8-2019

Evaluation of Data Collection Operations for Real-Time Influenza Surveillance during an Emergency

Yuwen Gu
Western Michigan University, heathergu85@yahoo.com

EVALUATION OF DATA COLLECTION OPERATIONS FOR REAL-TIME INFLUENZA SURVEILLANCE DURING AN EMERGENCY

by

Yuwen Gu

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Industrial and Entrepreneurial Engineering and Engineering Management
Western Michigan University
August 2019

Doctoral Committee:

Steven Butt, Ph.D., Chair
Diana Prieto, Ph.D.
Azim Houshyar, Ph.D.
Larry Mallak, Ph.D.
Elise de Doncker, Ph.D.

ACKNOWLEDGEMENTS

I would never have been able to finish my dissertation without the guidance of my committee, help from the professors and administrators, and support from friends and family.

I would like to extend deepest thanks and appreciation to my dissertation mentor, committee co-chair, Dr. Diana Prieto, for her excellent guidance, thought-provoking suggestions and encouragement throughout the process. I am very grateful for the research associate she offered me all the way from the very beginning to the completion of my dissertation. Dr. Prieto provided a great balance between guidance and freedom on my dissertation journey. Additionally, she also provided extreme patience in the face of numerous obstacles. I am truly fortunate to have had the opportunity to work with her. Without Dr. Prieto, the successful completion of my dissertation would not have been possible.

I would also like to thank my committee chair and department chair, Dr. Steven Butt, for his strong support all the time, especially when there were special changes during my Ph. D program. Besides, I am very grateful the opportunity of teaching associate he gave to me during the past four years, which enriched my teaching experiences in addition to financial support.

Sincerely thanks go to my committee members, Dr. Azim Houshyar and Dr. Elise de Doncker and Dr. Larry Mallak, for their participation and insight words of advice along the way. They were always supporting and encouraging me to pursue my dream. Without the encouragement from them, it would not be possible to make my journey to the doctoral degree.

I would also like to thank all the professors for their friendly guidance and input on my dissertation. I am especially thankful to my professors, Dr. Bob White, Dr. Lee Wells, Dr. Rajib

Paul, Dr. Naranjo, and Dr. Applegate for their professional courses and helpful recommendations on my research methodologies.

Thanks go to the administrators support staff at WMU Department of Industrial and Entrepreneurial Engineering and Engineering Management, who provided support and assistance—especially in the later, challenging phases. I extend additional appreciation to Mrs. Jerri Pursley for her support. In a similar vein, special thanks also go to Courtney Buckmaster, Lori Atwater, for their wonderful support.

I would also like to thank fellow students, James Burns, Shengfeng Chen, Rabia Emhamed Al mamlook, Megan Ashley Hammond, Kimberly Nagulpelli, Joan Yami Martinez, Gregory Ostroy, for their support, favors, friendship.

Last, but certainly not least, I would especially like to whole heartedly thank my family. I am so grateful for my parents and parents-in-law, Guoqing Gu, Boyan Huang, Shengchang Gao and Guifang Chen, for their unwavering support and love with their best wishes, for their help and share of household burdens for my family. I would like to give special thanks to my husband, Xingyuan Gao, who sacrificed innumerable time in shouldering far more than his fair share of parenting, encouraged me to pursue my dream and also provided guidance in my research. And many thanks to my cherished daughters, Mina Gao and Hannah Gao, who always believed in me, supported me and loved me, despite my many hours of distraction from family life over the past years.

Yuwen Gu

EVALUATION OF DATA COLLECTION OPERATIONS FOR REAL-TIME INFLUENZA SURVEILLANCE DURING AN EMERGENCY

Yuwen Gu, Ph.D.

Western Michigan University, 2019

It is unclear how data collection operations for surveillance alter the disease portrayal that influenza reported trends attempt to provide during an emergency. This study developed a model that simulates the collection and testing of influenza specimens after an outbreak is declared in Michigan. It performed simulation based optimization to understand which operational factors affect the biases between the growth rates of original and observed influenza incidence trends, and to quantify the predictive power of the influenza incidence trends at different points of data collection. The results show that emergency driven high risk perception increases the reporting, which leads to the reduction of biases in the growth rates. Therefore, a recently declared emergency is a potential opportunity to collect larger sample sizes.

This study also suggests that the growth rate that better predicts the original ILI growth rate, is the one estimated from the trend of specimens submitted to the Public Health Laboratories. State Health departments can benefit from the explanatory power of the submitted trend in their efforts to improve the epidemiological characterization of emergent influenza viral strains.

Several criteria under which Public Health Laboratories can order specimens for case testing and confirmation were tested. First come first serve outperforms other criteria as long as there is enough testing capacity. When the capacity was limited, collecting first come first serve for certain groups of interest (e.g., collecting until a sample per age group is completed) seems to be a better strategy.

TABLE OF CONTENTS

Table of Contents - continued

Table of Contents - continued

CHAPTER

Table of Contents - continued

CHAPTER

LIST OF TABLES

List of Tables - continued

LIST OF FIGURES

List of Figures - continued

List of Figures - continued

LIST OF ACRONYMS

AB: Agent-based

ABM: Agent-based model

CCD: Central Composite Design

CCI: Central Composite Inscribed Design

CDC: Center for Disease Control and Prevention

CL: Clinic laboratory

CPU: Central Process Unit

GPU: Graphic Process Unit

ED: Emergency department

ILI: Influenza-like illness

MARCC: Maryland Advanced Research Computing Center

MDOT: Michigan Department of Transportation

MDHHS: Michigan Department of Health and Human Services

MSSS: Michigan Syndromic Surveillance System

P&I: Pneumonia & Influenza

PHL: Public Health Laboratory

SEMCOG: Southeast Michigan Council of Governments

TMA: Transportation Management Areas

WHO: World Health Organization

CHAPTER I

INTRODUCTION

Background

Influenza epidemics and pandemics can create influenza emergencies that have seriously threaten human health and taken millions of lives over centuries. Every year, around 291,000 to 646,000 people all over the world die of seasonal influenza (Iuliano et al., 2018). In the United States, there have been an estimate of 25 – 49 million flu illnesses, 25,000 – 79,000 deaths, and 310,000 – 960,000 hospitalizations associated to influenza every year since 2012 ("Past Seasons Estimated Influenza Disease Burden | CDC", 2018). Among those infected people, prenatal women, little children (5-yr-old and under), elders (65-yr-old and above), and patients with chronic medical conditions are defined as high-risk population groups that are more likely to develop life-threatening complications ("People at High Risk For Flu Complications | CDC", 2018).

Influenza emergencies occur when a circulating virus present with unexpected features challenges the immune system of the susceptible population and increase the incidence alter healthcare seeking behaviors, and challenge the public health infrastructure created for disease surveillance (Briand, Mounts, & Chamberland, 2011). The recent flu season (2017-2018) in the United States, turned into an emergency with widespread Influenza-like-illness (ILI) activity, with 223,487 confirmed cases, and a total of 186 pediatric deaths (“Summary of the 2017-2018

Influenza Season | CDC", 2018). The emergency was likely the result of a flu vaccine with about 25% of effectiveness in generating immunity for the predominant A(H3N2) virus (Flannery et al., 2018).

Problem Statement

Although digital surveillance systems have proven beneficial for early warning of emerging outbreaks (e.g., Pro MED – mail, Health maps, and the Global Public Health Intelligence Network) (Magid, Gesser-Edelsburg, & Green, 2018), disease epidemiological characterization still occurs with official sources of data collection and the emergency management is yet as tricky as in the 2009 Pandemic Flu (Neumann, Noda, & Kawaoka, 2009).

As per WHO surveillance standards, data collection occurs passively when symptomatic individuals report their symptoms to the healthcare system ("WHO global technical consultation: global standards and tools for influenza surveillance", 2011; Morse, 2007). In Michigan, for example, emergency departments connected to the Michigan Syndromic Surveillance System (MSSS) report their daily cases with constitutional and respiratory symptoms, which are daily monitored by the local and regional public health departments across the entire state ("Emergency Department Syndromic Surveillance in Michigan", 2017). Primary/urgent care practitioners can report their weekly ILI cases to the sentinel network, but the reporting is optional and the system does not have the capabilities for real-time updating of the overall case trend.

When the pandemic outbreak was declared, it is likely that individuals who reported their disease to the MSSS and the sentinel network were influenced by the media coverage

about the disease. Consider the trend of cases with constitutional and respiratory symptoms that reported through the MSSS in Figure 1 and Figure 2 (data from MSSS database 09). Although the respiratory trend does not show any relevant pattern related to ILI progression, the constitutional trend shows that symptomatic cases seeking healthcare increased after April 21st (on Tue.), which was the date when the news media first declared pandemic outbreak in the US after two California children were infected with unusual swine flu ("2009 flu pandemic timeline", 2019). One might suspect that the trend carries the effect of reporting symptomatic cases that were in fear of the pandemic. Some insights about these effects are provided in Chapter IV.

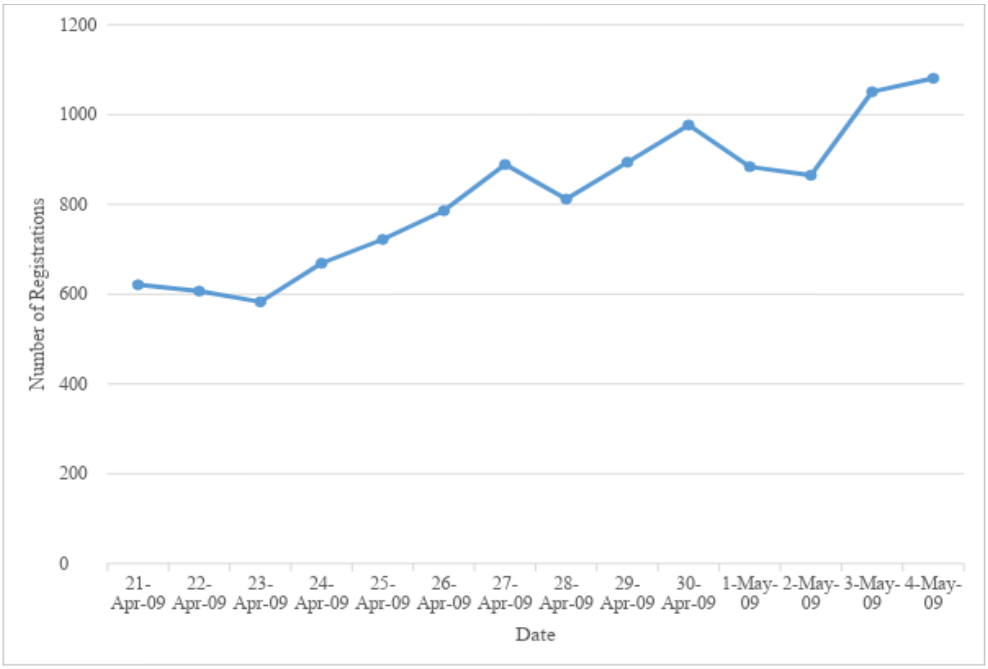
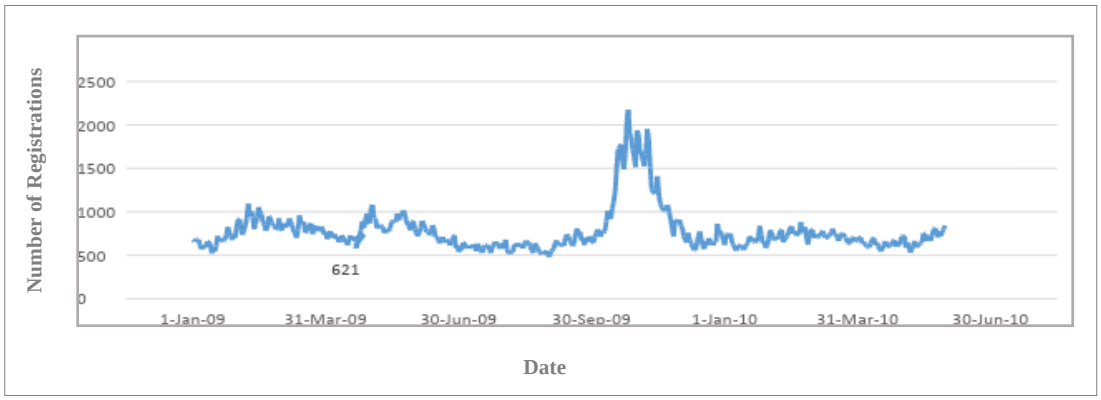


Figure 1. Number of Constitutional Registrations at Emergency Departments 2009 State of Michigan After Pandemic Declaration

a)



b)

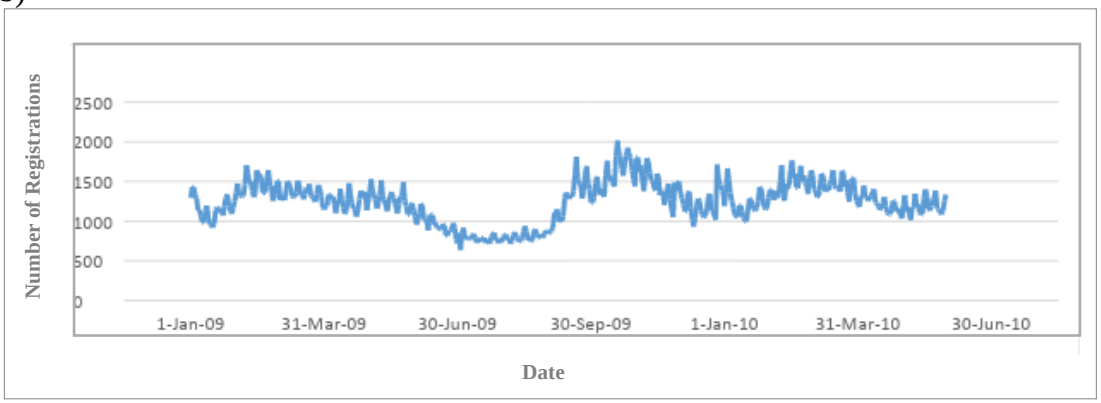


Figure 2. Number of Emergency Department Registrations 2009 State of Michigan. a) Constitutional Registration. b) Respiratory Registration.

Once infected cases seek healthcare, sentinels route a subset of the cases to the viral surveillance labs for further specimen testing and disease characterization. When a novel virus spreads and the outbreak is declared, the state initially recommends that not only sentinels but any healthcare provider sends suspicious cases to the surveillance labs. This policy, together with the limited lab capacities for testing, the labs' existing first-come-first-serve (FCFS) testing policy and the manual methods for receiving and processing the specimens, created overcrowding and delays of up to two weeks in the information reported (Prieto et al., 2012; Santillana et al., 2015; Santillana et al., 2016). State public health departments seek then to

control the specimen submission with sampling policies that consider the impact of the disease in the population (e.g., in the H1N1 pandemic outbreak, states like Michigan and Florida restricted specimen submission to children, elderly and pregnant women, which typically are at higher risk of flu related mortality). These sampling policies, while convenient for reducing the public health laboratory (PHL) workload, might produce biased trends that are not representative of the epidemiological features and the burden of influenza ("WHO Interim Global Epidemiological Surveillance Standards for Influenza", 2012).

Incidence trends are reported by the healthcare system in the form of weekly or daily ILI, or by the PHL in the form of daily cases confirmed with a virus subtype. Daily ILI and confirmed trends have proven useful to infer the total daily incidence of cases with an emergent (e.g., pandemic H1N1) virus subtype (Birrell et al., 2011). Daily and weekly ILI trends have been used in conjunction with other internet based data sources to produce ILI forecasts of weekly and daily resolution (Santillana et al., 2015; Shaman & Karspeck, 2012; Shaman, Karspeck, Yang, Tamerius, & Lipsitch, 2013). But these estimates assume that all trends are ready and available at the time of the inference or forecast, which is still unrealistic.

Even with complete data, ILI and confirmed case trends are noisy and lagged that might be biased by the reactive operational landscape for data collection. It has been found useful to select the surveillance trends that are closer to infection time or reported from severe symptoms, complimentary to a state-of-the-art Bayesian inference pandemic outbreak model (Birrell, Pebody, Charlett, Zhang, & De Angelis, 2017). However, their model provided less predictive power at the early stage of pandemic, when the collected data tend to be instable. The ability of

any method to predict unbiased incidence trends during the initial phase of an emergency is still an open question.

Therefore, there is a strong need to understand how the data collection operations for influenza surveillance during the initial phase of an emergency affect the biasedness in the officially observed incidence trends. Significant factors in influenza disease burden has been previously identified by means of a Bayesian Model fitted to medical claims data (Lee et al., 2018), but the effects of these factors on the biases of the observed incidence trends have not been explored.

Research Objectives

This study presents agent-based (AB) simulation to investigate the effect of the data collection operations in the biasedness between the real incidence trends and the observed incidence trends. Simulation models have been used in surveillance research to explore policies for spatial allocation of data providers (Souty & Boëlle, 2016; Scarpino, Dimitrov, & Meyers, 2012), and to test the sensitivity, specificity, and timeliness of univariate process monitoring control algorithms for detecting outbreaks of influenza (Cao et al., 2014) or other diseases (Antunes, Jensen, Halasa, & Toft, 2017; Dórea, Mcewen, Mcnab, Sanchez, & Revie, 2013; Dupuy et al., 2015). In contrast to these applications, this study used the simulation model to evaluate the biasedness of the existing data collection infrastructure for surveillance, but it can be also adapted to test the biasedness of other proposed systems for data collection. The simulation model allows further exploration of optimal operational conditions, which enhances understanding on existing sampling policies.

To prepare the simulation testbed, a baseline agent-based model (ABM) (Prieto & Das, 2016) for influenza co-circulating outbreak is refined with mitigation and containment (i.e., vaccination, antiviral treatment and self-induced absenteeism). The refined version models the influenza co-circulation across different regions of Michigan incorporating its demographic and geographic features grounded on real data (See Chapter III for detailed model description). The refined model was calibrated and analyzed with central composite designs and response surface optimization models for optimal and systematic exploration of the parameter space.

To summarize, the overall research objective of this study was to evaluate the biasedness of the data collection operations for influenza surveillance during an emerging influenza outbreak. There were four sub-objectives:

Objective 1 To prepare a high performance simulation testbed embedded with the realistic demographic, geographic, travel patterns in the state of Michigan in presence of mitigation and surveillance policies;

Objective 2 To estimate the simulation model parameter space by calibrating the model with the MSSS influenza incidence trends during the initial phase of the 2009 pandemic influenza outbreak in state of Michigan;

Objective 3 To develop the surveillance operations with the data collection processes in the Michigan simulation testbed;

Objective 4 To explore the effect of the data collection operational factors in the exponential growth rate biases between the real and observed incidence trends with the simulation based experiments.

Research Questions

Research question 1 Does the observed real-time trend of influenza incidence exclusively capture the disease behavior in presence of mitigation or is it also affected by other factors (e.g., public communication about the outbreak status)?

Research question 2 Which factors of the data collection operations have significant impact on biasedness between the real and observed incidence trends?

Research question 3 How do sampling criteria affect the biasedness between the real and observed incidence trends?

Research question 4 To what extent do the observed influenza incidence trends explain the real influenza incidence trends?

Significance of the Research

Daily trends of influenza incidence are broadly used to understand how to better prepare for influenza emergencies. Biased trends can result in inappropriate use of societal resources when incidence is under/over-estimated. These trends inform not only healthcare preparedness and operations but also many other entities that are forced to make decisions during emergencies (e.g., With the arousal of a pandemic outbreak, the provost at WMU must face the decision of whether to close the university). In addition, influenza incidence trends are intently used by researchers as inputs to their analyses. Therefore, the societal cost of inaccurate and untimely information is very high.

The present study uses the state of Michigan as the simulation testbed for evaluating the specimen collection and sampling processes under the existing surveillance systems. The

outcomes aim at understanding how data collection operations can distort the influenza incidence trends. The results of the study are relevant to public health as the State Departments and PHLs will receive recommendations to guide their data collection operations.

This study is not only useful in the evaluation of the existing data collection infrastructure for influenza surveillance, but also applicable for data collection processes in epidemic outbreak or other fields.

Organization of the Dissertation

This dissertation is organized into five chapters as described below.

CHAPTER I

INTRODUCTION

This chapter presents a brief discussion of the problems that frame the study, and a description of the research objectives and research questions.

CHAPTER II

REVIEW OF THE LITERATURE

This chapter presents a general review of the current influenza surveillance systems, plus a study of appropriate techniques that form the basis of the research methods. It includes 1) an introduction of influenza, 2) landscape of influenza surveillance systems and other novel surveillance tools, 3) limitations of influenza surveillance systems, 4) existing work to address those limitations, 5) research gap that contrasts this study to the existing work.

CHAPTER III

HIGH PERFORMANCE AGENT-BASED SIMULATION OF INFLUENZA CO-CIRCULATION IN THE STATE OF MICHIGAN

This chapter presents a baseline agent-based model of an influenza outbreak using the demographic, geographic, and travel features of the State of Michigan. It includes: 1) the baseline agent-based simulation testbed, 2) Michigan demographic, geographic and travel patterns together with containment and mitigation, 3) the parallelization of the simulation testbed.

CHAPTER IV

ESTIMATES OF EPIDEMIOLOGICAL CHARACTERISTICS AND MITIGATION LEVELS AFTER PANDEMIC DECLARATION IN MICHIGAN

This chapter addresses Research question 1. It includes 1) the simulation based optimization methods and 2) results and discussion.

CHAPTER V

EVALUATION OF DATA COLLECTION OPERATIONS FOR INFLUENZA SURVEILLANCE

This chapter addresses Research question 2, 3, and 4. It includes 1) flow diagrams of data collection processes for specimen collection and sampling, 2) methods of simulation based experiment and statistical analysis of the simulation outputs, 3) results and discussion.

CHAPTER VI

CONCLUSION

This chapter concludes the dissertation. It includes 1) a summary of the accomplished work and significant findings. 2) limitations and future work.

CHAPTER II

REVIEW OF THE LITERATURE

This chapter provides a review of relevant literature. It first introduces influenza and its viral strains. Then, it stresses the limitations in the existing influenza surveillance systems and discusses how previous studies regarding improving surveillance operations addressed the limitations. Next, this chapter also reviews methods and objectives for influenza outbreak modelling. At last, two major research gaps are identified as a result from the literature view.

Influenza

Over the centuries, humans have been through multiple influenza emergencies that were created by both seasonal and pandemic viruses. Humans present symptoms once they encounter novel virus strains for which they have not developed immunity. The major types of influenza and the unique features of seasonal influenza and pandemic influenza are summarized as follows.

Types of Influenza

There are three types of influenza viral strains that affect human beings: A, B, and C. Type A viruses naturally infect animals, and the typical hosts are birds and swine. However, Type A influenza can also infect humans when it combines with a human cell. Once type A viruses infect humans, they can further circulate among humans, with subtypes categorized by hemagglutinin (HA) and neuraminidase (NA), such as A(H1N1) and A(H3N2). Type B viruses

only circulate in humans, in most cases, having less severe symptoms. Type B viruses usually do not cause pandemic influenza but circulate as seasonal influenza viruses. Nowadays, B/Victoria and B/Yamagata are two major influenza circulating among humanity. Type C viruses infect only humans but with milder symptoms than Type A or type B and rarely cause pandemic influenza outbreaks ("Types of Influenza Viruses", 2017).

Influenza Emergencies

Influenza outbreaks have turned into emergencies when a circulating virus evolves as a result of changes in its genetic structure that is hard to predict (Carrat & Flahault, 2007).

Pandemic Influenza

Pandemic influenza occurs when an entirely novel influenza virus emerges, which is significantly distinct from its ancestors in its genetic structure (antigenic shift) (Carrat & Flahault, 2007). It is usually caused by an animal-hosted Type A virus initially interacting with a single human cell. The Type A virus combines with an essential human protein complex that regulates genes expression and mutates to a new virus strain transmitting among susceptible people (Fauci, 2006). Pandemic influenza can cause significant incidence and mortality since human immune systems cannot recognize such viruses and allow the infection to spread. For example, the 2009 pandemic influenza outbreak in Mexico and North America was caused by animal host influenza A(H1N1) that firstly circulated in swine, resulting in 80% of adult hospitalizations and 65% of child hospitalizations related with 2009 H1N1 in the US ("2009-2010 Influenza (Flu) Season", https://www.cdc.gov/flu/pastseasons/0910season.htm).

Seasonal Influenza

In contrast, seasonal influenza results from frequent and small genetic changes (antigenic drift) (Carrat & Flahault, 2007). Seasonal influenza is generally caused by re-emerging human-hosted influenza virus strains that can be identified by human immune systems already exposed to ancestor viruses or protected by vaccination ("Influenza virus infections in humans", 2014). Major recurrent seasonal influenza viral strains include Type A strains (i.e., H1N1, H3N2, H7N9) and Type B strains (i.e. Victoria and Yamagata). Seasonal influenza may still cause serious emergencies when vaccines are less effective for a particular virus strain (Flannery et al., 2018). Vaccine effectiveness can be altered during the process of vaccine cultivation when a vaccine virus mutates and no longer resembles the structure of the initially targeted virus (Flannery et al., 2018).

Influenza emergencies have significantly impacted human life and health. Since last century, multiple pandemic influenza strains have spread across humans. The notorious 1918 Spanish Flu was the worst influenza pandemic, infected 500 million people and took ~ 50-100 million lives around the world ("Spanish Flu", 2019). Forty years later, the 1957-1958 Asian Flu killed around 2 million people, and 70,000 of them were in the United States ("Influenza A virus subtype H2N2", 2019). This Asian Flu virus H2N2 was then shifted to H3N2, which caused the 1968-1969 Hongkong Flu with almost 1 million deaths ("Influenza pandemic", 2019). The 2009 H1N1 "swine flu" first silently affected 60% of La Gloria Population, a town with a population of 2,243 in Mexico, and spread rapidly over the US, ended up with over 1 million pandemic respiratory deaths ("2009 flu pandemic", 2019).

Seasonal influenza has also resulted in multiple emergencies. The 2005 A(H5N1) virus was a new strain with highly pathogenic features and considered as a pandemic threat (Mittal & Medhi, 2007). And the 2016 "Avian Flu" in China was a result of a new mutation of A(H7N9) virus strain, with 766 human infections in Asia ("Asian Lineage Avian Influenza A(H7N9) Virus | Avian Influenza (Flu)", 2018).

Traditional Influenza Surveillance

Landscape of Traditional Influenza Surveillance

Nowadays, World Health Organization (WHO) and many countries have made great efforts on influenza monitoring and prevention. In 1952, the WHO Global Influenza Surveillance and Response System (GISRS) was established ("Global Influenza Surveillance and Response System (GISRS)", 2019). Since then, GISRS has continuously collected information on the influenza viruses circulating in humans to update seasonal influenza vaccine compositions twice a year ("Influenza (Seasonal)", 2018).

GISRS is also a global network connecting National Influenza Centers (NICs) and WHO collaborating centers that make further research and analysis on the collected information. The influenza surveillance systems of the US Centers for Disease Control and Prevention (CDC) serve as one of WHO collaborating centers and are playing a significant role in national influenza surveillance in the United States ("CDC's World Health Organization (WHO) Collaborating Center for Surveillance, Epidemiology and Control of Influenza | CDC", 2017.).

The CDC influenza surveillance systems are categorized in five sub-systems with eight functional components that are deployed in 10 regions throughout all 50 states in the US as

presented in Figure 3 (created from “CDC's World Health Organization (WHO) Collaborating Center for Surveillance, Epidemiology and Control of Influenza | CDC”, 2017). Virological Surveillance detect newly emerging virus and collect data of confirmed and subtyped specimens through over 100 state-level public health laboratories (PHLs). Outpatient Illness Surveillance monitor influenza-like illness (ILI) visits in outpatient healthcare providers such as emergency departments (EDs), primary care providers, urgent care providers. Mortality Surveillance obtain pneumonia and influenza (P&I) related mortality statistics for both adults and pediatricians. Hospitalization Surveillance report the cases that are hospitalized due to influenza associated complications. Lastly, Summary of the Geographic Spread of Influenza watch the overall epidemic level of influenza (i.e., no activity, sporadic, local, regional, or widespread). These CDC surveillance components work together to contribute to seasonal influenza vaccines update, antiviral treatment, population-targeted intervention, detecting an emerging new virus and preparing for influenza pandemics (“Overview of Influenza Surveillance in the United States | CDC”, 2017).

Limitations in Virological Surveillance Data Collection

Virological Surveillance realizes epidemiological characteristics of circulating influenza by the public health laboratories (PHLs) (also called Bureau of Laboratories in Michigan), testing and confirming emergent virus types of suspected specimens that the sentinel Clinic Laboratories (CL) of both outpatient and in-hospital healthcare providers submit to the PHLs according to WHO and CDC’s guidance on specimen collection. For instance, in the state of Michigan, most virus positivity tests are completed through polymerase chain reaction (PCR)

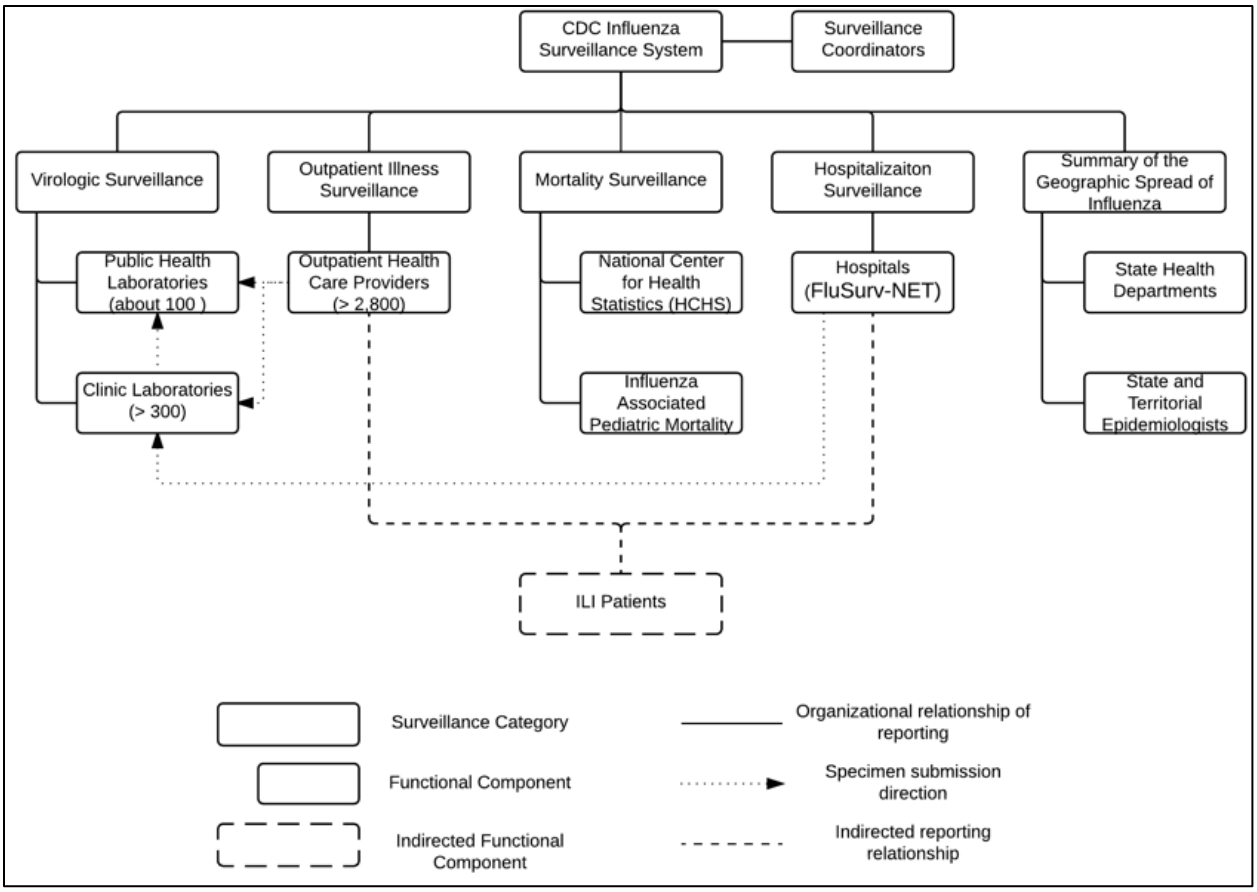


Figure 3. Organizational structure of CDC comprehensive influenza surveillance system and the reporting relationship

test in the CLs. Only sentinel CLs are required to submit 10 specimens monthly to the PHL for epidemiological characterization during normal flu seasons. Daily or weekly influenza confirmed incidence trends are collected through the virological surveillance scheme.

During an emergent outbreak, when the test kits for new virus are only available in the PHLs but have not been commercially distributed in the CLs, all healthcare providers, including those non-sentinel ones, send ILI specimens to the PHLs that test specimens using first-come-first-serve (FCFS) method. Numerous specimens submitted to the PHLs lead to their testing capacities quickly saturated, causing delay in data collected. Moreover, those specimens must be

tested within 3-4 days after being collected from suspected cases, to ensure the validity of test results. When the PHLs are deficient of test capacities, many specimens are never tested before expiration, or simply frozen until the end of flu season. The PHLs then restrict the target population to only high-risk groups (i.e., little children, elders, hospitalization, etc.). This compromised sampling method might have reduced the PHLs testing burden, but could lose the epidemiological characterization in healthy and low-risk populations, that are also susceptible with pandemic influenza.

Studies for Improving Virological Surveillance Data Collection

PHL Specimen Sampling Methods

The PHLs test the specimens using first-come-first-serve (FCFS) methods. During an emergency, the PHLs testing capacity is saturated by too many specimens. To meet high demand service, the PHLs restrict specimens to only high-risk population groups including little children (≤5 years old), elderly (≥65 years old), prenatal women and hospitalized cases (“("Lessons from a virus: Public health laboratories respond to the H1N1 pandemic ", 2011; Prieto et al., 2012). Under some circumstances, the PHLs prioritize the specimens that are submitted in the first week, without testing any specimen in later weeks. These sampling methods might be helpful to relieve the PHL testing workload, but it is unclear how much they have affected capture the epidemiological features.

Sample Size calculators

The Association of Public Health Laboratories (APHL) have issued a collection of sample size calculators that recommend how many specimens the PHL need to test to detect a specific level of influenza outbreak (seasonal or novel) ("Influenza Virologic Surveillance Right Size Roadmap", 2012). The calculators have some limitations in specimen sampling during influenza emergency: 1) The calculators require the user to assume a % of ILI cases seeking healthcare, 2) The calculators require the user to assume a % of the ILI cases that will test positive for flu in the PHLs. In both cases, percentages are assigned values with estimates from baseline data, but during an emergency, the percentages are the goals that need to be sought. In summary, the sample size calculators do not consider the outbreak dynamics, and require input that should be part of the output of virological surveillance process.

Optimize Specimen Collection Network

Some enhancement methods for specimen sampling have been proposed to model and optimize the network of specimen submission from sentinel providers to the PHLs that confirm and subtype influenza viruses. The optimization functions are proposed for maximizing geographic coverage (Polgreen et al., 2009), minimizing the total distance from each individual to each sentinel site using K-means allocation (Fairchild, Polgreen, Foster, Rushton, & Segre, 2013), or maximizing case count predictive ability with a proposed sentinel network model minimizing the difference between the real and obtained case count (Scarpino, Dimitrov, & Meyers, 2012). These improvements are based on the condition that the state health departments are responsible for sentinel recruitments. These sentinel providers are only volunteers that are not

pre-selected. Furthermore, during an emergency, the PHL capacity is so constrained that they fail to match the high demand testing services. In general, the focus of the research above model and optimize the specimen collection network, without considering the impact of human healthcare seeking behaviors and constrained PHL capacities.

Limitations in Novel Influenza Surveillance

Other than the traditional virological surveillance, data collection are also conducted by other surveillance schemes.

Michigan Syndromic Surveillance System (MSSS)

In the state of Michigan, the Emergency Departments (EDs) report daily numbers of constitutional registrations (i.e., fever, fatigue, chill, headache, body ache, diarrhea, etc.) and respiratory registrations (i.e., sore throats, sneeze, cough, congestion, etc.) to the Michigan Syndromic Surveillance System (MSSS) ("Clinical Signs and Symptoms of Influenza", 2019). In 2009, the MSSS collects data from 76 of 131 hospitals in Michigan (data from MSSS database 09). The MSSS trends serve as a real-time surveillance that informs the state level ILI dynamics with severe ILI symptoms. However, the trends are still dependent on human healthcare seeking behavior and contain information unrelated with influenza illness.

Internet-based Surveillance

Internet-based data collection tools have been created to monitor the influenza dynamics: internet search engine based tools such as Google flu trends (Ginsberg et al., 2009), internet

search queries such as Yahoo! (Polgreen, Chen, Pennock, & Nelson, 2008) and Baidu (Yuan et al., 2013), internet based microblogs such as Twitter messages (Signorini, Segre, & Polgreen, 2011), Wikipedia article views (McIver & Brownstein, 2014), on-line dinner reservations (Nsoesie, Buckeridge, & Brownstein, 2014), and healthcare practitioners search queries (Santillana, Nsoesie, Mekaru, Scales, & Brownstein, 2014). Although these real-time surveillance tools collect ILI data through a wider range of population, they might contain a lot of unexpected information (i.e, Google flu trend) (Lazer, Kennedy, King, & Vespignani, 2014).

Problems with Biased Surveillance Data

A previous study has used both daily ILI and confirmed incidence trends to infer the daily incidence of a particular subtype of emergent virus (e.g. H1N1) (Birrell et al., 2011). Some have used daily and weekly ILI trends combining with other internet based data sources to generate weekly and daily ILI forecasts (Santillana et al., 2015; Shaman et al., 2012; Shaman et al., 2013). However, these methods require all the trends are at hand for any future inference or forecast, which is still unrealistic. Even all the trends are complete, they still might be biased by the reactive data collection operations under current surveillance landscape.

A recent study (Birrell et al., 2017) has incorporated general practices data, serological data, hospitalization data, virological continued data and commuting data in their real-time pandemic outbreak model using Bayesian Inference and Markov Chain Monte Carlo (MCMC) methods (also used Sequential Monte Carlo as an alternative). Their parallel-region (PR) model has been proven to better and timely fit the epidemic data, but "lack the predictive power to forecast the spread of infection in the early stages of a pandemic" (Birrell et al., 2017) for the

sake of unstable and incomplete surveillance data, especially at the beginning of pandemic. Despite the fact that these methods have been proved to forecast ILI trends of weekly and daily resolution, they couldn't achieve optimal performance with biased surveillance data at the early stage of an emergent outbreak.

Research Gap

The PHL tested incidence trends is the unique way to know the disease epidemiological features but might be biased during the data collection operations. However, after an influenza emergency is declared, a growing number of symptomatic cases seek healthcare from any kind of healthcare providers. The MSSS trend report real-time influenza-like cases through ED registrations, but might also be affected by the effects other than disease characterization (i.e. the effect of "fear"). Primary/urgent care practitioners collect a proportion of specimens by their own judgement, which might lose the track of uncollected infected specimens. The tested trends in PHL might be distorted by the sampling criteria for selecting specimens due to constrained lab capacity. Although novel internet-based surveillance tools have been beneficial in early detection of emergent influenza, they are not able to define the disease epidemiological features of a specific virus subtype.

There have been many studies that used one or multiple influenza outbreak data from different surveillance schemes to model and predict influenza outbreak, but the performance of their methods depend on the accuracy of surveillance data. Therefore, it is necessary to evaluate the biasedness in the influenza incidence trends that several data collection schemes provide at

the initial phase of an emergent outbreak. To my knowledge, no study has been done for this objective.

