.23 per day. Mossong et al. (62) conducted a population-based prospective survey of mixing patterns in eight European countries (POLYMOD study) to quantify contact patterns relevant for infections transmitted by the respiratory and reported the mean number of household contacts per day was ranging from 8.9 to 17.7 based on the household size.

Towers and Chowell(63) used the contact data from Mossong et al. (62) and assumed the average number of child-to-child contacts per day was 8.9, child-to-adult was 5.5, adult-to-child was 1.9, and adult-to-adult was 9.3. Potter et al. (64) used the data from the POLYMOD study in eight European countries of social contact behavior and estimated the conditional probability of age-specific contact network for household. Wallinga et al. (65) used the data from a cross-Sectional survey conducted in the Netherlands to estimated the matrix of the number of age-specific social contacts among all age groups. Eames et al. (66) conducted internet-based survey for measuring dynamic social contact patterns during 2009 H1N1 influenza pandemic and reported age-specific matrix of number of conversational contacts and physical contacts during school term time and school holidays.

Table 9 summarizes the data sources and reference values for contact rate from reviewed articles.

Table 9. Summary of data sources and reference values for contact rate

Articles	Data sources	Reference values (number per day)
Mikolajczyk et al. (49)	Collected data from 24 primary schools in Germany in 2004 in form of questionnaire	Children: 25.1 Adults: 7.5
Chen et al. (61)	274 diary-based questionnaires collected from schools in Taiwan in 2010	Elementary: (9.44, 11.18) Middle and High: 5.66 (6.23)
Mossong et al. (62)	POLYMOD study: cross-Sectional surveys conducted in 8 European countries in 2005-06	Household: (8.9, 17.7)
Towers et al (63)	POLYMOD study: cross-Sectional surveys conducted in 8 European countries in 2005-06	Child-to-child: 8.9 Child-to-adult: 5.5 Adult-to-child: 1.9 Adult-to-adult: 9.3

Table 9 continued

Articles	Data sources	Reference values (number per day)
Potter et al. (64)	POLYMOD study: cross-Sectional surveys conducted in 8 European countries in 2005-06, but only used the Belgian data	Probability distribution of contact networks among children and adults
Wallinga et al. (65)	Cross-Sectional survey conducted in the Netherlands in 1986	See table 10
Eames et al. (66)	Internet-based social contact survey completed by a cohort of participants in 2009-10 in UK	See table 11

Table 10. Age-specific social contacts estimated by Wallinga et al. (65)

Age	1-5	6-12	13-19	20-39	40-59	60+
0-5	12.26	2.28	1.29	2.50	1.15	0.83
6-12	2.72	23.77	2.80	3.02	1.78	1.00
13-19	2.00	3.63	25.20	5.70	4.22	1.68
20-39	11.46	11.58	16.87	25.14	16.43	8.34
40-59	3.59	4.67	8.50	11.21	13.89	7.48
60+	1.94	1.95	2.54	4.25	5.59	9.19

Table 11. Age-specific social contacts estimated by Eames et al. (66)

Period	School term time				School holidays			
Contact	Conversational		Physical		Conversational		Physical	
	Children	Adults	Children	Adults	Children	Adults	Children	Adults
Children	34.9	19.1	12.3	7	20.9	19.4	8.5	7.5
Adults	5	24.5	1.8	5.5	4.1	23.2	1.5	6.1

2.2 REFERENCE PARAMETER VALUES FOR THE FRED INFLUENZA MODEL

When considering the reference values for parameters in FRED model from literature reviews, articles that used systematic review and meta-analysis methods were preferable. For example, the reference values for latent period and symptomatic rate were chosen from the results of systematic review by Lessler et al. (30) and Carrat et al. (27), respectively. When articles that used systematic review were not found, the reference value was chosen by the average value of reviewed articles (e.g., the probability of staying home when sick). Parameters with a range between 0 and 1 (e.g., symptomatic rate) will choose beta distribution, other parameters will choose truncated normal distribution based on their ranges.

For natural history parameters, the reference value of symptomatic rate follows a beta distribution with mean at 0.669 and standard deviation at 0.271(27). The latent period follows a normal distribution with a mean of 1.4 days and a standard deviation of 0.49(30). The symptomatic period follows a normal distribution with mean at 4.7 days and standard deviation at 0.9. Since asymptomatic period is difficult to measure and there were few studies about it in the literature, we assume that asymptomatic period has the same distribution as symptomatic period in this study. Therefore, the asymptomatic period follows a normal distribution with mean at 4.7 days and standard deviation at 0.9.

For transmission parameters, since there were very few articles that had assumptions for asymptomatic infectivity and those articles assumed an average of 50% for asymptomatic infectivity with a possible range between 0.33 and 0.67.

For health related behavior parameters, the reference value for the probability of staying home when sick follows a beta distribution with mean at 0.543 and standard deviation at 0.179.

For transmission probability, the definitions in reviewed papers are not very consistent, and transmission probability is associated with different location. In FRED influenza model, the transmission probability is set as a calibration number. Therefore, it may be difficult to use any reference value of transmission probability from reviewed papers, and we decide to keep the default value of transmission probability in FRED influenza model.

For contact rate, all the reviewed papers stated that it is associated with age, and it would very different between child-to-child and child-to-adult. In FRED influenza model, contact rate is set as a single value. Due to the structure conflict, contact rate will also use the default value in FRED influenza model.

Table 12 summarizes the reference values and distributions of parameters from literature review.

Table 12. Reference values and distributions of parameters from literature review

Parameter	Reference value(range)	Distribution
Latent period	1.4 (0-2) days	normal
Symptomatic period	4.7 (3-6) days	normal
Asymptomatic period	4.7 (3-6) days	normal
Symptomatic rate	0.669 (0-1)	beta
Asymptomatic infectivity	0.5 (0.33-0.67)	beta
Transmission probability	0.00218 – 0.25	Unknown
Probability of staying home when sick	0.543 (0-1)	beta
Contact rate	by age	Fixed value

2.3 DISCUSSIONS AND CONCLUSIONS

It is worth mentioning that there are interactions among above parameters. For example, Stilianakis et al. (33) stated that infected persons with clinical symptoms might show reduced contacts if they are sufficiently ill to be confined to bed. The probability of staying home when sick is associated with demographic background (gender, age etc.) and economical considerations (47, 51-55, 57-59). The number of social contacts varies by the day of the week (49, 62), and is associated with the contact age and household size (61-66).Studies also showed interactions among epidemiological parameters. For example, the clinical attack rate of influenza was affected by prior immunity and mixing patterns in population, and by the asymptomatic rate in population (35).

In summary, this chapter performed a literature review to identify important parameter values in the FRED influenza model: latent period, symptomatic period, symptomatic rate, asymptomatic infectivity, transmission probability, the probability of staying home when sick, and contact rates. The reference values and distributions concluded from this chapter will be used in the following chapters for sensitivity analysis.

CHAPTER 3 LOCAL SENSITIVITY ANALYSIS

This chapter gives an overview of the background on local sensitivity analysis (Section 3.1), proposes the details of performing local sensitivity analysis in FRED model (Section 3.2), develops algorithms and R programs that quickly create input values for some parameters with the input format of discrete cumulative distribution function (Section 3.3), then conducts local sensitivity analysis to ten selected parameters in FRED model (Section 3.4). The summary, discussions and conclusions are present in Section 3.5. This chapter uses the results of reference parameters values from Chapter 2 to perform local sensitivity analysis to identify high sensitivity parameters in FRED model. The results of this chapter will be used in Chapter 4 and 5.

3.1 LOCAL SENSITIVITY ANALYSIS

Generally, sensitivity analysis has been classified into local sensitivity analysis and global sensitivity analysis. In local sensitivity analysis, the model responses are obtained by changing the values of input parameters one at a time, while the remaining parameters are set to their default values. It also assumes that the relationships between input parameters and model outputs are linear. Since in local sensitivity analysis one parameter is changed at a time, it may not be able to detect interactions among parameters. In contrast, global sensitivity analysis deals with

the entire ranges of input parameters to understand the full range of model behavior. It can detect the impacts of both individual parameters and their interactions (22, 67).

The general procedure of performing sensitivity analysis involves the following steps (21): First, identify input parameters and define distributions to characterize the uncertain parameters. Second, use one of the sampling procedures (random sampling, stratified sampling, and Latin Hypercube sampling etc.) to generate the samples of the uncertain parameters. Third, propagate the samples to generate the responses from the model. Fourth, perform statistical methods such as regression analysis or analysis of variance to assess the impacts of the uncertain parameters on the model output.

In local sensitivity analysis, a sensitivity measure, S, is used to identify high sensitivity factors or variables in the model. The simplest sensitivity measure S is defined as the derivative of the model function.

For example, let $Y = f(X)$, where Y is the output of model, and $X = X_1,...X_k$ are k-independent inputs. Then the sensitivity measure S is defined as

$$S_i = \frac{\partial Y}{\partial X_i}$$

A simplified calculation of sensitivity measure is using the parameter's reference value (P), then vary it by dP (for example, $d = \pm 5\%$). Then the sensitivity measure S is

$$S = \frac{\triangle Y}{(dP / P)} = \frac{\triangle Y}{0.05}$$

Parameters with low sensitivity may not be important in the model and may be left out of the model from further research. Some parameters may have low sensitivity but high uncertainty, and may not be simply left out by a single sensitivity measure. Parameter uncertainty is defined as the exact values of parameters are unknown and are difficult to be controlled in models or

experiments. Therefore, it needs to pay more attention on parameters with high uncertainty in the model(67). In this chapter, we will determine the important parameters based on both uncertainty in the references and sensitivity measure from local sensitivity analysis.

3.2 PROCEDURES OF PERFORMING LOCAL SENSITIVITY ANALYSIS IN FRED MODEL

The following criteria and steps are used to evaluate the input parameters and compare the consequence outputs to identify important factors in FRED model.

Step 1: identify input parameters and define distributions to characterize input parameters, use the reference values from literature review for each parameter. In this step, ten important parameters will be used in local sensitivity analysis: latent period, symptomatic period, asymptomatic period, symptomatic rate, asymptomatic infectivity, the probability of staying home when sick, household contact rate, neighborhood contact rate, school contact rate and workplace contact rate. Those ten parameters will use the default values in FRED model or the reference values determined from Chapter 2.

Step 2: generate the responses from the model using the values from step 1, one at a time, vary the input values of parameters by *dP* (for example, $d = \pm 5\%$, but also depends on the range of the parameter) and generate the corresponding model simulations.

The expected model output is the attack rate. Attack rate is defined as the cumulative incidence of infection in a population observed over a period of time during an epidemic. There are also other output variables in FRED model, such as number of new cases by day, reproductive rate by day, etc. We choose attack rate as the model output because it is easy to

obtain the overall mean attack rates from the model and also easy to interpret for a simulated influenza pandemic. Besides, in Chapter 5, attack rate is used to calculate the cost-effectiveness of different school closure intervention strategies.

Some parameters are set as discrete cumulative distribution function in FRED model. To make small changes in these input values for the use of performing local sensitivity analysis, the algorithms described in Section 3.3 will be used.

Step 3: calculate sensitivity measure for each input parameter using the following formula:

$$S = \frac{\triangle Y}{(dP / P)} \text{ , e.g., } S = \frac{\triangle Y}{0.05}$$

Step 4: summarize the results and draw conclusions on the sensitivity of parameters. Parameters with low sensitivity might be left out of the model from further research.

3.3 ALGORITHMS FOR VARYING DISCRETE CUMULATIVE DISTRIBUTION FUNCTION INPUTS OF PARAMETERS

This Section addresses a technical issue in how to vary input parameters that are specified by distributions. In FRED model, three input parameters are specified as discrete cumulative distribution functions: latent period, symptomatic period and asymptomatic period. For example, the latent period ranges between 1 day and 2 days with assigned probability for each day; the symptomatic period and the asymptomatic period range from 3 days to 6 days with assigned probability for each day. In other words, when an infection occurs in FRED, each infected agent is independently assigned a latent period, a symptomatic period and an asymptomatic period

based on the discrete distributions of those parameters, but the overall means of those parameters are determined by the initial input values of the parameters. In this Section, we develop algorithms and R programs that quickly create input values of discrete cumulative distribution function from an assigned overall mean.

Assume the parameter follows the following discrete probability density function (PDF) and discrete cumulative distribution functions(CDF) with mean *u*:

Parameter value	d_1	d_2	…	d_a	d_{a+1}	…	d_{a+k}
PDF	p_1	p_2	…	p_a	p_{a+1}	...	p_{a+k}
CDF	p_1	p_1+p_2	…	$\sum_1^a p_i$	$\sum_1^{a+1} p_i$	…	$\sum_1^{a+k} p_i = 1$

while any $p_i > 0$, and the mean *u* is defined as $u = \sum_1^{a+k} p_i d_i$

Assume that $d_1, d_2, ..., d_a < u < d_{a+1}, ..., d_{a+k}$. To change the CDF and PDF slightly to achieve the mean at *m* while $m < u$, we will increase the probability of $d_1, d_2, ..., d_a$, and decrease the probability of $d_{a+1}, ..., d_{a+k}$. To make the changes easier in FRED model, we will change the same value of *p* to each p_i. In other words, the mean *m* is defined as

$$m = (p_1 + \frac{k}{a}p)d_1 + (p_2 + \frac{k}{a}p)d_2 + ... + (p_a + \frac{k}{a}p)d_a + (p_{a+1} - p)d_{a+1} + ... + (p_{a+k} - p)d_{a+k}$$

$$\text{while} (p_1 + \frac{k}{a}p) + (p_2 + \frac{k}{a}p) + ... + (p_a + \frac{k}{a}p) + (p_{a+1} - p) + ... + (p_{a+k} - p) = \sum_1^{a+k} p_i = 1$$

$$\text{then } m = \sum_1^{a+k} p_i d_i - p[(d_{a+1} + ... + d_{a+k}) - \frac{k}{a}(d_1 + ... + d_a)] = u - p(\sum_{a+1}^{a+k} d_i - \frac{k}{a}\sum_1^a d_i)$$

The small value *p* that will be used to change the PDF of the parameter is calculated as

$$p = \frac{u - m}{\sum_{a+1}^{a+k} d_i - \frac{k}{a}\sum_1^a d_i}$$

The new input values for the parameter follows the following discrete PDF and CDF:

Parameter value	d_1	d_2	$\cdots$	d_a	d_{a+1}	$\cdots$	d_{a+k}
PDF	$p_1+\frac{k}{a}p$	$p_2+\frac{k}{a}p$	$\cdots$	$p_a+\frac{k}{a}p$	$p_{a+1}-p$	$\cdots$	$p_{a+k}-p$
CDF	$p_1+\frac{k}{a}p$	$p_1+p_2+2\frac{k}{a}p$	$\cdots$	$\sum_1^a p_i+kp$	$\sum_1^{a+1} p_i+(k-1)p$	$\cdots$	$\sum_1^{a+k} p_i=1$

Similarly, to achieve the mean at *n* while $n>u$, the change of *p* is calculated as

$$p=\frac{n-u}{\sum_{a+1}^{a+k} d_i-\frac{k}{a}\sum_1^a d_i}$$

The new input values for the parameter follows the following discrete PDF and CDF:

Parameter value	d_1	d_2	$\cdots$	d_a	d_{a+1}	$\cdots$	d_{a+k}
PDF	$p_1-\frac{k}{a}p$	$p_2-\frac{k}{a}p$	$\cdots$	$p_a-\frac{k}{a}p$	$p_{a+1}+p$	$\cdots$	$p_{a+k}+p$
CDF	$p_1-\frac{k}{a}p$	$p_1+p_2-2\frac{k}{a}p$	$\cdots$	$\sum_1^a p_i-kp$	$\sum_1^{a+1} p_i-(k-1)p$	$\cdots$	$\sum_1^{a+k} p_i=1$

Note that, above algorithms and formulas are applied only to the positive p_i in the PDF. If $p_i=0$, we will keep the zero probability as it is, and calculate the small *p* based on the rest of positive p_i, and change the values of *a* and *k* based on the number of positive p_i.