CHAPTER III

HIGH PERFORMANCE AGENT-BASED SIMULATION OF INFLUENZA CO-CIRCULATION IN THE STATE OF MICHIGAN

Over the past two decades, disease modelers have proposed several approaches to inform influenza preparedness by estimating relevant disease parameters, evaluating mitigation strategies, and forecasting disease trends (Schlegelmilch et al., 2012; Nsoesie, Mararthe, & Brownstein, 2013; Chretien, George, Shaman, Chitale, & Mckenzie, 2014) and the models in their review are summarized as follows.

Approaches for these purposes include statistical models, compartment models, meta-population models and agent-based models. Statistical approaches include time series models, generalized linear models, Bayesian networks, classification methods, survival analysis, prediction market, and meteorology using analog methods, Statistics models have been used to model historical outbreaks but had limitation for novel viruses. For example, time series were good at modelling periodic changes and temporal related trends.

Compartmental models mainly apply differential equations to represent health states transitions (i.e., SIR model: susceptible – infectious – recovered) in the subpopulation. Several compartment models have been developed from the SIR model, such as susceptible– exposed– infectious–recovered (SEIR) model, and the SIR model with ensemble adjustment Kalman filter techniques (SIRS-EAKF) and further improved by integrating geographic distinctions (Network-

SIRS-EAKF). Compartmental models usually aggregate the population, which complicates its use in simulations where a coarser detail in the population is needed.

Meta-population models divide people into multi-population groups to enhance compartmental models but they might not be coarse enough to evaluate mitigation and data collection strategies. Nevertheless, meta-population models were still constrained in modeling suppositions and defining parameters for various situations.

Agent-based models (ABM) have been recognized as dynamic large-scale complex systems that support scientists and researchers to study microscopic behaviors dependent on certain micro-level attributes and criteria (Abar, Theodoropoulos, Lemarinier, & O'Hare, 2017). Distinct from other simulation models, ABMs are featured with numerous agents capable of automatic learning and self-adaptive to surroundings. ABMs have been largely applied in the studies of infectious diseases caused by transmissible pathogenic microorganisms circulating among individuals (Deangelis, & Grimm, 2014; Jit, & Brisson 2011; Kleef, Robotham, Jit, Deeny, & Edmunds, 2013). Influenza has been recognized as one of major infectious diseases studied with ABM on the purposes of controlling seasonal and pandemic influenza outbreaks (Willem, Verelst, Bilcke, Hens, & Beutels, 2017).

Flu MODELO 1.0

Flu MODELO 1.0 is an AB model that extends a baseline simulation of circulating influenza viruses (Prieto & Das, 2016).

In the baseline simulation, the outbreak is initiated by I_i infected cases with viral strain *i*, where $i = 1, \ldots, n$, and *n* corresponds to the number of co-circulating influenza viruses. Default

initial infection cases is set to 30 cases at day 0. Influenza viruses spread among humans through daily contacts in a closed urban population setting (i.e., no commuting or long distance travel was allowed), in the absence of mitigation policies. In Flu MODELO 1.0, the baseline is adapted to simulate influenza spread in the State of Michigan with more realistic assumptions for the geographic, demographic, traveling patterns and mitigation effects. In addition, the simulation was implemented with high performance parallel computing. These features are explained as follows.

Geographic

Flu MODELO 1.0 models flu spread among people in different geographical regions. For the State of Michigan, Flu MODELO 1.0 links four closed population settings to simulate Michigan characteristic regions: 1. Upper Peninsula Rural, 2. Northern Lower Peninsula, 3. Southeast Michigan Council of Governments, and 4. Southern Lower Peninsula.

Michigan Department of Transportation (MDOT) divided Michigan into Northern Lower Peninsula, Southeast Michigan Council of Governments (SEMCOG), Small Cities, Small Urban Model Area, Southern Lower Peninsula, Transportation Management Areas (TMA) and Upper Peninsula (see Figure 4 from "2009 Comprehensive Household Travel Data Collection Program MI Travel Counts II Comparison Report"). In Flu MODELO 1.0, Small Cities, Small Urban

Model Area and TMA were combined as a single region called Southern Lower Peninsula, since these areas only count a tiny part of regional population.

Demographic

Flu MODELO 1.0 is populated with the Michigan demographic features: 10 million people, 212,000 non-household locations (14 types, including home, workplace, school, store, and restaurant), and 4 million household locations (Yax, 2010; "2009 Comprehensive Household Travel Data Collection Program MI Travel Counts II Comparison Report", 2010).

Each region preserves its realistic size in population (created with household structures) and locations (see Table 1-4). Population distribution by each location type (e.g., percentage of people in a type of workplace per day) is provided for each region to create initial individual affiliation to a location (see Table 5). Michigan statewide household composition is used (see Table 6) (e.g., % of households with 2 adults and 1 children), and individual age distributions in Table 7 and Table 8 (e.g. % of individuals between 5 and 9), assuming the same household and age structures in all regions. The same data structures are used for these Michigan distributions as the baseline model.

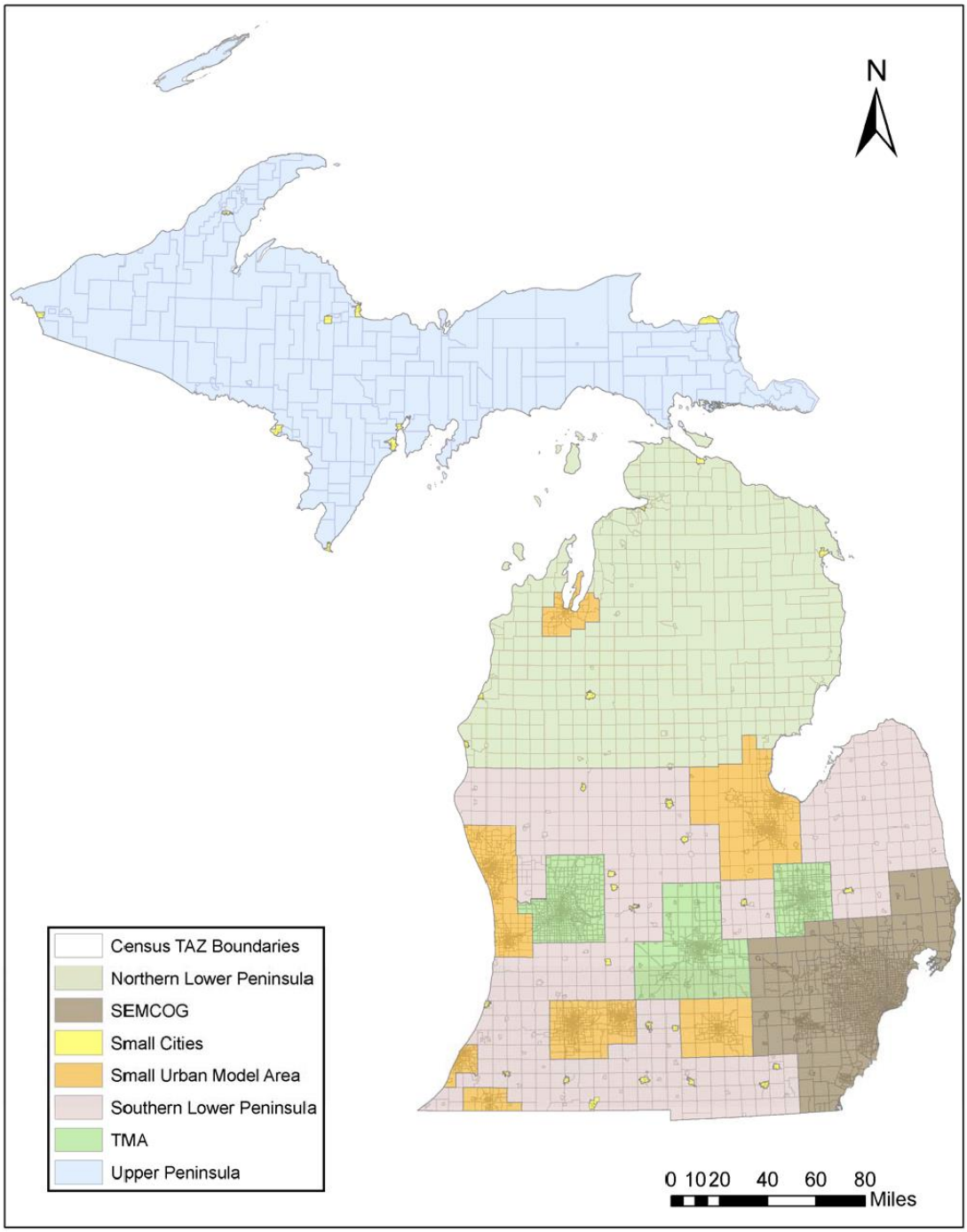


Figure 4. Michigan Geographic Regions Provided by Michigan Department of Transportation

Table 1 *Frequency Distributions of Location Types, in Region 1. Upper Peninsula Rural*

Location Type ID	Location Type	Number of Locations
0	home	90553
1	factory	8965
2	office	16544
3	pre-school	399
4	elementary school	687
5	middle school	548
6	high school	462
7	university	52
8	afterschool center	6279
9	grocery store	11067
10	other stores	0
11	restaurant	4646
12	entertainment	807
13	church	41

Table 2 *Frequency Distributions of Location Types, in Region 2. Northern Lower Peninsula*

Location Type ID	Location Type	Number of Locations
0	home	218238
1	factory	8987
2	office	16585
3	pre-school	400
4	elementary school	689
5	middle school	549
6	high school	463
7	university	52
8	afterschool center	6295
9	grocery store	11094
10	other stores	0
11	restaurant	4658
12	entertainment	809
13	church	41

Table 3 F*requency Distributions of Location Types, in Region 3. Southeast Michigan Council of Governments*

Location Type ID	Location Type	Number of Locations
0	home	2071786
1	factory	9923
2	office	18311
3	pre-school	441
4	elementary school	761
5	middle school	606
6	high school	511
7	university	57
8	afterschool center	6950
9	grocery store	12249
10	other stores	0
11	restaurant	5143
12	entertainment	894
13	church	46

Table 4 *Frequency Distributions of Location Types, in Region 4. Southern Lower Peninsula*

Location Type ID	Location Type	Number of Locations
0	home	1776549
1	factory	9734
2	office	17963
3	pre-school	433
4	elementary school	746
5	middle school	594
6	high school	502
7	university	56
8	afterschool center	6817
9	grocery store	12016
10	other stores	0
11	restaurant	5045
12	entertainment	877
13	church	45

Table 5 *Relative Population Distribution of Weekdays Worktime Errands, Weekdays After Work Errands and Weekends Errands, in Michigan*

Location Type ID	Weekday Errands Worktime	Weekday Errands After Work	Weekend Errands
0	0.412147901	0	0
1	0.073188396	0	0
2	0.133284601	0	0
3	0.006227524	0	0
4	0.080693333	0	0
5	0.04127468	0	0
6	0.058387304	0	0
7	0.035924159	0	0
8	0.058854716	0.003209243	0.002617801
9	0.060778756	0.300385109	0.323298429
10	0	0.33055199	0.27486911
11	0.034742409	0.154043646	0.181937173
12	0.004496221	0.166238768	0.197643979
13	0	0.045571245	0.019633508

Table 6 *Cumulative Distribution Function (CDF) by Household Composition, in Michigan*

Adults	Children	CDF
1	0	0.279
1	1	0.319
2	0	0.628
1	2	0.671
2	1	0.8
1	3	0.812
2	2	0.939
1	4	0.944
2	3	1

Table 7 *Cumulative Distribution Function (CDF) by Adult Age, in Michigan*

Ages	CDF
29	0.16
64	0.83
99	1

Table 8 *Cumulative Distribution Function (CDF) by Child Age, in Michigan*

Ages	CDF
5	0.24
9	0.47
14	0.72
17	0.85
22	1

Travel Patterns

In the baseline model, the simulation was limited to one population setting with intra-regional travel patterns. This assumption was not suitable for the State of Michigan since the travel frequency varies with the type of trip and destination (e.g., short distance errands, commuting, or long distance travel within or outside the State). As travel patterns affect the contact frequency, they also may influence the disease spread.

Flu MODELO 1.0 linked four AB models, simulating each of the four characteristic Michigan regions: 1. Upper Peninsula Rural, 2. Northern Lower Peninsula, 3. Southeast Michigan Council of Governments, 4. Southern Lower Peninsula. Individuals in Flu MODELO 1.0 can make three types of trips: intra-regional, inter-regional, and across the Michigan boundaries.

Intra-regional

These trips occur when individuals circulate within the region they live. According to the "2009 Comprehensive Household Travel Data Collection Program MI Travel Counts II Comparison Report", most contacts occur while individuals engage in intra-regional trips as they

are the most frequent (92% of the travels were short-distance travel < 60 mins). The same errand distribution was used (in Table 5) as the intra-regional travel. The same contact rates (see Table 9) in the baseline model.

Inter-regional

These trips propagate infection across regions when infected individuals leave their home regions. To simplify the work, long distance travelling (> 60 mins) was assumed as inter-regional travelling and individuals make inter-regional travelling among the four regions in Michigan with the correspondent daily trip frequency per person (see Table 10). Each traveler follows the probability in Table 11 to travel to a destined region. The inter-regional simulation combine both tables as the origin and destination for travelling on a specific day.

Table 9 *Frequency Distribution of Contact Rates, in Michigan*

Location type ID	Contact rate
0	3
1	3
2	3
3	2
4	2
5	3
6	3
7	2
8	2
9	2
10	0
11	2
12	2
13	2

Table 10 *Daily Long Distance Trips in Michigan, by Regions*

Originated Region ID	Name	Trips/Day/Person
1	Upper Peninsula Rural	0.02936
2	Northern Lower Peninsula	0.08844
3	SEMCOG	0.01245
4	Southern Lower Peninsula	0.02955

Table 11 *Relative Distribution by Regions as Destinations, in Michigan*

Destination	Cumulative Probability
Upper peninsula	0.0650
Northern lower peninsula	0.2790
SEMCOG	0.4426
Southern lower peninsula	0.5890
Other states or countries	1.0000

Across Michigan Boundaries

This feature allows infected travelers to leave or enter Michigan, and therefore reduce or increase the chance of disease propagation. The last row in Table 11 describes the probability of a destination to be outside of Michigan is 41.1% ("State long range transportation plan 2005-2030: Travel characteristics technical report", 2006). Daily number of arrivals in each region from outside of Michigan refers to Table 12 from Michigan airport report ("Scheduled Passenger Deplanements 2009 Airport Facility Year to Date Record", 2009) and were averaged to daily arrivals, assuming most people chose to travel by airplane when travelling from outside Michigan.

Table 12 *Arriving from Outside Michigan, by Region*

Region	Total arrivals yearly	Average arrivals per day
Upper Peninsula	105,408	289
Northern Lower Peninsula	213,149	584
SEMCOG	15,653,937	42,887
South Lower Peninsula	1,786,217	4,894

Mitigation

To incorporate the effects of mitigation, the baseline infection probability model is extended. The model considers the symptomatic, vaccination, and antiviral status of an infectious and a susceptible individual when they contact each other. In what follows these two types of individuals will be referred as "infectious" and "susceptible".

In Flu MODELO 1.0, the infectious individual has vaccination status j and antiviral status j', where $j = 1$ if the infectious is vaccinated and $j = 0$ otherwise; $j' = 1$ if the infectious received antiviral and $j' = 0$, otherwise. The susceptible individual has vaccination status k and antiviral status k', where $k = 1$ if the susceptible is vaccinated, and $k = 0$ otherwise; $k' = 1$ if the susceptible received antiviral, and $k' = 0$, otherwise. In the simulation model, if a contact between an infectious and a susceptible occurs between time t and $t + 1$, the susceptible gets influenza with probabilities of infection $P_{t,j,j',k,k',i}$ (in Equation1) and $P_{t,j,j',k,k',i}$ (in Equation 2) as follows:

When the infectious is symptomatic:

$$P_{t,j,j',k,k',i} = \theta_{i,V}^{k} \times \theta_{AV}^{k'} \times \frac{[\psi_{i,V}^{j} \times \psi_{AV}^{j'} \times \frac{\phi_{i,V}^{j} \times \phi_{AV}^{j'} \times E[\mathcal{R}_i]}{(1-\gamma)\pi+\gamma}]E[W_{t,i}]}{h_t + w_t + o_t} \qquad \text{Equation 1}$$

When the infectious is asymptomatic:

$$P_{t,j,j',k,k',i} = \theta_{i,V}^{k} \times \theta_{AV}^{k'} \times \frac{\gamma \frac{\phi_{i,V}^{j} \times \phi_{AV}^{j'} \times E[\mathcal{R}_i]}{(1-\gamma)\pi+\gamma} E[W_{t,i}]}{h_t + w_t + o_t} \quad \text{Equation 2}$$

Equations (1) and (2) are used with the following parameters and assumptions:

Epidemiological

$E[\mathcal{R}_i]$: Expected reproduction number for an *i* influenza viral strain. $E[\mathcal{R}_i]$ is interpreted as the expected number of cases infected by a case with virus *i*, and can be fitted from scenarios of exponential growth in the incidence (Wallinga & Lipsitch, 2007).

$E[W_{t,i}]$: Expected viral shedding profile of an infectious individual from time *t* to *t +1* infected with an influenza strain *i*. The viral shedding profile defines the pace at which infection will be shed by a case, and can be estimated by analyzing nasal swabs in human volunteers (Carrat et al., 2008).

γ: Reduction in infectiousness due to asymptomatic infection.

π: Percentage of symptomatic individuals in the population.

h_t, w_t, and o_t: Number of susceptible individuals contacted by an infected case with enough closeness and duration to be infection candidates in the household (h_t), workplace (w_t), and other places (o_t), respectively. These parameters update their values through the simulation with the progression of time t (Prieto & Das, 2016). Table 13 shows the values and feasible ranges used in the simulation implementation.

Table 13 *Epidemiological Parameters Used in the Simulation*

Parameter	Values	Reference
γ	0.5	(Basta, Halloran, Matrajt, & Longini, 2008)
π	0.67	(Basta, Halloran, Matrajt, & Longini, 2008)
$E[W_{t,i}]$	Gamma (2.3, 0.68)	(Carrat et al., 2008)
$E[\mathcal{R}_1]$	1.1 – 2.1	(Wallinga & Lipsitch, 2007)

Vaccination

$\theta^{k}_{i,V}$: Reduction factor in the transmission probability of a susceptible given exposure to infection. It is assumed that this reduction factor is equal to the risk ratio between the number of individuals infected after they received the vaccine for viral strain i and the number of individuals infected after they received a placebo. The risk ratio is calculated using experimental data from human volunteers (Basta, Halloran, Matrajt, & Longini, 2008).

$\phi^{j}_{i,V}$: Reduction in the infectiousness of an infectious. It is assumed that this reduction factor is equal to the ratio between the number of individuals shedding infection after they received the vaccine for viral strain i and the number of individuals shedding infection after they received a placebo. The risk ratio is calculated using experimental data from human volunteers (Basta, Halloran, Matrajt, & Longini, 2008).

$\psi^{j}_{i,V}$: Reduction in the infectiousness of an infectious symptomatic. It is assumed that this reduction factor is equal to the ratio between the number of individuals with symptom after they received the vaccine for viral strain i and the number of individuals with symptoms after they received a placebo. The risk ratio is calculated using experimental data from human volunteers (Basta, Halloran, Matrajt, & Longini, 2008).

Let s_i be vaccination status of viral strain i. If vaccination for viral strain i is in place, s_i = 1, and 0 otherwise. If vaccination for viral strain i is not in place, assume $\theta_V^1 = 1$, $\phi_V^1 = 1$, and $\psi_V^1 = 1$ for the that virus (see Table 14 for the values of the seasonal virus). The vaccine coverage rates are stratified per age and are consistent with typical vaccination estimates for the population in Michigan (see Table 15). Vaccinated individuals are selected at random.

Table 14 *Vaccination and Antiviral Parameters Incorporated in the Infection Probability Model*

Parameter	Value	Reference
θ_V^k	0.6 when k = 1, and 1 otherwise	(Basta, Halloran, Matrajt, & Longini, 2008)
ϕ_V^j	0.5 when j = 1, and 1 otherwise	(Basta, Halloran, Matrajt, & Longini, 2008)
ψ_V^j	0.17 when j = 1, and 1 otherwise	(Basta, Halloran, Matrajt, & Longini, 2008)
$\theta_{AV}^{k'}$	0.7 when k' = 1, and 1 otherwise	(Longini, Halloran, Nizam, & Yang, 2004)
$\phi_{AV}^{j'}$	0.2 when j' = 1, and 1 otherwise	(Longini, Halloran, Nizam, & Yang, 2004)
$\psi_{AV}^{j'}$	0.4 when j' = 1, and 1 otherwise	(Longini, Halloran, Nizam, & Yang, 2004)
s_i	1, when vaccination for viral strain i is in place, and 0 otherwise	

Table 15 *Vaccination Coverage in Michigan*

Age Interval	% of Vaccine Coverage in the Age Interval
0 – 5	43.4
6 – 14	27.7
15 – 17	19.9
18 – 29	10.85
30 – 64	23.15
65 –	39.85

Note. Data from the Michigan Care Improvement Registry 2014-2015 ("Influenza update", 2016).

Antivirals

$\theta_{AV}^{k'}$: Reduction in the transmission probability of an antiviral protected susceptible given exposure to infection. Similar interpretation and calculation as θ_{V}^{k} (Longini, Halloran, Nizam, & Yang, 2004).

$\phi_{AV}^{j'}$: Reduction in the infectiousness of an antiviral protected infectious. Similar interpretation and calculation as ϕ_{V}^{j} (Longini, Halloran, Nizam, & Yang, 2004).

$\psi_{AV}^{j'}$: Reduction in the infectiousness of an antiviral protected infectious symptomatic. Similar interpretation and calculation as ψ_{V}^{j} (Longini, Halloran, Nizam, & Yang, 2004).

Table 14 presents the values for the parameters. Only a proportion a of the mortality risk groups (i.e., individuals with ≤ 5y and ≥ 65y) will receive the antiviral course. Individuals are randomly selected for antiviral application.

Self-induced Absenteeism and Hospitalization

If absenteeism is in place, each day, an infected case will be randomly withdrawn from his/her usual schedule with the following probability in Equation 3:

$$u_{t,i} = E\left[W_{t,i}\right] \times \sigma, \qquad \text{Equation 3}$$

where σ is the probability that the case decides to withdraw from her usual activities due to symptoms (withdrawal probability). If a case is selected, then it is decided whether the patient will be randomly selected to stay at home or being hospitalized based on U.S. hospitalization rates (see Table 16).

Table 16 *Hospitalization Rates in Michigan*

Age interval	% hospitalization per day (upper bound)
0 – 4	0.001
5 – 9	0.000429
10 – 14	0.000429
15 – 17	0.000429
18 – 22	0.000429
22 – 29	0.000429
30 – 64	0.001
>64	0.005

Note. Data from the FLuView Influenza Hospitalization Surveillance Network 2013-2014 ("Laboratory-Confirmed Influenza Hospitalizations", n.d.)

Parallelization

Flu MODELO1.0 is an agent-based simulation of human influenza co-circulation equipped with travel and mitigation functions (Ostroy, Prieto, Gu, Dedoncker, & Paul, 2017). It is a C++ version refined from the baseline model (Prieto & Das, 2016) and further adapted to the demographic, geographic and travel patterns of the state of Michigan. The model creates about 10 million (based on Michigan population) individuals that are susceptible to influenza infection through hourly updated contact activities through more than 212,000 workplaces and 4 million households. Infectious probability is customized on individual level, regulated by model parameters related with self-immunity (i.e., co-infect or re-infect), epidemiological characteristics, and mitigation measures in place. Flu MODELO1.0 is useful to uncover influenza epidemic and propose essential containments to downgrade the outbreak.

This level of decision making requires as fast process as within one day (Prieto et al., 2012). However, the Flu MODELO1.0 simulation for Michigan takes more than 72 hours to

complete just a single simulation run (125 simulation days/run). With the growing number of experimental runs, it is even less practical to provide policy makers with timely recommendations. Several previous attempts have been made to accelerate the baseline model (Prieto & Das, 2016) using OpenMP（Soto-Ferrari, Holvenstot, Prieto, Doncker, Kapenga, 2013). Nvidia CUDA platform on single Graphic Process Unit (GPU) (Holvenstot, Prieto, & Doncker, 2014) and multiple GPUs (Shekh, Doncker, Prieto, 2015). These studies proposed general-purpose methods to reduce simulation runtime, but they were either based on the assumptions violating the demographic and geographic distributions of simulated area, or made idealistic but impractical changes on model behavior, missing the target of guiding influenza preparation for a real influenza outbreak.

This study employs OpenMP directives for Flu MODELO1.0 parallelization. OpenMP is parallel computing tool used as directives which partition a process into multiple threads over a number of Central Process Unit (CPU) cores. The use of OpenMP is usually more convenient for many researchers and scientists from various fields. Furthermore, by using CPU shared memory, OpenMP is favorable in models needing large memory which is very common in ABMs.

To achieve a better CPU usage with OpenMP implementation, the model was parallelized by optimizing model functions to fit parallel programming by solving data/task dependency problems. This is the key step to apply OpenMP in ≈95% of model functions. OpenMP directives were added which include “omp for”, static/dynamic scheduling, reduction, critical, atomic and random seeds for each thread. Some specific changes were made for vectors to improve the vector operation efficiency. Figure 5 shows the speed up results using 1, 2, 4, 8 and

16 CPU cores after the parallelization procedures were implemented. The resulting model was tested in Maryland Advanced Research Computing Center (MARCC).

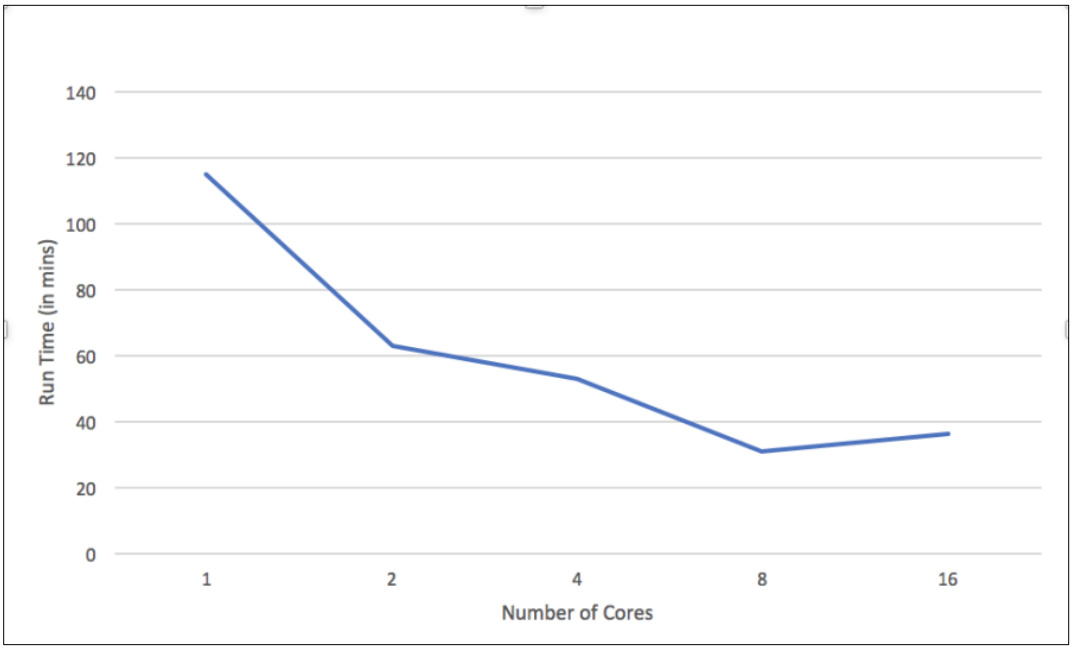


Figure 5. Speedup of Model with Parallelization

Model Demonstration

The refined model Flu MODELO 1.0 was demonstrated by running the simulation to produce the real influenza incidence trends based on the parameter default values in Table 17 and the Michigan demographic, geographic and travel features (see Flu MODELO 1.0 of this Chapter). Table 17 shows the simulated daily influenza incidence with the co-circulation of both pandemic and seasonal influenza for 125 days. With the presence of mitigation and containment, the influenza outbreak resulted in less daily influenza incidences and less peak time incidences. When all mitigation measures were in place, the influenza outbreak magnitude was reduced to the lowest.

Table 17 *A Summary of Model Parameter List*

Name	Description	Feasible Space	Default value
WITHDRAWAL_PROB	probability that a person is withdrawn from regular activities	0-1	0.5
PANDEMIC_VACCINATION	pandemic vaccination in place or not	0=not in place 1=in place	0
SEASONAL_VACCINATION	seasonal vaccination in place or not	0=not in place 1=in place	0
NUM_CITIES	number of regions in simulation	4	4
ALLOW_TRAVEL	allow people to travel among the regions	0=not allow 1=allow	1
ALLOW_OOS_TRAVEL	allow people to travel across state boundaries	0=not allow 1=allow	1
VACC_COVERAGE_0TO5	percentage of people vaccinated by age 0 to 5	0-1	0.434
VACC_COVERAGE_6TO14	percentage of people vaccinated by age 6 to 14	0-1	0.277
VACC_COVERAGE_15TO17	percentage of people vaccinated by age 15 to 17	0-1	0.199
VACC_COVERAGE_18TO29	percentage of people vaccinated by age 18 to 29	0-1	0.1085
VACC_COVERAGE_30TO64	percentage of people vaccinated by age 30 to 64	0-1	0.2315
VACC_COVERAGE_65PLUS	percentage of people vaccinated by age 65 and up	0-1	0.3985
VACCINE_COEFF_TRANSMISSION	modifier to the chance of transmitting virus when vaccinated	0-1	0.6
VACCINE_COEFF_INFECTIOUSNESS	modifier to the infectiousness of a virus transmitted by a vaccinated person	0-1	0.5
VACCINE_COEFF_SYMPTOMATIC	additional modifier for vaccinated symptomatic individuals	0-1	0.17
ANTIVIRAL_TREATMENT	antiviral treatment in place or not	0=not in place 1=in place	1
AV_TRANSMISSION	modifier to the chance of transmitting virus when on antiviral treatment	0-1	0.7
AV_INFECTIOUSNESS	modifier to the infectiousness of a virus transmitted by a person on antiviral treatment	0-1	0.2

Table 17 – continued

AV_SYMPTOMATIC	additional modifier for vaccinated symptomatic individuals on antiviral treatment	0-1	0.4
ANTIVIRAL_PERCENTAGE	the proportion of vulnerable people receiving antiviral treatment	0-1	0.5
ASYMP	factor by which the infectiousness profile of an asymptomatic is reduced.	0-1	0.5
epsilon_p	factor by which a pandemic infectious profile is affected given the influence of a seasonal profile that occurs (e.g. cross-immunity created by the seasonal virus)	0-1	1
epsilon_pr	factor by which a pandemic infectious profile is affected given the influence of a seasonal profile that has already passed (e.g. cross-immunity created by the seasonal virus)	0-1	1
epsilon_ps	factor by which a pandemic infectious profile is affected given the influence of a seasonal profile when both occur at the same time (e.g. cross-immunity created by the seasonal virus)	0-1	1
epsilon_s	factor by which a seasonal infectious profile is affected given the influence of a pandemic profile that also occurs (e.g. cross-immunity created by the pandemic virus)	0-1	1
epsilon_sr	factor by which a seasonal infectious profile is affected given the influence of a pandemic profile that has already passed (e.g. cross-immunity created by the pandemic virus)	0-1	1
epsilon_sp	factor by which a seasonal infectious profile is affected given the influence of a pandemic profile when both occur at the same time (e.g. cross-immunity created by the pandemic virus)	0-1	1
k_hh	proportion of duration and closeness of a household contact with repect to the household contact k_hh=1. (Mossong et al., 2006)	0-1	1
k_wp	proportion of duration and closeness of a workplace contact with respect to the household contact k_wp=0.67. (Mossong et al., 2006)	0-1	0.67
k_er	proportion of duration and closeness of an errand contact with respect to the household contact k_er=0.44. (Mossong et al., 2006)	0-1	0.44

Table 17 – continued

INI_PANDEMIC_INCIDENCE	number of initial infectiousness with pandemic virus	30-10000	30
INI_SEASONAL_INCIDENCE	number of initial infectiousness with seasonal virus	30-10000	30
PANDEMIC_RN	base reproduction number for pandemic influenza	1.1-2.5	1.8
SEASONAL_RN	base reproduction number for seasonal influenza	1.1-2.5	1.3
PERCENT_SYMPTOMATIC	Percentage of people present symptomatic infection.(Carrat et al., 2008)	0-1	0.669
ABSENTEEISM	level of self-induced absenteeism mitigation in place	0=no absenteeism 1=self-induced absenteeism 2=self-induced absenteeism with hospitalization	2
BEGIN_DAY	day of the week when the simulated outbreak begins	1=MON, 2=TUE, 3=WED, 4=THUR, 5=FRI, 6=SAT, 7=SUN	3

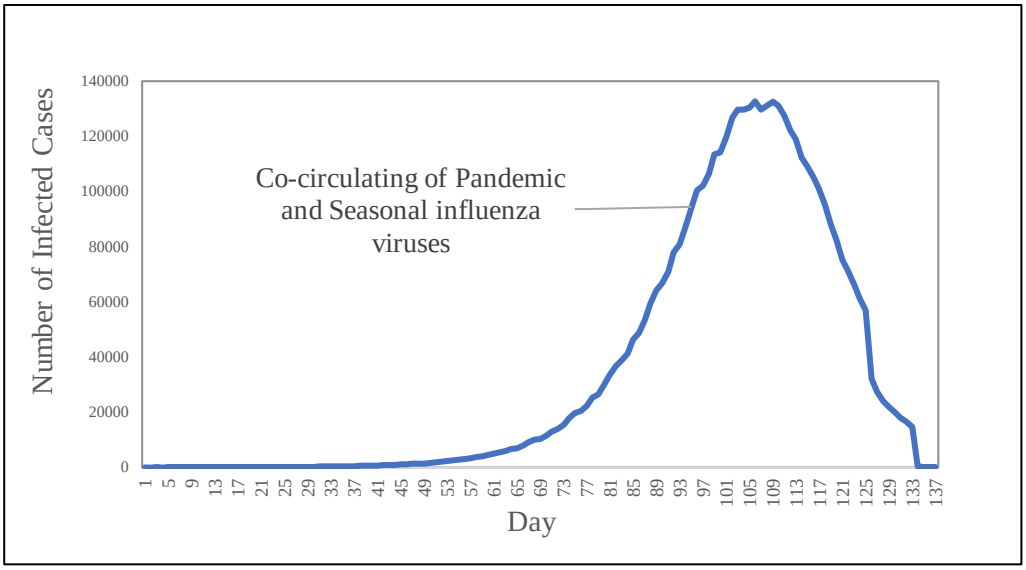


Figure 6. Simulated Daily Number of Influenza Incidences by Flu MODELO 1.0 Influenza Co-circulation in the State of Michigan

CHAPTER IV

ESTIMATES OF EPIDEMIOLOGICAL CHARACTERISTICS AND MITIGATION LEVELS AFTER PANDEMIC DECLARATION IN MICHIGAN

In this chapter, Flu MODELO 1.0 is used in a simulation based optimization approach to provide estimators for two types of parameters: 1) Epidemiological parameters describing the baseline spread behavior, and 2) Mitigation parameters attenuating the baseline spread behavior.

The estimators obtained address **Research question 1** by providing understanding of plausible scenarios under which influenza propagates after an official influenza emergency is declared in the US. The plausible scenarios are used as a testbed for the evaluation of data collection operations in Michigan in Chapter V.

Methodological Overview

As a testbed scenario, the growth phase after the H1N1 pandemic declaration in the State of Michigan was used. Simulation based optimization using response surface experiments and Derringer's desirability function were conducted with two optimization objectives: 1) to minimize the difference between the real and simulated growth rate estimate of new cases infected daily, 2) to minimize the difference between the real and simulated precision (or standard error) of growth rates of new cases infected daily. Formally, the optimization problem can be written as follows:

$$Minimize \quad z_1 = |r_1 - r_2| \qquad \text{Equation 4,}$$

$$Minimize \quad z_2 = |\eta_1 - \eta_2| \qquad \text{Equation 5,}$$

where r_1 is a constant growth rate from real data, and η_1 is the precision (or standard error) of $\hat{r}_1$. The parameter r_2 is the growth rate extracted from the output incidence trend in Flu MODELO 1.0, and η_2 is the precision (or standard error) of $\hat{r}_2$. To minimize the optimization function, these parameters were systematically adjusted to yield an r_2 and η_2 that closely matched $r_1$and η_1, respectively, constrained by the ranges of the parameters described in Table 18.

Real Data

Flu MODELO 1.0 was calibrated using aggregated daily time series of cases with constitutional symptoms visiting the 76 emergency departments (EDs) of a total of 131 hospitals in the state of Michigan. The ED data was provided by the Michigan Department of Health and Human Services (MDHHS) and is exempted from institutional review board (IRB). The exponential growth rate r_1=0.04447 was extracted from April 21 to May 4 in 2009 (14 days), since this period begins with the public declaration of the pandemic H1N1 outbreak by the U.S. media, and continues with the noticeable increase in symptomatic cases reported to the ERs (see Figure 7). The ED data was used for calibration as there exists evidence of the strong relationship between influenza incidence and ED crowdedness (Olshaker, & Rathlev, 2006).

Table 18 *Parameters As Predictor Variables*

Parameter	Description	Feasible Sample Space
Epidemiological Parameter		
PANDEMIC_RN	$E[\mathcal{R}_1]$. Expected reproduction number for a pandemic influenza viral strain, noted as "1"	1.6-2.1^{R+}
SEASONAL_RN	$E[\mathcal{R}_2]$. Expected reproduction number for a seasonal influenza viral strain, noted as "2"	1.1-1.6^{R+}
INI_PANDEMIC_INCIDENCE	I_1. Number of initial infectiousness with pandemic virus	30-1000^{I+}
INI_SEASONAL_INCIDENCE	I_2. Number of initial infectiousness with seasonal virus	30-1000^{I+}
Mitigation parameter		
WITHDRAWAL_PROB	σ. Probability that a person is withdrawn from regular activities	0.00001-0.99999^{R+}
ANTIVIRAL_PERCENTAGE	a. Probability that a vulnerable person receives antiviral treatment	0.00001-0.99999^{R+}
SEASONAL VACCINATION	s. Whether seasonal vaccination is in place or not (in place once the pandemic is declared) Whether self-induced	1
ABSENTEISM	q. Absenteeism is in place or not (in place once the pandemic is declared)	2

Note: $^{R+}$ positive real numbers, $^{I+}$ positive integers.

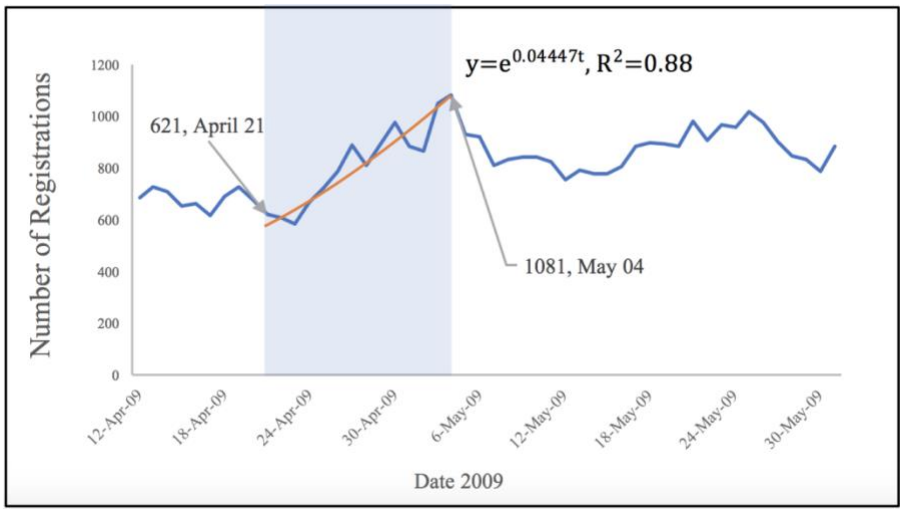


Figure 7. MSSS ERCR09 Data. Blue line: ED Daily Constitutional Syndromic Registrations in the state of Michigan During 4/12-5/31. Solid blue area: started with pandemic declaration date (4/21) to incidence peak date (5/04). Orange curve: fitted curve using exponential growth model.