In some circumstances, after calculated the small *p* from the formula, we might find out that p_i-p is less than zero, or $p_i-\frac{k}{a}p$ is less than zero. In such cases, we will convert the negative p_i-p or $p_i-\frac{k}{a}p$ to zero, and rearrange and divide the value of $|p_i-p|$ or $|p_i-\frac{k}{a}p|$

to its nearest non-negative $p_i - p$ or $p_i - \frac{k}{a}p$. Then repeat above procedure until all the values of probability become non-negative. The algorithm of this procedure is described below.

For $m < u$, assume that the new PDF is calculated as following table. Assume $p_{a+r} - p < 0$ for d_{a+r}, convert the probability of d_{a+r} to zero, then the new mean $m' = m + d_{a+r}(p - p_{a+r})$. The value of $p_{a+r} - p$ will be divided into the probabilities of d_{a+r-1} and d_{a+r+1}, which will satisfy the following equations:

$$\begin{cases} d_{a+r-1}p'_{a+r-1} + d_{a+r+1}p'_{a+r+1} = d_{a+r}(p_{a+r} - p) \\ p'_{a+r-1} + p'_{a+r+1} = p_{a+r} - p \end{cases}$$

while $p'_{a+r-1} = \frac{(d_{a+r+1} - d_{a+r})(p_{a+r} - p)}{d_{a+r+1} - d_{a+r-1}}$ and $p'_{a+r+1} = \frac{(d_{a+r} - d_{a+r-1})(p_{a+r} - p)}{d_{a+r+1} - d_{a+r-1}}$

Thus, the revised PDF with the zero probability for d_{a+r} is as follows:

Mean	d_1	…	d_a	d_{a+1}	…	d_{a+r-1}	d_{a+r}	d_{a+r+1}	…	d_{a+k}
m	$p_1 + \frac{k}{a}p$	…	$p_a + \frac{k}{a}p$	$p_{a+1} - p$	…	$p_{a+r-1} - p$	$p_{a+r} - p$	$p_{a+r+1} - p$	…	$p_{a+k} - p$
$m' = m - d_{a+r}(p_{a+r} - p)$	$p_1 + \frac{k}{a}p$	…	$p_a + \frac{k}{a}p$	$p_{a+1} - p$	…	$p_{a+r-1} - p$	0	$p_{a+r+1} - p$	…	$p_{a+k} - p$
m	$p_1 + \frac{k}{a}p$	…	$p_a + \frac{k}{a}p$	$p_{a+1} - p$	…	$p_{a+r-1} - p + p'_{a+r-1}$	0	$p_{a+r+1} - p + p'_{a+r+1}$	…	$p_{a+k} - p$

For $p_{a+k} - p < 0$, the probability of d_{a+k} will be converted to zero, and the value of $p_{a+k} - p$ will be divided into the probabilities of d_{a+k-1} and d_{a+k-2}, which will satisfy the following equations:

$$\begin{cases} d_{a+k-1}p'_{a+k-1} + d_{a+k-2}p'_{a+k-2} = d_{a+k}(p_{a+k} - p) \\ p'_{a+k-1} + p'_{a+k-2} = p_{a+k} - p \end{cases}$$

while $p'_{a+k-1} = \frac{(d_{a+k} - d_{a+k-2})(p_{a+k} - p)}{d_{a+k-1} - d_{a+k-2}}$ and $p'_{a+k-2} = \frac{(d_{a+k-1} - d_{a+k})(p_{a+k} - p)}{d_{a+k-1} - d_{a+k-2}}$

The revised PDF with the zero probability for d_{a+k} is as follows:

Mean	d_1	…	d_a	d_{a+1}	…	d_{a+k-2}	d_{a+k-1}	d_{a+k}
m	$p_1 + \frac{k}{a}p$	…	$p_a + \frac{k}{a}p$	$p_{a+1} - p$	…	$p_{a+k-2} - p + p'_{a+k-2}$	$p_{a+k-1} - p + p'_{a+k-1}$	0

Similarly, for $n > u$, assume that the new PDF is calculated as following table. Assume $p_r - \frac{k}{a}p < 0$ for d_r, convert the probability of d_r to zero, then the new mean $n' = n - d_r(p_r - \frac{k}{a}p)$.The value of $p_r - \frac{k}{a}p$ will be divided into the probabilities of d_{r-1} and d_{r+1}, which will satisfy the following equations:

$$\begin{cases} d_{r-1}p'_{r-1} + d_{r+1}p'_{r+1} = d_r(p_r - \frac{k}{a}p) \\ p'_{r-1} + p'_{r+1} = p_r - \frac{k}{a}p \end{cases}$$

while $p'_{r-1} = \frac{(d_{r+1} - d_r)(p_r - \frac{k}{a}p)}{d_{r+1} - d_{r-1}}$ and $p'_{r+1} = \frac{(d_r - d_{r-1})(p_r - \frac{k}{a}p)}{d_{r+1} - d_{r-1}}$

The revised PDF with the zero probability for d_r is as follows:

Mean	d_1	…	d_{r-1}	d_r	d_{r+1}	…	d_{a+1}	…	d_{a+k}
n	$p_1 - \frac{k}{a}p$	…	$p_{r-1} - \frac{k}{a}p$	$p_r - \frac{k}{a}p$	$p_{r+1} - \frac{k}{a}p$	…	$p_{a+1} + p$	…	$p_{a+k} + p$
$n' = n - d_r(p_r - \frac{k}{a}p)$	$p_1 - \frac{k}{a}p$	…	$p_{r-1} - \frac{k}{a}p$	0	$p_{r+1} - \frac{k}{a}p$	…	$p_{a+1} + p$	…	$p_{a+k} + p$
n	$p_1 - \frac{k}{a}p$	…	$p_{r-1} - \frac{k}{a}p + p'_{r-1}$	0	$p_{r+1} - \frac{k}{a}p + p'_{r-1}$	…	$p_{a+1} + p$	…	$p_{a+k} + p$

For $p_1 - \frac{k}{a}p < 0$, the probability of d_1 will be converted to zero, and the value of $p_1 - \frac{k}{a}p$ will be divided into the probabilities of d_2 and d_3, which will satisfy the following equations:

$$\begin{cases} d_2 p_2^{'} + d_3 p_3^{'} = d_1(p_1 - \frac{k}{a}p) \\ p_2^{'} + p_3^{'} = p_1 - \frac{k}{a}p \end{cases}$$

while $p_2^{'} = \frac{(d_1 - d_3)(p_1 - \frac{k}{a}p)}{d_2 - d_3}$ and $p_3^{'} = \frac{(d_2 - d_1)(p_1 - \frac{k}{a}p)}{d_2 - d_3}$

The revised PDF with zero probability for d_1 is as follows:

Mean	d_1	d_2	d_3	…	d_a	d_{a+1}	…	d_{a+k}
n	0	$p_2 - \frac{k}{a}p + p_2^{'}$	$p_3 - \frac{k}{a}p + p_3^{'}$	…	$p_a - \frac{k}{a}p$	$p_{a+1} + p$	…	$p_{a+k} + p$

The following is an example for applying above algorithms. For example, the following table is the default discrete cumulative distribution function for symptomatic period with a mean of 4.7 in FRED model.

Mean	Symptomatic period	0	1	2	3	4	5	6
4.7	CDF	0.0	0.0	0.0	0.1	0.4	0.8	1.0
	PDF	0.0	0.0	0.0	0.1	0.3	0.4	0.2

To achieve a mean of 5.6, $p = \frac{5.6 - 4.7}{(6) - \frac{1}{3}(3+4+5)} = 0.45$, the new PDF is as follows:

Mean	Symptomatic period	0	1	2	3	4	5	6
5.6	PDF	0.0	0.0	0.0	-0.05	0.15	0.25	0.65

Note that the probability of day 3 becomes negative in the PDF for the mean of 5.6. We will convert it to zero, and recalculate the probabilities of day 4 and day 5 as follows:

$$p_4^{'} = \frac{(d_3 - d_5)(p_3 - \frac{k}{a}p)}{d_4 - d_5} = \frac{(-2)(-0.05)}{(-1)} = -0.1$$

$$p_5^{'} = \frac{(d_4 - d_3)(p_3 - \frac{k}{a}p)}{d_4 - d_5} = \frac{(1)(-0.05)}{(-1)} = 0.05$$

The revised PDF and CDF are as follows:

Mean	Symptomatic period	0	1	2	3	4	5	6
5.6	PDF	0.0	0.0	0.0	-0.05	0.15	0.25	0.65
	Revised PDF	0.0	0.0	0.0	0.0	0.15-0.1=0.05	0.25+0.05=0.30	0.65
	CDF	0.0	0.0	0.0	0.0	0.05	0.35	1

3.4 MODEL OUTPUTS AND SENSITIVITY MEASURES

In this Section, ten parameters are selected for using in local sensitivity analysis. Those ten parameters are reviewed in Chapter 2 with reference values and distributions. Those ten parameters are also of interests and important in FRED model.

In FRED model, Allegheny County, PA is selected as the population for all the simulation runs in local sensitivity analysis. Each condition of input parameters will generate 20repeated simulation runs and the mean attack rate of those 20 simulation runs will be

presented as the model output of each condition of input parameters. The number of simulation runs is determined by a preliminary study in FRED model (please see Appendix A for more details).

3.4.1 Latent period

The parameter, "latent period", is defined as a discrete cumulative distribution function for the number of days between becoming exposed and becoming infectious. In FRED model, the default CDF of latent period is as follows:

Latent period	0	1	2
CDF	0.0	0.6	1.0
PDF	0.0	0.6	0.4

The mean of latent period is $0.0 \times 0 + 0.6 \times 1 + 0.4 \times 2 = 1.4$

To achieve about $\pm 7\%$ change of the reference value of latent period (1.3 and 1.5, respectively), the small change p is calculated as $p = \frac{0.1}{2 - (1)} = 0.1$

We choose $\pm 7\%$ change for latent period because in the literature review, latent period has a narrow range from 1.3 to 1.5, and it is easy to the calculate and display the discrete cumulative distribution functions for this narrow range.

The input values for the discrete cumulative distribution function are the follows:

Mean	Latent period	0	1	2
1.3	CDF	0.0	0.7	1
1.5	CDF	0.0	0.5	1

Table 13 shows the mean attack rates by the reference value of latent period (1.4) and by the $\pm 7\%$ of the reference value of latent period (1.3 and 1.5, respectively).

Table 13. Local sensitivity analysis results for latent period

	Latent period		
	1.3	1.4	1.5
Attack rate	55.607	55.594	55.5675

The sensitivity measure is calculated as follows:

$$S = \frac{55.594 - 55.607}{0.1/1.4} = -0.182 \text{ or } S = \frac{55.5675 - 55.594}{0.1/1.4} = -0.371$$

The results show that by decreasing or increasing 7% of input value of latent period, the corresponding sensitivity measures of mean attack rates are -0.182 and -0.371, respectively.

3.4.2 Symptomatic period

The parameter, "symptomatic period", is defined as a discrete cumulative distribution function for the number of days the agent is infectious and symptomatic. In FRED model, the default CDF of symptomatic period is as follows:

Symptomatic period	0	1	2	3	4	5	6
CDF	0.0	0.0	0.0	0.1	0.4	0.8	1.0
PDF	0.0	0.0	0.0	0.1	0.3	0.4	0.2

The mean of symptomatic period is $0.1 \times 3 + 0.3 \times 4 + 0.4 \times 5 + 0.2 \times 6 = 4.7$

To achieve $\pm 5\%$ change of the reference value of symptomatic period (4.465 and 4.935, respectively), the small change p is calculated as $p = \frac{0.235}{11 - (7)} = 0.05875$

The input values for the discrete cumulative distribution function are the follows:

Mean	Symptomatic period	0	1	2	3	4	5	6
4.465	CDF	0.0	0.0	0.0	0.15875	0.5175	0.85875	1
4.935	CDF	0.0	0.0	0.0	0.04125	0.2825	0.74125	1

Table 14 shows the mean attack rates by the reference value of symptomatic period (4.7) and by the $\pm 5\%$ of the reference value of symptomatic period (4.465 and 4.935, respectively).

Table 14. Local sensitivity analysis results for symptomatic period

	Symptomatic period		
	4.465	4.7	4.935
Attack rate	53.4335	55.594	57.705

The sensitivity measure is calculated as follows:

$$S = \frac{\triangle Y}{0.05} = \frac{55.594 - 53.4335}{0.05} = 43.21 \text{ or } S = \frac{\triangle Y}{0.05} = \frac{57.705 - 55.594}{0.05} = 42.22$$

The results show that by decreasing or increasing 5% of input value of symptomatic period, the corresponding sensitivity measures of mean attack rates are 43.21 and 42.22, respectively.

3.4.3 Asymptomatic period

The parameter, “asymptomatic period”, is defined as a discrete cumulative distribution function for the number of days the agent is infectious and asymptomatic. In FRED model, the default CDF of asymptomatic period is as follows:

Asymptomatic period	0	1	2	3	4	5	6
CDF	0.0	0.0	0.0	0.1	0.4	0.8	1.0
PDF	0.0	0.0	0.0	0.1	0.3	0.4	0.2

The mean of asymptomatic period is $0.1 \times 3 + 0.3 \times 4 + 0.4 \times 5 + 0.2 \times 6 = 4.7$

To achieve $\pm 5\%$ change of the reference value of asymptomatic period (4.465 and 4.935, respectively), the small change *p* is calculated as $p = \frac{0.235}{11 - (7)} = 0.05875$

Similar to the symptomatic period, the input values for the discrete cumulative distribution function of asymptomatic period are the follows:

Mean	Asymptomatic period	0	1	2	3	4	5	6
4.465	CDF	0.0	0.0	0.0	0.15875	0.5175	0.85875	1
4.935	CDF	0.0	0.0	0.0	0.04125	0.2825	0.74125	1

Table 15 shows the mean attack rates by the reference value of asymptomatic period (4.7) and by the $\pm 5\%$ of the reference value of asymptomatic period (4.465 and 4.935, respectively).

Table 15. Local sensitivity analysis results for asymptomatic period

	Asymptomatic period		
	4.465	4.7	4.935
Attack rate	54.5545	55.594	56.6275

The sensitivity measure is calculated as follows:

$$S = \frac{\triangle Y}{0.05} = \frac{55.594 - 54.5545}{0.05} = 20.79 \text{ or } S = \frac{\triangle Y}{0.05} = \frac{56.6275 - 55.594}{0.05} = 20.67$$

The results show that by decreasing or increasing 5% of input value of asymptomatic period, the corresponding sensitivity measures of mean attack rates are 20.79 and 20.67, respectively.

3.4.4 Symptomatic rate

Table 16 shows the mean attack rates by the reference value of symptomatic rate (0.669) and by the $\pm 5\%$ of the reference value of symptomatic rate (0.70245 and 0.63555, respectively).

Table 16. Local sensitivity analysis results for symptomatic rate

	Symptomatic rate		
	0.63555	0.669	0.70245
Attack rate	55.3745	55.594	55.7845

The sensitivity measure is calculated as follows:

$$S = \frac{\triangle Y}{0.05} = \frac{55.594 - 55.3745}{0.05} = 4.39 \text{ or } S = \frac{\triangle Y}{0.05} = \frac{55.7845 - 55.594}{0.05} = 3.81$$

The results show that by decreasing or increasing 5% of input value of symptomatic rate, the corresponding sensitivity measures of mean attack rates are 4.39 and 3.81, respectively.

3.4.5 Asymptomatic infectivity

Table 17 shows the mean attack rates by the reference value of asymptomatic infectivity (0.5) and by the $\pm 5\%$ of the reference value of asymptomatic infectivity (0.475 and 0.525, respectively).

Table 17. Local sensitivity analysis results for asymptomatic infectivity

	Asymptomatic infectivity		
	0.475	0.5	0.525
Attack rate	54.572	55.594	56.5775

The sensitivity measure is calculated as follows:

$$S = \frac{\triangle Y}{0.05} = \frac{55.594 - 54.572}{0.05} = 20.44 \text{ or } S = \frac{\triangle Y}{0.05} = \frac{56.5775 - 55.594}{0.05} = 19.67$$

The results show that by decreasing or increasing 5% of input value of asymptomatic infectivity, the corresponding sensitivity measures of mean attack rates are 20.44 and 19.67, respectively.