Simulated Data

The initial growth observed in the real data was simulated for Michigan using *Flu MODELO 1.0*. Figure 6 (in Chapter III – Demonstration) shows an example of a simulated incidence trend of influenza-like-illness, which combines the new cases infected with pandemic and seasonal H1N1 flu. The growth rate is extracted from the first 14 days of the simulation.

Estimation of Growth Rate and Its Precision

The growth rate r_1 and precision η_1 of the ED constitutional growth phase is estimated using an exponential growth model:

$$c_t = c_0 e^{r_1 t} \varepsilon_{1,t} \qquad \text{Equation 6}$$

where c_t denotes the number of constitutional registrations on day t, and c_0 is a constant

denoting the number of constitutional registrations on day 0 (see Figure 7). $\varepsilon_{1,t}$ is the random error on day t. The precision of growth rate estimate (standard error), η_1, is calculated with the maximum likelihood method when r_1 is estimated using the model (Equation 6).

Likewise, exponential growth model was fitted to the simulated data by selecting the first 14 days to match real data. Let r_2 be the growth rate estimated from simulated data using Equation 7:

$$n_t = n_0 e^{r_2 t} \varepsilon_{2,t} \qquad \text{Equation 7}$$

where n_t is number of new infected cases simulated on day t, and n_0 is constant denoting the number of infected cases simulated on day 0. $\varepsilon_{2,t}$is the random error on day t. The precision of growth rate estimate (standard error), η_2, is calculated with the maximum likelihood method when r_2 is estimated using the model (Equation 7).

A General Linear Regression model was fitted using the Log link function was specified (see Equation 8 and Equation 9), which assumes that, c_t and n_t , were both assumed to follow a Poisson distribution.

$$LN(c_t) = LN(c_0) + r_1 t + LN(\varepsilon_{1,t}) \qquad \text{Equation 8}$$

$$LN(n_t) = LN(n_0) + r_2 t + LN(\varepsilon_{2,t}) \qquad \text{Equation 9}$$

Experimental Strategy

Response Surface Methodology (RSM)

Response Surface Methodology (RSM) involves estimating and analyzing the form and parameters of a function, relating a response (yield variable) to one or more factors (stimulus

variables) that are assumed to influence the response, and possibly determining an optimal factor combination. When RSM is applied to simulation, the resulting analytical response model is a metamodel of the simulation model, which in turn is a metamodel of a real-life system.

Application of simulation-based optimization with RSM has been used to propose optimal resource allocation strategies during infectious outbreaks (Kasaie & Kelton, 2013; Paleshi, Bae, Evans, Heragu, 2017; Venkatraman & Selvagopal, 2018). Application of RSM for parameter estimation would require less experiment replications and computing resources.

Single or multi-responses have been used to evaluate the processes of influenza transmission and progression. Derringer's desirability function (Derringer & Suich, 1980) was used for simultaneous evaluation of multiple criteria for decision making. It is an objective function that searches for a solution that is within upper and lower limits, which then returns a desirability score from 0 (completely undesirable) to 1 (completely desirable). Numerically, responses are fitted with target, maximal, and minimal values. A gradient search algorithm is used to find optimal solutions within a user-defined value search intervals. For multiple responses, there is a desirability function for each individual response. The overall desirability or composite desirability of all responses is calculated using the geometric mean of each desirability function. To set the importance of the response in the overall desirability, investigators can give a weight (a scaling factor) to each individual response. In Minitab 18, the weight ranges from 0 to 10, where 0 means the response is of the least importance, and 10 means it is of the most importance. To the best of my knowledge, no work has been done to estimate the underlying influenza epidemiological characteristics using response surface optimization with multiple criteria for decision making.

Flu MODELO 1.0 provided us with a testbed to simulate incidence data and extract the simulated exponential growth rate at different scenario. Response surface methodology (RSM) (Gunst, Myers, & Montgomery, 1996) with central composite design (CCD) and Derringer's Desirability Function were used to (Jeong & Kim, 2009) to perform the optimization for estimating epidemiological and mitigation factors.

The real exponential growth rate estimate and its standard error have already been estimated from the ERCR09 data. The study aims to minimize the difference between the real and simulated growth rate estimate and standard error by employing a multi-response optimization procedure.

The following parameters were chosen for estimation (see Table 19): PANDEMIC_RN, SEASONAL_RN, INI_PANDEMIC_INCIDENCE, INI_SEASONAL_INCIDENCE , WITHDRAWAL_PROB, and ANTIVIRAL_PERCENTAGE. Since SEASONAL VACCINATION and ABSENTEEISM were already in place once the pandemic was declared, they were set to "1" and "2" to activate both mitigation alternatives. Other parameters use the default values in Table 17 (presented in Chapter III).

Note: $^{R+}$ positive real numbers, $^{I+}$ positive integers.

RSM and central composite design were employed to build the metamodel for growth rate mean r_2 and estimate precision η_2, respectively. The metamodel is a multiple linear regression model explicated with second-order polynomial equations:

$$r_2 = \beta_{1,0} + \sum_{i=1}^{k} \beta_{1,i} x_i + \sum_{j=i+1}^{k} \sum_{i=1}^{k-1} \beta_{1,ij} x_i x_j \qquad \text{Equation 10,}$$

$$\eta_2 = \beta_{2,0} + \sum_{i=1}^{k} \beta_{2,i} x_i + \sum_{j=i+1}^{k} \sum_{i=1}^{k-1} \beta_{2,ij} x_i x_j \qquad \text{Equation 11,}$$

Table 19 *Parameters As Predictor Variables*

Parameter	Description	Feasible Sample Space
Epidemiological Parameter		
PANDEMIC_RN	base reproduction number for pandemic influenza	$1.6\text{-}2.1^{R+}$
SEASONAL_RN	base reproduction number for seasonal influenza	$1.1\text{-}1.6^{R+}$
INI_PANDEMIC_INCIDENCE	number of initial infectiousness with pandemic virus	$30\text{-}1000^{I+}$
INI_SEASONAL_INCIDENCE	number of initial infectiousness with seasonal virus	$30\text{-}1000^{I+}$
Mitigation parameter		
WITHDRAWAL_PROB	probability that a person is withdrawn from regular activities	$0.00001\text{-}0.99999^{R+}$
ANTIVIRAL_PERCENTAGE	the proportion of vulnerable people receiving antiviral treatment	$0.00001\text{-}0.99999^{R+}$

where x_1 through x_k are k distinct input variables (see Table 19), x_i and x_j's are the i^{th} and j^{th} variables, β_1's are parameter coefficients for growth rate estimate r_2, and β_2's are parameter coefficients for estimate precision η_2. In this study, significance level 0.05 was used as significance level for all statistical analysis.

In the following sections, **Research question 1** with the corresponding four hypotheses are answered: Does the observed real-time trend of influenza incidence exclusively capture the disease behavior in presence of mitigation or is it also affected by other factors (e.g., public communication about the outbreak status)?

Hypothesis 1.

H_0: There is no main effect of the predictor variables (Table 18) on the real growth rate estimate r_2 and/or precision of growth rate estimate η_2.

H_1: There is main effect of the predictor variables (Table 18) on the real growth rate estimate r_2 and/or precision of growth rate estimate η_2.

Hypothesis 2.

H_0: There is no second-order interaction effect of the predictor variables (Table 18) on the real growth rate estimate r_2 and/or precision of growth rate estimate η_2.

H_1: There is second-order interaction effect of the predictor variables (Table 18) on the real growth rate estimater$_2$ and/or precision of growth rate estimate η_2.

Hypothesis 3.

H_0: There is no quadratic effect of the predictor variables (Table 18) on the real growth rate estimate r_2 and/or precision of growth rate estimate η_2.

H_1: There is quadratic effect of the predictor variables (Table 18) on the real growth rate estimater$_2$ and/or precision of growth rate estimate η_2.

Hypothesis 4.

H_0: There is no higher-order effect of the predictor variables (Table 18) on the r real growth rate estimate r_2 and/or precision of growth rate estimate η_2.

H_1: There is higher-order effect of the predictor variables (Table 18) on the real growth rate estimater$_2$ and/or precision of growth rate estimate η_2.

Experimental Design

Due to the large number of parameters in the simulation, it was inefficient to perform sensitivity analysis to estimate parameter values since sensitivity analysis tests one factor at a time (OFAT). A 2-level full fractional Central Composite Inscribed (CCI) design to was used to

investigate the impact of the predictor variables (Table 19) on the response variables and 2) to estimate a full second-order model with main, interaction and quadratic effects on response variables. Within the feasible range of each continuous variables, the factors' axial points as well as center points are decided as in Table 20. To choose appropriate factorial levels, $\alpha = 2.828$ was selected to achieve both rotatability and orthogonality design with a single block (i.e., minimize the variance of the linear model coefficients) (Myers, Vining, Giovannitti-Jensen, & Myers, 1992). Table 21 shows the Full CCI design matrix obtained through the software that includes 90 runs in total, 64 factorial points, 14 center points and 12 axial points. Simulated results of growth rates and standard errors of growth rates (precision) were appended at the columns "Simulated r_2" and "Simulated η_2" in Table 21.

Table 20 *Factor Levels with Full CCI Design for Estimation*

Predictor Variables	Name Code	Factorial Points uncoded value (coded value)		Axial Points uncoded value (coded value)		Center Point uncoded value (coded value)
PANDEMIC_RN	A	1.9384 (0.35355)	1.7616 (-0.35355)	2.1 (1)	1.6 (-1)	1.85 (0)
SEASONAL_RN	B	1.4384 (0.35355)	1.2616 (-0.35355)	1.6 (1)	1.1 (-1)	1.35 (0)
INI_PANDEMIC_INCIDENCE	C	686 (0.35355)	344 (-0.35355)	1000 (1)	30 (-1)	515 (0)
INI_SEASONAL_INCIDENCE	D	686 (0.35355)	344 (-0.35355)	1000 (1)	30 (-1)	515 (0)
WITHDRAWAL_PROB	E	0.67674 (0.35355)	0.32326 (-0.35355)	0.99999 (1)	0.00001 (-1)	0.5 (0)
ANTIVIRAL_PERCENTAGE	F	0.67674 (0.35355)	0.32326 (-0.35355)	0.99999 (1)	0.00001 (-1)	0.5 (0)

Table 21 *Design Matrix Using CCI for Parameter Estimation (Coded Matrix)*

Std. Order	Run Order	Point Type	A	B	C	D	E	F	Sim r_2	Sim η_2	Pred r_2	Pred η_2
1	30	1	-0.35355	-0.35355	-0.35355	-0.35355	-0.35355	-0.35355	0.01159526	0.00365445	0.0077103	0.0037018
2	76	1	0.35355	-0.35355	-0.35355	-0.35355	-0.35355	-0.35355	0.03856182	0.00342469 1	0.0405584	0.0033911
3	29	1	-0.35355	0.35355	-0.35355	-0.35355	-0.35355	-0.35355	0.01530713	0.00351263 1	0.0124884	0.0036189
4	10	1	0.35355	0.35355	-0.35355	-0.35355	-0.35355	-0.35355	0.03922272	0.00335206 2	0.0453365	0.0033151
5	46	1	-0.35355	-0.35355	0.35355	-0.35355	-0.35355	-0.35355	0.03264244	0.00279874	0.0292151	0.0028048
6	89	1	0.35355	-0.35355	0.35355	-0.35355	-0.35355	-0.35355	0.06286425	0.00257129 2	0.0620632	0.0025694
7	15	1	-0.35355	0.35355	0.35355	-0.35355	-0.35355	-0.35355	0.02665668	0.00283782 6	0.0339933	0.0027420
8	58	1	0.35355	0.35355	0.35355	-0.35355	-0.35355	-0.35355	0.07715996	0.00244974 2	0.0668414	0.0025118
9	31	1	-0.35355	-0.35355	-0.35355	0.35355	-0.35355	-0.35355	-0.02871957	0.00339548 5	-0.0209829	0.0033386
10	3	1	0.35355	-0.35355	-0.35355	0.35355	-0.35355	-0.35355	0.00638836 8	0.00309521 1	0.0118652	0.0030583
11	25	1	-0.35355	0.35355	-0.35355	0.35355	-0.35355	-0.35355	-0.01458508	0.00324620 4	-0.0162047	0.0032638
12	14	1	0.35355	0.35355	-0.35355	0.35355	-0.35355	-0.35355	0.02563591	0.00290924 4	0.0166434	0.0029899
13	43	1	-0.35355	-0.35355	0.35355	0.35355	-0.35355	-0.35355	0.01107608	0.00257599 3	0.0054542	0.0026426

Table 21 – continued

14	71	1	0.35355	-0.35355	0.35355	0.35355	-0.35355	-0.35355	0.03624097	0.002486559	0.0383023	0.0024208
15	40	1	-0.35355	0.35355	0.35355	0.35355	-0.35355	-0.35355	0.009388833	0.002555948	0.0102323	0.0025834
16	75	1	0.35355	0.35355	0.35355	0.35355	-0.35355	-0.35355	0.03991465	0.002389560	0.0430804	0.0023666
17	1	1	-0.35355	-0.35355	-0.35355	-0.35355	0.35355	-0.35355	-0.01088421	0.003870794	-0.0081875	0.0038654
18	39	1	0.35355	-0.35355	-0.35355	-0.35355	0.35355	-0.35355	0.0201319	0.003573238	0.0246606	0.0035409
19	88	1	-0.35355	0.35355	-0.35355	-0.35355	0.35355	-0.35355	0.004842231	0.003685755	-0.0034093	0.0037788
20	21	1	0.35355	0.35355	-0.35355	-0.35355	0.35355	-0.35355	0.02259274	0.003539419	0.0294388	0.0034617
21	23	1	-0.35355	-0.35355	0.35355	-0.35355	0.35355	-0.35355	0.01378574	0.002877518	0.0133173	0.0029288
22	62	1	0.35355	-0.35355	0.35355	-0.35355	0.35355	-0.35355	0.04800131	0.002618473	0.0461654	0.0026829
23	49	1	-0.35355	0.35355	0.35355	-0.35355	0.35355	-0.35355	0.01460808	0.002927412	0.0180955	0.0028632
24	59	1	0.35355	0.35355	0.35355	-0.35355	0.35355	-0.35355	0.04978878	0.002659512	0.0509436	0.0026229
25	65	1	-0.35355	-0.35355	-0.35355	0.35355	0.35355	-0.35355	-0.04100662	0.003534087	-0.0368806	0.0034861
26	19	1	0.35355	-0.35355	-0.35355	0.35355	0.35355	-0.35355	-0.00509196	0.003199508	-0.0040325	0.0031935
27	48	1	-0.35355	0.35355	-0.35355	0.35355	0.35355	-0.35355	-0.03113534	0.003426345	-0.0321025	0.0034081

Table 21 – continued

28	5	1	0.35355	0.35355	-0.35355	0.35355	0.35355	-0.35355	0.00136063	0.003090052	0.0007456	0.0031220
29	8	1	-0.35355	-0.35355	0.35355	0.35355	0.35355	-0.35355	-0.00925920 5	0.00277797 3	-0.0104436	0.0027594
30	51	1	0.35355	-0.35355	0.35355	0.35355	0.35355	-0.35355	0.02721676	0.00251117 9	0.0224045	0.0025278
31	60	1	-0.35355	0.35355	0.35355	0.35355	0.35355	-0.35355	-0.00891951 5	0.00270627 1	-0.0056654	0.0026976
32	70	1	0.35355	0.35355	0.35355	0.35355	0.35355	-0.35355	0.02333 23	0.00248250 8	0.0271827	0.0024712
33	73	1	-0.35355	-0.35355	-0.35355	-0.35355	-0.35355	0.35355	-0.02537337	0.00413327	-0.0190267	0.0040171
34	68	1	0.35355	-0.35355	-0.35355	-0.35355	-0.35355	0.35355	0.02434617	0.00358990 3	0.0138214	0.0036800
35	82	1	-0.35355	0.35355	-0.35355	-0.35355	-0.35355	0.35355	-0.00829877 9	0.00385907 2	-0.0142486	0.0039272
36	42	1	0.35355	0.35355	-0.35355	-0.35355	-0.35355	0.35355	0.01679158	0.00362812 4	0.0185995	0.0035976
37	17	1	-0.35355	-0.35355	0.35355	-0.35355	-0.35355	0.35355	-0.00379109	0.00313407 1	0.0024781	0.0030437
38	13	1	0.35355	-0.35355	0.35355	-0.35355	-0.35355	0.35355	0.03410381	0.00276511 9	0.0353262	0.0027883
39	79	1	-0.35355	0.35355	0.35355	-0.35355	-0.35355	0.35355	0.00177246 1	0.00306 24	0.0072563	0.0029756
40	16	1	0.35355	0.35355	0.35355	-0.35355	-0.35355	0.35355	0.03551709	0.00270867 4	0.0401044	0.0027258

Table 21 – continued

41	35	1	-0.35355	-0.35355	-0.35355	0.35355	-0.35355	0.35355	-0.04313974	0.003554611	-0.0477199	0.0036230
42	22	1	0.35355	-0.35355	-0.35355	0.35355	-0.35355	0.35355	-0.02563508	0.003422075	-0.0148718	0.0033189
43	27	1	-0.35355	0.35355	-0.35355	0.35355	-0.35355	0.35355	-0.04160196	0.00353423	-0.0429417	0.0035419
44	55	1	0.35355	0.35355	-0.35355	0.35355	-0.35355	0.35355	-0.000993099 5	0.00325749	-0.0100936	0.0032446
45	38	1	-0.35355	-0.35355	0.35355	0.35355	-0.35355	0.35355	-0.01529 88	0.002825323	-0.0212828	0.0028677
46	7	1	0.35355	-0.35355	0.35355	0.35355	-0.35355	0.35355	0.01396206	0.002614924	0.0115653	0.0026270
47	87	1	-0.35355	0.35355	0.35355	0.35355	-0.35355	0.35355	-0.01304767	0.002757787	-0.0165047	0.0028035
48	52	1	0.35355	0.35355	0.35355	0.35355	-0.35355	0.35355	0.02066797	0.002549553	0.0163434	0.0025682
49	61	1	-0.35355	-0.35355	-0.35355	-0.35355	0.35355	0.35355	-0.03766949	0.004305014	-0.0349245	0.0041947
50	18	1	0.35355	-0.35355	-0.35355	-0.35355	0.35355	0.35355	-0.000777406	0.003811956	-0.0020764	0.0038426
51	28	1	-0.35355	0.35355	-0.35355	-0.35355	0.35355	0.35355	-0.03308 66	0.004200165	-0.0301463	0.0041008
52	20	1	0.35355	0.35355	-0.35355	-0.35355	0.35355	0.35355	0.005987356	0.00367263	0.0027018	0.0037566
53	47	1	-0.35355	-0.35355	0.35355	-0.35355	0.35355	0.35355	-0.00701424 3	0.003118476	-0.0134197	0.0031783

Table 21 – continued

54	83	1	0.35355	-0.35355	0.35355	-0.35355	0.35355	0.35355	0.02562307	0.00284447	0.0194284	0.0029115
55	37	1	-0.35355	0.35355	0.35355	-0.35355	0.35355	0.35355	-0.01173438	0.00308447	-0.0086415	0.0031071
56	33	1	0.35355	0.35355	0.35355	-0.35355	0.35355	0.35355	0.02738972	0.00285832	0.0242066	0.0028463
57	9	1	-0.35355	-0.35355	-0.35355	0.35355	0.35355	0.35355	-0.05995864	0.00377388	-0.0636176	0.0037832
58	50	1	0.35355	-0.35355	-0.35355	0.35355	0.35355	0.35355	-0.02516215	0.00337478	-0.0307695	0.0034656
59	86	1	-0.35355	0.35355	-0.35355	0.35355	0.35355	0.35355	-0.05398995	0.00365098	-0.0588395	0.0036984
60	4	1	0.35355	0.35355	-0.35355	0.35355	0.35355	0.35355	-0.02922217	0.00346290	-0.0259914	0.0033880
61	45	1	-0.35355	-0.35355	0.35355	0.35355	0.35355	0.35355	-0.03433759	0.00296280	-0.0371806	0.0029945
62	53	1	0.35355	-0.35355	0.35355	0.35355	0.35355	0.35355	-0.01429684	0.00280501	-0.0043325	0.0027431
63	11	1	-0.35355	0.35355	0.35355	0.35355	0.35355	0.35355	-0.03304229	0.00293435	-0.0324024	0.0029274
64	34	1	0.35355	0.35355	0.35355	0.35355	0.35355	0.35355	-0.00639091	0.00273851	0.0004457	0.0026817
65	67	-1	-1	0	0	0	0	0	-0.04507972	0.00348491	-0.0404375	0.0034640
66	72	-1	1	0	0	0	0	0	0.05512134	0.00270259	0.0524710	0.0027032

Table 21 – continued

67	24	-1	0	-1	0	0	0	0	-0.00270516	0.00315397	-0.0007406	0.0031596
68	32	-1	0	1	0	0	0	0	0.01402772	0.00292697	0.0127741	0.0029636
69*	6	-1	0	0	-1	0	0	0	-0.1051691	0.005926885	NA	NA
70*	44	-1	0	0	1	0	0	0	0.01942918	0.00238575	NA	NA
71*	64	-1	0	0	0	-1	0	0	0.04537716	0.003554817	NA	NA
72*	54	-1	0	0	0	1	0	0	-0.001068 11	0.002686992	NA	NA
73	66	-1	0	0	0	0	-1	0	0.02755817	0.002877665	0.0284996	0.0028785
74	85	-1	0	0	0	0	1	0	-0.01387732	0.00323101	-0.0164661	0.0032531
75	90	-1	0	0	0	0	0	-1	0.04933736	0.002706725	0.0438285	0.0027259
76	77	-1	0	0	0	0	0	1	-0.03343374	0.003418195	-0.0317951	0.0034351
77	84	0	0	0	0	0	0	0	0.00837704	0.003013315	0.0060167	0.0030600
78	81	0	0	0	0	0	0	0	0.005228641	0.003073793	0.0060167	0.0030600
79	80	0	0	0	0	0	0	0	-0.00015937 6	0.003107918	0.0060167	0.0030600
80	63	0	0	0	0	0	0	0	0.004946115	0.003046515	0.0060167	0.0030600

Table 21 – continued

81	78	0	0	0	0	0	0	0	0.00284749 1	0.00308608	0.0060167	0.0030600
82	2	0	0	0	0	0	0	0	0.00527319 8	0.00306207 2	0.0060167	0.0030600
83	26	0	0	0	0	0	0	0	0.01002004	0.00299942 4	0.0060167	0.0030600
84	69	0	0	0	0	0	0	0	-0.00152618 2	0.00320046 7	0.0060167	0.0030600
85	57	0	0	0	0	0	0	0	0.01530 82	0.00305012 4	0.0060167	0.0030600
86	56	0	0	0	0	0	0	0	0.01346275	0.00301274 2	0.0060167	0.0030600
87	36	0	0	0	0	0	0	0	-0.00072683 4	0.00311306 1	0.0060167	0.0030600
88	74	0	0	0	0	0	0	0	0.00733566 1	0.00304008 7	0.0060167	0.0030600
89	12	0	0	0	0	0	0	0	0.01062752	0.00301885	0.0060167	0.0030600
90	41	0	0	0	0	0	0	0	0.00040047 8	0.00310645 8	0.0060167	0.0030600

Note: "*" means outlier observations that were not included in the prediction model.
Point type: "1" – Factorial Points, "-1" – Axial Points, "0" – Center Points.

Results

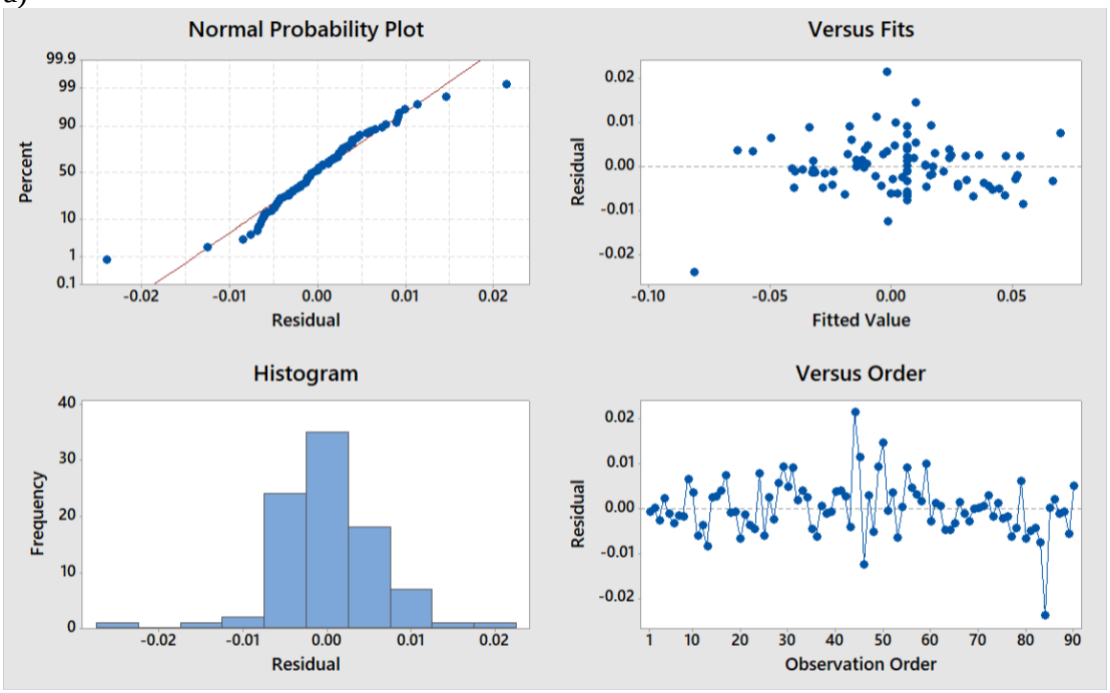
Model Diagnostics

Analysis of variance (ANOVA) was used to test the influence of predictor variables on the responsive growth rate estimate and estimate precision based on the simulation testbed. Significant factors for both models are found in Figure 10a and Figure 11a. Assumptions of normality, linearity, homoscedasticity and independency were checked with the fitted model residuals (see Figure 8a and Figure 9a). The predicted model for growth rate estimate met the assumption well except for a few outliers, but not the predicted model for the estimation precision. Therefore, a natural "e" based log transformation was made on η_2, to improve the model prediction accuracy (Figure 9b shows the model diagnostics), without changing the significant factors (Figure 11b).

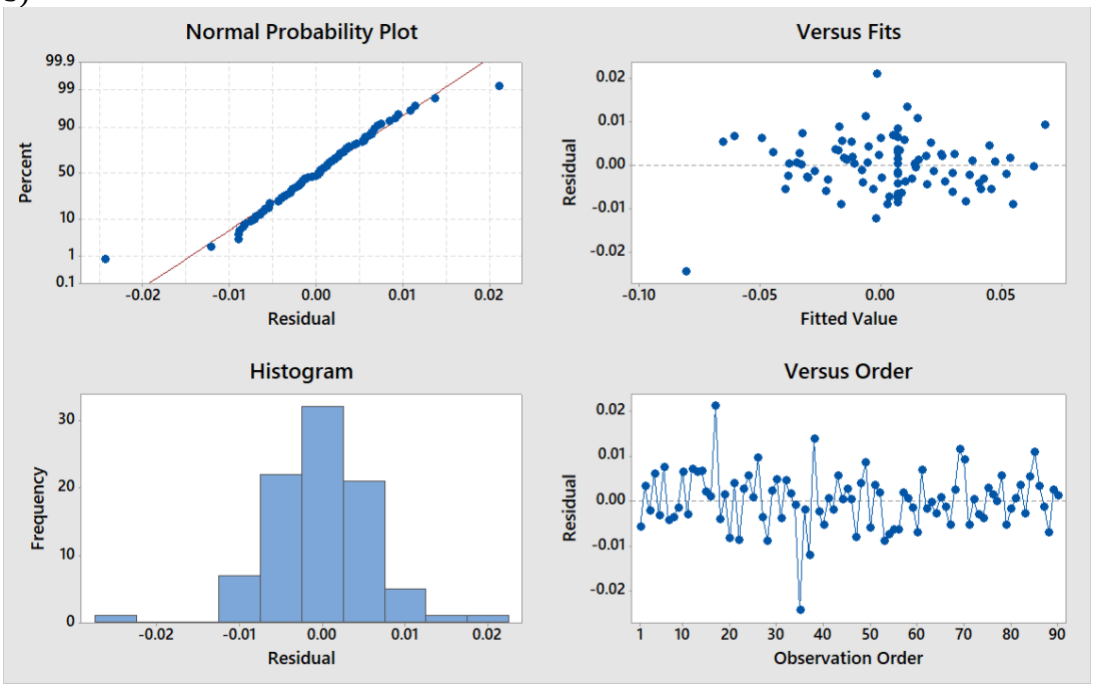
After reducing insignificant terms from the models but keeping terms if they were significant in one of the models (see Figure 10b and Figure 11c), both models turn out to be with outliers at the same points (see Figure 8b, Figure 9c and Table 22). Point 69 and 70 are axial points (± 1) of initial number of pandemic influenza incidences and point 71 and 72 are axial points (± 1) of initial number of seasonal influenza incidences. The outliers were detected by using Cook's Distance (D). The value 0.1 was set as the threshold for GR Estimate (r_2) and 0.05 for Precision LN(η_2), for most points are much smaller. Cook's D calculation of observation points greater than the threshold value were considered outliers. The existence of these outliers could be due to the simulation instability when using too low/high initial number of incidence values which have been explained in the baseline model \cite{prieto2016operational}. And the

model coefficients were estimated using maximum likelihood method, keeping outliers would distort the model estimation. These four outliers were then removed. The remaining 86 experimental runs could improve the prediction of the response variables within factorial levels of initial number of influenza incidences (353, 706). The model diagnostic of the final model is shown in Figure 8c and Figure 9d. The standardized effects of the final models are shown in Figure 10c and Figure 11d.

a)



b)



c)

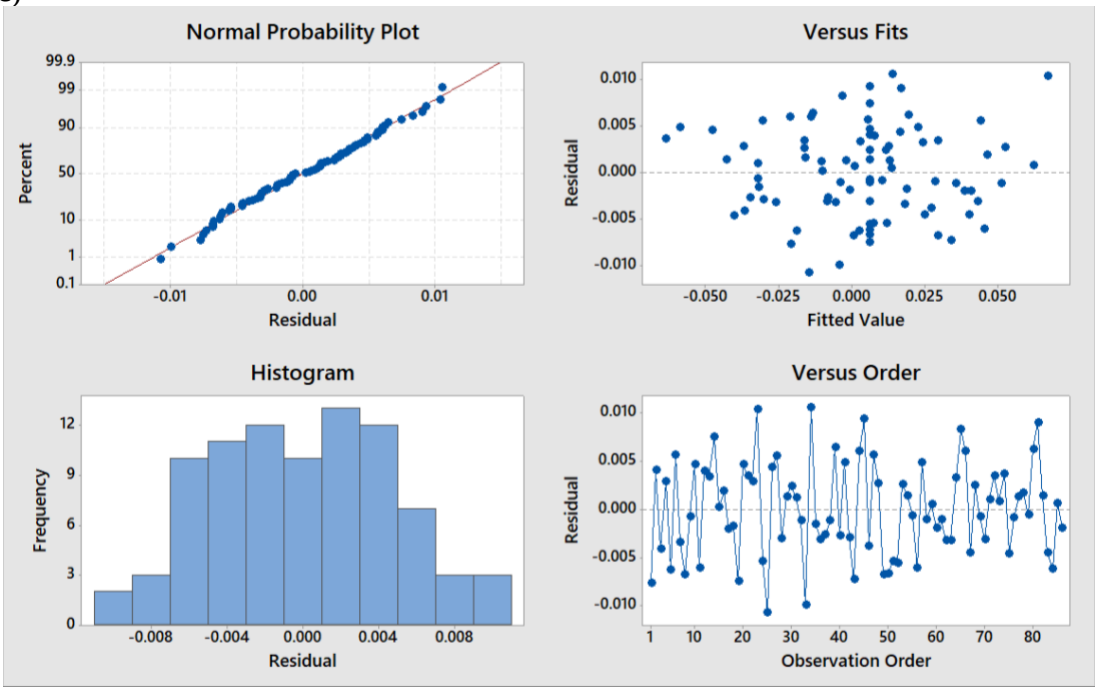
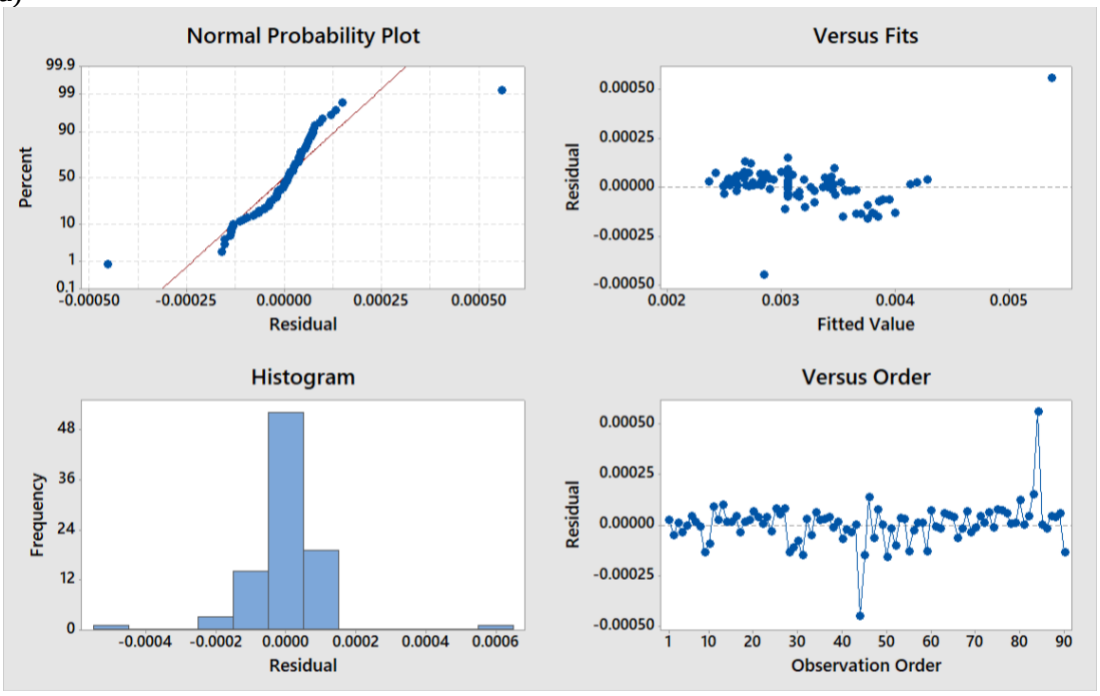
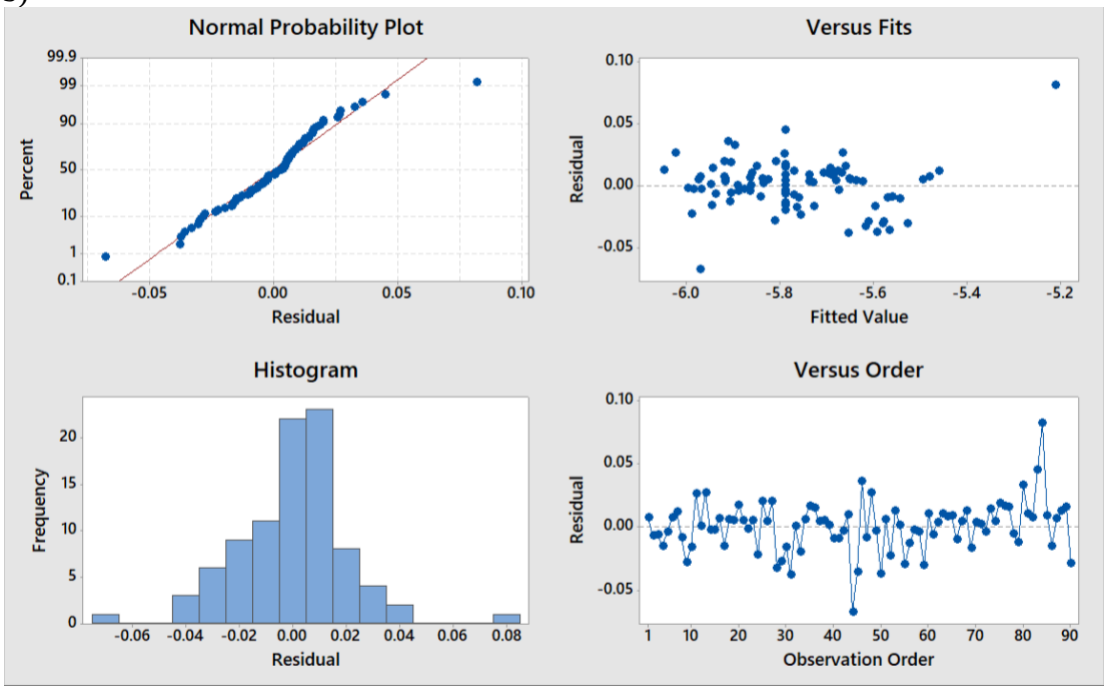


Figure 8. Model Fitting Diagnostics of Growth Rate Estimates. a) Full Model; b) Reduced Model; c) Final Model with Outliers Removed

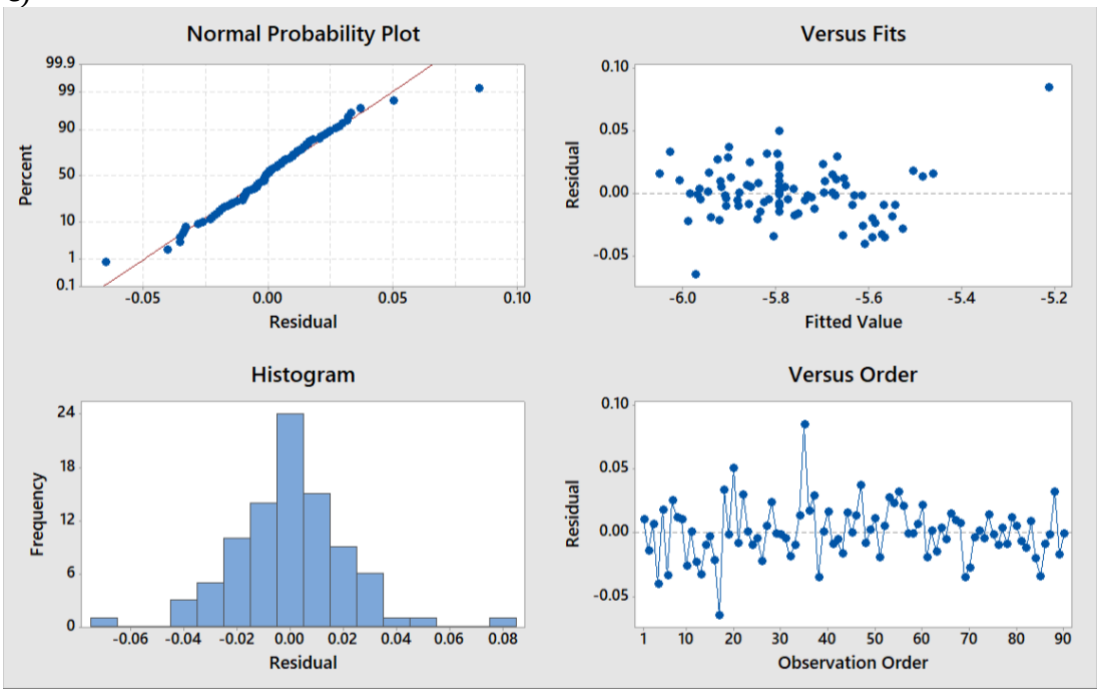
a)