3.4.6 Probability of staying home when sick

Table 18 shows the mean attack rates by the reference value of the probability of staying home when sick (0.543) and by the $\pm 5\%$ of the reference value (0.51585 and 0.57015, respectively).

Table 18. Local sensitivity analysis results for probability of staying home when sick

	Probability of staying home when sick		
	0.51585	0.543	0.57015
Attack rate	57.408	55.594	53.7965

The sensitivity measure is calculated as follows:

$$S = \frac{\triangle Y}{0.05} = \frac{55.594 - 57.408}{0.05} = -36.28 \text{ or } S = \frac{\triangle Y}{0.05} = \frac{53.7965 - 55.594}{0.05} = -35.956$$

The results show that by decreasing or increasing 5% of input value of probability of staying home when sick, the corresponding sensitivity measures of mean attack rates are -36.28 and -35.956, respectively.

3.4.7 Household contact rate

Table 19 shows the mean attack rates by the reference value of household contact rate (0.198) and by the $\pm 5\%$ of the reference value of household contact rate (0.188 and 0.208, respectively).

Table 19. Local sensitivity analysis results for household contact rate

	Household contact rate		
	0.188	0.198	0.208
Attack rate	54.692	55.594	56.4145

The sensitivity measure is calculated as follows:

$$S = \frac{\triangle Y}{0.05} = \frac{55.594 - 54.692}{0.05} = 18.04 \text{ or } S = \frac{\triangle Y}{0.05} = \frac{56.4145 - 55.594}{0.05} = 16.41$$

The results show that by decreasing or increasing 5% of input value of household contact rate, the corresponding sensitivity measures of mean attack rates are 18.04 and 16.41, respectively.

3.4.8 Neighborhood contact rate

Table 20 shows the mean attack rates by the reference value of neighborhood contact rate (42.479) and by the $\pm 5\%$ of the reference value of neighborhood contact rate (40.355 and 44.603, respectively).

Table 20. Local sensitivity analysis results for neighborhood contact rate

	Neighborhood contact rate		
	40.355	42.479	44.603
Attack rate	54.1025	55.594	57.056

The sensitivity measure is calculated as follows:

$$S = \frac{\triangle Y}{0.05} = \frac{55.594 - 54.1025}{0.05} = 29.83 \text{ or } S = \frac{\triangle Y}{0.05} = \frac{57.056 - 55.594}{0.05} = 29.24$$

The results show that by decreasing or increasing 5% of input value of neighborhood contact rate, the corresponding sensitivity measures of mean attack rates are 29.83 and 29.24, respectively.

3.4.9 School contact rate

Table 21 shows the mean attack rates by the reference value of school contact rate (14.320) and by the $\pm 5\%$ of the reference value of school contact rate (13.604 and 15.037, respectively).

Table 21. Local sensitivity analysis results for school contact rate

	School contact rate		
	13.604	14.320	15.037
Attack rate	55.361	55.594	55.8145

The sensitivity measure is calculated as follows:

$$S = \frac{\triangle Y}{0.05} = \frac{55.594 - 55.361}{0.05} = 4.66 \text{ or } S = \frac{\triangle Y}{0.05} = \frac{55.8145 - 55.594}{0.05} = 4.41$$

The results show that by decreasing or increasing 5% of input value of school contact rate, the corresponding sensitivity measures of mean attack rates are 4.66 and 4.41, respectively.

3.4.10 Workplace contact rate

Table 22 shows the mean attack rates by the reference value of workplace contact rate (1.589) and by the $\pm 5\%$ of the reference value of workplace contact rate (1.510 and 1.669, respectively).

Table 22. Local sensitivity analysis results for workplace contact rate

	Workplace contact rate		
	1.510	1.589	1.669
Attack rate	55.0215	55.594	56.108

The sensitivity measure is calculated as follows:

$$S = \frac{\triangle Y}{0.05} = \frac{55.594 - 55.0215}{0.05} = 11.45 \text{ or } S = \frac{\triangle Y}{0.05} = \frac{56.108 - 55.594}{0.05} = 10.28$$

The results show that by decreasing or increasing 5% of input value of workplace contact rate, the corresponding sensitivity measures of mean attack rates are 11.45 and 10.28, respectively.

3.5 DISCUSSIONS AND CONCLUSIONS

Table 23 summarizes above ten parameters considered above in increasingly order of their sensitivity measures. From these results, we can identify five parameters (asymptomatic infectivity, asymptomatic period, neighborhood contact rate, probability of staying home when

sick and symptomatic period) that have high sensitivity in the FRED model. The other five parameters have relatively low to medium sensitivity in the FRED model. We will consider the highest sensitive parameters (asymptomatic infectivity, asymptomatic period, neighborhood contact rate, probability of staying home when sick and symptomatic period) for further research.

Table 23. Local sensitivity analysis results for ten selected parameters

	Sensitivity Measure (Standardized)	
Parameter	-5%	+5%
Latent period	-0.182	-0.371
Symptomatic rate	4.39	3.81
School contact rate	4.66	4.41
Workplace contact rate	11.45	10.28
Household contact rate	18.04	16.41
Asymptomatic infectivity	20.44	19.67
Asymptomatic period	20.79	20.67
Neighborhood contact rate	29.83	29.24
Probability of staying home when sick	-36.28	-35.956
Symptomatic period	43.21	42.22

As mentioned in Chapter 2 Section 2.3, the reference value of contact rate is a format of matrix by different age groups. However, in FRED model, all the contact rates are assigned as fixed values. It will be difficult to convert the reference matrix values into FRED. Due to complicated input formats and the variety of the reference values, the contact rates in household, neighborhood, school, and workplace will use the default fixed values in FRED for further analysis.

The identification of high sensitivity input parameters provides important information for users of the FRED model, such as public health policy makers, about what factors may have the most influence in mitigating the spread of an infectious disease during an epidemic. For example, the probability of staying home when sick has high sensitivity in the FRED influenza model, suggesting that policy makers may wish to target this factor in their risk communications, encouraging symptomatic people to stay at home and help to limit the spread of disease. Our analysis also shows that reducing neighborhood or community contacts may also play a significant role to limit disease spread. This, the local sensitivity analysis may help policy makers to be more efficient for proper planning, monitoring, and decision-making.

In summary, this chapter uses local sensitivity analysis to identify high sensitivity parameters in FRED influenza model, and four parameters (asymptomatic infectivity, asymptomatic period, probability of staying home when sick and symptomatic period) will be used in Chapter 4 and 5 for probabilistic sensitivity analysis (uncertainty analysis).

CHAPTER 4 PROBABILISTIC SENSITIVITY ANALYSIS

This chapter introduces the background on global sensitivity analysis (Section 4.1), probabilistic sensitivity analysis (Section 4.2), and sampling procedures (Section 4.3), proposes the details of performing probabilistic sensitivity analysis in FRED model (Section 4.4), states the research hypotheses to compare two sampling procedures in FRED model (Section 4.5), then performs probabilistic sensitivity analysis with four selected parameters in FRED model and presents the results of statistical analysis (Section 4.6), followed by the discussions and conclusions in Section 4.7. This chapter uses the results from chapter 2 and 3 to perform probabilistic sensitivity analysis to compare two sampling procedures. The results of this chapter will be used for chapter 5 in an example of policy comparison question.

4.1 GLOBAL SENSITIVITY ANALYSIS

Global sensitivity analysis varies all parameters over their full ranges and looks at several different output measures to understand the full range of model behavior. It has advantages over local sensitivity analysis in that it can deal with the entire ranges of input parameters to understand the model behavior. Mishra et al. (68) applied stepwise rank regression analysis, mutual information analysis and classification tree analysis in conjunction with Monte Carlo simulation-based probabilistic analyses to identify key contributors of inputs to output variance

and determine the strength of input-output association. They stated that global sensitivity analysis performed better than local sensitivity analysis, because global analyses provided information not only regarding the sensitivity of each individual input parameters but also about which parameter was most influential to the model. However, global sensitivity analysis can be extremely time-consuming to run a vast number of combinations of parameter values. To overcome the time-consuming issue, one approach is using Monte Carlo simulation method to generate random samples from the predetermined ranges and distributions to evaluate the model for each element of the samples and estimate the effects of each factor on the model outputs.

There are two main procedures for performing global sensitivity and uncertainty analysis. One is sampling-based method. It obtains values of input parameters from selected ranges or distributions for model simulations and usually involves statistical analysis such as regression to estimate the impacts of input parameters on model outputs. Another procedure for global sensitivity analysis is ANOVA-like decomposition, which is a variance based method developed from Sobol' indices (69) to estimate the influence of individual variables or groups of variables on the model outputs.

One limitation of global sensitivity analysis is that the ranges of model outputs might be usually unrealistically wide if many input parameters are perturbed together. Such limitation can be overcome by treating the input parameters as random variables with specific distributions, which is known as probabilistic sensitivity analysis (uncertainty analysis)(70).

4.2 PROBABILISTIC SENSITIVITY ANALYSIS

Probabilistic sensitivity analysis is performed to quantify the impact of uncertainties in input parameters on the uncertainty in model outputs. The main task of probabilistic sensitivity analysis is to obtain the probability distribution of the model responses given the distributions of the input parameters (71). One of advantages of using probabilistic sensitivity analysis is that it can deal with a large set of random input parameters with the consideration of their correlations(72). It is especially useful when the relationships between the input parameters and the model outputs are nonlinear.

In probabilistic sensitivity analysis, one of the most challenging issues is the intensive computational demand for assessing the impact of probabilistic variations. Complex models are extremely time-consuming for computation and are impractical for use. An efficient approach to probabilistic sensitivity analysis is using optimized sampling procedures or statistical approach such as Bayesian method to improve the computational efficiency of sensitivity analysis(70, 71).

4.3 SAMPLING PROCEDURES

There are several sampling procedures that can be used to select the values of input parameters in probabilistic sensitivity analysis. First, systematic sampling could be used in the range of each input parameter. This procedure would be the simplest way to do sampling and may require a large number of simulation runs. Second, stratified random sampling can be used to generate a full factorial sampling for all input parameters. This procedure can evaluate the input parameters and their interactions to the model features. However, it may require a large number of

simulation runs when there are many strata for each input parameter. Third, Gaussian sampling is commonly used in simulation models(73). This procedure draws sample values near a point by using Gaussian distribution and biases the samples close to that point. Fourth, Latin Hypercube sampling procedure can be used. This procedure is a form of stratified sampling, can draw samples from equally space from all input parameters without replacement. Latin Hypercube sampling method was first introduced by McKay et al. (74) for computer experiment. This stratified sampling method ensures that all portions of sample space are sampled, and each of those components of sample space is represented in a fully stratified manner, no matter which components might turn out to be important.

The computational efficiency improvement of uncertainty analysis is highly dependent on the choice of sampling directions (75). Previous studies have proved that Monte Carlo simulations with the Latin Hypercube design produced more stable results and required fewer samples than random sampling method for the same accuracy for estimating statistics of a performance function (75, 76).

In this chapter, two sampling methods will be used and compared in the probabilistic sensitivity analysis: random sampling method and Latin Hypercube sampling method.

4.4 PROCEDURES OF PERFORMING PROBABILISTIC SENSITIVITY ANALYSIS IN FRED MODEL

The following criteria and steps are used to perform probabilistic sensitivity analysis in FRED model (Figure 6):

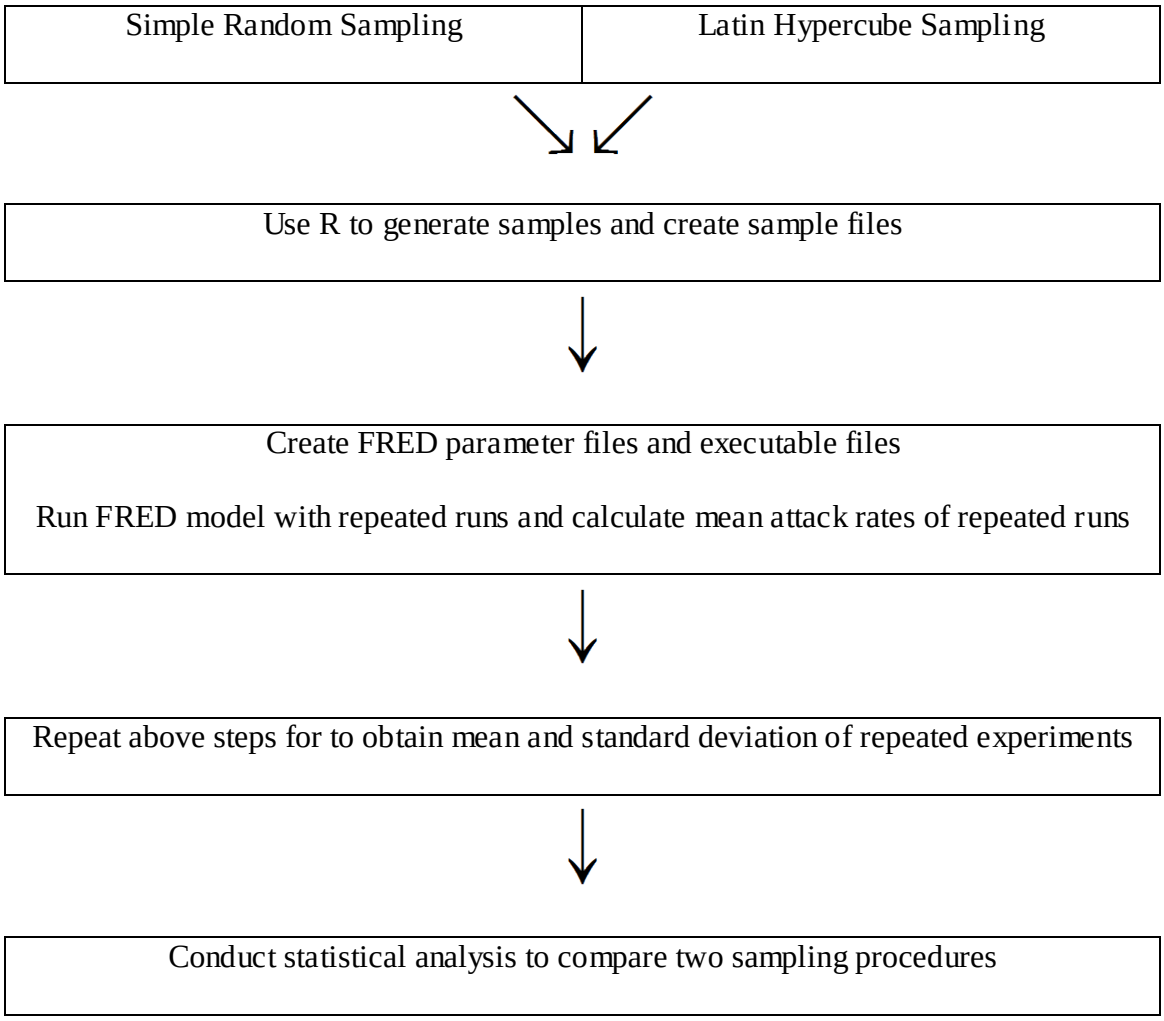

Figure 6. Procedures of performing probabilistic sensitivity analysis in FRED model

Step 1: perform simple random sampling procedure to select samples of the four parameters (asymptomatic infectivity, asymptomatic period, probability of staying home when sick and symptomatic period) from predetermined distributions in Chapter 2 and 3 (please see Appendix B.1 for R codes). In this sampling procedure, five sets of samples will be selected, each set contains four parameters and a sample size at 50, 100, 250, 500 and 1000, respectively.

Step 2: perform Latin Hypercube sampling procedure to select samples of the four parameters (asymptomatic infectivity, asymptomatic period, probability of staying home when sick and symptomatic period) from predetermined distributions in Chapter 2 and 3 (please see

Appendix B.2 for R codes). Each parameter will be divided into same number of regions with equal probability, and one sample will be randomly selected from each region. The sample size for each parameter is the total number of regions in its sample space. In this sampling procedure, four sets of samples will be selected, each set contains four parameters and a sample size at 50, 100, 250 and 500, respectively.