Normal Probability Plot
Percent
99.9
99
90
50
10
1
0.1
-0.00050
-0.00025
0.00000
0.00025
0.00050
Residual
Versus Fits
Residual
0.00050
0.00025
0.00000
-0.00025
-0.00050
0.002
0.003
0.004
0.005
Fitted Value
Histogram
Frequency
48
36
24
12
0
-0.0004
-0.0002
0.0000
0.0002
0.0004
0.0006
Residual
Versus Order
Residual
0.00050
0.00025
0.00000
-0.00025
-0.00050
1
10
20
30
40
50
60
70
80
90
Observation Order

b)


Normal Probability Plot
Percent
99.9
99
90
50
10
1
0.1
-0.05
0.00
0.05
0.10
Residual
Versus Fits
Residual
0.10
0.05
0.00
-0.05
-6.0
-5.8
-5.6
-5.4
-5.2
Fitted Value
Histogram
Frequency
20
15
10
5
0
-0.06
-0.04
-0.02
0.00
0.02
0.04
0.06
0.08
Residual
Versus Order
Residual
0.10
0.05
0.00
-0.05
1
10
20
30
40
50
60
70
80
90
Observation Order

c)



d)

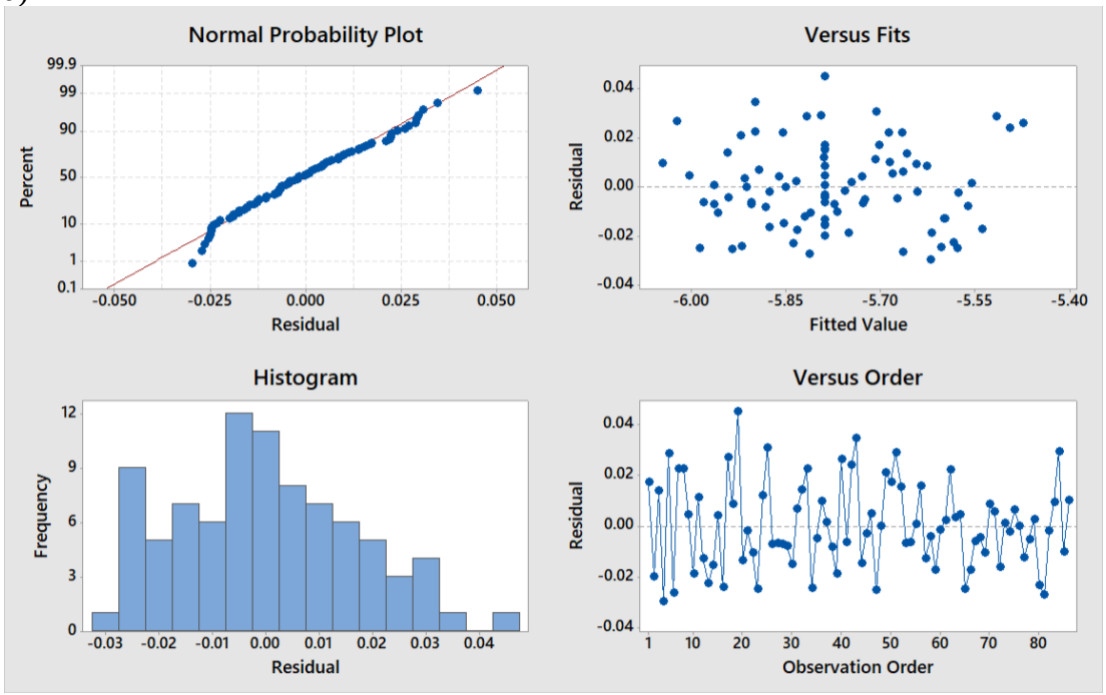
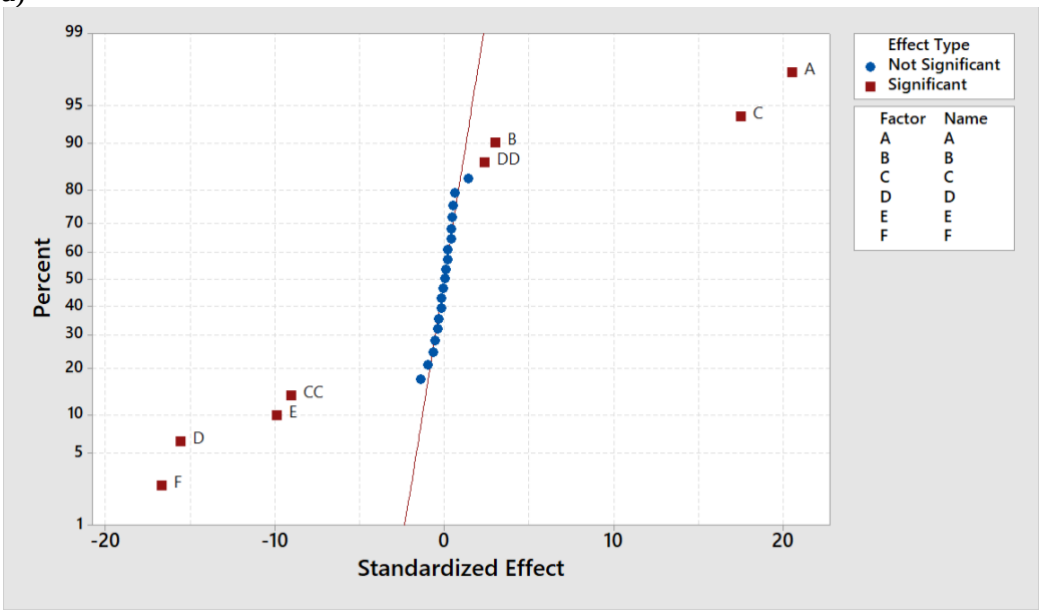
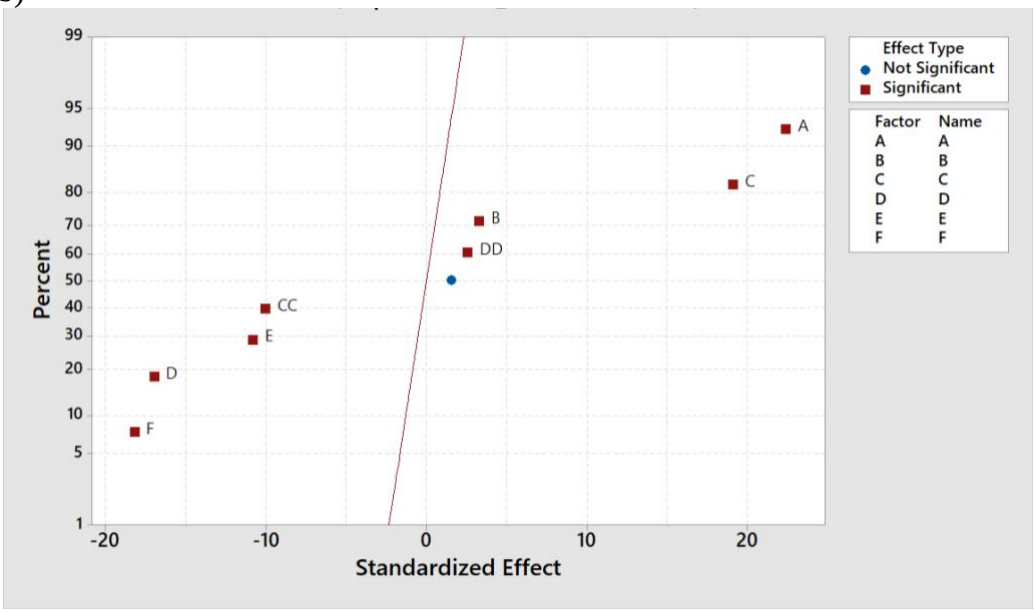


Figure 9. Model Fitting Diagnostics of Standard Errors of Growth Rate Estimates. a) Full Model. b) Log-transformed Standard Errors of Growth Rate Estimates. c) Reduced Model After Log-transformed Standard Errors of Growth Rate Estimates. d) Final Model after Log-transformed Standard Errors of Growth Rate Estimates with Removed Outliers

a)


Effect Type
Not Significant
Significant
Factor Name
A A
B B
C C
D D
E E
F F
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
A
C
B
DD
CC
E
D
F
-20
-10
0
10
20
Standardized Effect

b)


Effect Type
Not Significant
Significant
Factor Name
A A
B B
C C
D D
E E
F F
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
A
C
B
DD
CC
E
D
F
-20
-10
0
10
20
Standardized Effect

c)

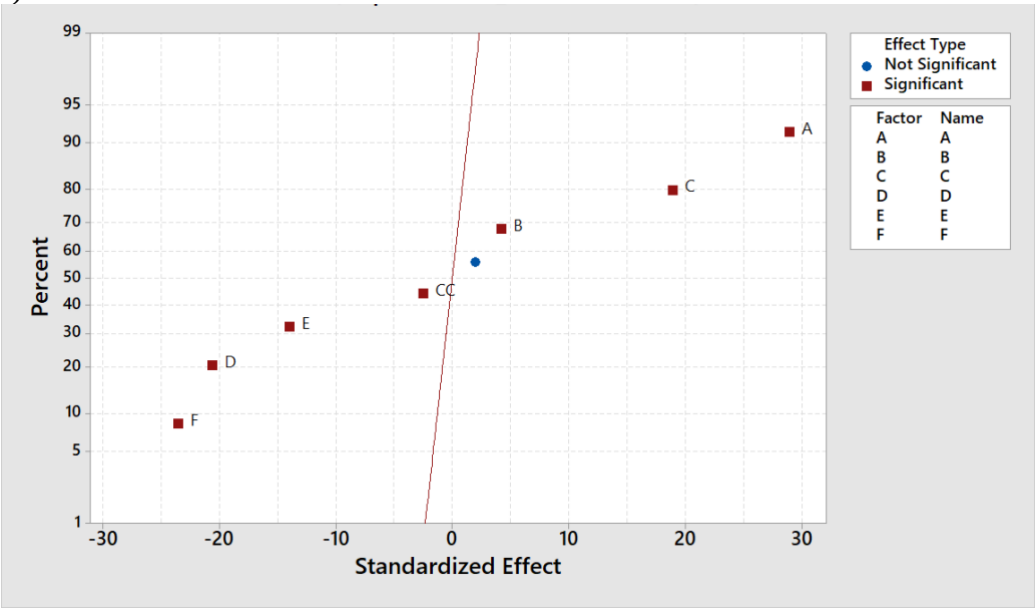
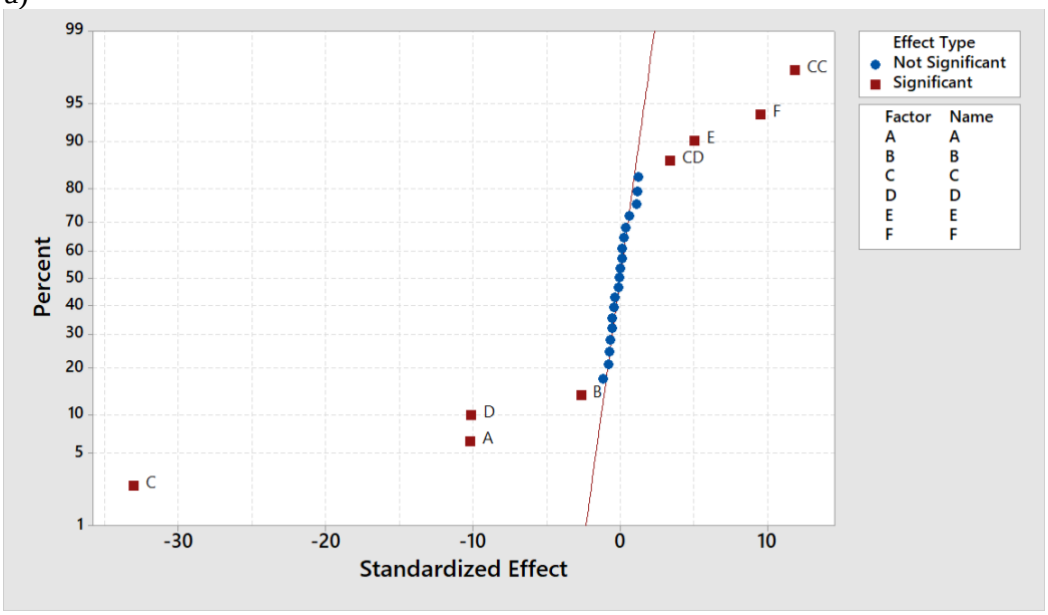
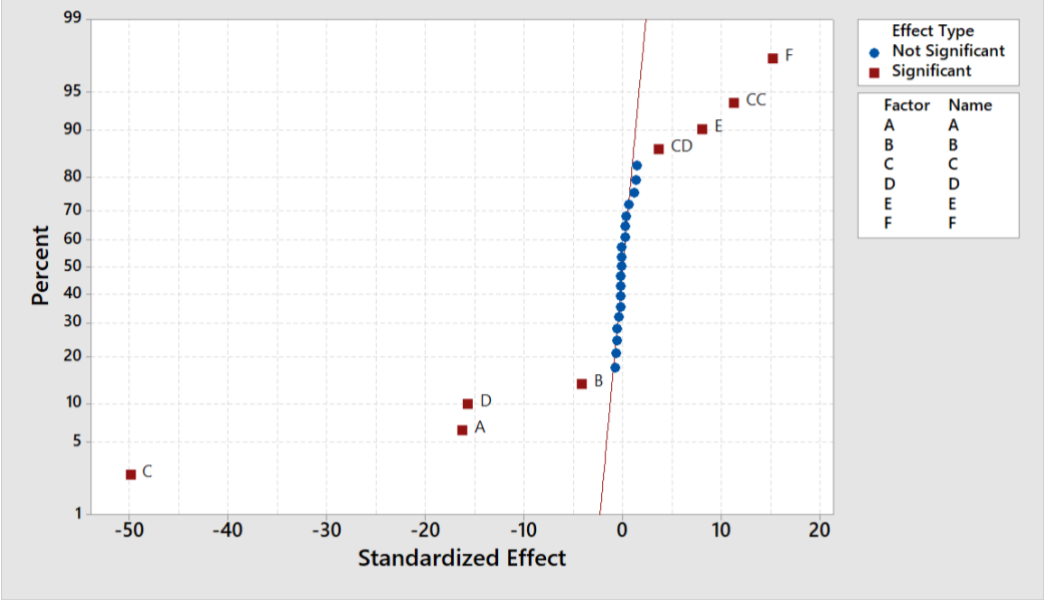


Figure 10. Standardized Effects on Growth Rate Estimate. a) Full Model. b) Reduced Model. c) Final Model with Outliers Removed

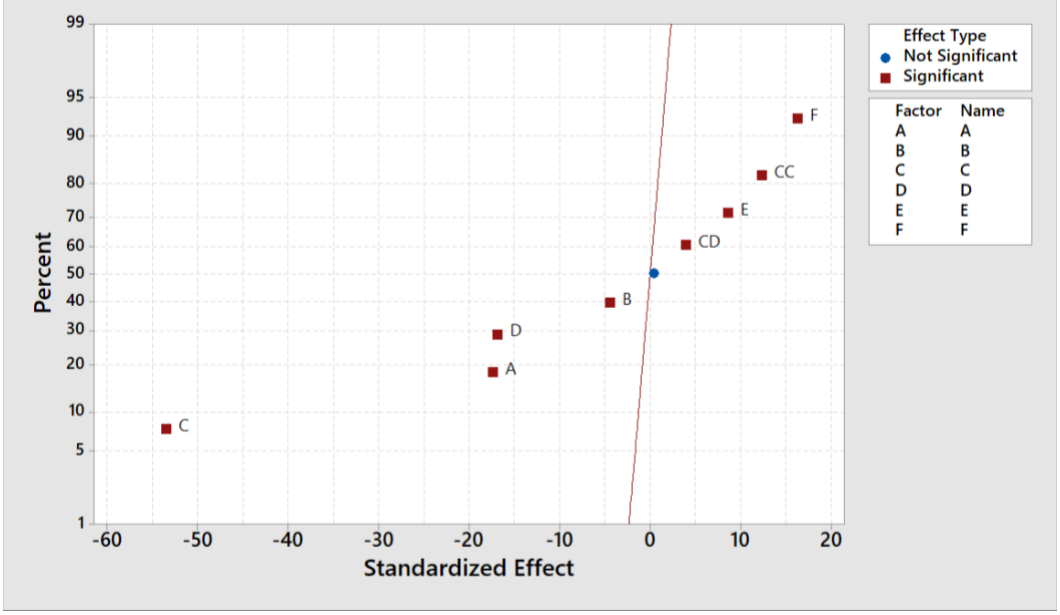
a)



b)


Percent
Standardized Effect
Effect Type
Not Significant
Significant
Factor Name
A A
B B
C C
D D
E E
F F
F
CC
E
CD
B
D
A
C

c)


Percent
Standardized Effect
Effect Type
Not Significant
Significant
Factor Name
A A
B B
C C
D D
E E
F F
F
CC
E
CD
B
D
A
C

d)

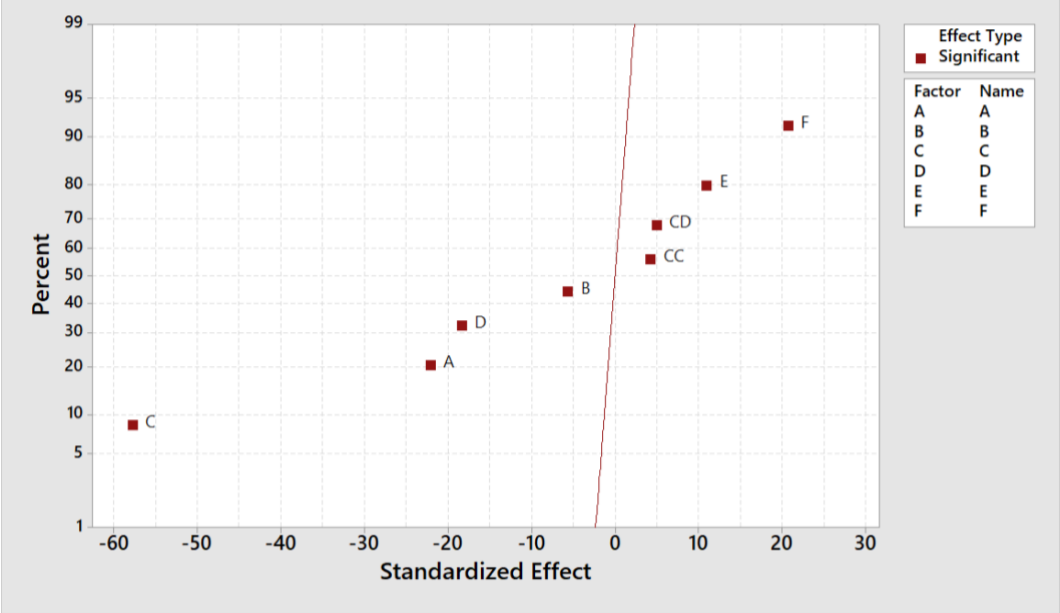


Figure 11. Standard Effects on Standard Error of Growth Rate Estimate. a) Full Model. b) Log-transformed Model. c) Reduced Model After Log-transformed d) Final Model After Log-transformed with Outliers Removed.

Table 22 *Diagnostics for Outlier Observations Using Standardized Residuals and Cook's Distance (D)*

	Growth Rate Estimate			Precision of Growth Rate Estimate		
Obs	Simulated	Fit	Cook's D	Simulated	Fit	Cook's D
69	-0.10517	-0.08077	3.33831	-5.12826	-5.21262	3.40664
70	0.01943	-0.00162	2.48444	-6.03824	-5.97306	2.03334
71	0.04538	0.05430	0.44677	-5.63945	-5.66888	0.41456
72	-0.01068	-0.01626	0.17432	-5.91933	-5.90909	0.05022

Model selection criteria were based on the measure of lack-of-fit and adjusted R^2, and prediction R^2 (Table 23). Insignificant lack-of-fit ($p>0.05$) means that model is equipped with sufficient predictor terms, and the higher the better. Adjusted R^2 stands for the relative variance

being explained by the model and prediction R^2 is the variance explained when a next response to be predicted. Higher values of both R^2 are preferred. Here, the final models were selected for both response variables present insignificant and higher lack-of-fit values, higher adjusted R^2 and higher Prediction R^2. Results of the analysis using the final models will be discussed in the next section.

Table 23 *Model Comparison*

Response	Model	Lack-of-Fit	R-square (adjust)	R-square (pred)
Growth Rate Estimate	Full (see Figure 10a)	0.084	94.13%	87.55%
	Reduced (see Figure 10b)	0.171	95.07%	89.71%
	Final* (see Figure 10c)	0.647	96.57%	96.14%
Precision of Growth Rate Estimate	Full (see Figure 11a)	0.001	94.61%	86.12%
	Log (see Figure 11b)	0.067	97.48%	94.80%
	Reduced (see Figure 11c)	0.122	97.81%	95.60%
	Final* (see Figure 11d)	0.505	98.25%	98.03%

Note: * means the best model that was selected for further analysis.

Factors Affecting Growth Rate Estimate

Figure 10c shows that all main factors have statistically significant impact on growth rate estimate. The only significant quadratic effect on the growth rate estimate is INI_PANDEMIC_INCIDENCE. No interaction effects have been found to be significant. ANOVA results suggest that the resulted regression model is significant ($P<0.000$). The Lack-of-Fit term is not significant ($P=0.647$) suggesting that including additional terms would not improve this model prediction. The goodness-of-fit (adjusted $R^2 = 96.14\%$) shows that this second-order model also explains most of the total variance.

Factors Affecting Precision of Growth Rate Estimate

It has been observed from the Figure 11d that all main factors have statistically significant impact on estimate precision of growth rate estimate. The interaction between INI_PANDEMIC_INCIDENCE and INI_SEASONAL_INCIDENCE had statistically significant impact on estimate precision of growth rate. Besides, the quadratic effect of INI_PANDEMICE_INCIDENCE had significant effect on estimate precision of growth rate. ANOVA results suggest that the resulted regression model is significant (P<0.000). The Lack-of-Fit term is not significant (P=0.505) suggesting that including additional terms would not improve this model prediction. The goodness-of-fit (adjusted R^2 = 98.03%) shows that this second-order model also explains most of the total variance.

Optimization Model

Although the interaction of INI_PANDEMICE_INCIDENCE and INI_SEASONAL_INCIDENCE had no statistically significant effect on the growth rate mean estimate r_2, this term was used for both optimization equations so all the model terms were consistently used. The resulted multiple responses surface models with coded predictor variable values are stated in the following equations (Equation 12, Equation 13):

$$r_2 = 0.00602 + 0.04645\,(\text{PANDEMIC\_RN}) + 0.00676\,(\text{SEASONAL\_RN}) + 0.03390\,(\text{INI\_PANDEMIC\_INCIDENCE}) - 0.03709\,(\text{INI\_SEASONAL\_INCIDENCE}) - 0.02248\,(\text{WITHDRAWAL\_PROB}) - 0.03781\,(\text{ANTIVIRAL\_PERCENTAGE}) - 0.0254\,(\text{INI\_PANDEMIC\_INCIDENCE})^2 + 0.00986\,(\text{INI\_PANDEMIC\_INCIDENCE})*(\text{INI\_SEASONAL\_INCIDENCE}) \quad \text{Equation 12}$$

$\eta_2 =$ -5.78933 - 0.12398 (PANDEMIC_RN) - 0.03202 (SEASONAL_RN) Equation 13
- 0.36153 (INI_PANDEMIC_INCIDENCE)
- 0.11514 (INI_SEASONAL_INCIDENCE)+ 0.06117 (WITHDRAWAL_PROB)
+ 0.11563 (ANTIVIRAL_PERCENTAGE)
+ 0.1462(INI_PANDEMIC_INCIDENCE)^2
+ 0.0874(INI_PANDEMIC_INCIDENCE)*(INI_SEASONAL_INCIDENCE)

Simulated VS Predicted

The resulted response surface optimization models (Equation 12 and Equation 13) were used to predict the growth rate estimates and precision of growth rate estimate except the outliers in observation 69-72 (see Table 21 Predicted r_2 and Predicted η_2). Then a paired-t test was used to compare the simulated responses with predicted responses. There is no sufficient evidence to conclude that the stimulated responses are statistically different from the predicted responses (r_2 Estimate: t(85)~0, p~1.0, η_2 Estimate: t(85)=0.08, p=0.934).

Optimal Solutions and Verification

With this multi-response optimization model, optimal scenarios were obtained with Desirability Function (Minitab) to achieve the minimal absolute difference between real growth rate estimate and simulated growth rate estimate (goal 1: set the target value of r_2 to 0.04447), and 2) achieve the minimal absolute difference of the precisions of the growth rate estimates between the real and simulated trends (goal 2: set the target value of η_2 to 0.002342). Therefore, in the Minitab Desirability Function, the two optimization goals were calculated simultaneously by setting the weight of goal 1) to 10, and setting the weight of goal 2) to 1 so that growth rate estimate could take more effect in the resulted overall desirability. The value search intervals of

both goals used the default settings by Minitab. As a result, two solutions turned out to have a 100% composite desirability and were selected as the optimal scenarios (Table 24).

Table 24 *Preliminary Optimal Solutions*

Solution	PANDEMIC _RN	SEASONAL _RN	INI_PANDEMIC _ INCIDENCE	INI_SEASONAL _INCIDENCE	WITHDRAWAL_ PROB	ANTIVIRAL_ PERCENTAGE
1	2.0965	1.6000	686	686	0.99999	0.34864
2	1.9606	1.6000	658	686	0.99999	0.00001

Both solutions were rational based on the understanding of the initial phase of influenza emergency. To verify the solutions, 50 simulation replicates were run for each solution. The mean and confidence intervals of the simulated growth rate estimates and SE growth rate estimates were calculated. Solution 1 (mean estimate of $r_2 = 0.03820$, CI = (0.03683, 0.03957)) and solution 2 (mean estimate of $r_2 = 0.04019$, CI = (0.03904, 0.04137)) turned out to have a statistically significantly lower estimate of growth rate than the targeted value (estimate of $r_1 = 0.04447$).

First, slight adjustment was made for solution 1. To eliminate the noise introduced by the adjustment, parameter ANTIVIRAL PERCENTAGE was tuned up leaving other parameters unchanged. Table 25 solution 1 shows the optimal solution after tuning this parameter. Although the mean SE of estimate of r_2 in solution 1 is still a little lower than targeted value (SE of estimate of $r_1 = 0.002342$), no further adjustment was made because 1) mean estimates of r_2 for both solutions have achieved the targeted value, 2) the estimate of r_2 has higher priority than SE, 3) the upper bound of mean SE of estimate of r_2 has been very closed to targeted value.

For solution 2 (Table 25), parameter ANTIVIRAL PERCENTAGE has already reached to minimum value, so tuning this parameter could not increase the growth rate estimate. No

further adjustment was made on other parameters with highest WITHDRAWAL PROBABILITY consistent in both solutions.

Therefore, this two optimal parameter sets were rational solutions to model parameters estimation, and this parameter set demonstrated that Flu MODELO 1.0 could generate the targeted growth rate estimate that was extracted from real data.

Table 25 *Final Optimal Solutions for Parameter Estimation*

Solution	PANDEMIC_RN	SEASONAL_RN	INI_PANDEMIC_INCIDENCE	INI_SEASONAL_INCIDENCE	WITHDRAWAL_PROB	ANTIVIRAL_PERCENTAGE	Mean estimate of r_2 (CI)	Mean SE of r_2 (CI)
1	2.0965	1.6	686	686	0.99999	0.26	0.0449 (0.0437, 0.046)	0.00232 (0.00231, 0.00233)
2	1.9606	1.6	658	686	0.99999	0.00001	0.0402 (0.039,0.0414)	0.00234 (0.00233, 0.00235)

Discussion

In this Chapter, two optimal scenarios for parameters in Table 19 were inferred from the ED constitutional data after the 2009 pandemic declaration. Both scenarios give us a picture of the plausible initial conditions of the pandemic outbreak in Michigan. High pandemic reproduction number, moderate seasonal reproduction number, high withdrawal percentage, and low antiviral usage coverage. An interesting result was obtained for the estimated initial pool of infected cases by the time of pandemic declaration: Over 650 pandemic cases and over 650 seasonal cases were infected at the beginning of the influenza emergency. These results indicate that disease progression should have started before the pandemic declaration. This behavior, however, is not observed in the ED trends which only show an exponential growth after pandemic declaration. Presumably the declaration influenced the reported number of ED

constitutional cases as a behavioral response due to fear of severe consequences from the disease. In light of this inference, the surge of ED visit counts seems to be not only related to the evolution of the disease itself, but also related with the timing of first media declaration.

Some literature discussed the relationship between media impact and the timing of flu declaration. “Early announcement” has been recommended by WHO as one of best practices for outbreak communication ("Outbreak Communication: Best practices for communicating with the public during an outbreak", 2004). In terms of outbreak communication, the media play an important role in uncovering the uncertainty (Glik, 2007) and broadcasting risk information in comprehensive ways (Berry, 2004). Once aware of possible risks, lay people may take preventive measures and seek health care as soon as there is onset of pandemic flu symptoms.

However, if the media communicates too early, it can be acting as a “crying wolf” if the outbreak is very mild to be alarmed (Bjørkdahl & Carlsen, 2017). As a result, the media tend to put off the pandemic announcement until it has been fully confirmed by health authorities. There is a long-standing need for public health care authorities, epidemiological experts and the news media to collaborate in such a way that they can declare the outbreak in an appropriate time.

It is important to understand what and how the epidemic messages spread to the public for each influenza emergency. The public should have enough awareness of the potential disease consequence, pandemic symptoms and outbreak dynamics. However, early declaration is especially difficult in situations where disease progression is uncertain (i.e. the number of infected cases with a particular virus is unknown). It is suggested the news media declare the pandemic at intervals (weekly updates or daily updates), and with neutral information about the pandemic. Besides, it is suggested that the news media transfer information about how to seek

medical help and the importance of seeing a doctor as soon as possible. To further mitigate the propagation of influenza, the declaration should not only serve as a crying wolf that only bring agony and fear to people, but encourage people to report their cases and get appropriate treatment.

CHAPTER V

EVALUATION OF DATA COLLECTION OPERATIONS FOR INFLUENZA SURVEILLANCE

In this chapter, a simulation-based evaluation of the surveillance operations for collection and testing of specimens during an emerging influenza in the State of Michigan is presented. This chapter aims to answer research questions 2, 3, and 4 (see Chapter I – Research Questions) by identifying the operational factors that significantly contribute to biases in the incidence trends and quantifying the explanatory power of the incidence trends.

Flu MODELO 1.0 (see Chapter III – Flu MODELO 1.0) was expanded in this chapter to allow further exploration of optimal operational conditions, which enhances understanding on existing sampling policies. The simulation is calibrated and analyzed with central composite designs and response surface optimization models for optimal and systematic exploration of the parameter space.

Figure 12 presents the main building blocks of the simulation model. Two co-circulating influenza viruses spread through a network of regions simulating human contact patterns in the State of Michigan (~10,000,000 inhabitants). The Michigan network connects its 4 characteristics regions: Upper Peninsula, Northern Lower Peninsula, Southeast Michigan Council of Governments, and Southern Lower Peninsula (see Figure 12a). Real data sources were used to build the geographic, and demographic features, as well as the travel patterns within

and between the regions, and across the Michigan boundaries (see Chapter III – Geographic, Demographic, Travel Patterns in Flu MODELO 1.0).

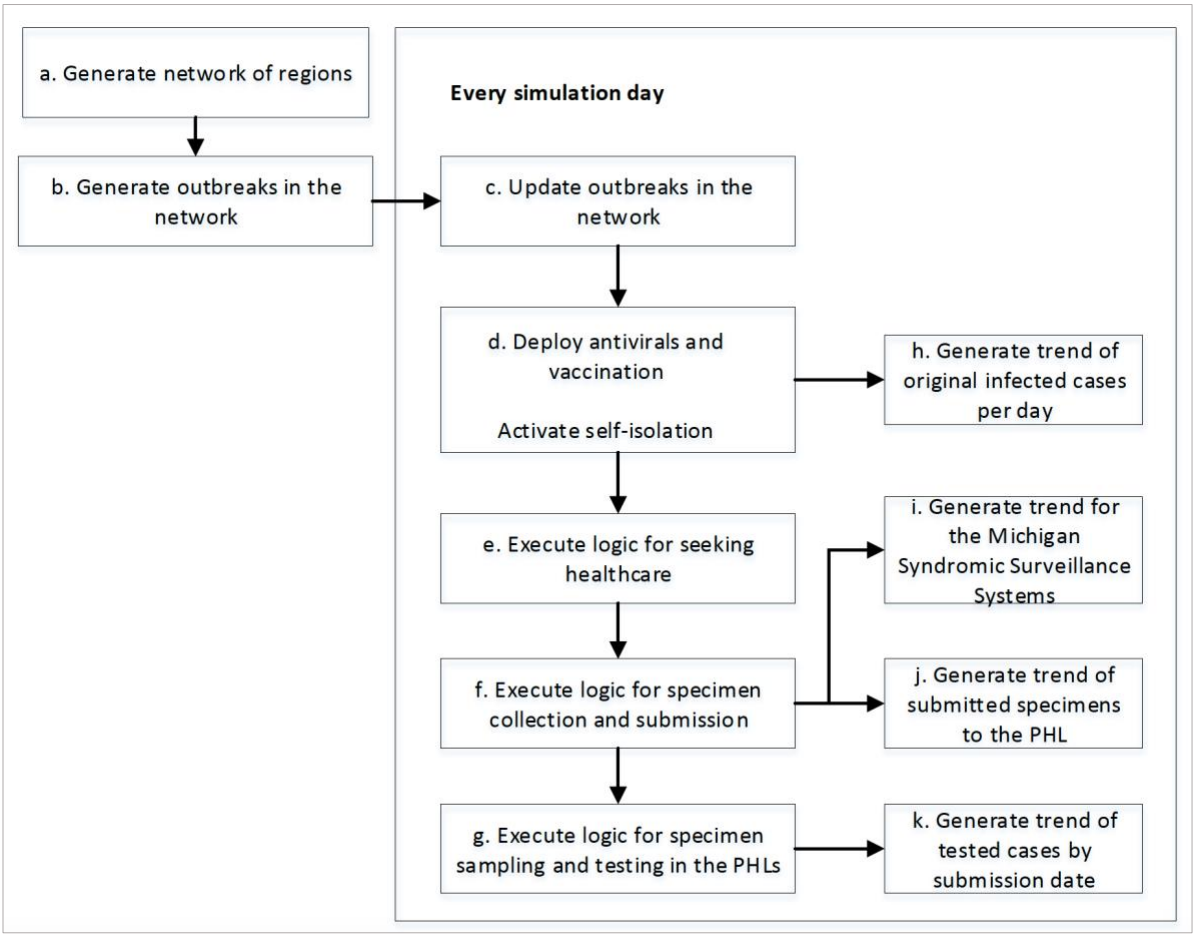


Figure 12. Overview of the simulation model. a) The State of Michigan is simulated by connecting 4 independent regional models through a network of travelers; b) Two influenza outbreaks are seeded in the network; c) Every day, the outbreaks are updated by adding new cases according to the infection probability model of the simulation; d) Vaccination, antivirals and self-isolation criteria are applied as in the beginning of a Pandemic outbreak. Case incidence per viral strain is collected after mitigation and self-isolation are imposed in box h; e) Symptomatic cases will seek healthcare; f) Cases going to the ED submit specimen data to the MSSS, generating daily incidence trends for constitutional and respiratory symptoms in box i. Also, during pandemic emergencies, any healthcare provider can submit to the PHLs, and the daily number of submitted cases is recorded in box j; g) Sampling and testing will take place in the PHL, and the daily number of tested cases by submission date is recorded in box k.

The two co-circulating viruses were seeded to recreate an emergent, high transmissibility virus that interacts with a low transmissibility strain already infecting the population (see Figure 12b). More specifically, the model used the reproduction number and viral shedding profiles of the pandemic and seasonal H1N1 2009 strains, where the seasonal strain represents the existing circulation of milder influenza viruses (Prieto & Das, 2016). This feature enables the further sampling and testing of influenza viruses in the surveillance system. The viruses are seeded in the Southern Lower Peninsula, and spread through contacts made by infected individuals while traveling within and between regions. Infected individuals are also allowed to enter and leave Michigan by traveling across the state boundaries (see Figure 12c).

Vaccination and Antiviral (see Figure 12d) effects have been modeled as factors that reduce the probability of infecting a susceptible individual, using existing estimates of influenza antiviral and vaccine efficacy (see Chapter III – Mitigation for Flu MODELO 1.0). Some sick individuals withdraw from their usual schedules and reduce the population of infectious individuals in the community (see Figure 12d). A fraction of these individuals will be hospitalized based on Michigan Hospitalization rates (see Chapter III – Mitigation for Flu MODELO 1.0). The previous contact and mitigation structure generates realistic variability in the observed daily influenza incidence, and provides a robust simulation test-bed for a wide variety of scenarios (see Figure 12h). The initial growth phase of the simulated outbreak was fitted to the growth phase of the observed data beginning from the pandemic declaration.

The surveillance system is simulated in 3 stages (see this chapter – Data Collection Processes). In the first stage, individuals seek healthcare (Figure 12e), based on a set of disease, behavioral, and financial factors that have been previously established in the literature as

influencing healthcare seeking behavior and introducing delay in the disease reporting (see section Data Collection Processes - Stage 1). In the second stage, healthcare providers report their collected specimens to the Michigan Syndromic Surveillance System (MSSS) and/or submit influenza specimens to the PHLs (see Figure 12f), based on a screening factors described in section Data Collection Processes Stage 2. The second stage generates the *trend of daily constitutional cases that are reported to the MSSS (see Figure 12i), and the trend of cases that are submitted to the PHLs* (see Figure 12j). In the third stage, cases are prioritized for testing and processed (see Figure 12g) based on the parameters in section Data Collection Processes Stage 3. It is then generated *the trend of confirmed cases with the emerging virus by date of submission to the PHL* (see Figure 12k). All the operational factors from the three data collection stages are listed in Table 26.

The chapter is organized as follows. In the second section, the data collection processes are described and implemented using the process flow charts with the choice of related operational factors. The third section states detailed dependent and independent variables for study and the simulation based methods for experimental design and statistical analysis. The fourth and last sections present and discuss the results of the simulation study.

Data Collection Processes

Processes regarding case reporting, specimen collecting and testing are presented in three stages: 1) process on healthcare seeking, 2) process on specimen collection by healthcare providers, 3) process on specimen sampling and testing in PHL. The important operational factors and events in each stage are presented as in the flow diagrams, where the letter D

represents operational factors and the letter E represents events. These processes were implemented in the calibrated Flu MODELO 1.0 in Chapter IV.

Stage 1. Process of Seeking Healthcare

Specimen collection is initiated by symptomatic cases seeking the health care providers that collect and ship specimens to the PHL for further subtyping and confirmation.

The process (Figure 1) starts with a case presenting flu-like symptoms (E1), and ends up with the case choosing and utilizing a type of health care provider (ED, primary care or urgent care). The process steps are explained as follows.