Step 3: in FRED model, use scripts to create parameter files for each sampling procedure from the data files obtained from Step 1 and Step 2 and create executable files to run the simulations for each sampling procedure under each condition of parameter files (please see Appendix B.3 and B.4 for the details of these scripts). Each condition of parameter files will have 20 repeated runs. Calculate the mean attack rates of 20 repeated runs under each condition of parameter files for each sampling procedure at each sample size and save the outputs of mean attack rates into a data file.

Step 4: repeat Step 1, 2 and 3 for n times (for example, n=20) to obtain the experiment errors of mean attack rates for both random sampling and Latin Hypercube sampling.

Step 5: summarize and report the results from step 3. Calculate the mean and standard deviation for each sampling procedure at each sample size. Use statistical methods to compare the results between two sampling procedures as well as the means among all sample sizes. Make conclusions based on the statistical results and choose the most efficient sampling procedure and sample size for further research.

4.5 RESEARCH HYPOTHESES

To compare random sampling method and Latin Hypercube sampling method, three hypotheses will be tested by two-sample t test:

1. Latin Hypercube sampling procedure produces smaller standard deviation than simple random sampling.

2. Latin Hypercube sampling procedure produces unbiased estimates as the simple random sampling.

3. The 20 repeated experiments in each sampling procedure at each sample size procedure 20 unbiased estimates as a large sample size (the gold standard, for example, N=10000).

4.6 RESULTS AND STATISTICAL ANALYSIS

To obtain the simulation results quickly, we choose Washington County, PA with 200,000 people as the population in FRED model. The time of running simulation in Washington County, PA takes much less time than Allegheny County, PA with about 1.2 millions people. Each condition of samples will take 20 repeated runs. For each sampling procedure and each sample size, we will generate 20 independent sample sets and run those 20 sample sets in FRED model. Therefore, the outputs for each sampling procedure and each sample size will be the mean attack rates and standard deviations of 20 repeated experiments from 20 independent sample sets.

Table 24 summarized the means and standard deviations of attack rates of 20 repeated experiments for each sampling procedure at each sample size.

Table 24. Means and variances of attack rates for each sampling procedure by sample size

Sample size	Sampling Procedure	
	Random mean (standard deviation)	Latin Hypercube mean (standard deviation)
50	53.134 (1.602)	52.505 (1.542)
100	52.801 (1.085)	53.106 (1.191)
250	52.893 (0.725)	52.829 (0.745)
500	52.786 (0.610)	52.920 (0.404)
1000	52.952 (0.437)	-

For hypothesis 1, the standard deviations between random sampling procedure and Latin Hypercube sampling procedure at each sample size was compared by two-sample t test. The null hypothesis is $H_0 : \sigma_{\text{random}} = \sigma_{\text{LatinHypercube}}$. The alternative hypothesis is $H_1 : \sigma_{\text{random}} \neq \sigma_{\text{LatinHypercube}}$.

$$H_0 : \sigma_{\text{random}} = \sigma_{\text{LatinHypercube}}$$

$$H_1 : \sigma_{\text{random}} \neq \sigma_{\text{LatinHypercube}}$$

Table 25 and Figure 7 summarized the results from two-sample t tests.

Table 25. Two-sample t test results of comparing standard deviations of attack rates between sampling procedure at each sample size

Sample size	Sampling Procedure		p value
	Random	Latin Hypercube	
50	1.602	1.542	0.7907
100	1.085	1.191	0.3539
250	0.725	0.745	0.6777
500	0.610	0.404	0.0001
1000	0.437	(compared with 0.404)	0.0464

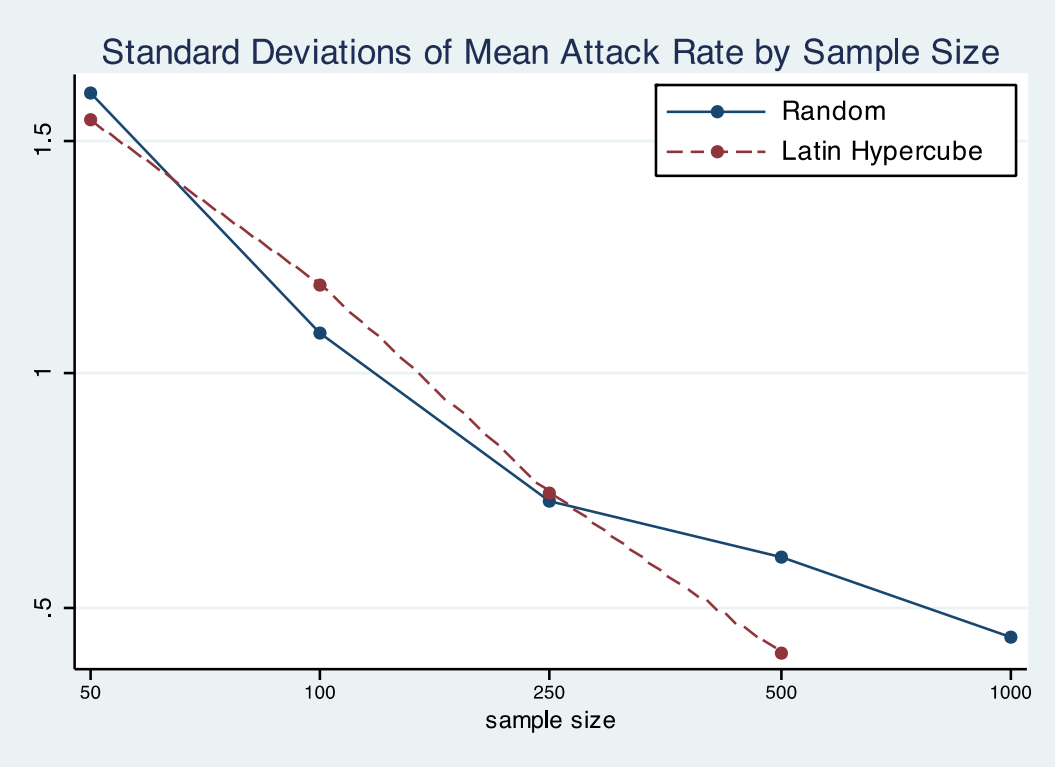

Figure 7. Standard deviation of mean attack rate by sample size

From the test results, there are no statistical differences in standard deviation between random sampling and Latin Hypercube sampling at sample size 50, 100 and 250. However, Latin Hypercube sampling at sample size 500 produces significantly smaller standard deviation than random sampling at sample size 500 ($p<0.0001$) and 1000 ($p<0.05$). It suggests that at sample size 500, Latin Hypercube sampling is preferable to random sampling using both 500 and 1000 samples.

For hypothesis 2, the mean attack rates between random sampling procedure and Latin Hypercube sampling procedure at each sample size was compared by two-sample t test. The null hypothesis is $H_0 : \mu_{\text{random}} = \mu_{\text{LatinHypercube}}$, the alternative hypothesis is $H_1 : \mu_{\text{random}} \neq \mu_{\text{LatinHypercube}}$.

$$H_0 : \mu_{\text{random}} = \mu_{\text{LatinHypercube}}$$

$$H_1 : \mu_{\text{random}} \neq \mu_{\text{LatinHypercube}}$$

Table 26 summarized the results from two-sample t tests

Table 26. Two-sample t test results of comparing mean attack rates between sampling procedure at each sample size

Sample size	Sampling Procedure		p value
	Random mean (standard deviation)	Latin Hypercube mean (standard deviation)	
50	53.134 (1.602)	52.505 (1.542)	0.0480
100	52.801 (1.085)	53.106 (1.191)	0.0595
250	52.893 (0.725)	52.829 (0.745)	0.3330
500	52.786 (0.610)	52.920 (0.404)	0.0001

From the test results, there is a borderline significant difference between the two sampling procedures at sample size 50 ($p=0.048$), and significant difference at sample size 500 ($p<0.0001$). However, there are no significant differences between two sampling procedures at sample size 100 and 250.

For hypothesis 3, a two-sample t test was used to compare the mean from each experiment in each sampling procedure at each sample size to the mean of a larger sample size (the gold standard). For each sampling procedure and sample size, there were 20 experiment means that were each compared with the gold standard. Here we choose a sample size of 10000 from random sampling as the gold standard for comparisons. The null hypothesis for each sampling procedure is $H_0 : \mu_{ij} = \mu_{10000}$, the alternative hypothesis is $H_1 : \mu_{ij} \neq \mu_{10000}$, while i= 50, 100, 250, 500 and 1000, j=1, 2, …, 20.

$H_0 : \mu_i = \mu_{10000}$ versus $H_1 : \mu_i \neq \mu_{10000}$ (i= 50, 100, 250, 500 and 1000; j=1, 2, …, 20)

The overall mean and standard deviation from the gold standard are 52.928 and 12.950, respectively. Table 27 summarizes the proportions of unbiased mean attack rates from above results, showing the percent of the 20 experiments for which the experimental mean was significantly different from the mean of the gold standard. The results show that one of 20 means from 20 experiments is statistically significantly different from the gold standard at sample size 50 and 100 for both two sampling procedures. At sample size 250 and 500, none of the 20 means from 20 experiments in Latin Hypercube sampling are significantly different from the gold standard. However, for random sampling at sample size 250 and 500, one of 20 means from 20 experiments is significantly different from the gold standard. At sample size 1000, all the 20 means from 20 experiments in random sampling are not significantly different from the gold standard.

Table 27. Certainty of unbiased mean attack rates as the gold standard (10000 samples from random sampling)

Sample size	Sampling Procedure	
	Random	Latin Hypercube
50	95%	95%
100	95%	95%
250	95%	100%
500	95%	100%
1000	100%	-

The above results show the probability of producing the same unbiased mean attack rates as the gold standard among 20 repeated experiments for each sampling procedure at each sample size. For Latin Hypercube sampling, there are no statistical differences in mean attack rates

among repeated 20 experiments when the sample size is equal to 250 or higher. But for random sampling, the minimum sample size to produce 100% of unbiased mean attack rates is 1000. The results may suggest that the minimum sample size we can choose for further study is 250 from Latin Hypercube sampling.

4.7 DISCUSSIONS AND CONCLUSIONS

Since FRED is a stochastic simulation model, in this chapter, we considered two orders of uncertainty in the results via probabilistic sensitivity analysis. The first-order uncertainty was addressed by repeated simulation runs for each condition of sample sets. Here, we used 20 simulation runs. The second order uncertainty was addressed through repeated experiments for each sample size. We generated 20 independent sample sets (20 experiments) to obtain and analyze experiment errors. In other words, the means and standard deviations in Table 24 were obtained from 20 repeated experiments for each sampling method (random sampling and Latin Hypercube sampling) at each sample size, and each experiment included 20 repeated simulation runs for each component of sample sets. The two-order control in uncertainty in stochastic simulation model by repeated measures can help to reduce the experiment errors and obtain unbiased estimates.

As shown in Section 4.6, Latin Hypercube sampling can produce significant smaller experiment errors in estimating the mean attack rate than random sampling in the FRED model. The certainty analysis using the unbiased mean attack rates using 10000 samples from random sampling as the gold standard shows that Latin Hypercube sampling can produce 100% unbiased mean attack rates at a smaller sample size (250) than random sampling (1000). From the above

results we will choose 250 samples from Latin Hypercube sampling for next chapter in an intervention comparison study.

There are several factors that may affect the choice of minimum sample size. For example, a smaller number of initial infected cases could result in more variation in the outputs of the FRED model. The demographic features associated with the initial infected cases (such as age) could also affect the outputs of the FRED model when the number of initial infected cases is very small (see Appendix A for more details). Therefore, it is recommended to follow the procedures in Section 4.4 and use consistent initial settings to determine the minimum sample size for different initial settings before conducting further research in intervention strategies in the FRED model.

The results in this chapter may have an important impact on policy makers who use simulation models for decision-making. All the results in here are about the computational improvement in running probabilistic sensitivity analyses on the FRED simulation model. In practice, these improvements of computational efficiency mean that it will now be feasible to perform uncertainly analysis for our models more quickly, thereby improving the quality of the results in the simulation model. We believe that more extensive uncertainty analysis will increase the trust-worthiness of simulation models, and will improve the advice provided to policy makers. For example, rather than provided a single estimate of a model's output, we will now be able to describe the likely range of outputs in a systematic way.

In summary, this chapter provides a detailed procedure to perform probabilistic sensitivity analysis and to compare simple random sampling and Latin Hypercube sampling with repeated measures for controlling uncertainty in stochastic model. The procedure is used to determine the optimal sample size and sampling procedure for the uncertainty analysis of the

FRED system. The procedure and relative conclusions provide the foundation for the next chapter, where we will apply probabilistic sensitivity analysis to a sample policy comparison question.

CHAPTER 5 UNCERTAINTY ANALYSIS OF AN EXAMPLE OF SCHOOL CLOSURE INTERVENTION STRATEGIES

This chapter gives an overview of the background on uncertainty analysis and school closure intervention strategies in pandemic (Section 5.1), proposes the procedure of uncertainty analysis in school closure intervention strategies in FRED model (Section 5.2), states the research hypotheses to compare different school closure intervention strategies in FRED model (Section 5.3), then presents the results of uncertainty analysis and statistical analysis for comparing 5 different school closure intervention strategies in FRED model (Section 5.4), calculate the cost-effectiveness for different school closure intervention strategies (Section 5.5), followed by the discussions and conclusions in Section 5.6. This chapter uses all the results from previous chapters (Chapter 2-4) to perform uncertainty analysis to compare different school closure intervention strategies and calculate the cost-effectiveness for decision-making.

5.1 INTRODUCTION

Uncertainty analysis is defined as quantifying the uncertainties of input parameters and making technical contributions to decision-making problems in models or experiments. It is usually addressed through sensitivity analysis (for example, probabilistic sensitivity analysis). In literature, probabilistic sensitivity analysis and uncertainty analysis and are closely related but

with different main tasks: probabilistic sensitivity analysis is the process of obtaining the probability distribution of the responses of a model given the probability distributions of the input parameters of the model, while uncertainty analysis is the process of obtaining the probability associated with a policy decision based on a model given the probability distributions of the input parameters of the model. In this chapter, uncertainty analysis is performed through probabilistic sensitivity analysis to access the uncertainty of the length of school closure intervention strategy.

School closure is defined as closing of a school and sending of all the children and staff home. It is different from class dismissal while a school remains open with administrative staff but most children stay home. The trigger for closure is defined as the number of cases that triggers a school to be closed, and it is very crucial. Since school can be closed for many reasons, in this study, we only consider school closure as a mitigation strategy for an epidemic.

School closure could limit community spread of the disease, protect particularly vulnerable students, and respond to staff shortages or student absenteeism. In 2009 H1N1 influenza pandemic, school closures were used as mitigation policies in communities or schools. Cauchemez et al. (77) estimated the impact of school closure on influenza transmission from sentinel data and showed that holidays prevent 18-21% of seasonal influenza cases in children and prolonged school closure during a pandemic might reduce the cumulative number of cases by 18-23% in children and peak attack rates by 47-52% in children. Heymann et al. (78) examed 1.86 millions children aged between 6 and 12 years old in Israel in Maccabi Healthcare Services and found out a significant drop in respiratory infection during a 2-week school closure period from 44.0/1000 infected children to 24.7/1000 infected children, and an increase to 30.8/1000 infected children after school closure period. Lee et al. (15) discussed that entire school system

closures were not more effective than individual school closures, and relatively short school closures (i.e., 2 weeks or less) may slightly increase the overall attack rate because susceptible students were sent back into schools in the middle of the epidemic.