Presence of Symptoms (E1). A large proportion of Influenza cases present flu symptoms including fever, cough, sore throat, runny or stuffy nose, muscle or body aches, headaches, fatigue, vomiting and diarrhea ("Flu Symptoms & Complications", 2017). It serves as the first trigger for a case to seek health care from medical professionals (Blendon, Benson, DesRoches, Raleigh, & Taylor-Clark, 2004; Bish & Michie, 2010).

Severity of Symptoms (D1). Severe symptoms may shorten delay of illness appraisal, leading to immediate utilization of health care providers (Safer, Tharps, Jackson, Leventhal, 1979). It is also found that patients with mild symptoms tend to wait and apply self-monitoring/self-treatment (Burman, 1996). A severe case is more likely to see a health care provider.

Perceived Pandemic Influenza (D2, E2). Consider this situation: When pandemic influenza is declared, people tend to see providers right away even if they are not severely affected. Therefore, “Perceived to be pandemic influenza” is a combined factor with two aspects:

"Perceived to be ill" and "Pandemic declarations". Several studies have shown that people are likely to take health action when they believe they are susceptible to be ill (Kegeles, 1963; Wouterse & Tankari, 2015). Also, "pandemic declarations" has impact on people's "attitudes on health (O'Meara et al., 2014)", knowledge of disease consequence (Kegeles, 1963) and curability (Safer et al., 1979), which may shorten delay in health care seeking actions. If a case has mild symptoms, there is still a possibility for the case to seek health care when she becomes suspicious of symptoms. If the case doesn't perceive to have pandemic influenza, she will not seek health care for that day (E2).

Weekday VS Weekends (D3). In US, the majority of primary care providers/physicians have their office hours during weekdays (Mon. through Fri.). It is assumed that at weekdays, sick cases see a physician at a regular time (weekday) (E4) or go to the ED (E3), but at weekends, they go to urgent care (E5) or go to the ED (E3).

Primary Care Accessibility (D4). Primary care accessibility includes the impact of "affordability", "geographical accessibility", "acceptability" and "appointment wait times". Some research presented that household income (Kegeles, 1963; Wouterse & Tankari, 2015), asset (Wouterse & Tankari, 2015), and cost (O'Meara et al., 2014) affect the choice of providers. It is also found that geographical access or people distribution affect the choice of providers (Wouterse & Tankari, 2015; O'Meara et al., 2014). In addition, gender (Wouterse & Tankari, 2015), age (Wouterse & Tankari, 2015; Neill, Roland, Jones, Thompson, & Lakhanpaul, 2015) and education (Kegeles, 1963) are believed to be significant factors leading to different choice of providers. At last, appointment wait time has been pointed out to be the most significant factor in changing care-seeking decisions (Ahmed & Fincham, 2010). For the purpose of simplicity in

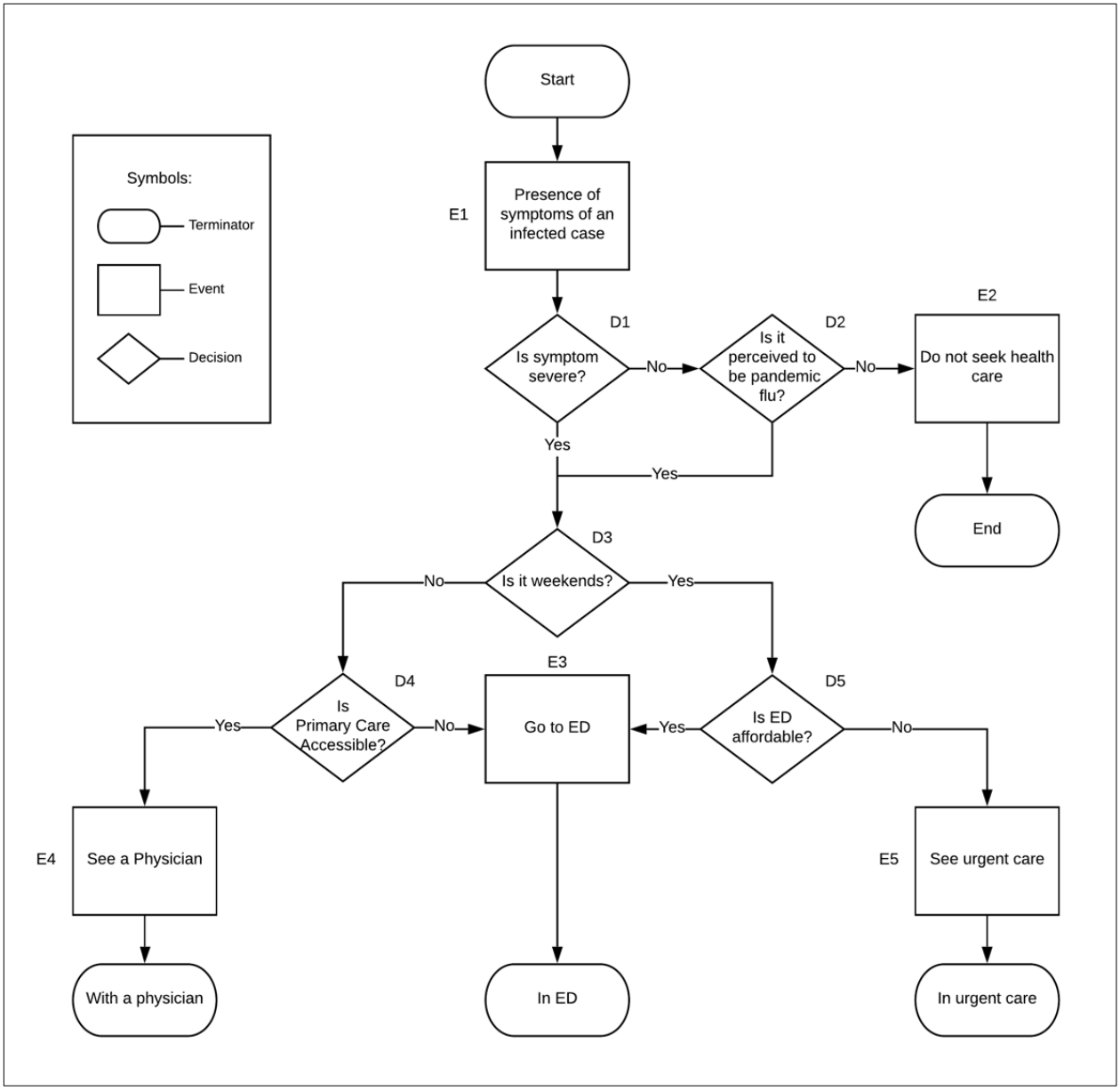


Figure 13. Flow Diagram of Process on Healthcare Seeking

model realization, these effects were merged into one major factor because all these effects will decide whether a case can access primary care or not. If no primary care is accessible, a case goes to visit ED immediately (E3) on weekdays. Otherwise, it goes to see a primary care physician (E4).

ED VS Urgent Care (D5). In US, ED is on 24/7 schedule and always available. However, it is expensive. Whether to go to the ED or the urgent care is a major influential factor that result in the choice of other health care providers instead of ED. Otherwise, the case goes to ED (E3).

Models for influenza surveillance should account for "the delay between onset of symptoms and infection reporting's", which is necessary to support real-time assessment of epidemiological parameters. The implementation considers the delay effect by conducting daily health care seeking behavior, which naturally reflects the growth rate in simulated time series.

Stage 2. Process on Specimen Collection by Healthcare Providers

Stage 2 begins with a case already seen by a health care providers, Each type of healthcare provider has a different strategy to report and/or submit the case's specimen to the PHL. This process (see Figure 14) was designed in consultation with the director of Department of Infection Prevention of Bronson Methodist Hospital. The model assumed that the same process should be applied to all health care providers in the state of Michigan. The process steps are described below.

Report to Michigan Syndromic Surveillance System (MSSS) (D1, E1). In the state of Michigan, EDs update simultaneously the Michigan Syndromic Surveillance System (MSSS) ("Michigan Syndromic Surveillance System Emergent/Urgent Data Submission Guide", 2015) which tracks daily number of ED visits, as an alternative way to monitor statewide influenza activity. It is assumed that 70% of the ED cases report to MSSS. If a case visits ED in this stage,

the case starts in ED (T1) and its syndrome is likely to be reported to MSSS (E1). Other types of providers do not submit this information.

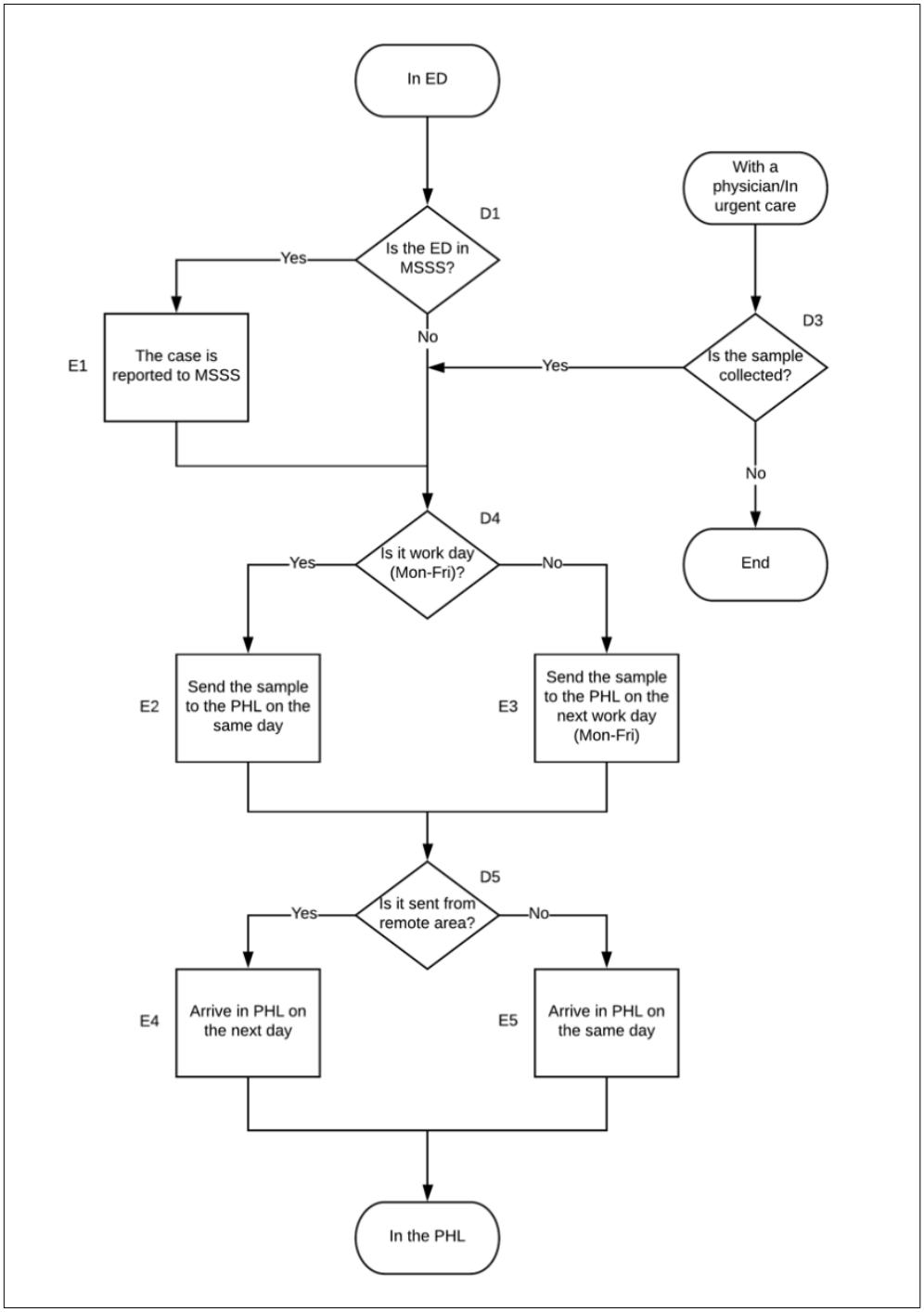


Figure 14. Flow Diagram of Process on Specimen Collection by Healthcare Providers

Collect Sample (D3). There is a difference in the process of sample collection between EDs and other type of providers. In EDs, all samples are collected from each visitor with ILI symptoms. But in primary care and urgent care, samples are collected based on the judgement of medical practitioners (physicians/nurses) and the needs of patients (willingness to pay for the test, attitude on health, etc.). Therefore, it is possible that the sample is not collected by a primary care or urgent care providers.

Shipping of Specimens (Weekday VS. Weekend) (D4, E2, E3). Once a sample is collected, it needs to be shipped to the PHL (In Michigan, the PHL is called the Michigan Bureau of Laboratories). The transportation happens once a day during weekdays (Mon. thru. Fri.) (E2), but shipment is avoided during weekends (Sun. & Sat.) and holidays. Samples collected during weekends are refrigerated and cannot be shipped until the upcoming business day (usually Monday) (E3).

Shipping from Remote areas (D5). Samples collected in normal areas arrive at the PHL on the same day (E4). Remote areas which are far from the PHL site (Lansing, Michigan), such as Upper Peninsula in Michigan. Samples collected from remote area are assumed to arrive at the PHL one day later than normal areas since samples are currently transported using ground services (cars, vans, etc.) (E5).

Stage 3. Process on Specimen Sampling and Testing in PHL

All samples submitted to the PHL that are PCR tested further viral characterization (i.e., sub-typing), and reported the CDC on a weekly basis, as is described in Figure 15. During an influenza emergency, massive samples arrive at the PHL. Due to lack of testing capabilities,

many samples are delayed in viral testing and some are frozen before expiration or discarded because of limited storage space. In this process, the factors related with the predictive accuracy based on the PHL viral testing results were focused. The factors for this process are described below.

PHL Work on Weekend (D1, E1). Whether the PHL works on weekend (Saturday/Sunday) is a decision variable telling if the submitted samples can be tested on weekends day or not. If it's on weekend, the testing of samples will be put off to the day (E1). Otherwise, the PHL work one that day and choose a sampling method (D2)

Sampling Criteria (D2). The PHL uses first-come-first-serve (FCFS, E2) sampling to test all available samples. During an emergency, due to limited testing capacity, the PHL still tests using FCFS, but allowing only samples from high-risk groups (Restricted FCFS, E3), such as children at 5-year-old or below, elders at 65-year-old and above, pregnant women, etc. These high-risk-group patients are more likely to be infected and develop flu-related complications. The current simulation test-bed assumes high-risk groups include children ($\leq$5) and elders($\geq$65) as the high-risk groups. Three additional sampling criteria were also tested: Hybrid (E4), 1WK FCFS (E5), and 1WK Restricted FCFS (E6). Hybrid (E4) test specimens with FCFS at the beginning. When the PHL capacity is fully used, they change to Restricted FCFS sampling to test the remaining submitted samples. 1WK FCFS (E5) uses FCFS to test the specimens submitted in the first week and doesn't test specimens during the second week. It assumes that the truncated 1 week PHL test trend is enough to portray the disease although the outbreak lasts 2 weeks. Similarly, 1WK Restricted FCFS (E6) uses Restricted FCFS (E3) but only test the specimens submitted in the first week.

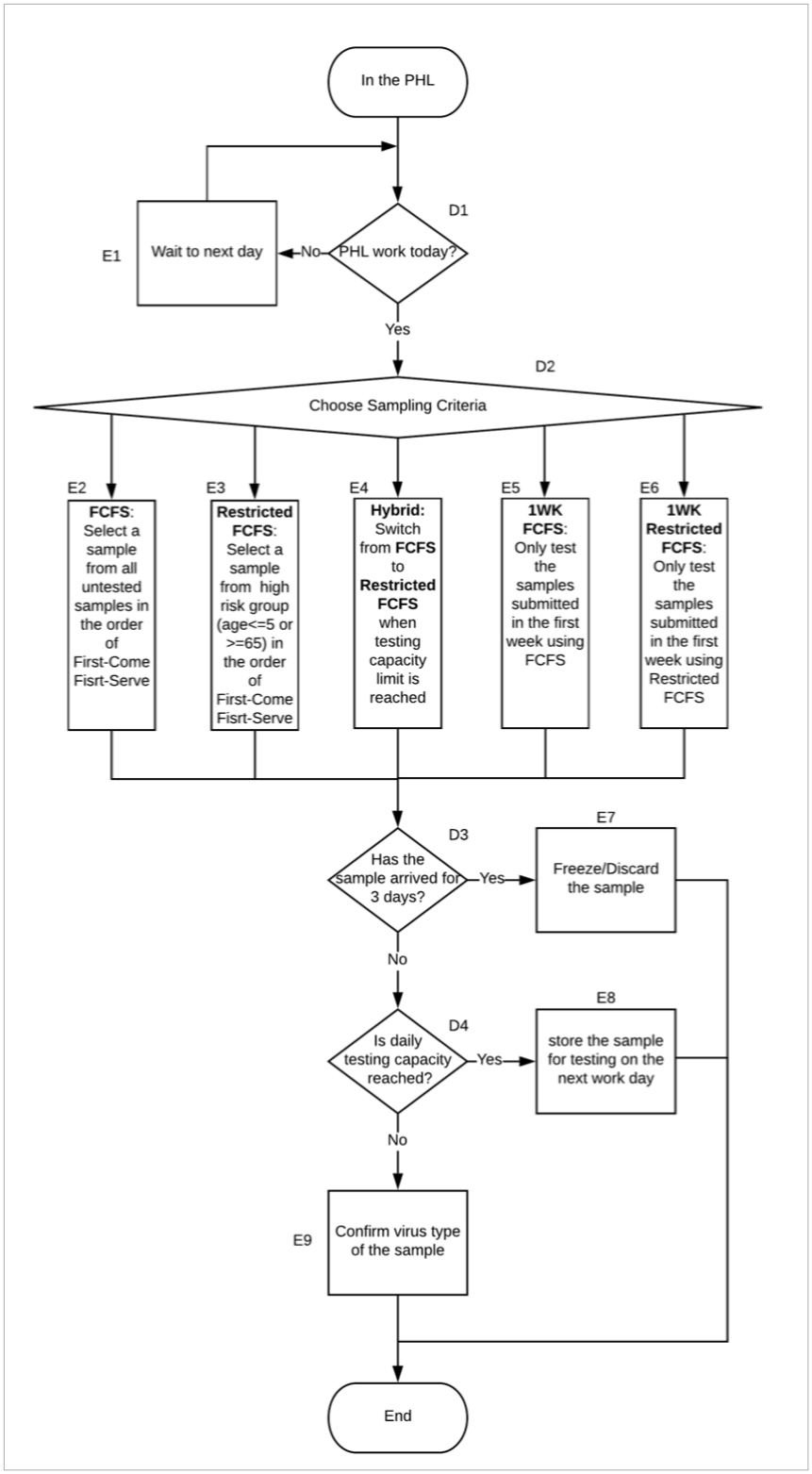


Figure 15. Flow Diagram of Process on Specimen Sampling and Testing in PHL

Freeze/discard Specimens (D3, E7). After samples are collected from patients using a sampling criterium, they are stored with refrigerant gel packs (or at 4oc) for shipment to the PHL. If delivery delays for 3-4 days, specimen samples must be frozen at -70^{o}c. During an influenza emergency, frozen samples are never tested due to constrained lab capabilities and some of them have to be discarded due to insufficient storage capacity. Therefore, it is assumed that if a sample has been waiting for 3 days after being collected, it will be frozen/discarded without being tested.

PHL Capacity (D4, E8). Maximum daily lab capacity is assumed as no more than 1000 in Michigan PHL. Once the number of tested samples reaches the maximum lab capacity, remaining samples are refrigerated and should be tested on the next day.

Influenza Virus Characterization (E9). When PHL testing capacity allows, samples are performed PCR testing for viral subtyping (influenza A and B). Viral culture should be ordered for negative cases to test novel viruses. All final testing results are reported to CDC via Michigan Department of Human and Health Services (MDHHS) on a weekly basis. This model assumes that novel viruses have been already detected which means viral confirmation can be performed with just PCR, meaning that viral testing takes only one day for each sample.

Simulation Input

Summarized from the processes at three stages, the model parameters (in Table 26) were implemented in the two optimal scenarios that calibrated Flu MODELO 1.0 to the real constitutional MSSS data (see Table 27 below and Chapter IV – Results for a review of the procedure).

Table 26 *Model Parameters for Data Collection Processes*

Parameter	Description	Parameter Space	Default Value
Poverty_Level	Percentage of low welfare people in the outbreak area	0-1	0.5
MSSS_Percentage	Percentage of emergency deparments reporting to MSSS	0-1	0.7
Perceived_Pandemic_Prob	Probability of a person perceived to pandemic influenza	0-1	0.5
Collect_Sample_Prob	Probability of a person specimen collected by healthcare providers	0-1	0.5
Primary_Care_Accessibility	Probablility of a person can accesss to primary care providers	0-1	0.5
PHL_Capacity	Maximum number of specimens tested in a PHL	1-INF	500
Number_PHLs	Number of PHLs in the outbreak area	1-10	1
PHL_Work_Weekend	PHL work at weekend (Sat. & Sun) or not	0=not work 1=work	1
Sampling_Criteria	Criteria for selecting the submitted specimens for testing and confirmation	0=surveillance not in place 1=FCFS 2=Restricted FCFS 3=Hybrid 4=1wk FCFS 5=1wk Restricted FCFS	1

Table 27 *Two Influenza Outbreak Scenarios Obtained from Chapter IV*

Solution	PANDEMIC _RN	SEASONAL _RN	INI_PANDEMIC _ INCIDENCE	INI_SEASONAL _INCIDENCE	WITHDRAWAL_ PROB	ANTIVIRAL_ PERCENTAGE
1	2.0965	1.6000	686	686	0.99999	0.26
2	1.9606	1.6000	658	686	0.99999	0.00001

Simulation Output

The current simulation testbed generates multiple data trends: 1) The daily trend of original infected cases per day, representing the "real" unobserved trend, which can disaggregated into pandemic and seasonal incidence (Figure 12h); 2) The trend of the Michigan Syndromic Surveillance System (Figure 12i), 3) the trends of daily submitted specimens to the PHL (Figure 12j), and 4) the trends of daily number of pandemic specimens tested and confirmed in the PHL (Figure 12k). Figure 16 and Figure 17 show the simulation examples of the initial growth phase of a pandemic influenza outbreak (Day 0 to day 13), and the output trends are described as above.

Methods

The objective is to evaluate the data collection biasedness in reporting the observed influenza trends under the current surveillance schemes. The measured biases are described in the following section.

Growth Rate Biases

In this study, exponential growth rate is used as an assessment of epidemiological parameter of the trends of daily influenza incidence. Table 28 described the 4 different growth rate based metrics that were used to assess the biases between observed and unobserved trends. The effect of the data collection processes in the biasedness of surveillance trends was assessed with these 4 metrics.

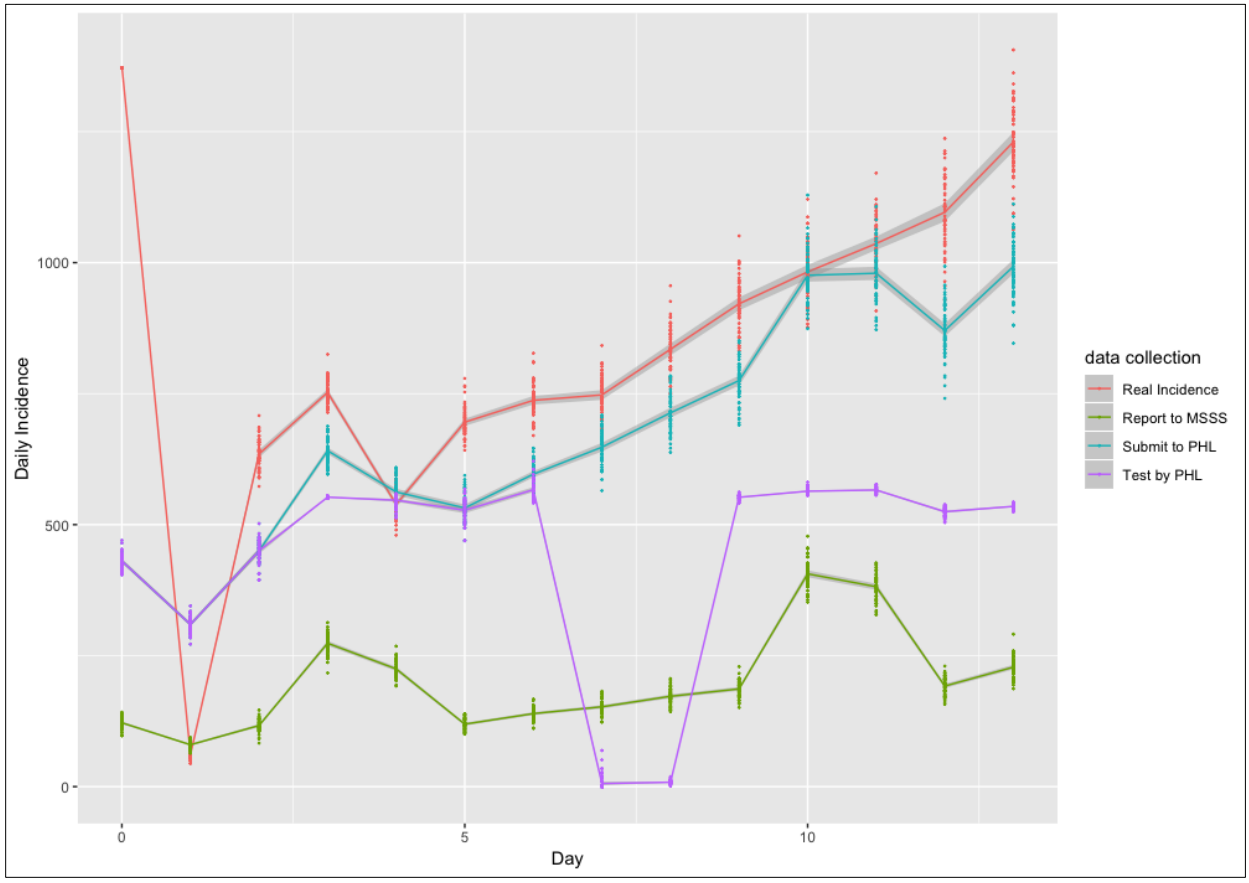


Figure 16. Simulated Surveillance Trends Under Influenza Outbreak Scenario 1. Simulated Data Collection Trends from 60 replicates using Basic reproduction numbers of 2.0965 and 1.6 for the pandemic and seasonal outbreaks, respectively; 686 initial infected cases for the pandemic outbreak and 686 initial infected cases for the seasonal outbreak; High disease severity (0.99 in the scale from 0 to 1); and 26% chance that a vulnerable person receives antiviral treatment; Scenario 1 has the following surveillance system parameters: POVERTY_LEVEL=0.5, PERCEIVED_PANDEMIC_PROB=0.5, PRIMARY_CARE_ACCESSIBILITY=0.5, COLLECT_SAMPLE_PROB=0.5, PHL_CAPACITY=550, PHL_WORK_WEEKEND=0 (not work on Saturday/Sunday), and SAMPLING_CRITERIA=1 (FIFO). Scenario 1 trials are are run until data collection is finished for 14-day co-circulating outbreak.

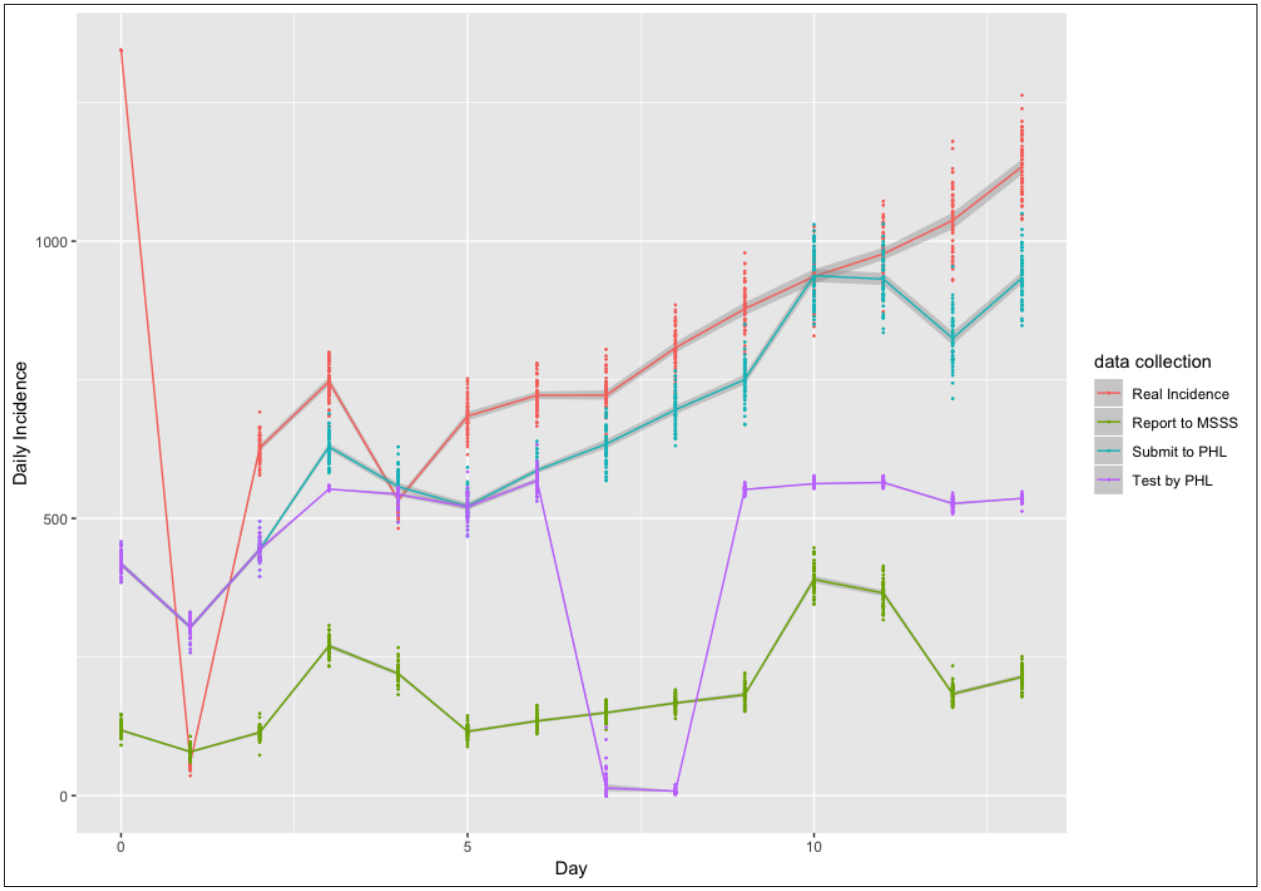


Figure 17. Simulated Surveillance Trends Under Influenza Outbreak Scenario 2. Simulated Data Collection Trends from 60 replicates using Basic reproduction numbers of 1.96 and 1.6 for the pandemic and seasonal outbreaks, respectively; 658 initial infected cases for the pandemic outbreak and 686 initial infected cases for the seasonal outbreak; High disease severity (0.99 in the scale from 0 to 1); and 0.001% chance that a vulnerable person receives antiviral treatment. Scenario 2 has the same surveillance system parameters: POVERTY_LEVEL=0.5, PERCEIVED_PANDEMIC_PROB=0.5, PRIMARY_CARE_ACCESSIBILITY=0.5, COLLECT_SAMPLE_PROB=0.5, PHL_CAPACITY=550, PHL_WORK_WEEKEND=0 (not work on Saturday/Sunday), and SAMPLING_CRITERIA=1 (FIFO). Scenario 2 trials are run until data collection is finished for 14-day co-circulating outbreak.

Factors Influencing Growth Rate Biases

To understand how the processes were related with the biases, 7 independent variables (see Table 29) were analyzed from the parameters defined in the data collection processes in

Table 28 *Growth Rate Biases As Dependent Variables*

Name	Description
B_MSSS_T	the growth rate difference between the **Real Overall** trends and the **MSSS** trends
B_Submit_T	the growth rate difference between the **Real Overall** trends and the **Submit** trends
B_Test_T	the growth rate difference between the **Real Overall** trends and the **PHL Overall** trends
B_Test_P	the growth rate difference between the **Real Pandemic** trends and the **PHL Pandemic** trends

Table 25. Number_PHL was set to 1 since there is only one PHL in Michigan. Parameter MSSS_Percentage was assumed to 0.7. Note that for Sampling_Criteria 4 - 1 WK FCFS, and 5 - 1 WK Restricted FCFS, the 1 week PHL test trends were compared with the 2 week real incidence trends to investigate how different the growth rate biases would be when 2 week confirmed trends are truncated to only 1 week as guided by CDC.

Table 29 *Data Collection Operational Factors As Independent Variables*

Name	Value Range	Involved in Which Surveillance Trends
Poverty_Level	$0\text{-}1^{R+}$	MSSS, Submit, PHL Overall test, PHL Pandemic test
Perceive_Pandemic_Prob	$0\text{-}1^{R+}$	MSSS, Submit, PHL Overall test, PHL Pandemic test
Primary_Care_Accessibility	$0\text{-}1^{R+}$	MSSS, Submit, PHL Overall test, PHL Pandemic test
Collect_Sample_Prob	$0\text{-}1^{R+}$	Submit, PHL Overall test PHL Pandemic test
PHL_Capacity	$0\text{-}1000^{I+}$	PHL Overall test PHL Pandemic test
PHL_Work_Weekend	0 – not work 1 – work	PHL Overall test PHL Pandemic test
Sampling_Criteria	1 – FCFS 2 – Restricted FCFS 3 – Hybrid 4 – 1 WK FCFS 5 – 1 WK Restricted FCFS	PHL Overall test PHL Pandemic test

Simulation Based Experiments and Statistic Analysis

To investigate effect of the operational factors in Table 29 on each bias in Table 28, simulation-based experiments and statistical analysis were conducted following the three steps. First, the experiment was designed, and the Michigan testbed was used to obtain surveillance trends (see Figure 12, Figure 16 and Figure 17) for each experimental run. Second, the growth rates were estimated from the simulated trends by fitting an Exponential Growth Model to the trends. The growth rate biases were then calculated (see Table 28 for the related trends for each bias). Lastly, the relationship between each growth rate bias and the involved operational factors were investigated using General Linear Model and ANOVA.

Experimental Design

The simulation based experimental design used a Half Factorial Central Composite Inscribed Design (CCI) to analyze the effects on the growth rate biases. The design used one block design and chose $\alpha=2$ to obtain a rotatable and orthogonal design. According to the value range of each of the seven operational factors, five of them were defined as continuous variables (see Table 30), and two were defined as categorical variables (see Table 31). The continuous variables in Table 30 were assigned with 2-level factorial points, 2-level axial points and 1 center points. The categorical variables in Table 31 were assigned groups, two groups for PHL_Work_Weekend and five groups for Sampling_Criteria. According to the half fractional CCI design, an experiment design of 32 single runs was created for each group of the two categorical variables (i.e., 32 single runs x 2 groups for PHL_Work_Weekend x 5 groups for Sampling_Criteria). With 10 replicates for each run, the experiment required 3200 runs (320

single run x10 replicates) . Table 32 shows the design matrix of 32 single runs under one group of categorical variables (when PHL_Work_Weekend was set to 0 and Sampling_Criteria was set to 1). See Appendix Table A1 for the full design matrix with 3200 runs.

Table 30 *Factor Levels of Continuous Variables for Surveillance*

Continuous Variable	Name Code	Factorial Points uncoded value (coded value)		Axial Points uncoded value (coded value)		Center point uncoded value (coded value)
Poverty_Level	A	0.75	0.25	0.99999	0.00001	0.5
		(0.5)	(-0.5)	(1)	(-1)	(0)
Perceive_Pandemic_Prob	B	0.75	0.25	0.99999	0.00001	0.5
		(0.5)	(-0.5)	(1)	(-1)	(0)
Collect_Sample_Prob	C	0.75	0.25	0.99999	0.00001	0.5
		(0.5)	(-0.5)	(1)	(-1)	(0)
Primary_Care_Accessibility	D	0.75	0.25	0.99999	0.00001	0.5
		(0.5)	(-0.5)	(1)	(-1)	(0)
PHL_Capacity	E	0.75	0.25	1000	100	550
		(0.5)	(-0.5)	(1)	(-1)	(0)

Table 31 *Factor Levels of Categorical Variables for Surveillance*

Categorical Variable	Label	Group				
PHL_Work_Weekend	F	0	1			
Sampling_Criteria	G	1	2	3	4	5

Note: 1. For PHL_Work_Weekend, 0 is coded as “not work at weekends” and 1 is coded as “work at weekends”;
2. For Sampling_Criteria, 1 is coded as “FCFS”, 2 is coded as “Restricted FCFS”, 3 is coded as “Hybrid”, 4 is coded as “1 WK FCFS”, and 5 is coded as “1 WK Restricted FCFS”.

Table 32 *Partial Half Fractional CCI Design Matrix with 32 Combinations and Single Run (Coded Matrix)*

Std Order	Run Order	PtType	Blocks	Poverty Level	Perceive Pandemic Prob	Collect Sample Prob	Primary Care Accessibility	PHL Capacity	PHL Work Weekend	Sampling Criteria
1	1226	1	1	-0.5	-0.5	-0.5	-0.5	0.5	0	1
2	454	1	1	0.5	-0.5	-0.5	-0.5	-0.5	0	1
3	1119	1	1	-0.5	0.5	-0.5	-0.5	-0.5	0	1
4	639	1	1	0.5	0.5	-0.5	-0.5	0.5	0	1
5	2726	1	1	-0.5	-0.5	0.5	-0.5	-0.5	0	1
6	2560	1	1	0.5	-0.5	0.5	-0.5	0.5	0	1
7	1141	1	1	-0.5	0.5	0.5	-0.5	0.5	0	1
8	74	1	1	0.5	0.5	0.5	-0.5	-0.5	0	1
9	1114	1	1	-0.5	-0.5	-0.5	0.5	-0.5	0	1
10	508	1	1	0.5	-0.5	-0.5	0.5	0.5	0	1
11	1419	1	1	-0.5	0.5	-0.5	0.5	0.5	0	1
12	1362	1	1	0.5	0.5	-0.5	0.5	-0.5	0	1
13	3032	1	1	-0.5	-0.5	0.5	0.5	0.5	0	1
14	2330	1	1	0.5	-0.5	0.5	0.5	-0.5	0	1
15	2605	1	1	-0.5	0.5	0.5	0.5	-0.5	0	1
16	2763	1	1	0.5	0.5	0.5	0.5	0.5	0	1
17	1220	−1	1	−1	0	0	0	0	0	1
18	407	−1	1	1	0	0	0	0	0	1
19	2583	−1	1	0	−1	0	0	0	0	1
20	701	−1	1	0	1	0	0	0	0	1
21	2030	−1	1	0	0	−1	0	0	0	1
22	2057	−1	1	0	0	1	0	0	0	1
23	3078	−1	1	0	0	0	−1	0	0	1
24	384	−1	1	0	0	0	1	0	0	1
25	326	−1	1	0	0	0	0	−1	0	1
26	2949	−1	1	0	0	0	0	1	0	1
27	2628	0	1	0	0	0	0	0	0	1
28	3017	0	1	0	0	0	0	0	0	1
29	581	0	1	0	0	0	0	0	0	1
30	2831	0	1	0	0	0	0	0	0	1
31	1116	0	1	0	0	0	0	0	0	1
32	212	0	1	0	0	0	0	0	0	1

Calculating the Growth Rate Biases

Under each outbreak scenario (see Table 27), an experiment of 3200 runs was conducted, and, 4 growth rate biases (see Table 28) were calculated for each of the runs in each scenario, yielding 8 growth rate biases in total. Let S1 denote Scenario 1, S2 denote Scenario 2, the four growth rate biases in Scenario 1 are then named as B_MSSS_T_S1, B_Submit_T_S1, B_Test_T_S1, and B_Test_P_S1. The four growth rate biases in Scenario 2 are named as B_MSSS_T_S2, B_Submit_T_S2, B_Test_T_S2, and B_Test_P_S2.

Statistical Analysis

To answer the research questions in this chapter, General Linear Model (GLM) was used to analyze the main, interaction and quadratic effects of the operational factors on each of the growth rate biases that were obtained from the simulation based experiments. Rstudio 1.1.456 and Minitab 18 were used for the data processing, DOE and statistical analysis. The results are regarded as statistically significant with a p-value less than a significant level of 0.05.

Equation 14 shows the GLM for the growth rate biases regarding each surveillance trends. Let B_[Trends] denote the identification of growth rate biases regarding each surveillance trends ([MSSS_T]= MSSS trends, [Submit_T]=Submit trends, [Test_T]=PHL Overall trends, [Test_P]=PHL Pandemic trends). Let S denote the outbreak scenarios (S1=Scenario 1, S2=Scenario 2).

$$B\_[Trends]\_S = \beta_{Trends,S,0} + \sum_{i=1}^{k} \beta_{Trends,S,i} x_i + \sum_{j=i+1}^{k} \sum_{i=1}^{k} \beta_{Trends,S,ij} x_i x_j + \sum_{j=1}^{k} \beta_{Trends,S,j} {x_j}^2$$

Equation 14,

where, x_i and x_j are the i^{th} and j^{th} operational factors corresponding to particular surveillance trends (see Table 28). $\beta_{Trends,S,0}$ is the intercept for B_[Trends] under Scenario S. $\beta_{Trends,S,i}$'s are the parameter coefficients of first-order term for B_[Trends] under Scenario S. $\beta_{Trends,S,ij}$'s are the parameter coefficients of the second-order interaction terms for B_[Trends] under Scenario S. $\beta_{Trends,S,j}{}'s$ are the parameter coefficients of the second-order quadratic terms for B_[Trends] under Scenario S.

The GLM polynomial equations were used to answer the below research questions one by one.

Research question 2. Which factors of the data collection operations have significant impacts on biasedness between the real and observed incidence trends?

In this question, the parameter coefficients of the first-order model terms (see Equation 14) were used to replace sensitivity analysis of each involved operational factors as in (Zhang & Brown, 2014) These factors were hypothesized to have significant impacts on each growth rate biases by checking their main effects, interaction effects and quadratic effects in the model.

Research question 3. How do Sampling Criteria affect the biasedness between the real and observed incidence trends?

In this question, the main effects of Sampling Criteria in Equation 14 were analyzed on the growth rate biases in the PHL Overall/Pandemic test trends. Besides, the interaction effect of Sampling_Criteria and each of other operational factors were also analyzed. Since the current model (see Equation 14) includes the quadratic terms, all of the analyses were made when controlling the curvature effect in the model.

Research question 4. To what extent do the observed influenza incidence trends explain the real influenza incidence trends?

In this question, Simple Linear Regression model was fitted to determine the explanatory power that a growth rate of one surveillance trend had in the growth rate of the real influenza incidence tends. For each outbreak scenario, the 3200 run simulation experiment consist of 320 combinations with 10 replicates for each combination. Therefore, each of the 320 combinations get 10 growth rates of each real incidence trends (Real Overall, Real Pandemic) and each surveillance trends (MSSS, Submit, PHL Overall, PHL Pandemic). The R-square values of model fitting was checked to compare which growth rate of the surveillance trends is explained more variance of the growth rate of the real incidence trends. The models with highest R^2 were tallied for each surveillance trends. Finally, the average explanatory power of the best surveillance treads were given.

Results of Research Question 2

In the DOE, a Half Factorial CCI design was developed with the parameter levels in Table 30 and Table 31 and conducted in two influenza outbreak scenarios. The results from the experiment are described below based on the research questions in this chapter.

Factors Affecting Biases in MSSS Trends (B_MSSS_T)

Cases reported to MSSS were related with stage 1 (health care seeking stage) and stage 2 (specimen submission by health care providers). Three factors were analyzed for MSSS report: Poverty_Level (A), Perceive_Pandemic_Prob (B), and Primary_Care_Accessibility (D). Under

both influenza outbreak scenarios, the impact of these three factors were investigated on the growth rate biases between the simulated real overall influenza incidence trends and the simulated MSSS trends (see Table 28).

Outliers

For both scenarios, major outliers were detected on the combinations where Poverty_Level was set to 0.99999 when very few people went to ED due to poor status. In this case, a couple of cases sought healthcare in the ERs. In both scenarios, there were outliers identified due to the same problems.

For Scenario 1, two major outliers from simulation run 306 and 2418 were detected (Figure 18a). The growths rate of the MSSS trends under the two situations were overestimated using the Exponential Growth Model. The two outliers were then replaced by 0.038166, which is the average growth rate of the replicates that are not outliers. With the previous replacement, the graph of residuals (Figure 18b) still show three more outliers under this situation: 2194, 2546, 2770. The five outliers were then replaced by 0.044253, which is the average growth rate of the replicates that are not outliers. The final model met the assumptions of normal distribution, independence and equal variance (Figure 19a).

For Scenario 2, similarly, one major outlier from simulation run 370 was detected by the residuals plot (Figure 18c). This outlier was then mutated with 0.037423 by taking the average of the growth rates from the rest replicates of the same situation. The residuals check of model fitting with mutated data (Figure 18d) still show three more outliers under this situation: 114, 2034, 2610. The average of the rest growths 0.039436 was used to mutate all these four outliers.

The final model met the assumptions of normal distribution, independence and equal variance (Figure 19b).

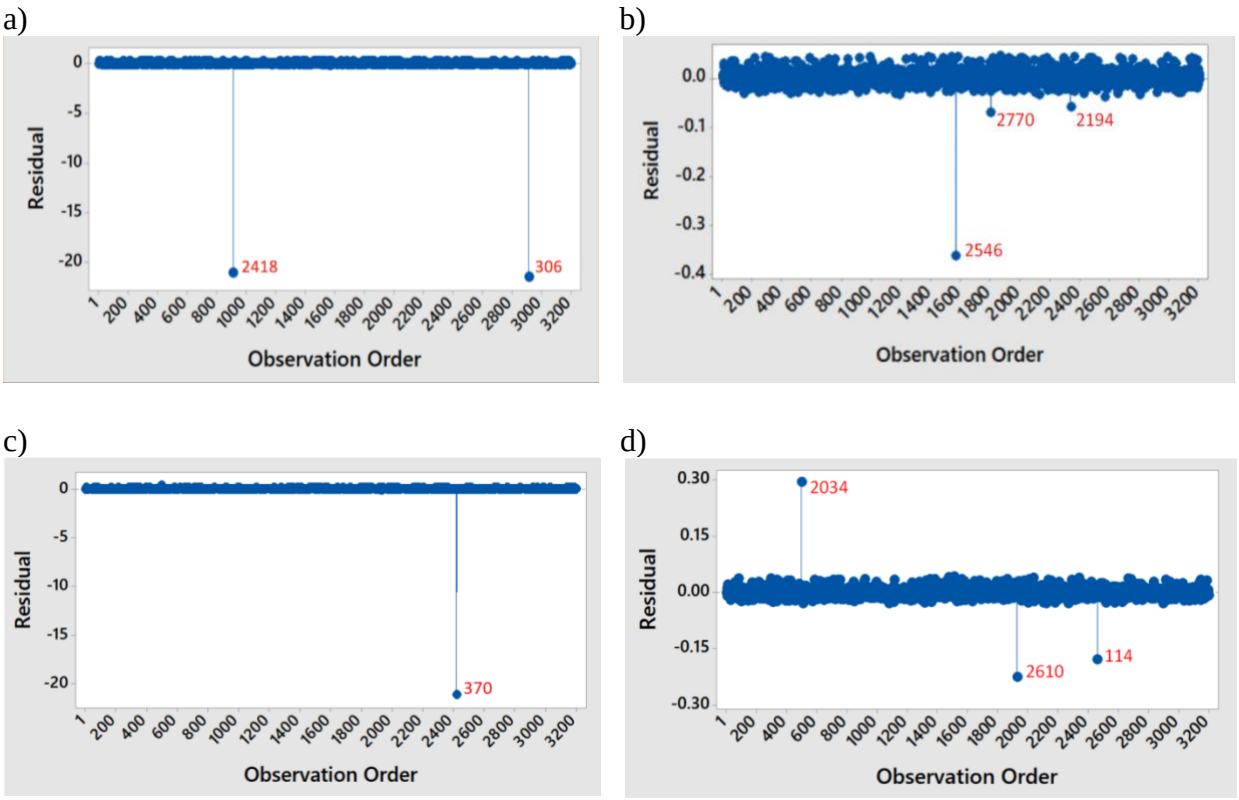
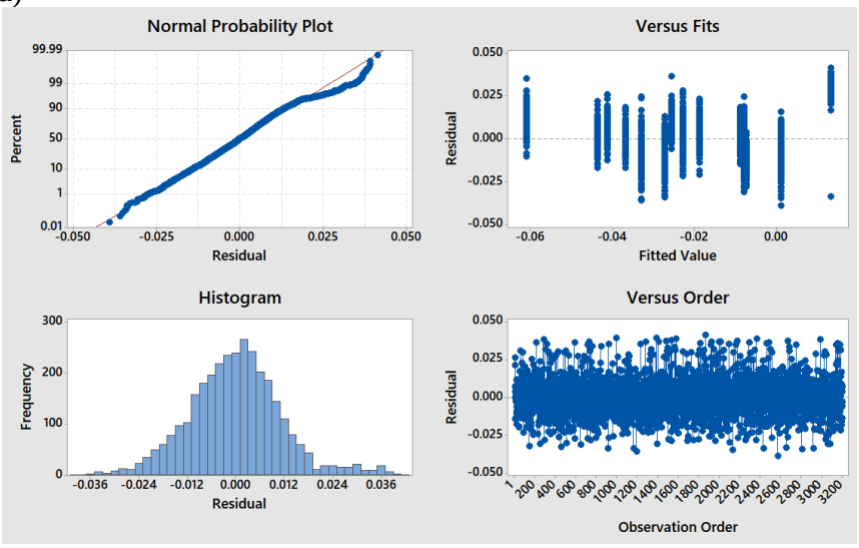


Figure 18. Residuals Versus Observations Plot Detecting Outliers for Biases Regarding MSSS Trends Under Two Scenarios. a) B_MSSS_T_S1 Step 1. b) B_MSSS_T_S1 Step 2. c) B_MSSS_T_S2 Step 1. d) B_MSSS_T_S2 Step 2.

a)



b)

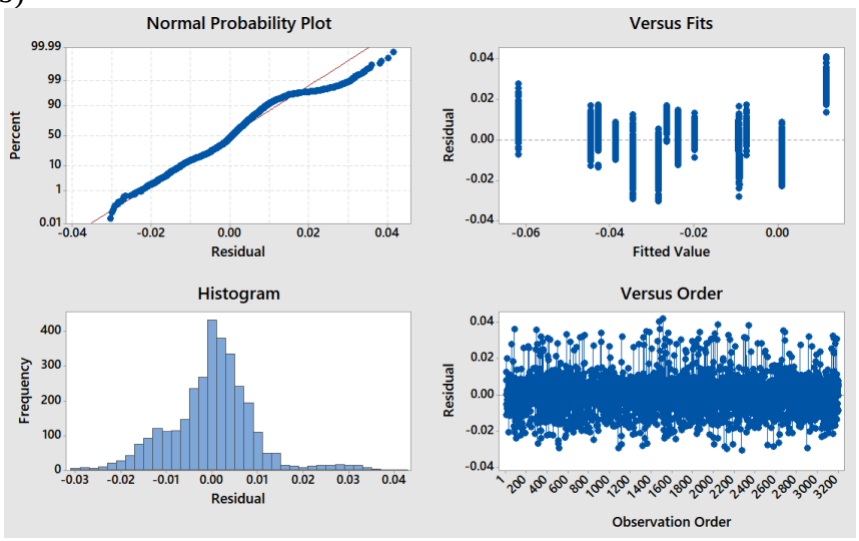


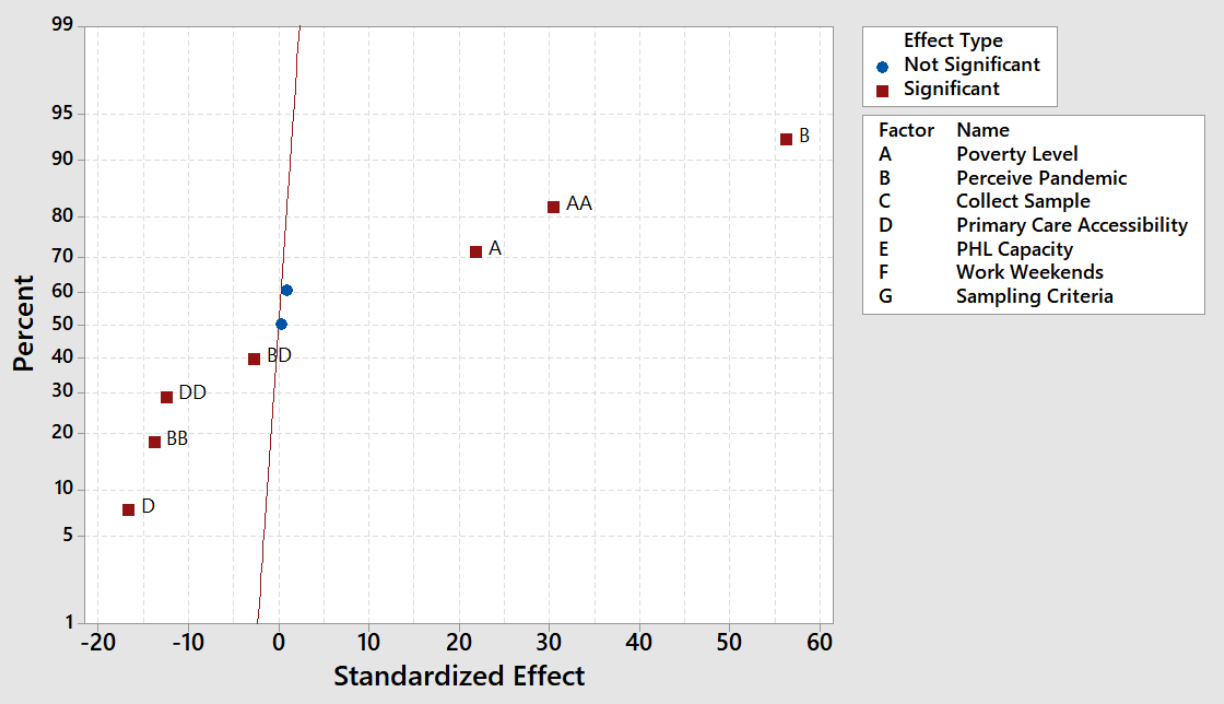
Figure 19. Final Model Diagnostics for Biases Regarding MSSS Trends Under Two Scenarios. a) B_MSSS_T_S1. b) B_MSSS_T_S2.

A multiple linear regression was employed to predict Poverty_Level (A), Perceived_Pandemic_Prob (B) and Primary_Care_Accessibility (D) on the growth rate biases between the real overall influenza incidence trends and the MSSS trends under each scenario (B_MSSS_T_S1, B_MSSS_T_S2). For scenario 1, a significant regression was found

(F(9,3190)= 588.54, P <0.001), with an R^2 of 62.41%, R^2(adj) of 64.31%. For scenario 2, a significant regression was found (F(9,3190)= 896.47, P <0.001), with an R^2 of 71.67%, R^2(adj) of 71.59%.

Both scenarios yielded similar ANOVA results. There are statistically significant main and quadratic effects of Poverty_Level (A), Perceived_Pandemic_Prob (B) and Primary_Care Accessibility (D) on the growth rate biases regarding MSSS trends (Figure 20). There was statistically significant interaction effect of Perceived_Pandemic_Prob (B) and Primary_Care_Accessibility (D). The other interaction has no statistically significant effect. There is no obvious multicollinearity of any terms (VIF=1.00-1.01).

a)



b)

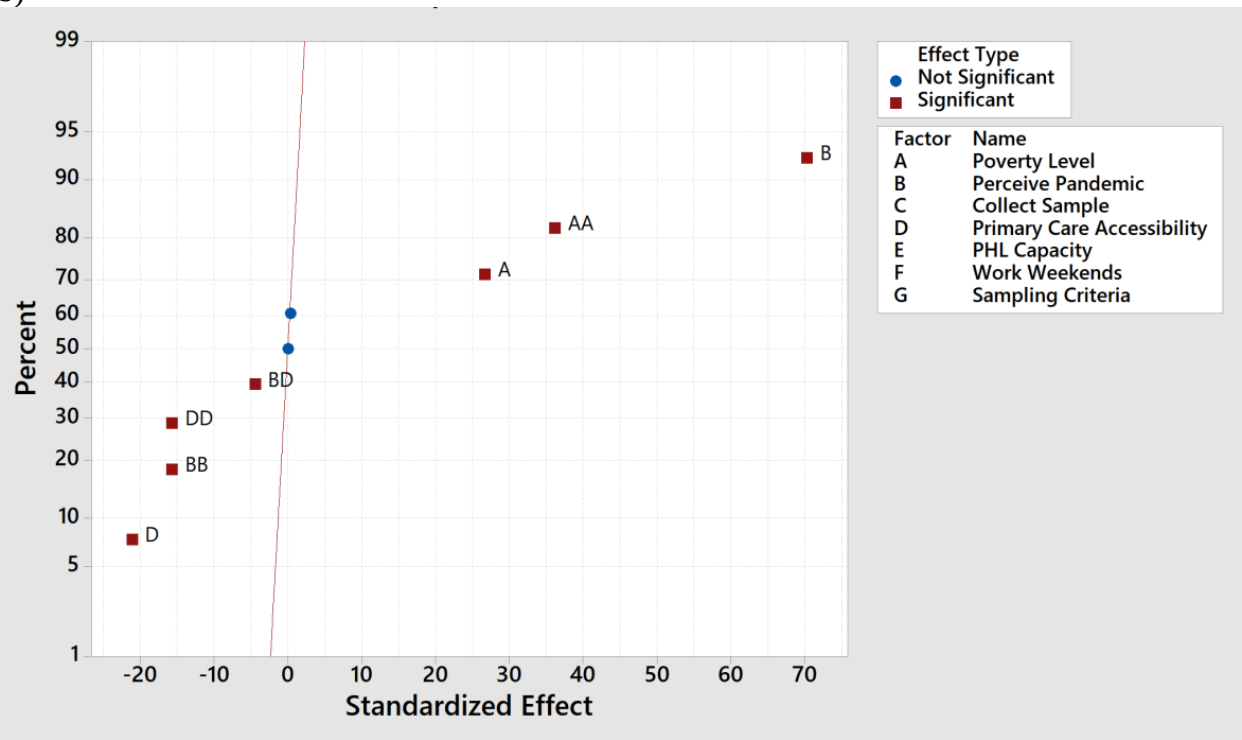


Figure 20. Standardized Effects of Factors Related with Biases Regarding MSSS Trends Under Two Scenarios. a) B_MSSS_T_S1. b) B_MSSS_T_S2.

The polynomial regression equations in uncoded units (Equation 15, Equation 16) for Scenario 1 and Scenario 2 were obtain after the insignificant terms were removed from the original regression model:

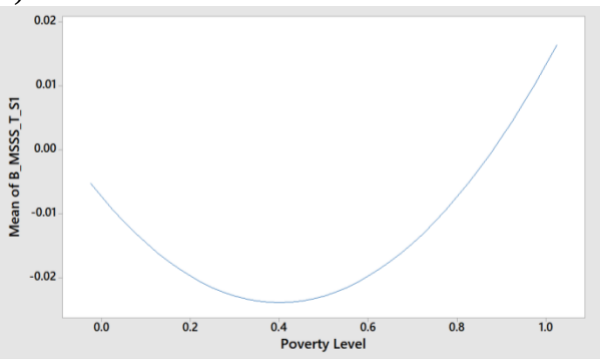
$$\begin{aligned} B\_MSSS\_T\_S1 = {} & -0.05163 - 0.08248\ Poverty\_Level + 0.10637\ Perceived\_Pandemic\_Prob \\ & + 0.03323\ Primary\_Care\_Accessibility + 0.10309\ (Poverty\ Level)^2 \\ & - 0.04687\ (Perceived\_Pandemic\_Prob)^2 \\ & - 0.04251\ (Primary\_Care\_Accessibility)^2 \\ & - 0.01289\ (Perceived\_Pandemic\_Prob)*(Primary\_Care\_Accessibility) \end{aligned} \quad \text{Equation 15}$$

$$\begin{aligned} B\_MSSS\_T\_S2 = {} & -0.05411 - 0.07983\ Poverty\_Level + 0.10631\ Perceived\_Pandemic\_Prob \\ & + 0.03566\ Primary\_Care\_Accessibility + 0.10043\ (Poverty\_Level)^2 \\ & - 0.04358\ (Perceived\_Pandemic\_Prob)^2 \\ & - 0.04353\ (Primary\_Care\_Accessibility)^2 \\ & - 0.01686\ (Perceived\_Pandemic\_Prob)*(Primary\_Care\_Accessibility) \end{aligned} \quad \text{Equation 16}$$

The mean biases of growth rates regarding MSSS trends for each scenario is -0.052 and -0.054 when Poverty_Level, Perceived_Pandemic_Prob and Primary_Care_Accessibility equal 0.00001.

The main effect plot from Scenario 1 (Figure 21a) and Scenario 2 (Figure 21b) shows that within the feasible region of these factors, Poverty_Level (A) is positively associated with B_MSSS_T when Poverty_Level (A) is greater than 0.4, and negatively associated with B_MSSS_T when Poverty_Level (A) is less than 0.4, after other independent variables were controlled. Poverty_Level lower than 0.4 cannot reduce the growth rate biases regarding MSSS trends. Poverty_Level above 0.4 is more likely to related with smaller absolute mean growth rate biases regarding MSSS trends.

a)



b)

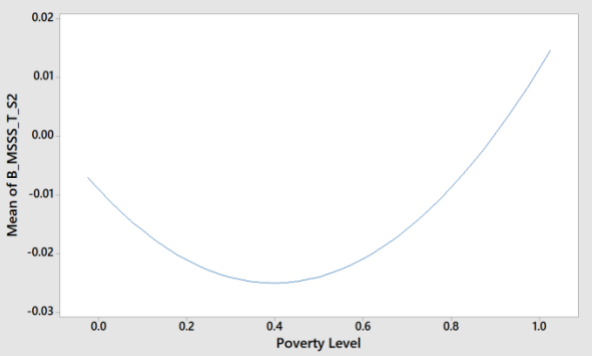
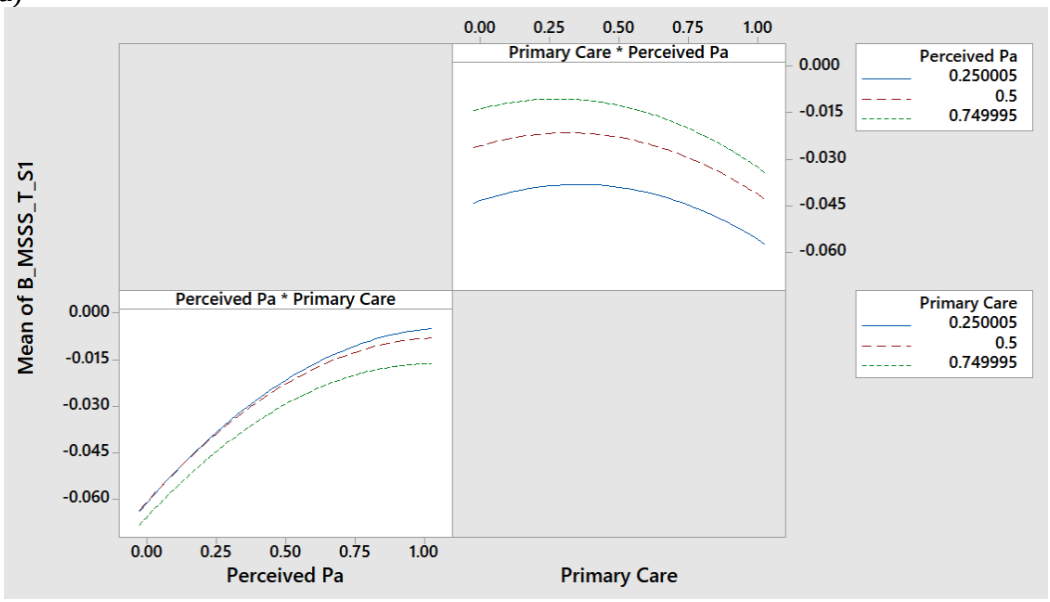


Figure 21. Main Effects Plot on Biases Regarding MSSS Trends Under Two Scenarios Respectively. a) B_MSSS_T_S1. b) B_MSSS_T_S2.

The interaction effect of Perceived_Pandemic_Prob (B) and Primary_Care_Accessibility (D) (Figure 22) shows that there is additive effect of the interaction on the growth rate biases regarding MSSS trends within the feasible range 0-1. At lower level of Perceived_Pandemic_Prob, Primary_Care_Accessibility, larger absolute mean growth rate biases

are obtained regarding MSSS trends. Therefore, the increase of Perceived_Pandemic_Prob is more likely to reduce the absolute growth rate biases regarding MSSS trends.

a)



b)

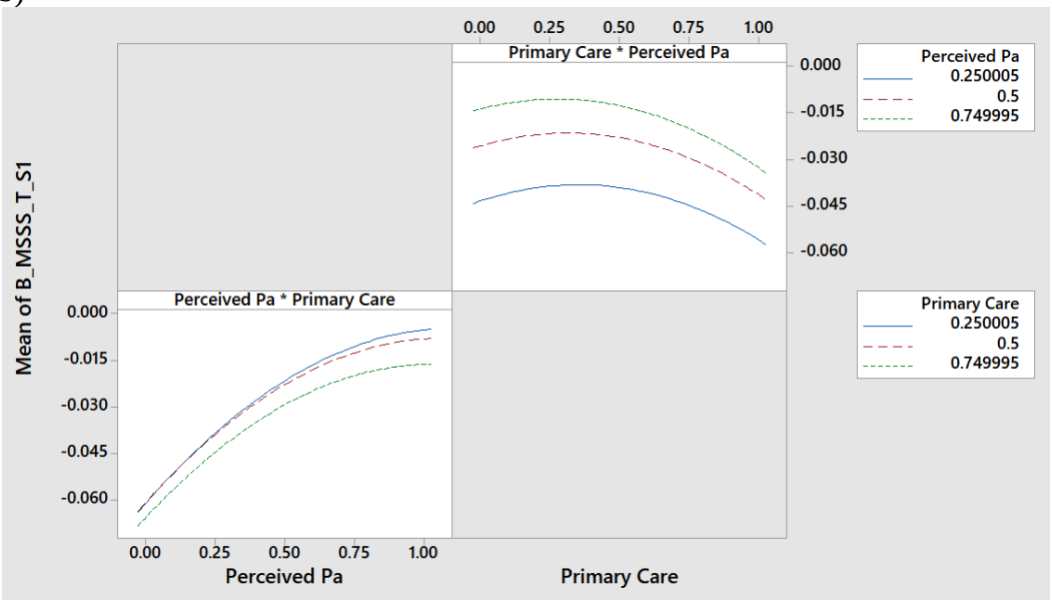
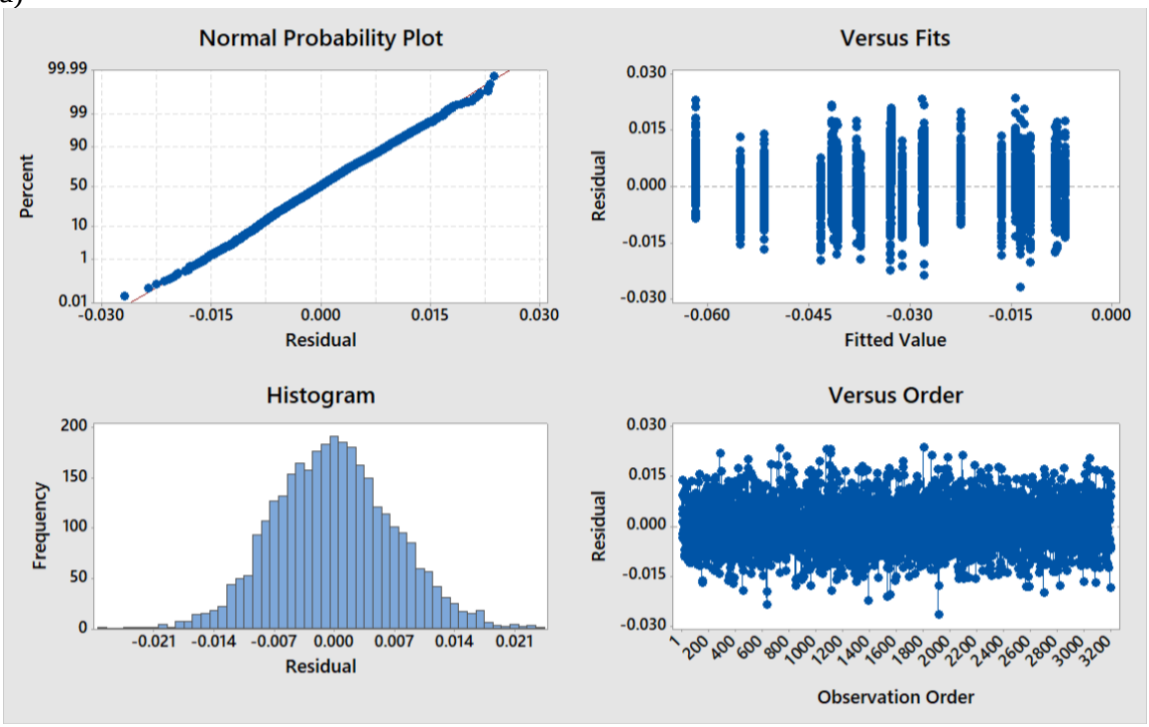


Figure 22. Interaction Effects on Biases Regarding MSSS Trends Under Two Scenarios. a) B_MSSS_T_S1. b) B_MSSS_T_S2.

Factors Affecting Biases in Submit Trends (B_Submit_T)

Specimens submitted to the PHL were also related with the stage 1 and 2 (health care seeking stage and specimen collection by health care providers). All the four factors in the first two stages: Poverty_Level (A), Perceived_Pandemic_Prob (B), Collect_Sample_Prob (C) and Primary_Care_Accessibility (D) were related with the process for specimen submission to the PHL. The impact of these four factors were investigated on the growth rates biases between the simulated real overall influenza incidence trends and the simulated trends of specimen submitted to the PHL (B_Submit_T). The residuals check of model fitting met the assumptions of normal distribution, independence and equal variance (Figure 23).

a)



b)

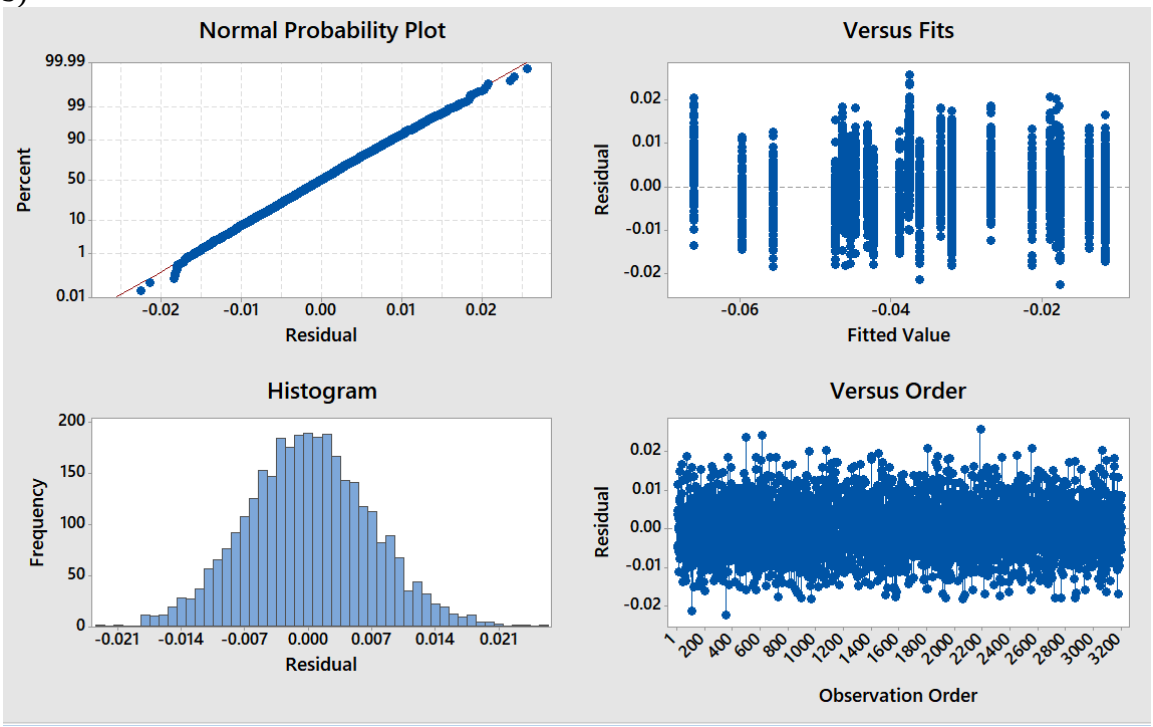


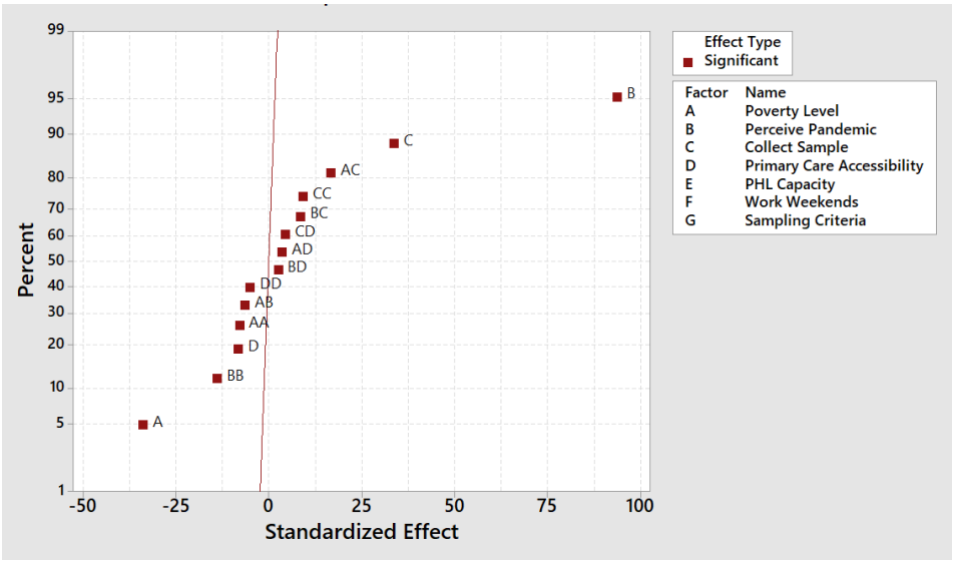
Figure 23. Model Diagnostics for Biases Regarding Submit Trends Under Two Scenarios. a) B_Submit_T_S1. b) B_Submit_T_S2.

A multiple linear regression was calculated to predict Poverty_Level (A), Perceived Pandemic_Prob (B), Collect_Sample_Prob (C) and Primary_Care_Accessibility (D) on the biases of growth rates regarding Submit trends for each scenario. For Scenario 1, a significant regression was found ($F(14,3185)=853.19$, $P<0.001$), with an R^2 of 78.95%. For Scenario 2, a significant regression was found ($F(14,3185)=871.57$, $P<0.001$), with an R^2 of 79.30%, and R^2(adj) of 79.21%.

There are statistically significant main and quadratic effects of all factors: Poverty_Level (A), Perceived_Pandemic_Prob (B), Collect_Sample_Prob (C), and Primary_Care_Accessibility (D) (Figure 24). All the interaction terms have statistically significant effect on B_Submit_T_S1. All the interaction effects except Poverty Level and Primary_Care_Accessibility (AD) have

statistically significant effect on B_Submit_T_S2. There is no obvious multicollinearity of any terms (VIF=1.00-1.02). The difference is very slight between the two scenarios and were mostly like to be caused by random variations from simulation experiments.

a)



b)

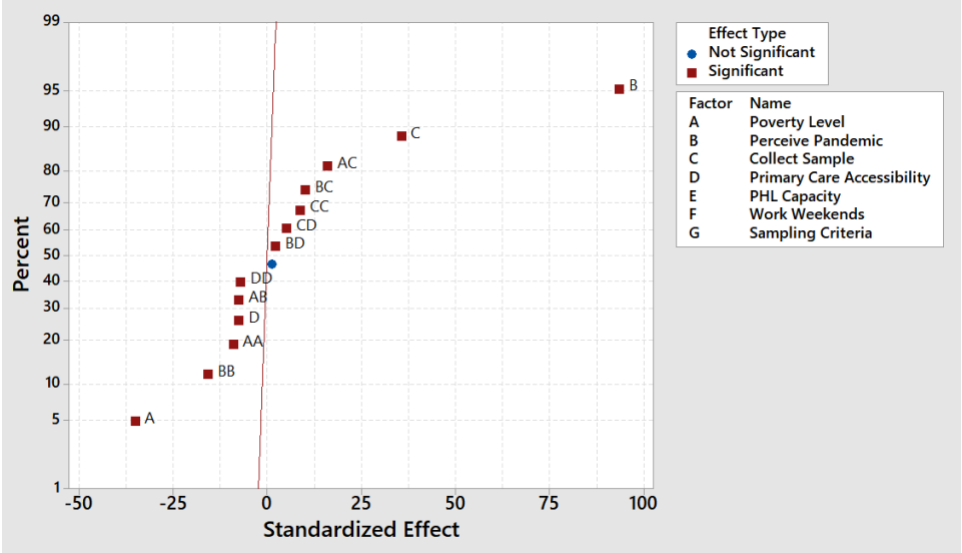


Figure 24. Standardized Effects of Factors for Biases Regarding Submit Trends Under Two Scenarios. a) B_Submit_T_S1. b) B_Submit_T_S2.