The effect of school closure is associated with some factors of influenza pandemic, such as the transmission rate, the reproductive rate R_0, and the contact rate etc. For example, Cauchemez et al. (79) showed that school closures may have an important effect on disease spread if 50% of transmissions occur in schools, however, if only 20% of transmissions happen in schools, school closure may have a much smaller effect. Vynnycky and Edmunds (80) studied the 1957 Asian influenza pandemic in the United Kingdom and showed that if the reproductive rate R_0 is high (e.g. 2.5 or 3.5), school closure can reduce the epidemic size by a very small amount (<10%); and if the reproductive rate R_0 is low (e.g. 1.8), school closure can reduce the epidemic size by a modest amount (e.g. 22%). Their studies indicated that R_0 is an important factor in an epidemic. Halder et al. (81) reported that 8-week school closure intervention could reduce the illness attack rate from 33% to 19% in the community for an influenza pandemic with R_0 at 1.5; for R_0at 2.0 or higher, 8-week school closure intervention would be less effective (from 50% to 41%).Studies also showed that individuals made substantially fewer contacts when they were ill than when they were well or in holidays and weekend than during weekdays, suggesting that school closure can have a substantial impact on the spread of an infectious disease transmitted via close (non sexual) contacts (82, 83).

The values and impacts of school closures for mitigating influenza pandemic are associated with some related issues, such as economic costs, ethical concerns and operational issues etc. For example, Klaiman et al. (84) showed that school closure is relevant to its timing, nature, and duration. The unclear rationales for closure can challenge the effectiveness of school

closure intervention. Brown et al. (5) reported that closing schools resulted in substantially higher net costs than not closing schools. The median cost per influenza case averted for 8-week school closure would be as high as $53,461 for R_0 at 2.0.Sander et al. (85) estimated a high economic cost of school closure about $2.7 million per 1000 population per work (6% of GDP). Berkman (86) showed that school closures may raise serious ethical concerns such as lack of evidence of the efficacy of school closures, the impacts of school closures to vulnerable populations, cost-benefit of school closures, and the equitable distribution of intervention etc. Johnson et al. (87) conducted a questionnaire survey to 201 households in North Carolina and found that short-term school closure would not cause substantial hardship for parents. Similarly, Gift et al. (88) surveyed 214 households in Pennsylvania after a 1-week elementary school closure because of the 2009 H1N1 influenza pandemic and found that 79% of households reported that adults missed no days of work to watch children.

In a systematic review study conducted by Jackson et al. (89) for reviewing the effects of school closures on pandemic and seasonal influenza outbreaks, the authors concluded that school closures played an important role in reducing influenza transmission, but the optimal strategy for school closure such as ideal length and timing of closure was still unclear, and other implications of school closure such as ethical and economic considerations should also be considered in policy decisions.

To address the question of the optimal strategy for school closure such as ideal length, the uncertainty analysis could be used to compare different length of school closure intervention strategies as well as their cost-effectiveness of different length of school closure intervention strategies. In this chapter, uncertainty analysis is performed via probabilistic sensitivity analysis.

5.2 PROCEDURES OF PERFORMING UNCERTAINTY ANALYSIS TO COMPARE DIFFERENT SCHOOL CLOSURE INTERVENTION STRATEGIES IN FRED MODEL

In this section, Washington County, PA with about 200,000 people is used as the population in FRED model. Five scenarios of school closure intervention strategies will be used in uncertainty analysis: none, 2-week, 4-week, 6-week, and 8-week. The number of triggers for triggering a school closure intervention strategy is 10. In other words, once there are 10 cases of infected persons in school in a day of epidemic, the FRED model will start the school closure intervention strategy on that day, then simulate and calculate the consequence results. The default reproductive rate R_0 is 1.5 in the FRED model. The Latin Hypercube sampling procedure with a sample size 250 will be used in this chapter. The outputs are the mean attack rates for each scenario of school closure intervention strategies.

The following criteria and steps are used to perform uncertainty analysis to compare different school closure intervention strategies in FRED model (Figure 8):

Step 1: perform Latin Hypercube sampling procedure to select samples of the four parameters (asymptomatic infectivity, asymptomatic period, probability of staying home when sick, and symptomatic period) from predetermined distributions specified in Chapter 2 and 3 (please see Appendix B.2 for R codes). Each parameter will be divided into same number of regions with equal probability, and one sample will be randomly selected from each region. The sample size for each parameter is 250.

Step 2: in FRED model, create parameter files for each sampling procedure from the data files obtained from Step 1 and create executable files to run the simulations under 5 different scenarios of school closure intervention strategies: none, 2-week, 4-week, 6-week, and 8-week (please see Appendix B.3 and B.4 for the details of above scripts). Each scenario of school

closure intervention strategies includes 250 samples to run in the FRED model. Then calculate the mean attack rates for each condition of parameter files and save the outputs of mean attack rates into a data file.

Step 3: summarize and report the results from Step 2. Calculate the mean and variance for each scenario of school closure intervention strategies. Use statistical methods to compare the results among all scenarios of school closure intervention strategies. Make conclusions based on the statistical results for uncertainty analysis of school closure intervention strategies comparison.

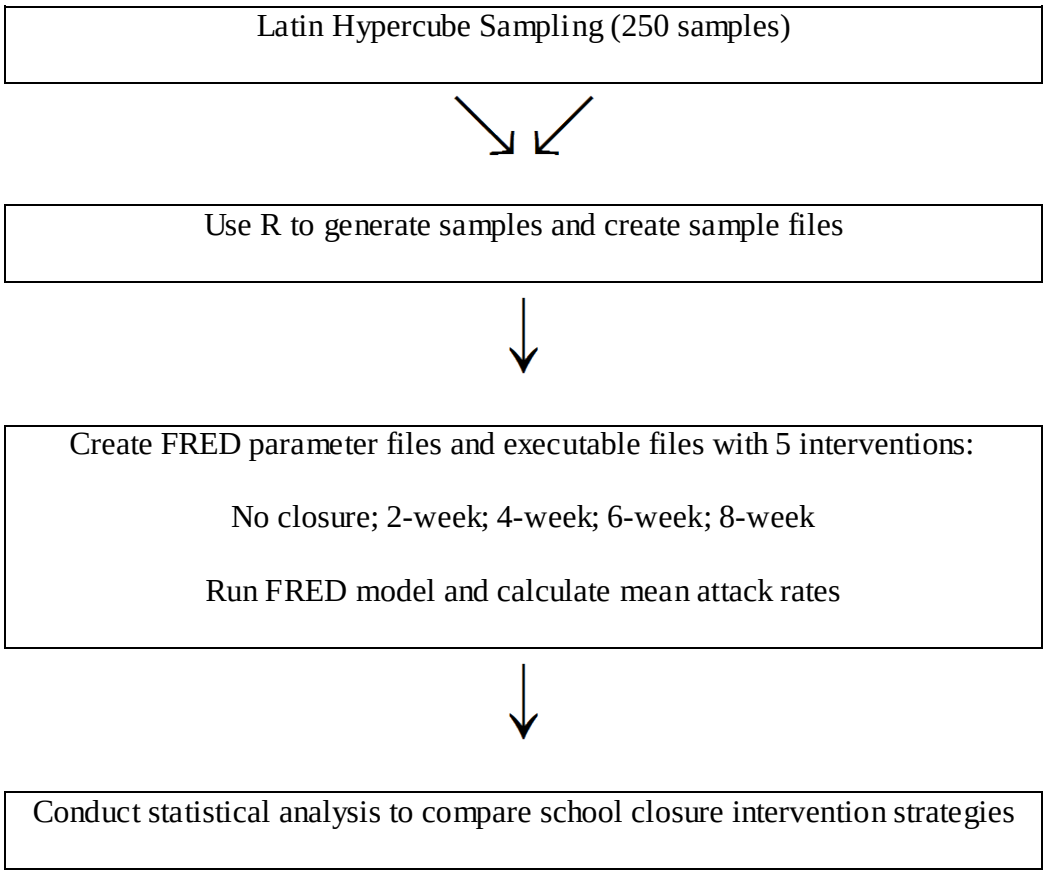

Figure 8. Procedures of performing uncertainty analysis to compare different school closure intervention strategies in FRED model

5.3 RESEARCH HYPOTHESES

To compare all scenarios of school closure intervention strategies, two directions of hypotheses will be tested by using paired two-sample test:

1. Compare to the no school closure intervention strategy, school closure intervention strategies reduce the mean attack rates.

2. Compare to the longest school closure intervention strategy (8-week), shorter duration of school closure intervention strategy can produce the same reduction in mean attack rates.

5.4 RESULTS AND STATISTICAL ANALYSIS

Table 28 shows the proportions of reduced mean attack rates from longer duration of school closure to shorter duration of school closure. For example, in all 250 samples, 99.6% of mean attack rates in 4-week school closure intervention strategy are lower than the mean attack rates in 2-week school closure intervention strategy. The proportions shown in Table 28 reflect the reliability of FRED model in running simulations under different intervention strategies using the same sample set. For example, using the same sample set, 99.6% of mean attack rates are lower in 8-week school closure intervention strategy than in 6-week school closure intervention strategy. The overall proportions for all school closure intervention strategies are 99.6% and higher, which showing reasonable reliability of FRED model in running simulations under different conditions by using same sample set.

Table 28. Proportions of reduced mean attack rates in longer duration of school closure compared to shorter duration of school closure

School Closure	School closure				
	None	2-week	4-week	6-week	8-week
2-week	100.0%	-			
4-week	100.0%	99.6%	-		
6-week	100.0%	100.0%	100.0%	-	
8-week	100.0%	100.0%	100.0%	99.6%	-

For hypothesis 1, use the no school closure intervention strategy as the baseline, the null hypothesis is $H_0 : \mu_{\text{no closure}} = \mu_{\text{school closure}}$, the alternative hypothesis is $H_1 : \mu_{\text{no closure}} \neq \mu_{\text{school closure}}$.

Table 29 shows the results from paired two-sample t test between no school closure intervention strategy and school closure intervention strategies.

Table 29. Paired two-sample t test results of comparing mean attack rates among school closure intervention strategies

School closure	Attack rate	p value
No closure	53.170 (12.738)	baseline
2-week	50.086 (13.223)	0.0082
4-week	47.577 (12.784)	0.0001
6-week	45.986 (12.398)	0.0001
8-week	45.033 (12.499)	0.0001

From the results, there are significant differences in mean attack rates between no school closure intervention strategy and school closure intervention strategies. In other words, school closure intervention strategies can reduce the mean attack rates significantly as short as 2-week closure. This conclusion is not consistent with all previous studies. For example, Heymann et al.

(78) found out a significant drop in respiratory infection during a 2-week school closure period and an increase in infection after school closure period. However, Lee et al. (15) reported that relatively short school closures (i.e., 2 weeks or less) may slightly increase the overall attack rate because susceptible students were sent back into schools in the middle of the epidemic. One possible reason is that school closure intervention strategy is associated with many factors during an epidemic, the different environment settings (for example, different infectious disease settings, population settings) in the model could lead to slight different results. However, the FRED influenza model is stable in simulation results for the same environment.

For hypothesis 2, use the longest school closure intervention strategy (8-week) as the baseline, the null hypothesis is $H_0 : \mu_{\text{8-week closure}} = \mu_{\text{less than 8-week closure}}$, the alternative hypothesis is $H_1 : \mu_{\text{8-week closure}} \neq \mu_{\text{less than 8-week closure}}$.

$$H_0 : \mu_{\text{8-week closure}} = \mu_{\text{6-week closure}} \text{ or } \mu_{\text{8-week closure}} = \mu_{\text{4-week closure}} \text{ or } \mu_{\text{8-week closure}} = \mu_{\text{2-week closure}}$$

$$H_1 : \mu_{\text{8-week closure}} \neq \mu_{\text{6-week closure}} \text{ or } \mu_{\text{8-week closure}} \neq \mu_{\text{4-week closure}} \text{ or } \mu_{\text{8-week closure}} \neq \mu_{\text{2-week closure}}$$

Table 30 shows the results from paired two-sample t test among school closure intervention strategies.

Table 30. Paired two-sample t test results of comparing mean attack rates among school closure intervention strategies

School closure	Attack rate	p value
2-week	50.086 (13.223)	0.0001
4-week	47.577 (12.784)	0.0249
6-week	45.986 (12.398)	0.3925
8-week	45.033 (12.499)	baseline

From the results, there are statistically significant differences in mean attack rates between 8-week closure and 2-week, and 8-week and 4-week, respectively. However, there is no statistically significant difference between 6-week and 8-week, indicating that longer duration of school closure intervention strategy (8-week) might not reduce the mean attack rates significantly lower than shorter duration of closure (6-week).

Some other factors, such as the conditions that trigger a school to closure, may also play a role in school closure intervention strategies. For example, the number of triggers that triggers a school to be closed. In this section, we used 10 triggers for all school closure intervention strategies. However, school closure intervention strategies may produce different model outputs in mean attack rates depending on the settings of number of triggers. Table 31 shows the mean attack rates assuming different number of cases as triggers and different durations of school closure intervention strategies while other input parameters remain their default settings in Allegheny County, PA in FRED model.

Table 31. Mean attack rates by number of triggers and duration of school closure intervention strategies

	Duration (week)					
Triggers	None	1	2	4	6	8
1	33.001	32.53	32.298	31.856	31.527	31.26
5	33.001	32.31	31.958	31.242	30.73	30.412
10	33.001	32.022	31.417	30.376	29.696	29.465

The regression results show that there is significant difference in mean attack rates by different number of triggers ($p<0.001$) especially for longer duration of school closure intervention strategies.

5.5 COST-EFFECTIVENESS OF SCHOOL CLOSURE INTERVENTION STRATEGIES

In this section, a simple cost-effectiveness calculation is provided to access the uncertainty of the different scenarios of school closure intervention strategy in terms of cost-effectiveness. The cost-effectiveness of school closure intervention strategies is defined as follows (5):

$$Cost_{\text{per case averted}} = \frac{\Delta Cost}{\Delta Case} = \frac{Cost_{\text{school closure strategy A}} - Cost_{\text{school closure strategy B}}}{Case_{\text{school closure strategy B}} - Case_{\text{school closure strategy A}}}$$

Assume the weekly cost for school closure is *W*, the total population is *P*, then the cost per influenza case averted is defined as:

$$Cost_{\text{per case averted}} = \frac{W \times (A_{\text{number of weeks of school closure}} - B_{\text{number of weeks of school closure}})}{P \times (\text{Attack Rate}_{\text{school closure strategy B}} - \text{Attack Rate}_{\text{school closure strategy A}})}$$

For example, the cost-effectiveness of school closure between 8-week and no closure is

$$Cost_{\text{per case averted}} = \frac{W \times (8-0)}{P \times (\text{Attack Rate}_{\text{8-week}} - \text{Attack Rate}_{\text{no closure}})} = \frac{8W}{P(53.17-45.03)} = \frac{8W}{8.14P} = 0.983\frac{W}{P}$$

Table 32 shows the calculation results of overall cost-per-case-averted of school closure intervention strategies.

Table 32. Cost-per-case-averted of school closure intervention strategies

School closure	2-week	4-week	6-week	8-week
Cost-per-case-averted	$0.649\frac{W}{P}$	$0.716\frac{W}{P}$	$0.836\frac{W}{P}$	$0.983\frac{W}{P}$

From the cost-per-case-averted calculation results, we found that longer duration of school closure strategy is not cost-effective compared with shorter duration of school closure strategy.

To assess the uncertainty of the cost-effectiveness in school closure intervention strategies, we calculated the cost-per-case-averted for all 250 samples for each school closure intervention strategy. We compare the cost-per-case-averted between two school closure intervention strategies. Table 31 shows the proportions of the cost-per-case-averted of shorter duration of school closure is lower than the cost-per-case-averted of longer duration of school closure. For example, among all 250 samples, 90.0% of the cost-per-case-averted of 2-week school closure is lower than the cost-per-case-averted of 8-week school closure. The proportions shown in table 33 can provide the information about how certainly one intervention strategy is more cost-effective than another one.

Table 33. Certainty of the cost-per-case-averted of shorter duration of school closure (labeled at the top of each column) is lower than the cost-per-case-averted of longer duration of school closure (labeled in each row)

School closure	2-week	4-week	6-week	8-week
4-week	66.4%			
6-week	76.8%	99.6%		
8-week	90.0%	100.0%	96.8%	-

From the results, only 66.4% of the cost-per-case-averted of 2-week school closure is lower than the cost-per-case-averted of 4-week school closure, while the proportions are much higher from 4-week of school closure to 6-week of school closure (99.6%), and from 6-week of school closure to 8-week of school closure (96.8%).

Figure 9 displays the cost-effectiveness acceptability curve for the cost-per-case-averted of different school closure intervention strategies among all 250 samples in term of dollar cost.

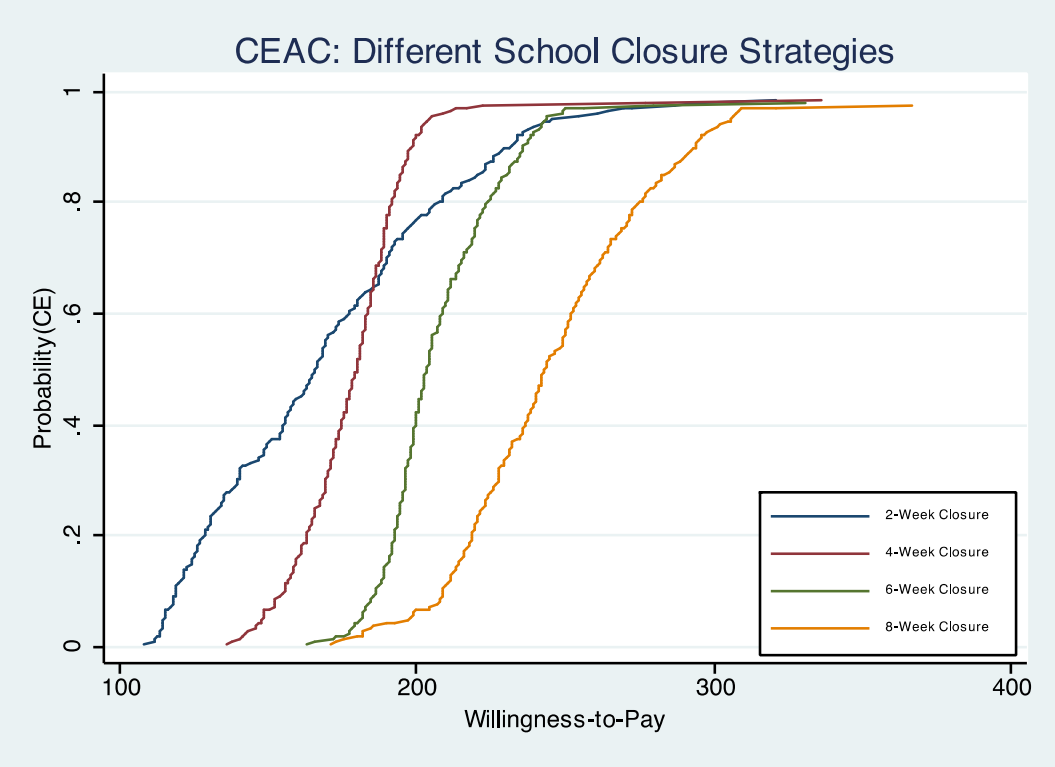

Figure 9. Cost-effectiveness acceptability curve for different school closure strategies

In the cost-effectiveness acceptability curve, we assume the weekly net cost for school closure is about $50 million, as estimated in Brown et al. (5) with adjustments based on population for Washington County, PA. The total population in Washington County, PA is 0.2 million. The overall means of cost-per-case-averted for 2-week, 4-week, 6-week and 8-week of school closure are $162.25, $179, $209 and $245.75, respectively.

The cost-effectiveness results are not exactly consistent from previous studies. For example, in a similar study conducted by Brown et al. (5), they used an earlier version of the FRED model to evaluate different school closure intervention strategies in Allegheny County, PA, and found that the cost-per-case-averted increased at 2-week school closure compared with 1-week school closure and was most cost-effective at 8-week school closure. A possible reason

for the differences may be that we use a simplified calculation for the weekly cost of school closure. On the other hand, in Brown et al.'s (5) study, they only used a single point estimates for the input parameters in FRED model while we sampled a wider range of input parameters from their probability distributions. Therefore, one could expect some differences in the two sets of simulation results.

5.6 DISCUSSIONS AND CONCLUSIONS

This chapter provides a detailed procedure for applying previous results from Chapter 4 to compare intervention strategies in the FRED model. Uncertainty analysis is performed through probabilistic sensitivity analysis to systematically investigate the effects of variation in uncertain input parameters and access the uncertainty of intervention comparison through probabilistic sensitivity analysis, statistical analysis and cost-effectiveness analysis. From the results, school closure strategies significantly reduce the mean attack rates compared with no school closure strategy, and longer duration of school closure intervention strategy (8-week school closure) can reduce the mean attack rates significantly lower than shorter duration of school closure intervention strategy (except 6-week school closure). However, cost-effectiveness analysis showed that longer duration of school closure is not cost-effective compared to shorter duration of school closure.

Certainty results between two school closure intervention strategies are provided in both mean attack rates and cost-effectiveness calculation. The proportions of reduced mean attack rates and proportions of lower cost-effectiveness between two school closure intervention strategies provide information for the reliability of the FRED model in running the same sample

set under different intervention strategies and the certainty of one intervention is more cost-effective than another, respectively.

This chapter provides policy makers with the intervention comparison results in FRED model. All the analyses are performed based on the solid foundations and computational optimizations from previous chapters. Therefore, the results illustrate that uncertainty analysis is feasible even in large-scale agent-based simulation models such a FRED. Simulations can be used not only to provide timely responses to intervention comparison questions, but can also quantify the degree of certainty associated with the comparison. We believe that providing this additional information about the expected range of results will help to improve the confidence in the modeling process among policy makers. In public health practice, the procedures and methods developed here will help provide efficient tools for policy makers to make decision when there is a crisis.

CHAPTER 6 DISCUSSIONS AND CONCLUSIONS

This chapter presents the contributions and implications of this study (Section 6.1) and limitations of this study (Section 6.2), summarizes the conclusions with two perspectives: (1) how other FRED model users apply our study results in future research (Section 6.3); (2) how policy makers benefit from our study results (Section 6.4).Possible future work that should be considered is also discussed in Section 6.5.

6.1 CONTRIBUTIONS AND IMPLICATIONS

The purpose of this thesis is to develop appropriate statistical methods and procedures for dealing with parameter uncertainty and improving the efficiency of sensitivity analysis in a large-scale agent-based model of infectious disease, and to apply those procedures in an example of policy evaluation question. This study emphasizes the important role of sensitivity analysis, uncertainty analysis and statistical analysis in computer simulation models and the use of those methods in such models to improve the computational efficiency of sensitivity analysis, and make great use of simulation results for decision-making. As stated in Chapter 1, the FRED model could computationally be very intensive when dealing with large sets of samples from input parameters. Time-consuming is a very important issue because longer time of simulation

runs could be very expensive, especially in large-scale simulation models. Therefore, it is importance to improve the computational efficiency of analysis.

The intensive computational issue is addressed through the approach of probabilistic sensitivity analysis and optimal sampling procedure. Before conduct probabilistic sensitivity analysis, literature review (Chapter 2) and local sensitivity analysis (Chapter 3) are performed to identify key parameters in FRED model. The purpose of literature review is to identify the reference values and distributions of important parameters in FRED model, and use those results for all later chapters. In literature review, systematic review papers with results of meta-analysis are preferable to identify the reference values of parameters (for example, latent period, symptomatic period, asymptomatic period, and symptomatic rate). When there are no systematic review papers available, the reference value will be chosen based on the average of reviewed papers (for example, the probability of staying home when sick). Some parameters (for example, contact rates and transmission probability) have complicated structure in FRED model, lack sufficient evidence from research for use as reference values in FRED model.

The purpose of local sensitivity analysis is to identify high sensitivity parameters in FRED model, and use those selected high sensitivity parameters in probabilistic sensitivity analysis. A detailed procedure for performing local sensitivity analysis and calculating sensitivity measure in FRED model is provided. The sensitivity measure is used to high sensitivity parameters in FRED model, and four parameters (asymptomatic infectivity, symptomatic period, probability of staying home when sick, and asymptomatic period) are selected to use in probabilistic sensitivity analysis. Some detailed algorithms are developed for varying discrete cumulative distribution functions for some parameters in the format of discrete cumulative distribution functions in the FRED model. Those algorithms are implemented as R program and

used as part of procedure for easily and automate computing samples for use in both local sensitivity analysis and probabilistic sensitivity analysis in FRED model.

The results and conclusions from literature review and local sensitivity analysis are in preparation to perform probabilistic sensitivity analysis (Chapter 4). A detailed procedure for performing probabilistic sensitivity analysis is provided to compare random sampling and Latin Hypercube sampling in FRED model. The procedure of performing probabilistic sensitivity analysis is implemented as an automated probabilistic sensitivity analysis script in FRED model. In the sampling comparison results, Latin Hypercube sampling produces significant smaller standard deviation than random sampling at sample size 500, and has 100% of certainty of producing unbiased mean attack rates as the gold standard (10000 samples from random sampling). The results suggest a minimum sample size of 250 from Latin Hypercube sampling for the same accuracy in estimates as gold standard. In this chapter, the computational efficiency of sensitivity analysis is addressed through the sampling comparison. An efficient sampling procedure and a reduced sample size are selected for future research with strong confidence from repeated experiments.

All the results and conclusions from literature review, local sensitivity analysis and probabilistic sensitivity analysis are in preparation to a school closure intervention comparison study (Chapter 5). The results show that school closure strategies can significantly reduce the mean attack rates as short as 2-week closure, and 8-week closure can reduce the mean attack rate significantly lower than 2-week closure and 4-week closure. However, longer duration of school closure strategy is less cost-effective than shorter duration of school closure strategy. In this chapter, we not only provide the comparison results, but also provide certainty results on the

reliability of FRED model in running the same sample set under different intervention strategies and the certainty of one intervention is more cost-effective than another.

6.2 LIMITATIONS

There are some possible limitations of this study. In local sensitivity analysis (Chapter 3), some parameters (for example, neighborhood contact rate) with high sensitivity were not included in the probability sensitivity analysis and uncertainty analysis of school closure intervention comparisons due to their complicated input formats in FRED model. Those high sensitivity parameters should not be neglected since they have high influence on the model outputs. In probabilistic sensitivity analysis (Chapter 4), due to a limited computation time, we only generated 5 sets of different sample size for random sampling and 4 sets of different sample size for Latin Hypercube sampling. For a better understanding of the differences in producing unbiased means and reduced standard deviations of repeated measures between random sampling and Latin Hypercube sampling, more sets of sample size should be considered and compared. In addition, we selected 10000 samples from random sampling as the gold standard, while a larger sample size (for example, 100,000) may be preferable. In the school closure intervention comparisons (Chapter 5), we conducted a simplified calculation of cost-effectiveness, which may not reflect all real situations of school closure and may lead to different conclusions compared with previous studies. Some of above limitations may be overcome by future work (see Section 6.5 for more details).

6.3 HOW OTHER FRED MODEL USERS APPLY OUR STUDY RESULTS IN FUTURE RESEARCH

In this study, a FRED influenza model is used to perform all the analyses. However, there are many different FRED models under the FRED framework depending on specific sets of the parameters. The methods and procedures that developed in this study can be generalized to all kinds of FRED models under the FRED framework.

All the procedures developed in this study are implemented as R program and automated probabilistic sensitivity analysis scripts in FRED model. Therefore it is easy for other FRED model users to follow and use those scripts to run in other FRED models. If other FRED model users are using the same FRED influenza model as in this study, they can just apply all the computational results (for example, optimal sample size and sampling procedure) in this study to different intervention comparison questions. However, if other FRED model users are using different FRED model under the FRED framework, they may need to follow all the procedures and the automated scripts provided in this study to determine the optimal methods in their FRED model to improve the computational efficiency of simulation runs.

6.4 HOW POLICY MAKERS BENEFIT FROM OUR STUDY RESULTS

Policy makers and researchers work in very different environments. Policy makers are focused on the practical solutions to particular policy issues. They need accessible information to help them make decisions or evaluate policy choices. For researchers, it would be a challenge to

translate their research findings into policy useful material, and support policy makers in identifying appropriate solutions to problems.

Policy makers may be skeptical of models and the simulation results from those models. To increase the trust in the model simulation results in this study, we emphasize the importance of uncertainty analysis, address the uncertainty analysis through probabilistic sensitivity analysis, repeated measures, and statistical analysis to ensure a strong confidence of the results, and provide certainty results for the reliability of FRED simulation model. Our results illustrate that uncertainty analysis is feasible even in large-scale agent-based simulation models such a FRED. The results from uncertainty analysis not only quantify the degree of certainty associated with the comparison, but also help to improve the confidence in the modeling process among policy makers. For example, in probabilistic sensitivity analysis (Chapter 4), rather than provide a single estimate of a model's output, we will now be able to describe the likely range of outputs in a systematic way, which will increase the trust-worthiness of simulation models and improve the advice provided to policy makers.

On the other hand, we also interpret some results with consideration in public health practice. For example, in local sensitivity analysis (Chapter 3), policy makers may pay more attention to those factors that may have the most influence in mitigating the spread of an infectious disease during an epidemic, and make proper solutions to control those parameters in practice for limiting the disease spread. In school closure intervention comparison study (Chapter 5), policy makers may use the comparison results in both mean attack rates and cost-effectiveness to make timely responses for a crisis. The information provided from model outputs may help policy makers to be more efficient for proper planning,monitoring, and decision-making.

In summary, the methods and procedures developed from this study in FRED model will help provide efficient tools for policy makers to make decision when there is a public health crisis. It also provides important information for public health policy makers about how certainly the FRED framework can provide reliable results of intervention strategy comparisons and cost-effectiveness comparisons for decision-making.

6.5 FUTURE WORK

Though the objectives of this study were met, the work has led to a number of new questions that should be considered in future work in this area. Future studies might address the inclusion of other high sensitivity parameters, such as contact rates associated with different social activities, in the probabilistic sensitivity analysis and uncertainty analysis. These contact rates are currently being estimated through other projects within the Public Health Dynamics Laboratory. It is also the subject of future work to extend the developed methods and procedures to guide the assessment of the full FRED model for other diseases, where the common parameters across different FRED models and correlation of model outputs will need to be addressed.

APPENDIX A

A PRELIMINARY STUDY

This section presents the results from a preliminary study for determining the minimum number of seed (initial cases) and simulation runs in FRED influenza model. Table 34 shows the mean attack rates by number of seed (initial cases) and age of seed (initial cases) in Allegheny County, PA with other input parameters set as default in FRED model. Table 35 shows the regression results to compare the differences among different number of seed (initial cases) by age.

Table 34. Mean attack rates by number of seed (initial cases) and age of seed (initial cases)

	Age				
Seed	0-4	5-17	18-64	65+	all
1	22.4454	39.9244	17.2336	18.1422	22.503
5	42.8116	45.092	42.042	39.279	44.0706
10	44.4274	45.0918	44.4506	42.6024	44.27
20	45.0748	45.0922	45.0992	45.1014	45.0988
50	45.0904	45.0896	45.0956	45.0814	45.0944
100	45.0834	45.1176	45.1036	45.1034	45.1006
1000	45.1128	45.1038	45.1364	45.1474	45.1276

Table 35. Regression results for number of seed (initial cases) and age

Seed	Coefficients	Standard Error	p value
1	-0.0878966	0.033851	0.009
5	-0.0056637	0.0149841	0.705
10	-0.00649	0.0094023	0.490
20	0.0001749	0.0017413	0.920
50	0.0000292	0.0017311	0.987
100	0.0000345	0.0017424	0.984
1000	0.0002281	0.0017294	0.895

Table 36 shows the mean attack rates by number of simulation runs and number of seed (initial cases) in Allegheny County, PA with other input parameters set as default in FRED model. Regressions are performed to compare if there is any difference among different number of simulation runs.