The polynomial regression equations in uncoded units (Equation 17, Equaiton 17) for Scenario 1 and Scenario 2 were obtain after the insignificant terms were removed from the original regression model:

$$\begin{aligned}
B_Submit_T_S1 = {} & -0.04130 - 0.02182\ Poverty_Level + 0.07584\ Perceived_Pandemic_Prob \\
& - 0.04074\ Collect_Sample_Prob - 0.00866\ Primary_Care_Accessibility \\
& - 0.01639\ (Poverty_Level)^2 \\
& - 0.02858\ (Perceived_Pandemic_Prob)^2 + 0.01891\ (Collect_Sample_Prob)^2 \\
& - 0.01046\ (Primary_Care_Accessibility)^2 \\
& - 0.01818\ (Poverty_Level)*(Perceived_Pandemic_Prob) \\
& + 0.04627\ (Poverty_Level)*(Collect_Sample_Prob) \\
& + 0.00973\ (Poverty_Level)*(Primary_Care_Accessibility) \\
& + 0.02358\ (Perceived_Pandemic_Prob)*(Collect_Sample_Prob) \\
& + 0.00689\ (Perceived_Pandemic_Prob)*(Primary_Care_Accessibility) \\
& + 0.01222\ (Collect_Sample_Prob)*(Primary_Care_Accessibility)
\end{aligned} \qquad \text{Equation 17}$$

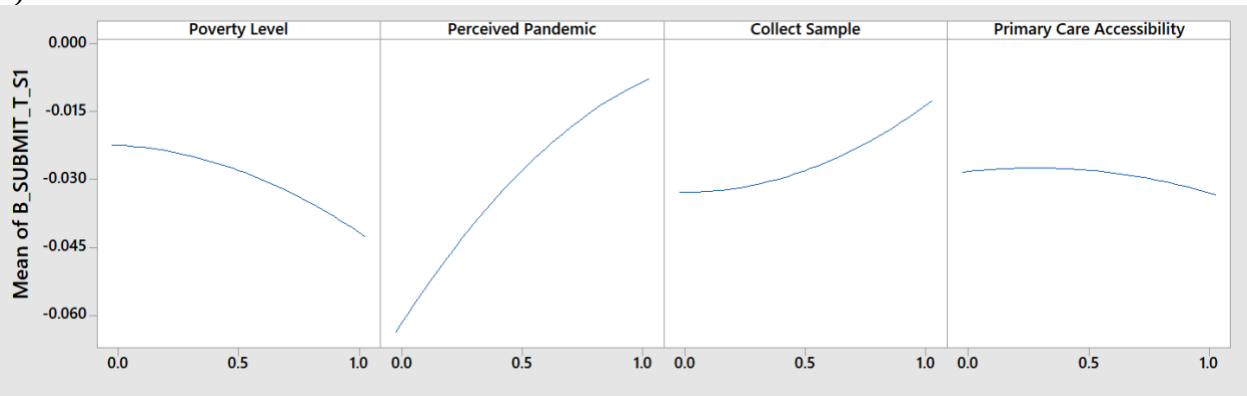
$$\begin{aligned}
B_Submit_T_S2 = {} & -0.04940 - 0.01466\ Poverty_Level + 0.07782\ Perceived_Pandemic_Prob \\
& - 0.04061\ Collect_Sample_Prob - 0.00155\ Primary_Care_Accessibility \\
& - 0.01837\ (Poverty_Level)^2 \\
& - 0.03190\ (Perceived_Pandemic_Prob)^2 + 0.01771\ (Collect_Sample_Prob)^2 \\
& - 0.01436\ (Primary_Care_Accessibility)^2 \\
& - 0.02074\ (Poverty_Level)*(Perceived_Pandemic_Prob) \\
& + 0.04390\ (Poverty_Level)*(Collect_Sample_Prob) \\
& + 0.00345\ (Poverty_Level)*(Primary_Care_Accessibility) \\
& + 0.02783\ (Perceived_Pandemic_Prob)*(Collect_Sample_Prob) \\
& + 0.00587\ (Perceived_Pandemic_Prob)*(Primary_Care_Accessibility) \\
& + 0.01397\ (Collect_Sample_Prob)*(Primary_Care_Accessibility)
\end{aligned} \qquad \text{Equation 18}$$

The mean biases of growth rates regarding Submit trends for each scenario is -0.041 and -0.049 when Poverty_Level (A), Perceived_Pandemic_Prob (B), Collect_Sample_Prob (C) and Primary_Care_Accessibility (D) equal 0.00001.

The main effect plot (Figure 25) shows that within the feasible region of these factors (0.00001-0.99999 for all these four factors), Perceive_Pandemic_Prob (B) and Collect_Sample_Prob (C) are positively associated with the mean growth rate bias regarding the Submit trends. Poverty_Level (A) and Primary_Care_Accessibility (D) are negatively associated

with the mean growth rate bias. Perceived_Pandemic_Prob (B) is more related with the growth rate biases regarding the Submit trends.

a)



b)

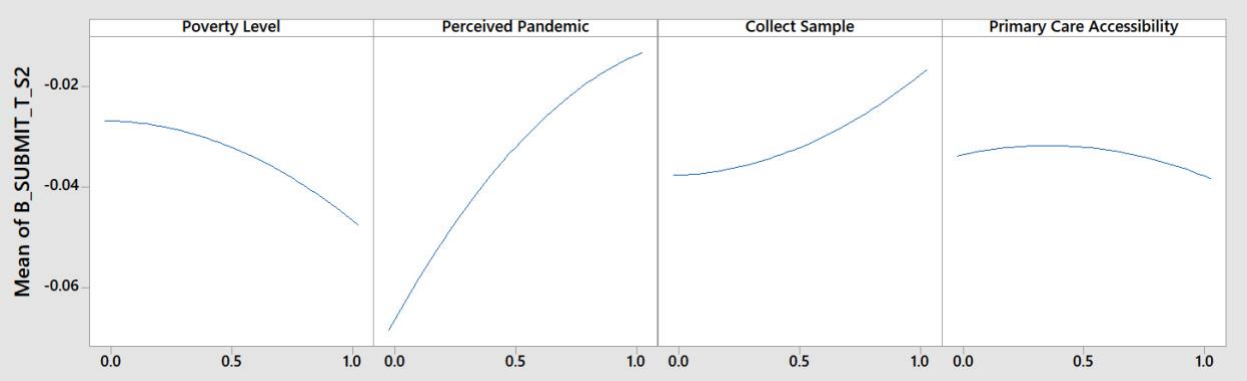


Figure 25. Main Effects on Biases Regarding Submit Trends Under Two Scenarios. a) B_Submit_T_S1. b) B_Submit_T_S2.

The interaction effects plot for Scenario 1 (Figure 26a) and Scenario 2 (Figure 26b) show that within the feasible range 0-1, there is additive effect of all these interaction terms on the growth rate biases regarding the Submit trends, and only the interaction of Poverty_Level and Collect_Sample_Prob (AC) has a little cross effect. The interaction effects don't change the direction of the main effects.

For the interaction of Poverty_Level and Perceived_Pandemic_Prob (AB), at higher level of Perceived_Pandemic_Prob, Primary_Care_Accessibility has the smaller absolute mean growth rate biases regarding the Submit trends.

For the interaction of Poverty_Level and Collect_Sample_Prob (AC), when Poverty_Level is greater than 0.25, lager Collect_Sample_Prob has yielded smaller absolute mean growth rate biases regarding the Submit trends. When Poverty_Level is less than 0.25, Collect_Sample_Prob has similar results over different values, presenting a little larger absolute mean growth biases with lower level of Collect_Sample_Prob (<0.25) values than higher level (>0.5).

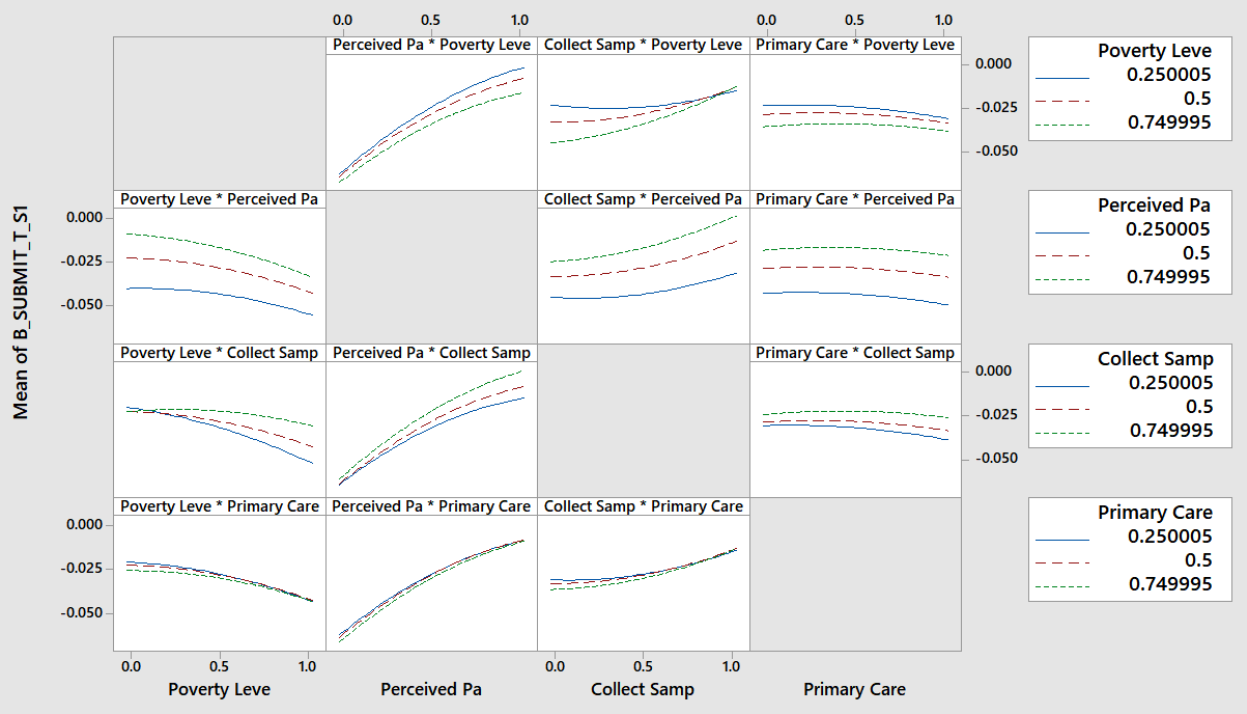
For the interaction of Poverty_Level and Primary_Care_Accessibility (AD), at lower level of Poverty_Level, Primary_Care_Accessibility has the smaller absolute mean growth rate biases regarding the Submit trends.

For the interaction of Perceived_Pandemic_Prob and Collect_Sample_Prob (BC), at higher level of Perceived_Pandemic_Prob, Collect_Sample_Prob has the smaller absolute mean growth rate biases regarding the Submit trends.

For the interaction of Perceived_Pandemic_Prob and Primary_Care_Accessibility (BD), at higher level of Perceived_Pandemic_Prob, Primary_Care_Accessibility has the smaller absolute mean growth rate biases regarding the Submit trends.

For the interaction of Collect_Sample_Prob and Primary_Care_Accessibility (CD), at higher level of Collect_Sample_Prob, Primary_Care_Accessibility has the smaller absolute mean growth rate biases regarding the Submit trends.

a)



b)

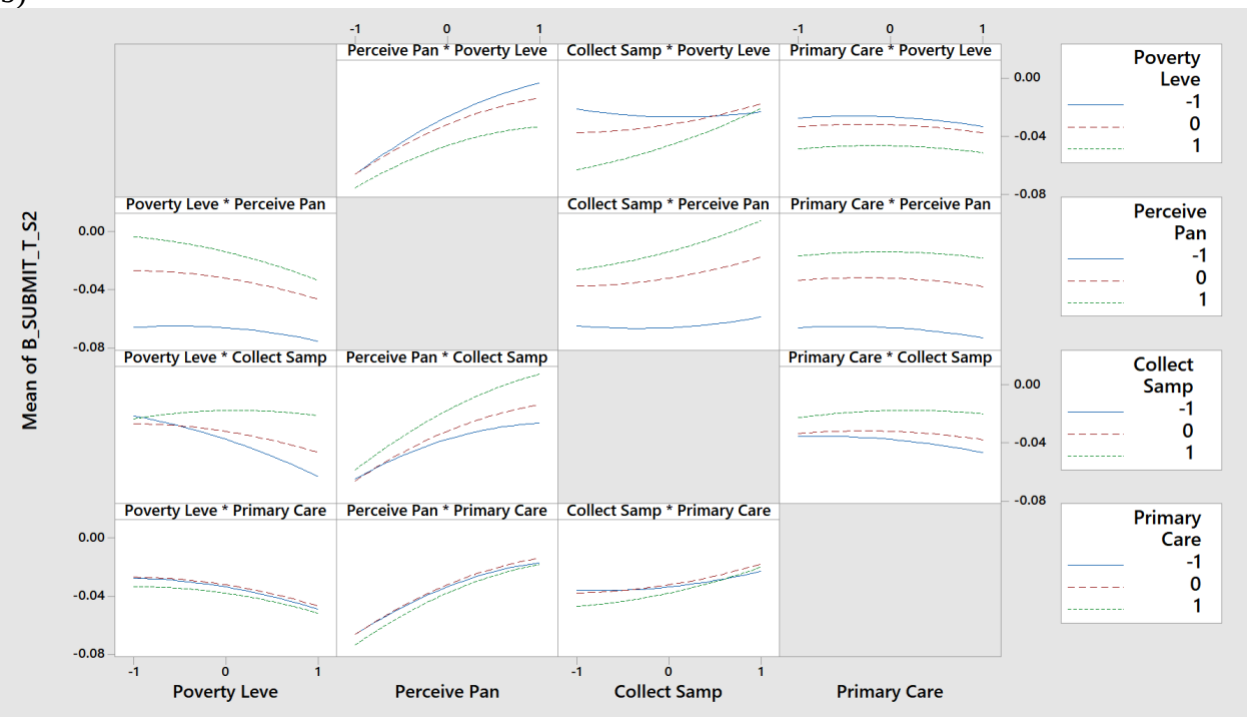
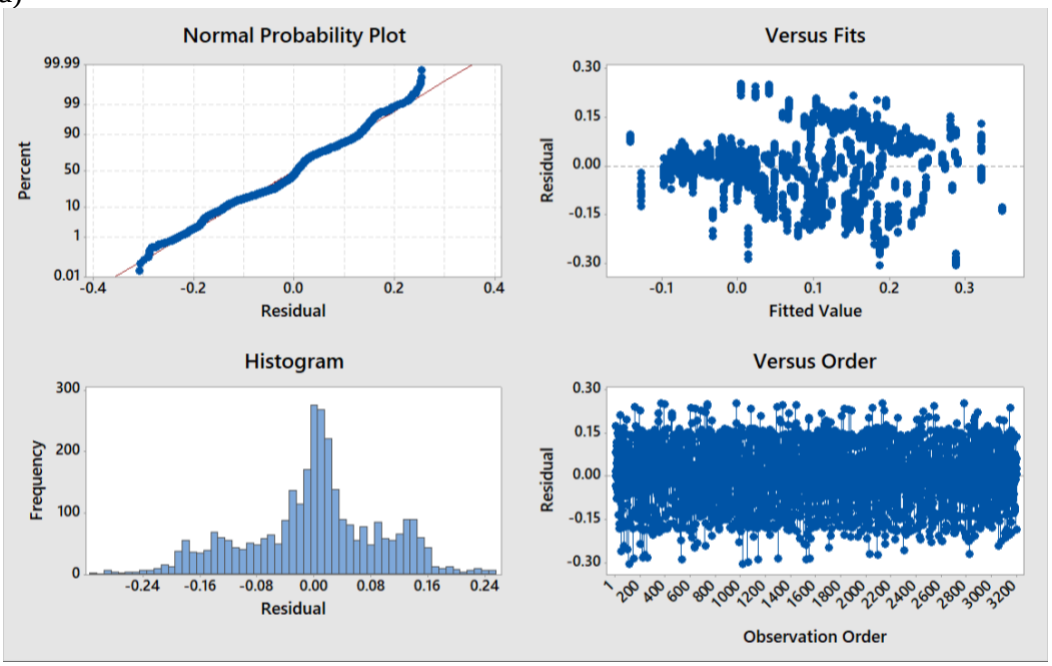


Figure 26. Interaction Effects on Biases Regarding Submit Trends Under Two Scenarios. a) B_Submit_T_S1. b) B_Submit_T_S2

Factors Affecting Biases in PHL Overall Test Trends (B_Test_T)

Specimen tested in the PHL were related with the all 3 stages (health care seeking stage, specimen collection by health care providers, and specimen tested by the PHL). All the seven factors in the three stages: Poverty_Level (A), Perceived_Pandemic_Prob (B), Collect_Sample_Prob (C), and Primary_Care_Accessibility (D), PHL_Capacity (E) PHL_Work_Weekend (F), Sampling_Criteria (G) were related with the process for specimen testing in the PHL. Under both scenarios, the impacts of these seven factors were investigated on the biases of growth rates between the simulated real overall influenza incidence trends and the simulated overall specimen tested trends in the PHL. The residuals check of model fitting met the assumptions of normal distribution, independence and equal variance (Figure 27).

a)



b)

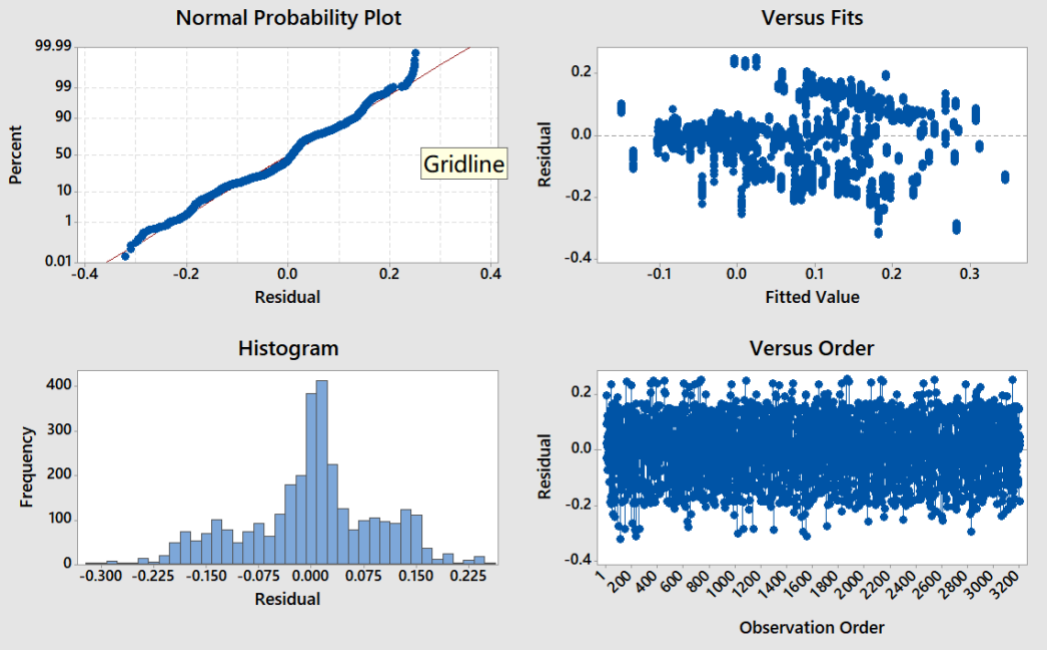


Figure 27. Model Diagnostics for Biases Regarding PHL Overall Test Trends Under Two Scenarios. a) B_Test_T_S1. b) B_Test_T_S2.

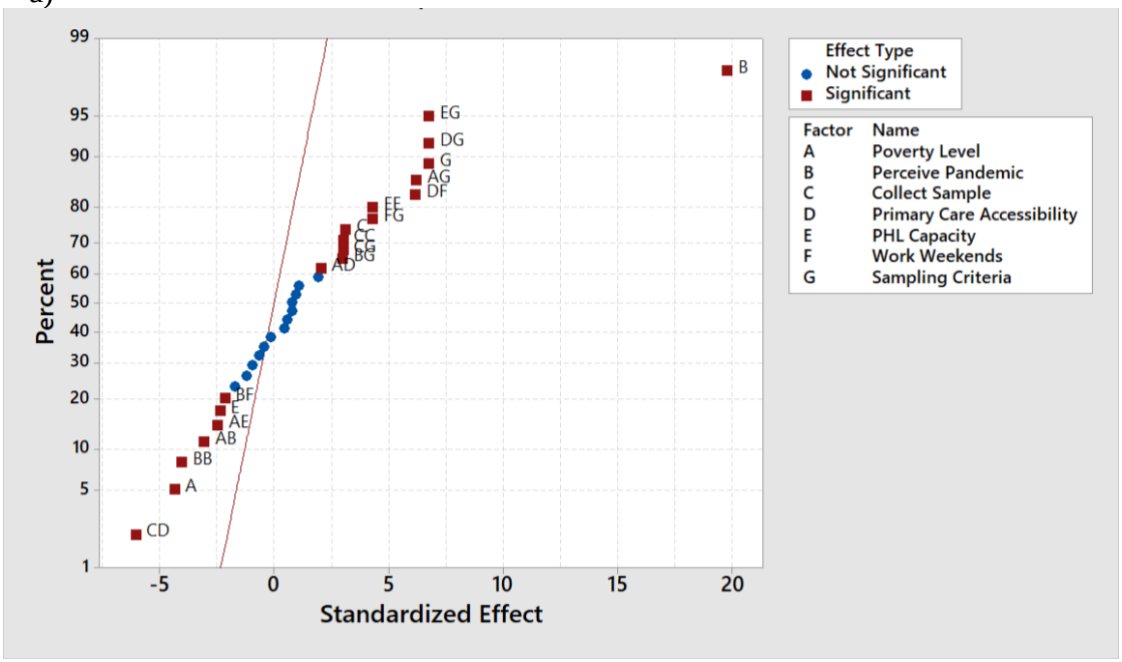
A multiple linear regression was calculated to predict Poverty_Level (A), Perceived Pandemic_Prob (B), Collect_Sample_Prob (C), and Primary_Care_Accessibility (D), PHL_Capacity (E) PHL_Work_Weekend (F), and Sampling_Criteria (G) on the biases of growth rate using the simulated overall influenza specimen tested trends in the PHL. For Scenario 1, a significant regression equation was found ($F(54,3145)=63.06$, $P<0.001$), with an R^2 of 51.99%. For Scenario 2, a significant regression was found ($F(54,3145)=61.94$, $P<0.001$), with an R^2 of 51.54%, R^2(adj) of 50.71%.

The significance of all the terms are shown in Figure 28. For both scenarios, there are statistically significant main effects of Poverty_Level (A), Perceived_Pandemic_Prob (B), Collect_Sample_Prob (C), PHL_Capacity (E) and Sampling_Criteria (G) on the growth rate

biases regarding the overall influenza incidence trends tested by the PHL (Figure 28). Perceived_Pandemic_Prob (B), Collect_Sample_Prob (C), and PHL_Capacity (E) also have statistically significant quadratic effects on the growth rate biases regarding the PHL Overall Test trends.

For both scenarios, there are statistically significant interaction effects of Poverty_Level and Perceived_Pandemic_Prob (AB), Poverty_Level and Primary_Care_Accessibility (AD), Collect_Sample_Prob and Primary_Care_Accessibility (CD), Poverty_Level and PHL_Capacity (AE), Perceived_Pandemic_Prob and PHL_Work_Weekend (BF), Primary_Care_Accessibility and PHL_Work_Weekend (DF), Poverty_Level and Sampling_Criteria (AG), Perceived_Pandemic_Prob, and Sampling_Criteria (BG), Collect_Sample_Prob and

a)



b)

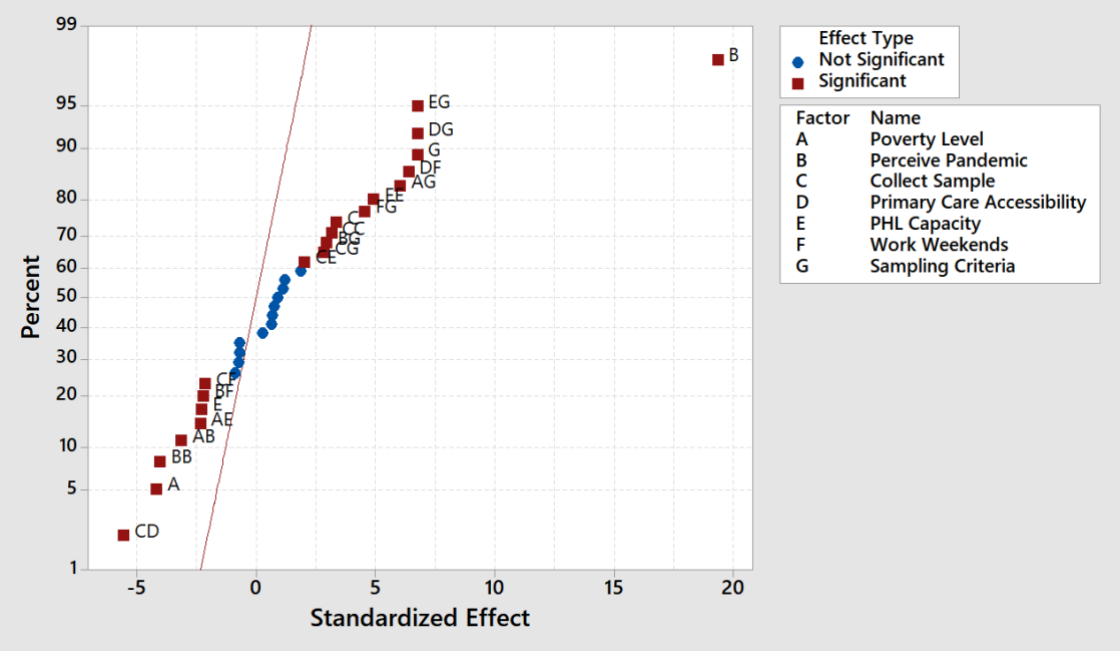


Figure 28. Standardized Effects on Biases Regarding PHL Overall Test Trends Under Two Scenarios. a) B_Test_T_S1. b) B_Test_T_S2.

Sampling_Criteria (CG), Primary_Care_Accessibility and Sampling_Criteria (DG), PHL_Capacity and Sampling_Criteria (EG), PHL_Work_Weekend and Sampling_Criteria (FG). Besides, one more significant interaction effect found in Scenario 2 is Collect_Sample_Prob (C) and PHL_Work_Weekend (CF). The other interaction effects are not statistically significant.

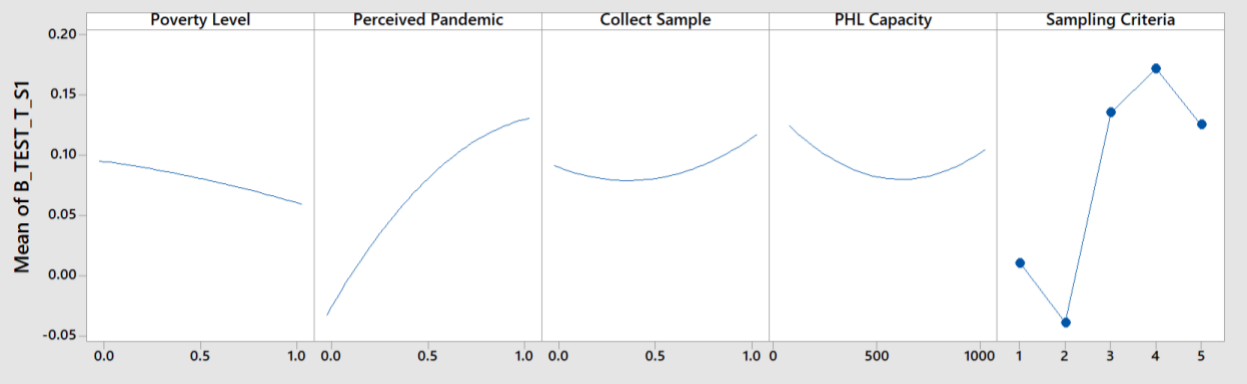
For both scenarios, the main effects plot (Figure 29) shows that Poverty_Level (A) is negatively associated with the growth rate biases regarding the overall Test trends in the PHL within its feasible range 0.00001-0.99999. Perceived_Pandemic_Prob (B) is a positively associated with the growth rate biases regarding the overall Test trends in the PHL within the feasible range 0.00001-0.99999. Collect_Sample_Prob (C) is positively associated with the growth rate biases regarding the overall Test trends within 0.5-0.99999, and negatively

associated within0.00001-0.5. PHL_Capacity (E) is positively associated with the growth rate biases regarding the overall Test trends within 550-1000, and negatively associated within 100-550. The main effect of Sampling_Criteria and the second-order interaction effect between Sampling_Criteria and other factors are presented in next sections.

Factors Affecting Biases in PHL Pandemic Test Trends (B_Test_P)

All the three stages were involved in the tested trends of pandemic influenza in the PHL (health care seeking stage, specimen collection by health care providers, and specimen tested by the PHL) with all the seven factors: Poverty_Level (A), Perceived_Pandemic_Prob (B), Collect_Sample_Prob (C), and Primary_Care_Accessibility (D), PHL_Capacity (E) PHL_Work_Weekend (F), Sampling_Criteria (G). The impacts of these seven factors on the biases of growth rate between the simulated real pandemic influenza incidence trends and

a)



b)

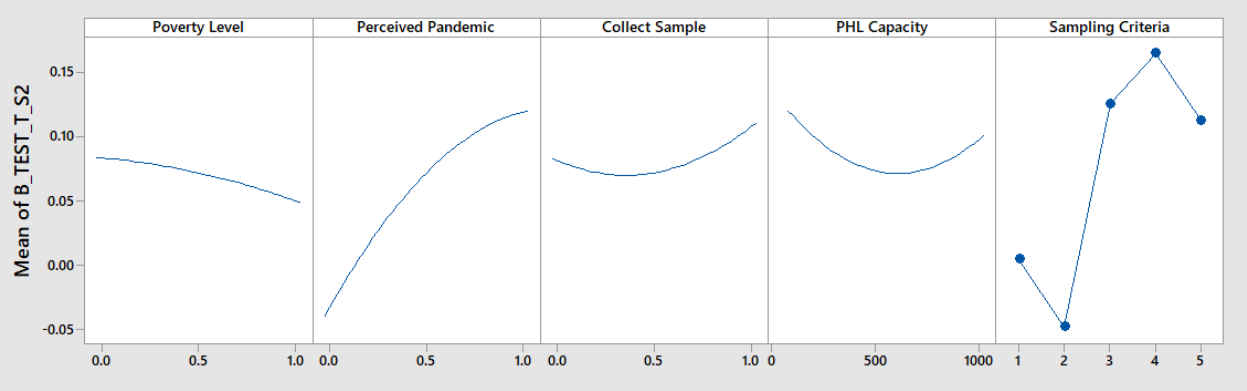
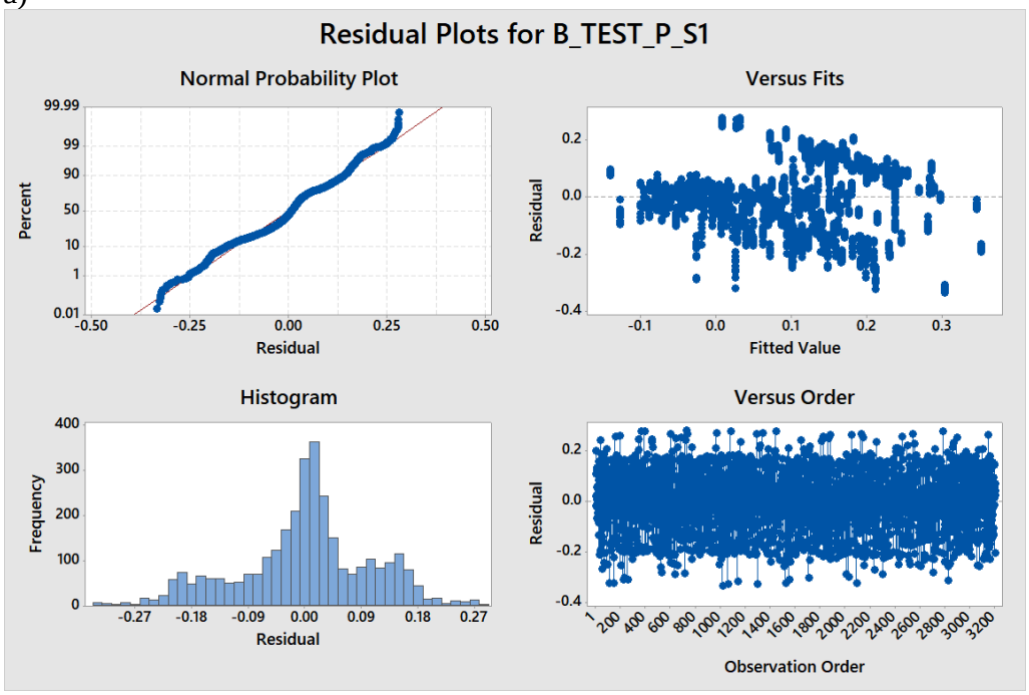


Figure 29. Main Effect on Biases Regarding PHL Overall Test Trends Under Two Scenarios. a) B_Test_T_S1. b) B_Test_T_S2.

investigated under both outbreak scenarios. For both scenarios, the residuals check of model fitting met the assumptions of normal distribution, independence and equal variance (Figure 30).

A multiple linear regression was calculated to predict Poverty_Level (A), Perceived_Pandemic_Prob (B), Collect_Sample_Prob (C), and Primary_Care_Accessibility (D), PHL_Capacity (E) PHL_Work_Weekend (F), Sampling_Criteria (G) on the biases of growth rate regarding the simulated pandemic influenza specimen tested trends in the PHL under both influenza outbreak scenarios. For scenario 1, a significant regression was found ($F(54,3145)=52.94$, $P<0.001$), with an R^2 of 47.62%, and R^2(adj) of 46.72%. For

a)



b)

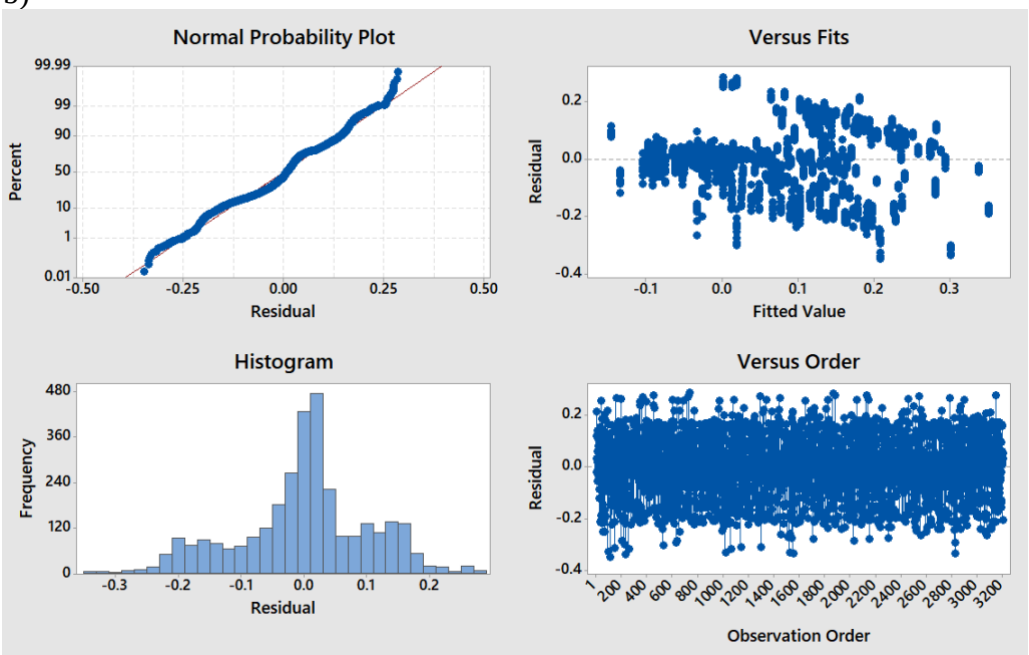


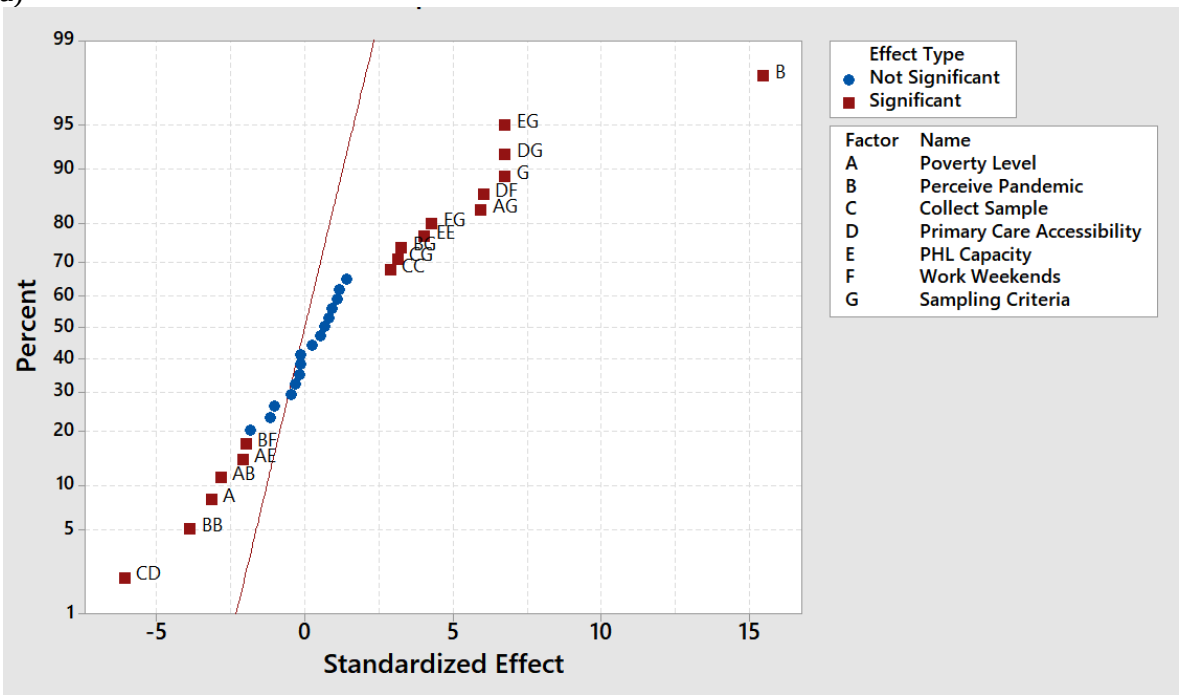
Figure 30. Model Diagnostics for Biases Regarding PHL Pandemic Test Trends Under Two Scenarios. a) B_Test_P_S1. b) B_Test_P_S2.

scenario 2, a significant regression was found ($F(54,3145)=53.51$, $P<0.001$), with an R^2 of 47.88%.

The significance of all the terms are shown in Figure 31. For both scenarios, there are statistically significant main effects of Poverty_Level (A), Perceived_Pandemic_Prob (B), and Sampling_Criteria (G). There are significant quadratic effects of Perceived_Pandemic_Prob (B), Collect_Sample_Prob (C), PHL_Capacity (E).