Table 36. Mean attack rates by number of simulation runs and number of seed (initial cases)

	Simulation Runs					Regression
Seed	10	20	30	40	50	p value
5	32.925	32.976	31.8873	29.6868	27.7052	0.008
10	33.004	33.008	33.0083	32.9985	32.992	0.379
20	33.02	32.996	33.0083	33.0085	33.0122	0.719
50	32.976	32.9925	32.9967	32.9972	33.0012	0.448
100	33.008	33.006	32.995	32.9922	32.9864	0.210
1000	33.075	33.076	33.0657	33.072	33.0672	0.695

Above results show that there is no difference if the number of seed (initial cases) is 10 or higher, and there is no difference in mean attack rates among different number of simulation runs. It suggests that the minimum number of seed (initial cases) is 10 and the minimum number of simulation runs is 10.

APPENDIX B

R CODES AND FRED SCRIPTS

B.1 R CODES FOR RANDOM SAMPLING PROCEDURE

```
### Install R packages
install.packages("msm")
install.packages("mvtnorm")
install.packages("lhs")
install.packages("MASS")
install.packages("Matrix")

### Load packages
library(msm)
library(lhs)
n<-50

### Generate random samples from truncated normal distributions and beta
distributions for 5 parameters

### asymptomatic infectivity
asympinf<-rtnorm(n, mean=0.5, sd=0.1, lower=0.33, upper=0.67)

### symptomatic period
daysymp<-rtnorm(n, mean=4.7, sd=0.9, lower=3.9, upper=5.1)

### asymptomatic period
dayasymp<-rtnorm(n, mean=4.7, sd=0.9, lower=3.9, upper=5.1)

### probability of stay home
prob<-rbeta(n, 3.662424, 3.082372, ncp = 0)

RSoutput<-data.frame(asympinf, daysymp, dayasymp, prob)
write.table(RSoutput, file = "/Users/Xiaozhi/downloads/RandomSamplingOutput",
row.names=FALSE,
col.names=c("AsympInf","DaysSymp","DaysAsymp","ProbStayHome"), sep = " ")
```

```
### Compute CDF of symptomatic period from generated samples

x<-c(3,4,5,6)
f<-c(0.1,0.3,0.4,0.2)
mu<-sum(x*f)
x1<-ifelse(x<mu,x,0)
x2<-ifelse(x<mu,0,x)
sum1<-sum(x1)
sum2<-sum(x2)
a<-4-length(x1[x1==0])
k<-4-length(x2[x2==0])
p<-ifelse(daysymp<=mu, (mu-daysymp)/(sum2-sum1*k/a), -(daysymp-mu)/(sum2-
sum1*k/a))
f3<-c(f[1]+p*k/a)
f4<-c(f[2]+p*k/a+f3)
f5<-c(f[3]-p+f4)
daysympCDF<-data.frame(0, 0, 0, f3, f4, f5, 1)
write.table(daysympCDF, file = "/Users/Xiaozhi/downloads/sympCDF-
RandomSampling", row.names=FALSE,
col.names=c("day0","day1","day2","day3","day4","day5","day6"), sep = " ")

### Compute CDF of asymptomatic period from generated samples

x<-c(3,4,5,6)
f<-c(0.1,0.3,0.4,0.2)
mu<-sum(x*f)
x1<-ifelse(x<mu,x,0)
x2<-ifelse(x<mu,0,x)
sum1<-sum(x1)
sum2<-sum(x2)
a<-4-length(x1[x1==0])
k<-4-length(x2[x2==0])
p<-ifelse(dayasymp<=mu, (mu-dayasymp)/(sum2-sum1*k/a), -(dayasymp-mu)/(sum2-
sum1*k/a))
f3<-c(f[1]+p*k/a)
f4<-c(f[2]+p*k/a+f3)
f5<-c(f[3]-p+f4)
dayasympCDF<-data.frame(0, 0, 0, f3, f4, f5, 1)
write.table(dayasympCDF, file = "/Users/Xiaozhi/downloads/asympCDF-
RandomSampling", row.names=FALSE,
col.names=c("day0","day1","day2","day3","day4","day5","day6"), sep = " ")
```

B.2 R CODES FOR LATIN HYPERCUBE SAMPLING PROCEDURE

```
### Install R packages
install.packages("msm")
install.packages("mvtnorm")
install.packages("lhs")
install.packages("MASS")
install.packages("Matrix")
```

```
### Load packages
library(msm)
library(lhs)
n<-50

### Latin Hypercube Sampling from 5 parameters

require(lhs)
x<-randomLHS(n, 4)
y<-x
y[,1]<-rtnorm(x[,1], mean=0.5, sd=0.1, lower=0.33, upper=0.67)
y[,2]<-rtnorm(x[,2], mean=4.7, sd=0.9, lower=3.9, upper=5.1)
y[,3]<-rtnorm(x[,3], mean=4.7, sd=0.9, lower=3.9, upper=5.1)
y[,4]<-rbeta(x[,4], 3.662424, 3.082372, ncp = 0)
write.table(y, file =
"/Users/Xiaozhi/downloads/LatinHypercubeSamplingOutput", row.names=FALSE,
col.names=c("AsympInf","DaysSymp","DaysAsymp","ProbStayHome"), sep = " ")

### Compute CDF of symptomatic period from generated samples

x<-c(3,4,5,6)
f<-c(0.1,0.3,0.4,0.2)
mu<-sum(x*f)
x1<-ifelse(x<mu,x,0)
x2<-ifelse(x<mu,0,x)
sum1<-sum(x1)
sum2<-sum(x2)
a<-4-length(x1[x1==0])
k<-4-length(x2[x2==0])
p<-ifelse(y[,2]<=mu, (mu-y[,2])/(sum2-sum1*k/a), -(y[,2]-mu)/(sum2-sum1*k/a))
f3<-c(f[1]+p*k/a)
f4<-c(f[2]+p*k/a+f3)
f5<-c(f[3]-p+f4)
daysympCDF<-data.frame(0, 0, 0, f3, f4, f5, 1)
write.table(daysympCDF, file = "/Users/Xiaozhi/downloads/sympCDF-
LatinHypercubeSampling", row.names=FALSE,
col.names=c("day0","day1","day2","day3","day4","day5","day6"), sep = " ")

### Compute CDF of asymptomatic period from generated samples

x<-c(3,4,5,6)
f<-c(0.1,0.3,0.4,0.2)
mu<-sum(x*f)
x1<-ifelse(x<mu,x,0)
x2<-ifelse(x<mu,0,x)
sum1<-sum(x1)
sum2<-sum(x2)
a<-4-length(x1[x1==0])
k<-4-length(x2[x2==0])
p<-ifelse(y[,3]<=mu, (mu-y[,3])/(sum2-sum1*k/a), -(y[,3]-mu)/(sum2-sum1*k/a))
f3<-c(f[1]+p*k/a)
f4<-c(f[2]+p*k/a+f3)
f5<-c(f[3]-p+f4)
dayasympCDF<-data.frame(0, 0, 0, f3, f4, f5, 1)
write.table(dayasympCDF, file = "/Users/Xiaozhi/downloads/asympCDF-
LatinHypercubeSampling", row.names=FALSE,
col.names=c("day0","day1","day2","day3","day4","day5","day6"), sep = " ")
```

B.3 FRED SCRIPTS FOR CREATING PARAMETER FILES FROM THE OUTPUT FILES OF B.1 AND B.2

```
#!/usr/bin/perl
use strict;
use warnings;
my ($dir, $sampling) = @ARGV;
my $outfile = "$dir/$sampling" . "SamplingOutput";
my $asympfile = "$dir/asympCDF-$sampling" . "Sampling";
my $sympfile = "$dir/sympCDF-$sampling" . "Sampling";
open FH, $outfile or die "ERROR: Can't open file $outfile\n";
open AS, $asympfile or die "ERROR: Can't open file $asympfile\n";
open SY, $sympfile or die "ERROR: Can't open file $sympfile\n";
my $line = <FH>;
$line = <AS>;
$line = <SY>;
my $n = 0;
while ($line = <FH>) {
  $n++;
my $paramsdir = "$dir/PARAMS-$sampling";
mkdir $paramsdir if not -d $paramsdir;
my $paramsfile = "$paramsdir/params.$sampling-$n";
chomp $line;
my ($asympinf, $x1, $x2, $sick_day_prob) = split " ", $line;
my $days_asymp = <AS>;
my $days_symp = <SY>;
open PAR, ">$paramsfile" or die "Can't write to file $paramsfile\n";
print PAR "days_asymp[0] = 7 $days_asymp";
print PAR "days_symp[0] = 7 $days_symp";
print PAR "sick_day_prob = $sick_day_prob\n";
print PAR "asymp_infectivity[0] = $asympinf\n";
close PAR;
}
close FH;
close AS;
close SY;
```

B.4 FRED SCRIPTS FOR CREATING EXECUTABLE FILES FOR RUNNING PROBABILISTIC SENSITIVITY ANALYSIS

```
#!/usr/bin/perl
use strict;
use warnings;
useEnv;
useGetopt::Std;
```

```
# File: fred_psa
# Author: John Grefenstette
# Created: Dec 11, 2012
my $FRED = $ENV{FRED_HOME};
die "$0: Please set environmental variable FRED_HOME to location of FRED home
directory\n" if not $FRED;
my $usage = "usage: $0 [-f config_file | -h | -c ]\n";

# create RESULTS directory if needed
my $FREDRESULTS = $ENV{FRED_RESULTS};
$FREDRESULTS = $ENV{FRED_HOME} if not $FREDRESULTS;
my $fred_results = "$FREDRESULTS/RESULTS";
if (not -d $fred_results) {
mkdir "$fred_results" or die "Can't create RESULTS directory
$fred_results\n";
}

# create PARAMS directory if needed
if (not -d "PARAMS_DIR") {
mkdir "PARAMS_DIR" or die "Can't create PARAMS_DIR directory\n";
}

# get command line arguments
my %options = ();
getopts("hcf:", \%options);
if (exists $options{h}) {
print "$usage";
print "Try:\n$0 -c\nto get a sample configuration file.\n";
exit;
}

# print sample file and exit if given -s
if (exists $options{c}) {
print_sample_file();
exit;
}

# readconfig file
my $config_file = $options{f};
die $usage if (not $config_file) or (not -e $config_file);
open FH, $config_file or die "Can't open $config_file\n";
my $line;

# find the title line
get_next_line(1);
die "$0: expecting title line\n" if $line !~ /FRED Probabilistic Sensitivity
Analysis Title:/;
my ($title) = $line =~ /Title:\s*(.*)/;
die "$0: bad title format\n" if not $title;
$title =~ s/\s/_/g;

# create baseline params file
my $paramsbase = "PARAMS_DIR/params.psa-$title-base";
open PAR, ">$paramsbase";
print PAR "# Probabilistic Sensitivity Analysis Title: $title\n";
print PAR "# Created: ", scalar localtime, "\n\n";
```

```
# get the fixed parameters
get_next_line(1);
die "$0: expecting \"Fixed Parameters:\"\n" if $line !~ /Fixed Parameters:/;
print PAR "# Fixed Parameters:\n";
get_next_line(1);
while ($line !~ /Sampling Directory:/) {
print PAR "$line\n";
get_next_line(1);
}

# get the sampling directory
my $sampling_dir;
get_next_line(1);
while ($line !~ /Sampling Method:/) {
chomp $line;
  ($sampling_dir) = $line =~ /Sampling Directory = (\S+)/;
get_next_line(1);
  # print "line = |$line|\n";
}

# get the sampling method
my $sampling;
get_next_line(1);
while ($line !~ /Output Variables:/) {
chomp $line;
  ($sampling) = $line =~ /Sampling Method = (\S+)/;
get_next_line(1);
  # print "line = |$line|\n";
}

# get the output variables
# print PAR "\n# Output Variables:\n";
my @outvars = ();
get_next_line(1);
while ($line !~ /Number of Samples:/) {
chomp $line;
push @outvars, $line if $line;
  # print PAR "# $line\n";
get_next_line(1);
}
# print "@outvars\n";
print PAR "\n# Control  Parameters:\n";
close PAR;

# get the number of samples
my ($N) = $line =~ /Samples: (\d+)/;
die "$0:Bad number of samples\n" if not $N > 0;

get_next_line(1);
die "$0: expecting \"Experimental Conditions:\"\n" if $line !~ /Experimental
Conditions:/;

get_next_line(1);
die "$0: expecting \"Condition Name:\"\n" if $line !~ /Condition Name:/;

my @condnames = ();
my %condition = ();
```

```
while ($line =~ /Condition Name:/) {
my ($name) = $line =~ /Condition Name:\s*(\S+)/;
die "$0: bad Condition Name on line:\n$line\n" if not $name;
push @condnames, $name;
  # print "\n@condnames\n";
  $condition{$name} = "";

  # get the control parameters
get_next_line(0);
while ($line !~ /Condition Name:/) {
chomp $line;
    $condition{$name} .= "$line;";
    # print "cond: #$condition{$name}#\n";
get_next_line(0);
last if ($line =~ /EOF/);
  }
last if ($line =~ /EOF/);
}
close FH;

print "OK\n";

# runmake_params
print "fred_psa_make_params $sampling_dir/$N-samples $sampling\n";
system "fred_psa_make_params $sampling_dir/$N-samples $sampling";

# create parameter files
for my $paramset (1..$N) {
my $paramsfile = "PARAMS_DIR/params.psa-$title-$paramset";
system "cp $paramsbase $paramsfile";
my $sampling_params = "$sampling_dir/$N-samples/Params-
$sampling/params.$sampling-$paramset";
die "Can't find sampling params $sampling_params\n" if (not -e
$sampling_params);
system "cat $sampling_params>> $paramsfile";

  # create a parameter file for each condition
for my $i (0..$#condnames) {
my $condname = $condnames[$i];
my $condfile = "PARAMS_DIR/params.psa-$title-$condname-$sampling-$paramset";
system "cp $paramsfile $condfile";
my $parlist = $condition{$condname};
my @pars = split ";", $parlist;
for my $par (@pars) {
      # print "cond = $condnameparlist = |$parlist| par = |$par|\n";
my ($name, $value) = $par =~ /(\S+)\s*=\s*(.*\S)/;
      # print "name = |$name| value = |$value|\n";
system "ch \'$name\' \'$value\' $condfile";
    }
system "echo >> $condfile";
  }
unlink $paramsfile;
}

# make script to run sensitivity analysis
my $exec = "run_psa.$title-$sampling";
open EXEC, ">$exec";
```

```
my $date = scalar localtime;
my $n = 20;
my $m = 4;
print EXEC <<EOF;
#!/bin/sh
# File: $exec
# Created: $date
# run a job for each parameter set

EOF
for my $i (0..$#condnames) {
my $condname = $condnames[$i];
for my $paramset (1..$N) {
my $key = "psa-$title-$condname-$sampling-$paramset";
my $paramsfile = "PARAMS_DIR/params.$key";
print EXEC <<EOF2;
date
fred_delete -f -k $key
fred_job -k $key -p $paramsfile -n $n -m $m
fred_AR -k $key
fred_delete -f -k $key

EOF2
  }
  # compare all condition pairs
for (my $j = 0; $j < $i; $j++) {
for my $outvar (@outvars) {
      print EXEC "fred_compare_jobs -j psa-$title-$condnames[$j] -k psa-
$title-$condname -N $N -v $outvar\n\n";
      }
  }
}
print EXEC "date\n";
close EXEC;

# run script
system "chmod +x $exec";
# system "$exec >&psa_$title.out";
exit;

subget_next_line {
my $no_eof = shift;
  $line = "";
while (not $line) {
    $line = <FH>;
if (not defined $line) {
if ($no_eof> 0) {
      die "$0: unexpected end of file\n";
      }
else {
      $line = "EOF";
      }
    }
chomp $line;
    $line = "" if $line =~ /^#/;
  }
}
```

BIBLIOGRAPHY

1. Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals: Princeton University Press; 2007.

2. Brauer F, Castillo-Chávez C. Mathematical Models in Population Biology and Epidemiology. NY: Springer; 2001.

3. Gilbert N. Agent-Based Models: Sage Publications; 2007.

4. Brown DG. Agent-based models. In: Geist H, editor. The Earth's Changing Land: An Encyclopedia of Land-Use and Land-Cover Change. Westport CT: Greenwood Publishing Group; 2006. p. 7-13.

5. Brown ST, Tai JH, Bailey RR, Cooley PC, Wheaton WD, Potter MA, et al. Would school closure for the 2009 H1N1 influenza epidemic have been worth the cost?: a computational simulation of Pennsylvania. BMC Public Health 2011;11:353.

6. Epstein JM. Modelling to contain pandemics. Nature 2009;460(7256):687.

7. Burke DS, Epstein JM, Cummings DA, Parker JI, Cline KC, Singa RM, et al. Agent-based computational modeling of smallpox epidemic control strategies. Acad Emerg Med 2006;13(11):1142-9.

8. Longini IM, Jr., Halloran ME, Nizam A, Yang Y, Xu S, Burke DS, et al. Containing a large bioterrorist smallpox attack: a computer simulation approach. Int J Infect Dis 2007;11(2):98-108.

9. Halloran ME, Ferguson NM, Eubank S, Longini IM, Jr., Cummings DA, Lewis B, et al. Modeling targeted layered containment of an influenza pandemic in the United States. Proc Natl Acad Sci U S A 2008;105(12):4639-44.

10. Lee BY, Ercius AK, Smith KJ. A predictive model of the economic effects of an influenza vaccine adjuvant for the older adult (age 65 and over) population. Vaccine 2009;27(16):2251-7.

11. Lee BY, Tai JH, Bailey RR, Smith KJ. The timing of influenza vaccination for older adults (65 years and older). Vaccine 2009;27(50):7110-5.

12. Shi P, Keskinocak P, Swann JL, Lee BY. Modelling seasonality and viral mutation to predict the course of an influenza pandemic. Epidemiol Infect 2010;138(10):1472-81.

13. Lee BY, Brown ST, Korch GW, Cooley PC, Zimmerman RK, Wheaton WD, et al. A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 H1N1 influenza pandemic. Vaccine 2010;28(31):4875-9.

14. Lee BY, Brown ST, Cooley PC, Zimmerman RK, Wheaton WD, Zimmer SM, et al. A computer simulation of employee vaccination to mitigate an influenza epidemic. Am J Prev Med 2010;38(3):247-57.

15. Lee BY, Brown ST, Cooley P, Potter MA, Wheaton WD, Voorhees RE, et al. Simulating school closure strategies to mitigate an influenza epidemic. J Public Health Manag Pract 2010;16(3):252-61.

16. Grefenstette JJ, Brown ST, Rosenfeld R, Depasse J, Stone NT, Cooley PC, et al. FRED (A Framework for Reconstructing Epidemic Dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations. BMC Public Health 2013;13(1):940.

17. Marino S, Hogue IB, Ray CJ, Kirschner DE. A methodology for performing global uncertainty and sensitivity analysis in systems biology. J Theor Biol 2008;254(1):178-96.

18. Kleijnen JPC. Verification and validation of simulation models. European Journal of Operational Research 1995;82(1):145-162.

19. Frey HC, Patil SR. Identification and review of sensitivity analysis methods. Risk Anal 2002;22(3):553-78.

20. Ward MP, Carpenter TE. Simulation modeling of the effect of climatic factors on Bluetongue virus infection in Australian Cattle Herds - 1. Model formulation, verification and validation. Preventive Veterinary Medicine 1996;27(12):1-12.

21. Cacuci DG, Ionescu-Bujor M. A comparative review of sensitivity and uncertainty analysis of large-scale systems - II: Statistical methods. Nuclear Science and Engineering 2004;147(3):204-217.

22. Patelli E , Pradlwarter HJ , Schue¨ller GI. Global sensitivity of structural variability by random sampling. Computer Physics Communications 2010;181:2072-2081.

23. Cullen AC, Frey HC. Probabilistic techniques in exposure assessment. New York: Plenum Press; 1999.

24. Hosmer DW, Lemeshow S. Applied logistic regression. 2 ed. New York: Wiley-Interscience Publication; 2000.

25. Winter BJ, Brown DR, Michels KM. Statistical principles in experimental design. 3 ed. New York: McGraw-Hill; 1991.

26. Lohr SL. Sampling: Design and Analysis. 2 ed: Duxbury Press; 2009.

27. Carrat F, Vergu E, Ferguson NM, Lemaitre M, Cauchemez S, Leach S, et al. Time lines of infection and disease in human influenza: a review of volunteer challenge studies. Am J Epidemiol 2008;167(7):775-85.

28. Cori A, Valleron AJ, Carrat F, Scalia Tomba G, Thomas G, Boelle PY. Estimating influenza latency and infectious period durations using viral excretion data. Epidemics 2012;4(3):132-8.

29. Tuite AR, Greer AL, Whelan M, Winter AL, Lee B, Yan P, et al. Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza. CMAJ 2010;182(2):131-6.

30. Lessler J, Reich NG, Brookmeyer R, Perl TM, Nelson KE, Cummings DA. Incubation periods of acute respiratory viral infections: a systematic review. Lancet Infect Dis 2009;9(5):291-300.

31. Baguelin M, Hoek AJ, Jit M, Flasche S, White PJ, Edmunds WJ. Vaccination against pandemic influenza A/H1N1v in England: a real-time economic evaluation. Vaccine 2010;28(12):2370-84.

32. Longini IM, Jr., Halloran ME, Nizam A, Yang Y. Containing pandemic influenza with antiviral agents. Am J Epidemiol 2004;159(7):623-33.

33. Stilianakis NI, Perelson AS, Hayden FG. Emergence of drug resistance during an influenza epidemic: insights from a mathematical model. J Infect Dis 1998;177(4):863-73.

34. Papenburg J, Baz M, Hamelin ME, Rheaume C, Carbonneau J, Ouakki M, et al. Household transmission of the 2009 pandemic A/H1N1 influenza virus: elevated laboratory-confirmed secondary attack rates and evidence of asymptomatic infections. Clin Infect Dis 2010;51(9):1033-41.

35. Mathews JD, McCaw CT, McVernon J, McBryde ES, McCaw JM. A biological model for influenza transmission: pandemic planning implications of asymptomatic infection and immunity. PLoS One 2007;2(11):e1220.

36. Kuster SP, Shah PS, Coleman BL, Lam PP, Tong A, Wormsbecker A, et al. Incidence of influenza in healthy adults and healthcare workers: a systematic review and meta-analysis. PLoS One 2011;6(10):e26239.

37. Yang Y, Halloran ME, Longini IM, Jr. A Bayesian model for evaluating influenza antiviral efficacy in household studies with asymptomatic infections. Biostatistics 2009;10(2):390-403.

38. Patrozou E, Mermel LA. Does influenza transmission occur from asymptomatic infection or prior to symptom onset? Public Health Rep 2009;124(2):193-6.

39. Longini IM, Jr., Koopman JS. Household and community transmission parameters from final distributions of infections in households. Biometrics 1982;38(1):115-26.

40. Longini IM, Jr., Koopman JS, Monto AS, Fox JP. Estimating household and community transmission parameters for influenza. Am J Epidemiol 1982;115(5):736-51.

41. Sugiyama H. Some statistical contributions to the health sciences. Osaka City Medical Journal 1960;6:141-158.

42. Reed C, Biggerstaff M, Finelli L, Koonin LM, Beauvais D, Uzicanin A, et al. Novel Framework for Assessing Epidemiologic Effects of Influenza Epidemics and Pandemics. Emerging Infectious Diseases 2013;19(1):85-91.

43. Piso RJ, Albrecht Y, Handschin P, Bassetti S. Low transmission rate of 2009 H1N1 Influenza during a long-distance bus trip. Infection 2011;39(2):149-53.

44. Klick B, Nishiura H, Ng S, Fang VJ, Leung GM, Peiris JS, et al. Transmissibility of seasonal and pandemic influenza in a cohort of households in Hong Kong in 2009. Epidemiology 2011;22(6):793-6.

45. Merl D, Johnson LR, Gramacy RB, Mangel M. A statistical framework for the adaptive management of epidemiological interventions. PLoS One 2009;4(6):e5807.

46. Murray JD. Mathematical Biology I: An Introduction. 3 ed: Springer; 2007.

47. Chan SS. Does vaccinating ED health care workers against influenza reduce sickness absenteeism? Am J Emerg Med 2007;25(7):808-11.

48. Schanzer DL, Zheng H, Gilmore J. Statistical estimates of absenteeism attributable to seasonal and pandemic influenza from the Canadian Labour Force Survey. BMC Infect Dis 2011;11:90.

49. Mikolajczyk RT, Akmatov MK, Rastin S, Kretzschmar M. Social contacts of school children and the transmission of respiratory-spread pathogens. Epidemiol Infect 2008;136(6):813-22.

50. Lau CH, Springston EE, Sohn MW, Mason I, Gadola E, Damitz M, et al. Hand hygiene instruction decreases illness-related absenteeism in elementary schools: a prospective cohort study. BMC Pediatr 2012;12:52.

51. Considine J, Shaban RZ, Patrick J, Holzhauser K, Aitken P, Clark M, et al. Pandemic (H1N1) 2009 Influenza in Australia: Absenteeism and redeployment of emergency medicine and nursing staff. Emerg Med Australas 2011;23(5):615-23.

52. Campbell DS, Rumley MH. Cost-effectiveness of the influenza vaccine in a healthy, working-age population. J Occup Environ Med 1997;39(5):408-14.

53. Neuzil KM, Hohlbein C, Zhu Y. Illness among schoolchildren during influenza season: effect on school absenteeism, parental absenteeism from work, and secondary illness in families. Arch Pediatr Adolesc Med 2002;156(10):986-91.

54. Rousculp MD, Johnston SS, Palmer LA, Chu BC, Mahadevia PJ, Nichol KL. Attending work while sick: implication of flexible sick leave policies. J Occup Environ Med 2010;52(10):1009-13.

55. Turnberg W, Daniell W, Duchin J. Influenza vaccination and sick leave practices and perceptions reported by health care workers in ambulatory care settings. Am J Infect Control 2010;38(6):486-8.

56. Tora I, Delclos J, Martinez JM, Tobias A, Jardi J, Alberti C, et al. [Comparison in Andalusia and Catalonia of sickness absence for influenza (H1N1) 2009]. Rev Esp Salud Publica 2011;85(1):89-95.

57. Nichol KL, Lind A, Margolis KL, Murdoch M, McFadden R, Hauge M, et al. The effectiveness of vaccination against influenza in healthy, working adults. N Engl J Med 1995;333(14):889-93.

58. Bridges CB, Thompson WW, Meltzer MI, Reeve GR, Talamonti WJ, Cox NJ, et al. Effectiveness and cost-benefit of influenza vaccination of healthy working adults: A randomized controlled trial. JAMA 2000;284(13):1655-63.

59. Seale H, Leask J, Po K, MacIntyre CR. "Will they just pack up and leave?" - attitudes and intended behaviour of hospital health care workers during an influenza pandemic. BMC Health Serv Res 2009;9:30.

60. Blendon RJ, Koonin LM, Benson JM, Cetron MS, Pollard WE, Mitchell EW, et al. Public response to community mitigation measures for pandemic influenza. Emerg Infect Dis 2008;14(5):778-86.

61. Chen SC, You SH, Ling MP, Chio CP, Liao CM. Use of seasonal influenza virus titer and respiratory symptom score to estimate effective human contact rates. J Epidemiol 2012;22(4):353-63.

62. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 2008;5(3):e74.

63. Towers S, Chowell G. Impact of weekday social contact patterns on the modeling of influenza transmission, and determination of the influenza latent period. J Theor Biol 2012;312C:87-95.

64. Potter GE, Handcock MS, Longini IM, Jr., Halloran ME. Estimating within-Household Contact Networks from Egocentric Data. Ann Appl Stat 2011;5(3):1816-1838.

65. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. Am J Epidemiol 2006;164(10):936-44.

66. Eames KTD, Tilston NL, Brooks-Pollock E, Edmunds WJ. Measured Dynamic Social Contact Patterns Explain the Spread of H1N1v Influenza. PLoS Computational Biology 2012;8(3):1-8.

67. Railsback SF, Grimm V. Agent-Based and Agent-based Modeling: a Practical Introduction: Princeton University Press; 2011.

68. Mishra S, Deeds N, Ruskauff G. Global sensitivity analysis techniques for probabilistic ground water modeling. Ground Water 2009;47(5):730-747.

69. Sobol I. Sensitivity estimates for nonlinear mathematical models. Mathematical Modeling and Computational Experiment 1993;1:407-414.

70. Oakley JE, O'Hagan A. Probabilistic sensitivity analysis of complex models: a Bayesian approach. Journal of the Royal Statistical Society: Series B 2004;66(3):751-769.

71. Sudjianto A, Du X, Chen W. Probabilistic Sensitivity Analysis in Engineering Design using Uniform Sampling and Saddlepoint Approximation. SAE 2005 Transactions Journal of Passenger Cars: Mechanical Systems 2005.

72. Baio G, Dawid AP. Probabilistic sensitivity analysis in health economics. Stat Methods Med Res 2011.

73. Ellis N, Maitra R. Multivariate Gaussian Simulation Outside Arbitrary Ellipsoids. Journal of Computational and Graphical Statistics 2007;16(3):692-708.

74. Mckay MD, Beckman RJ, Conover WJ. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 1979;21(2):239-245.

75. Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering & System Safety 2003;81(1):23-69.

76. Olsson A, Sandberg G, Dahlblom O. On Latin hypercube sampling for structural reliability analysis. Structural Safety 2003;25:47-68.

77. Cauchemez S, Valleron AJ, Boelle PY, Flahault A, Ferguson NM. Estimating the impact of school closure on influenza transmission from Sentinel data. Nature 2008;452(7188):750-4.

78. Heymann A, Chodick G, Reichman B, Kokia E, Laufer J. Influence of school closure on the incidence of viral respiratory diseases among children and on health care utilization. Pediatr Infect Dis J 2004;23(7):675-7.

79. Cauchemez S, Ferguson NM, Wachtel C, Tegnell A, Saour G, Duncan B, et al. Closure of schools during an influenza pandemic. Lancet Infect Dis 2009;9(8):473-81.

80. Vynnycky E, Edmunds WJ. Analyses of the 1957 (Asian) influenza pandemic in the United Kingdom and the impact of school closures. Epidemiol Infect 2008;136(2):166-79.

81. Halder N, Kelso JK, Milne GJ. Developing guidelines for school closure interventions to be used during a future influenza pandemic. BMC Infect Dis 2010;10:221.

82. Hens N, Ayele GM, Goeyvaerts N, Aerts M, Mossong J, Edmunds JW, et al. Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. BMC Infect Dis 2009;9:187.

83. Eames KT, Tilston NL, White PJ, Adams E, Edmunds WJ. The impact of illness and the impact of school closure on social contact patterns. Health Technol Assess 2010;14(34):267-312.

84. Klaiman T, Kraemer JD, Stoto MA. Variability in school closure decisions in response to 2009 H1N1: a qualitative systems improvement analysis. BMC Public Health 2011;11:73.

85. Sander B, Nizam A, Garrison LP, Jr., Postma MJ, Halloran ME, Longini IM, Jr. Economic evaluation of influenza pandemic mitigation strategies in the United States using a stochastic microsimulation transmission model. Value Health 2009;12(2):226-33.

86. Berkman BE. Mitigating pandemic influenza: the ethics of implementing a school closure policy. J Public Health Manag Pract 2008;14(4):372-8.

87. Johnson AJ, Moore ZS, Edelson PJ, Kinnane L, Davies M, Shay DK, et al. Household responses to school closure resulting from outbreak of influenza B, North Carolina. Emerg Infect Dis 2008;14(7):1024-30.

88. Gift TL, Palekar RS, Sodha SV, Kent CK, Fagan RP, Archer WR, et al. Household effects of school closure during pandemic (H1N1) 2009, Pennsylvania, USA. Emerg Infect Dis 2010;16(8):1315-7.

89. Jackson C, Vynnycky E, Hawker J, Olowokure B, Mangtani P. School closures and influenza: systematic review of epidemiological studies. BMJ Open 2013;3(2).