For both scenarios, there were significant interaction effects of Poverty_Level and Perceived_Pandemic_Prob (AB), Collect_Sample_Prob and Primary_Care_Accessibility (CD), Poverty_Level and PHL_Capacity (AE), Primary_Care_Accessibility and PHL_Work_Weekend (DF), Poverty_Level and Sampling_Criteria (AG), Perceived_Pandemic_Prob and Sampling_Criteria (BG), Collect_Sample_Prob and Sampling_Criteria (CG), Primary_Care_Accessibility and Sampling_Criteria (DG), PHL_Capacity and Sampling_Criteria (EG), and PHL_Work_Weekend and Sampling_Criteria (FG). In addition to the common significant interaction terms, for scenario 1, Perceived_Pandemic_Prob and PHL_Work_Weekend (BF) is significant, and for scenario 2, Collect_Sample_Prob and PHL_Work_Weekend (CF). Similarly, these two interaction terms have very small effect size. The other interaction effects are not statistically significant.

a)



b)

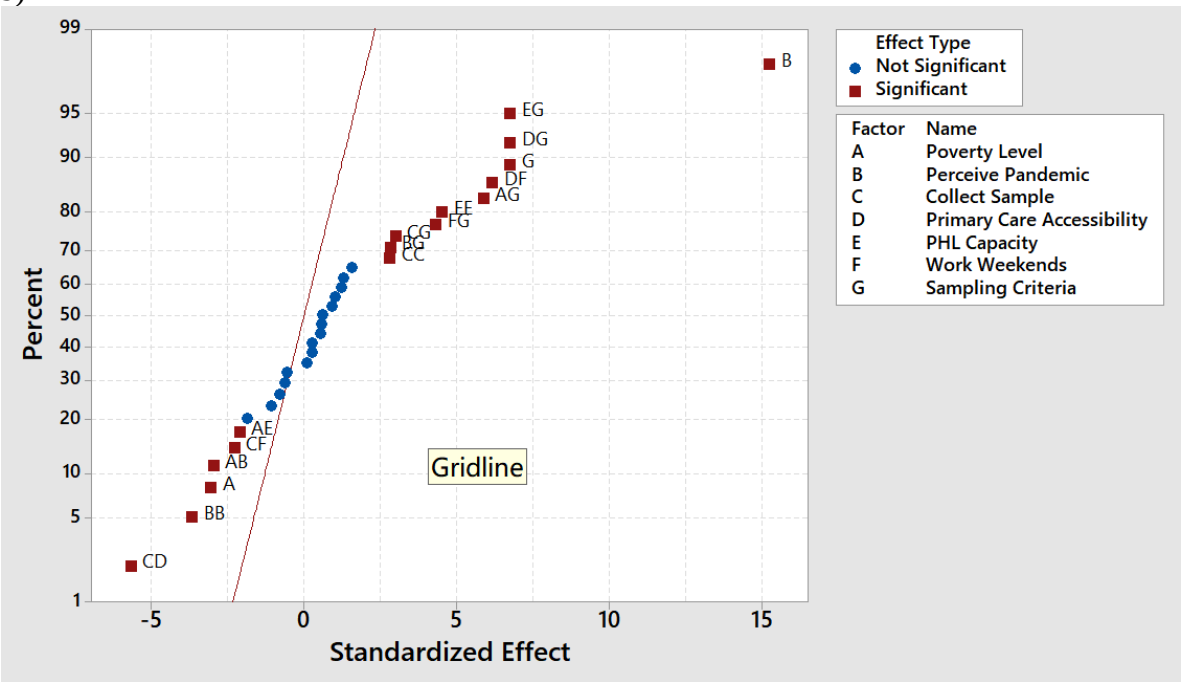
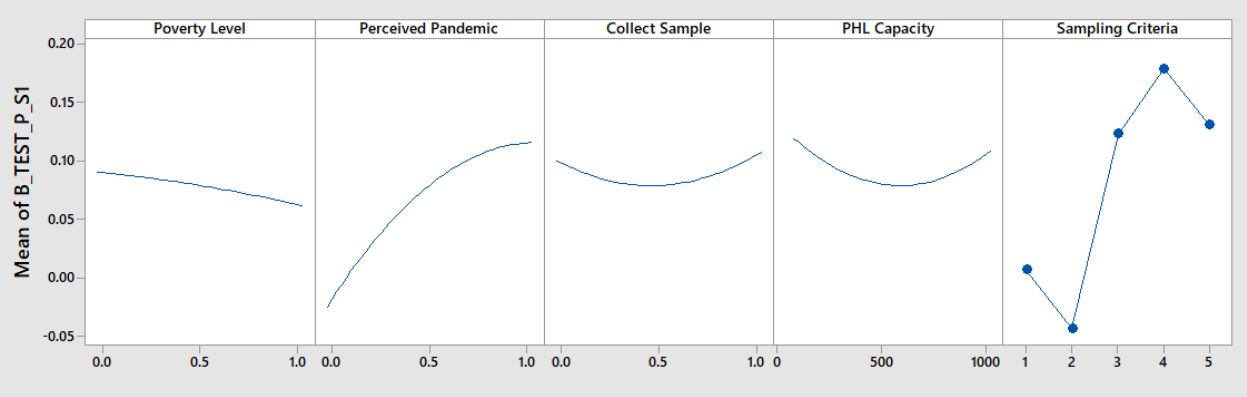


Figure 31. Standardized Effects on Biases Regarding PHL Pandemic Test Trends Under Two Scenarios. a) B_Test_P_S1. b) B_Test_P_S2.

Under both scenarios, the main effect plots (Figure 32) show that within the feasible region of these factors, Poverty_Level (A) is negatively associated with B_Test_P in the range of 0.00001 and 0.99999. There is a positive relationship between Perceived_Pandemic_Prob (B) and B_Test_P in the range of 0.00001 and 0.9999. There is a positive relationship between Collect_Sample_Prob (C) and B_Test_P in the range of 0.5 – 0.99999, otherwise negative. There is a positive relationship between PHL_Capacity (E) and B_Test_P in the range of 0.5 – 0.99999, otherwise negative. The main effect of Sampling_Criteria and the second-order interaction effect between Sampling_Criteria and other factors are presented in next sections.

a)



b)

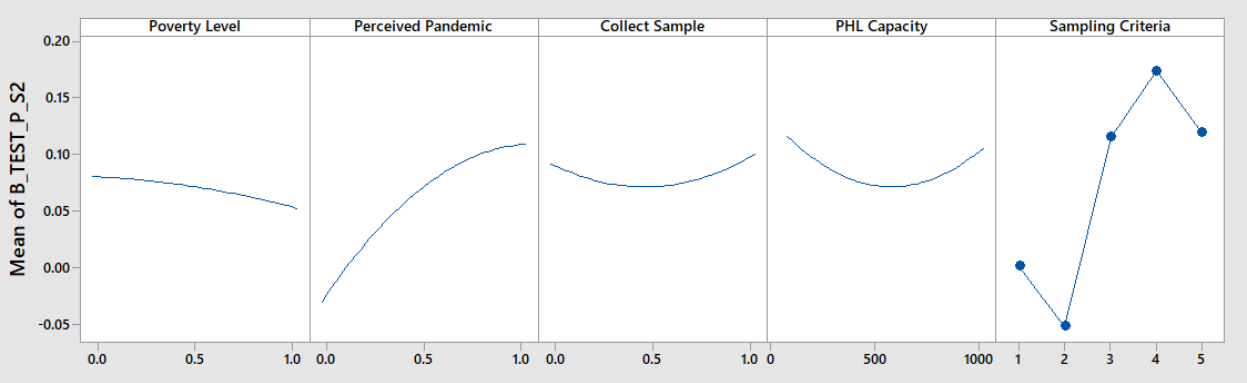


Figure 32. Main Effect on Biases Regarding PHL Pandemic Test Trends Under Two Scenarios. a) B_Test_P_S1. b) B_Test_P_S2.

Sensitivity Analysis

The sensitivity analysis were replaced by analyzing the main effect of each operational factor on the corresponding growth rate bias using GLM. From the results, all the involved factors have significant main effects on the growth rate biases regarding the MSSS trends and the Submit trends. For the PHL test trends (both Overall and Pandemic), all the factors except for Primary_Care_Accessiblity and PHL_Work_Week have significant main effects. However, since the interaction effects show that both factors affect the effect of other factors on the growth rate biases, they are still significant in the current model. All these factors are appropriately included in the simulation and statistically analysis for the growth rate biases in each surveillance trends.

Results of Research Question 3

Influence of Sampling_Criteria

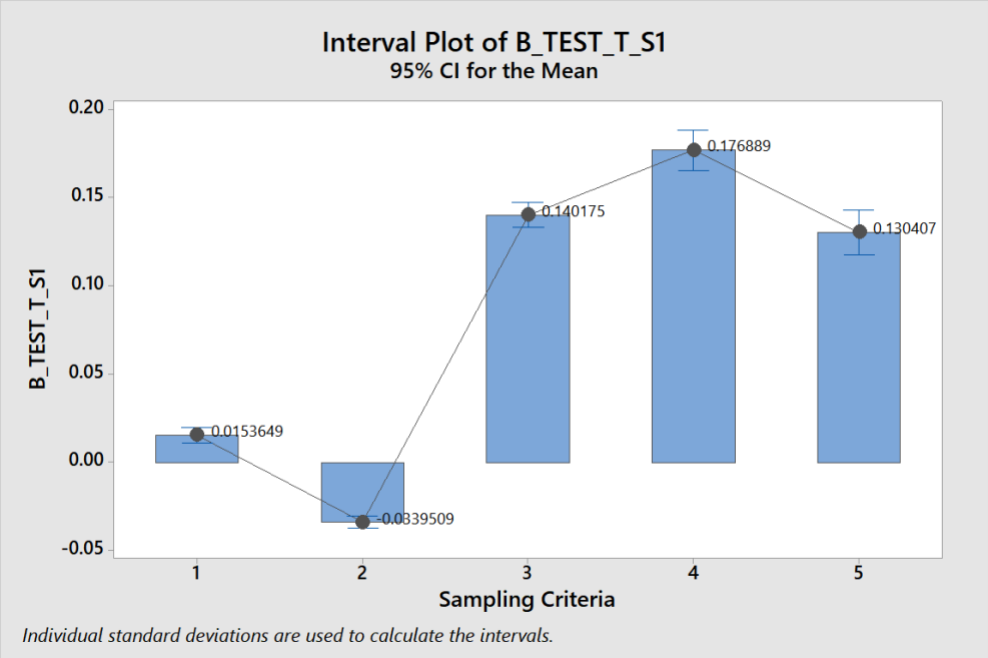
Five sampling criteria that the PHL used to select and test the specimens are 1) FCFS: using first-come-first-serve (FCFS) to test all submitted specimens, 2) Restricted FCFS: using FCFS to test only the submitted specimens from high-risk population groups, 3) Hybrid: switching FCFS to Restricted FCFS when the PHL_Capacity is saturated, 4) 1 WK FCFS: only testing the specimens submitted in the first week using FCFS, 5) 1WK Restricted FCFS: only testing the specimens submitted in the first week using Restricted FCFS (1WK Restricted FCFS).

The influence of the five sampling criteria was investigated on the biases between the growth rates of the simulated real overall/pandemic influenza incidence trends and the simulated tested overall/pandemic influenza specimen trends in PHL. Figure 33 shows the means and

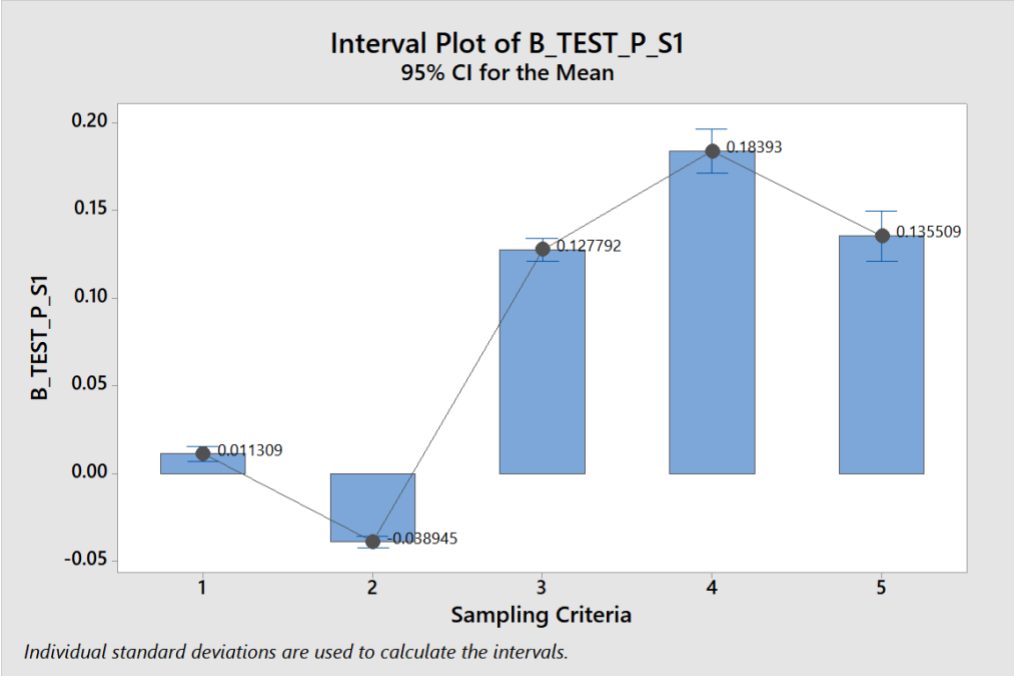
confidence intervals of the growth rate biases regarding PHL Overall Test trends and PHL Pandemic Test trends under two influenza outbreak scenarios.

All the sampling criteria resulted the means of growth rate biases that were statistically different from zero, meaning that none of these sampling criteria eliminated the biases between the growth rates. The absolute means obtained by the sampling criteria are ordered from smallest to largest: FCFS, Restricted FCFS, Hybrid and 1WK Restricted FCFS, sampling criteria 4 1WK FCFS. Note that there is no statistical difference between Hybrid and WK Restricted FCFS. The smallest mean growth rates biases were obtained by FCFS. Besides, except for Restricted FCFS that yielded larger growth rates of the tested trends in the PHL, all the other sampling criteria yielded smaller growth rates of the PHL tested trends. Under the current specimen collection and sampling criteria, the growth rate estimates are either over-estimated by Restricted FCFS or under-estimated by all the other sampling criteria. The results are similar for both scenarios.

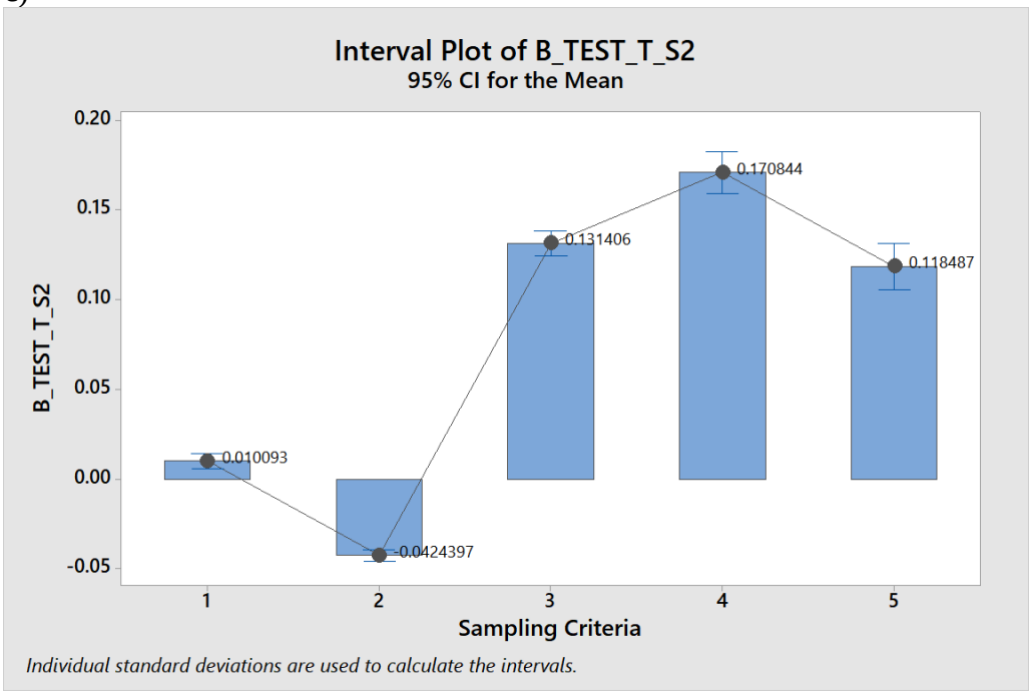
a)



b)


Interval Plot of B_TEST_P_S1
95% CI for the Mean
B_TEST_P_S1
0.20
0.15
0.10
0.05
0.00
-0.05
0.18393
0.135509
0.127792
0.011309
-0.038945
1
2
3
4
5
Sampling Criteria
Individual standard deviations are used to calculate the intervals.

c)


Interval Plot of B_TEST_T_S2
95% CI for the Mean
B_TEST_T_S2
0.20
0.15
0.10
0.05
0.00
-0.05
0.170844
0.131406
0.118487
0.010093
-0.0424397
1
2
3
4
5
Sampling Criteria
Individual standard deviations are used to calculate the intervals.

d)

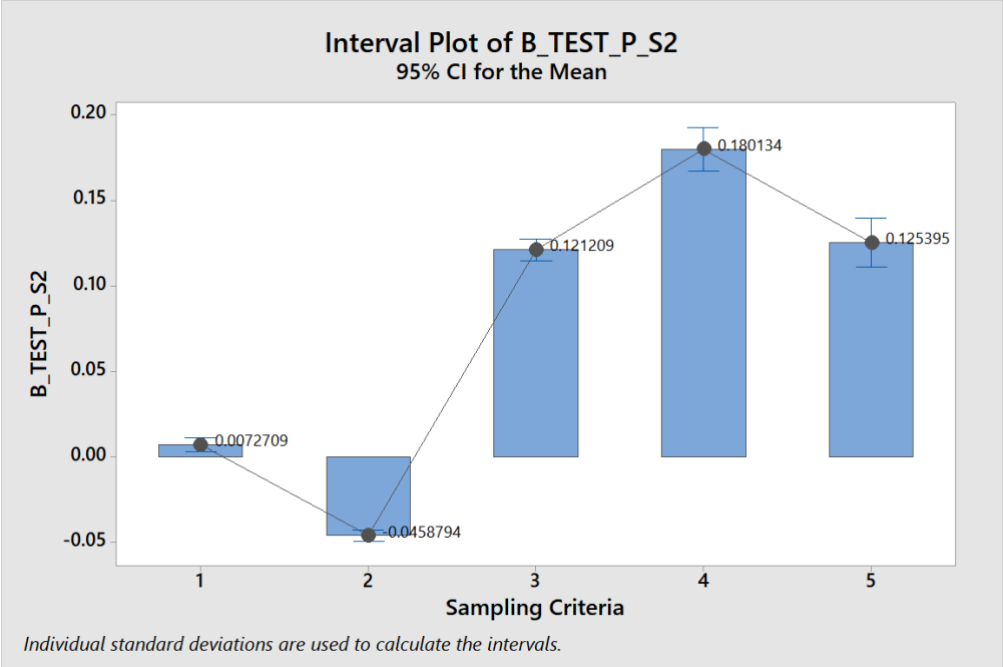


Figure 33. Interval Plot of Mean Bar and Confidence Interval versus Sampling_Criteria on Biases Regarding PHL Overall/Pandeimic Test Trends. Note: 1=FCFS, 2=Restricted FCFS, 3=Hybrid, 4=1WK FCFS, 5=1WK Restricted FCFS. a) B_Test_T_S1. b) B_Test_P_S1; c) B_Test_T_S2; d) B_Test_P_S2.

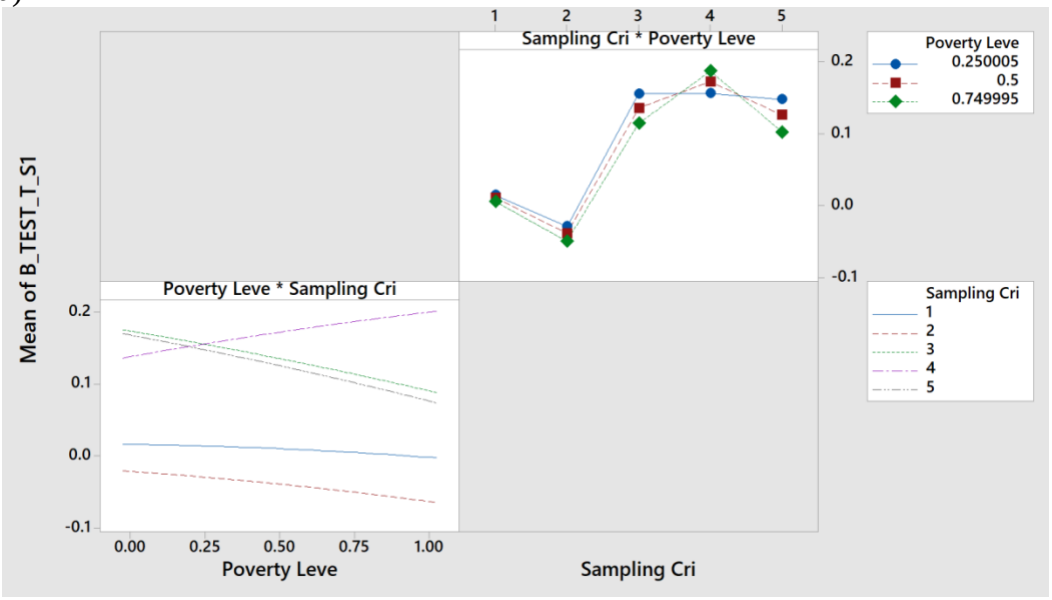
Influence of Sampling_Criteria When Interacting with Other Variables

Poverty_Level

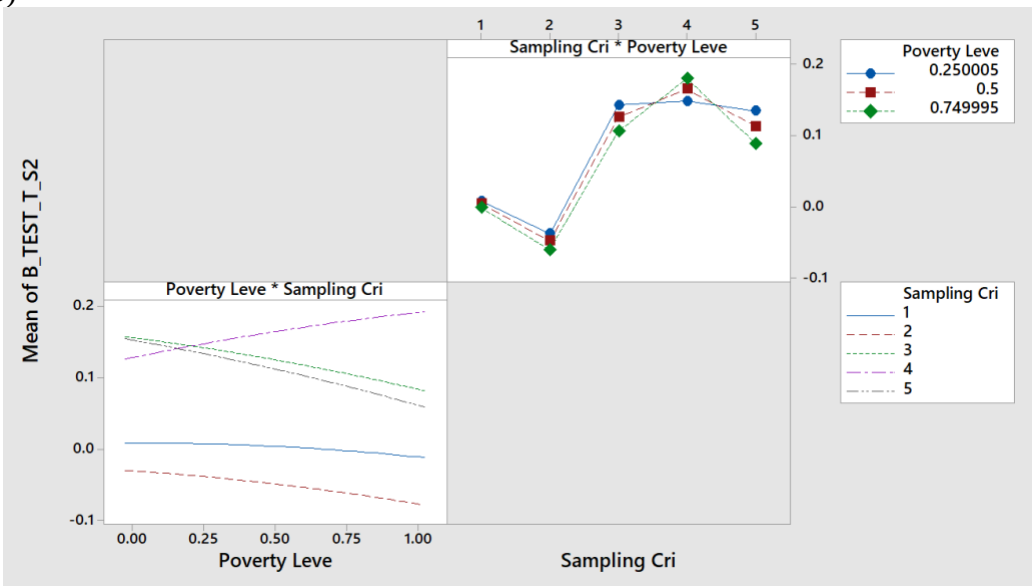
From Figure 34, when Poverty_Level was from 0.00001 to 0.99999, Sampling_Criteria 1 (FCFS) obtained the smallest absolute means growth rate biases for all these four response variables. When Poverty_Level is at 0.25, Sampling_Criteria 2 (Restricted FCFS) also achieved as small mean growth rate biases as Sampling_Criteria 1 (FCFS). The mean growth rate bias obtained by Sampling_Criteria 4 (1WK FCFS) is not statistically different from Sampling_Criteria 3 (Hybrid) or Sampling_Criteria 5 (1WK Restricted FCFS). However, when

Poverty_Level is between 0.00001 and 0.25, Sampling_Criteria 4 (1WK FCFS) yielded smaller mean growth rate biases than Sampling_Criteria 3 and 5.

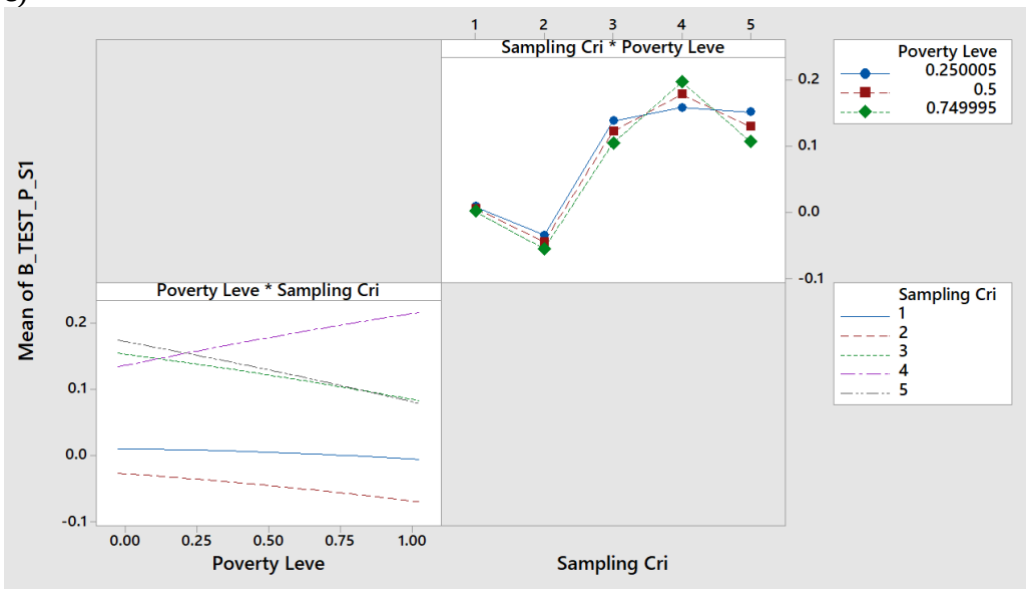
a)



b)



c)



d)

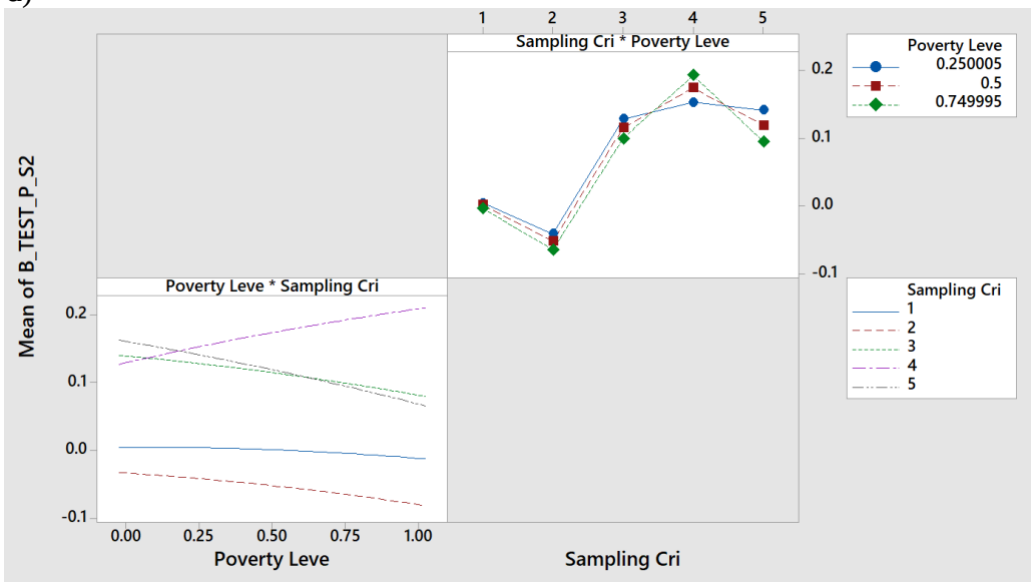
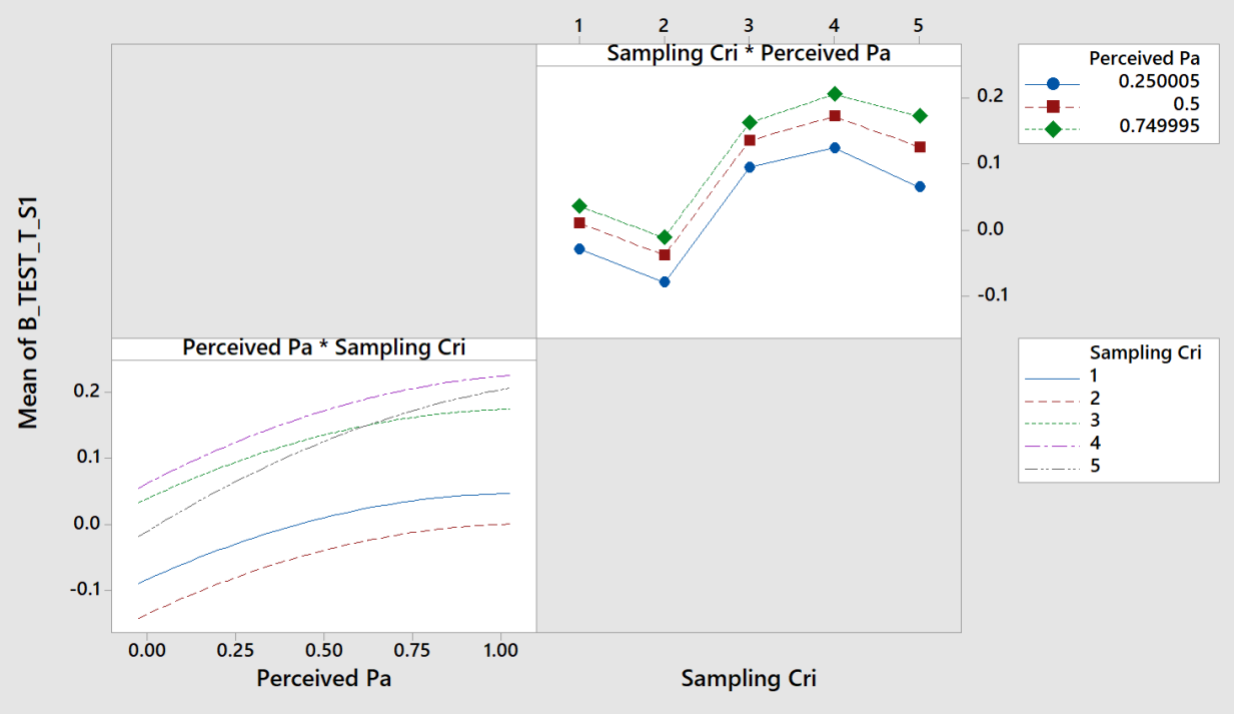


Figure 34. Interaction Plot of Sampling_Criteria and Poverty_Level on Biases Regarding PHL Overall/Pandemic Test Trends. a) B_Test_T_S1. b) B_Test_T_S2. c) B_Test_P_S1. d) B_Test_P_S2.

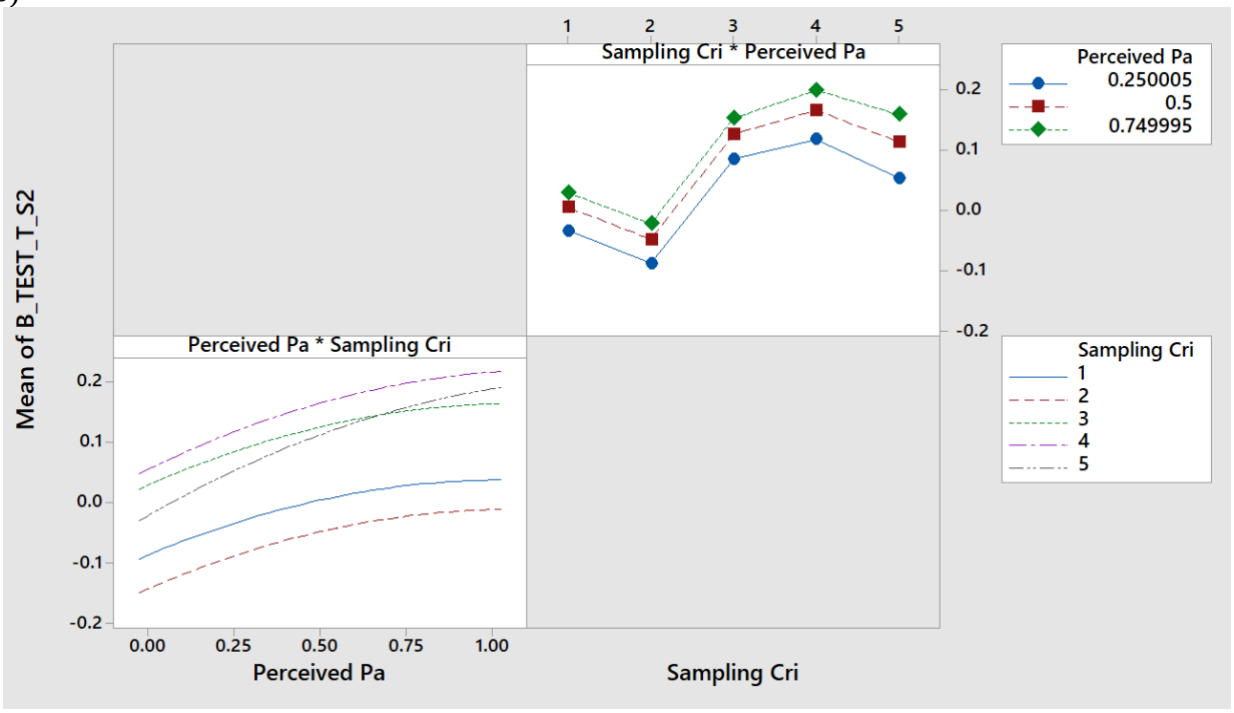
Perceived_Pandemic_Prob

From Figure 35, when Perceived_Pandemic_Prob is between 0.25 and 0.75, the absolute mean growth rate biases obtained by Sampling_Criteria 1 (FCFS) smallest. When Perceived_Pandemic_Prob is between 0.00001 and 0.25, Sampling_Criteria 5 (1WK Restricted FCFS) obtained smallest absolute mean growth rate biases. When Perceived_Pandemic_Prob is between 0.75 and 0.99999, the absolute mean growth rate bias obtained by Sampling_Criteria 2 (Restricted FCFS) are smallest. The relationship between Perceived_Pandemic_Prob and mean growth rate biases change faster when using Sampling_Criteria 3. Within 0.75-0.99999 of Perceived_Pandemic_Prob, Sampling_Criteria 5 (1WK Restricted FCFS) obtained larger absolute mean growth rate biases than Sampling_Criteria 3 (Hybrid).

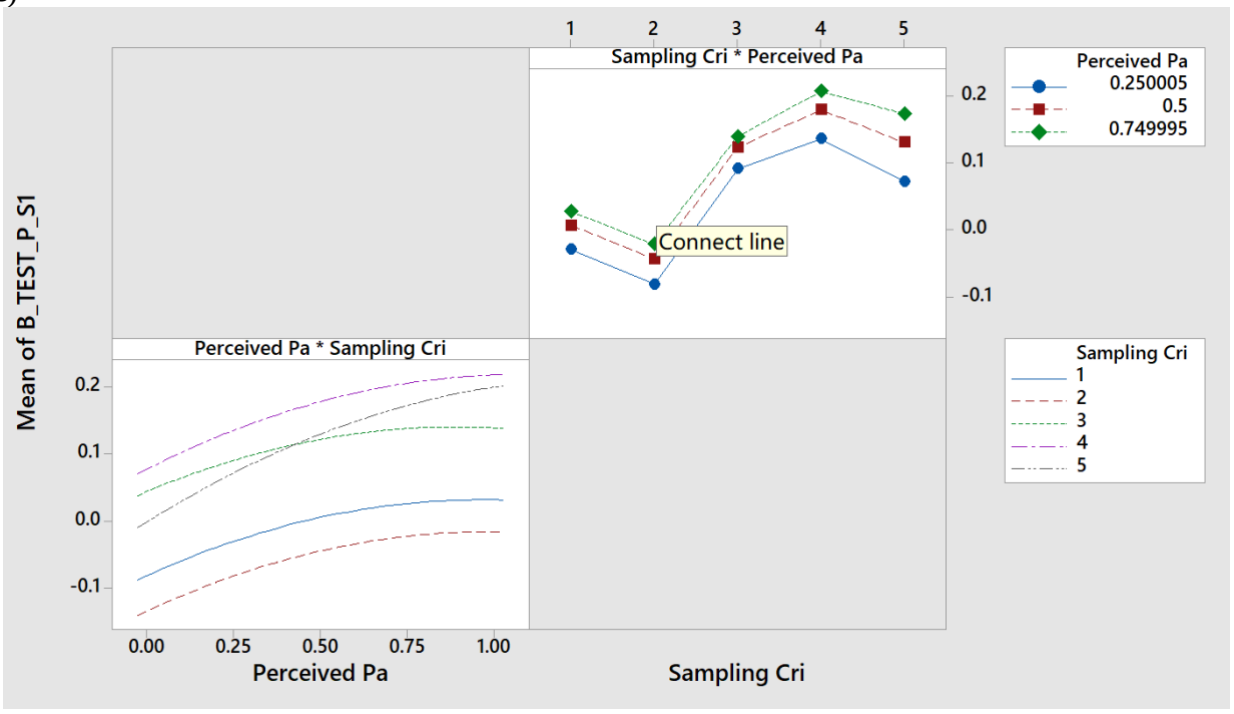
a)



b)


Sampling Cri * Perceived Pa
Perceived Pa * Sampling Cri
Mean of B_TEST_T_S2
Perceived Pa
Sampling Cri
Perceived Pa
0.250005
0.5
0.749995
Sampling Cri
1
2
3
4
5

c)


Sampling Cri * Perceived Pa
Perceived Pa * Sampling Cri
Mean of B_TEST_P_S1
Connect line
Perceived Pa
Sampling Cri
Perceived Pa
0.250005
0.5
0.749995
Sampling Cri
1
2
3
4
5

d)

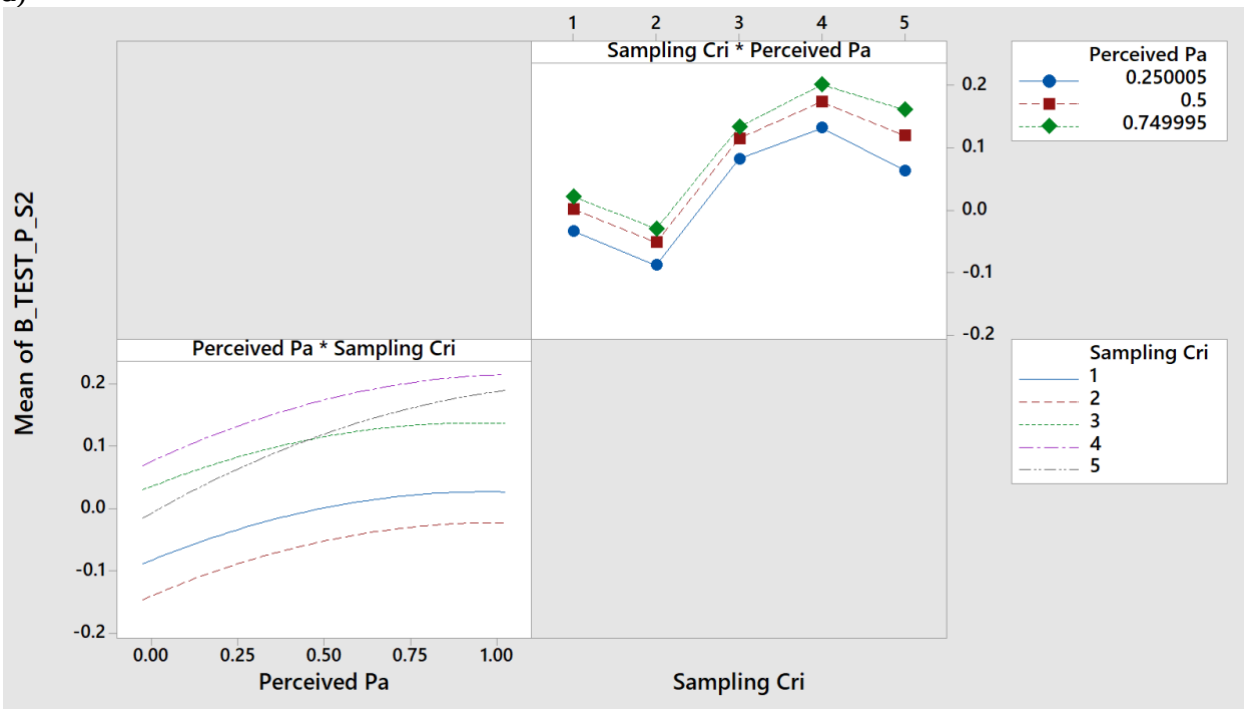


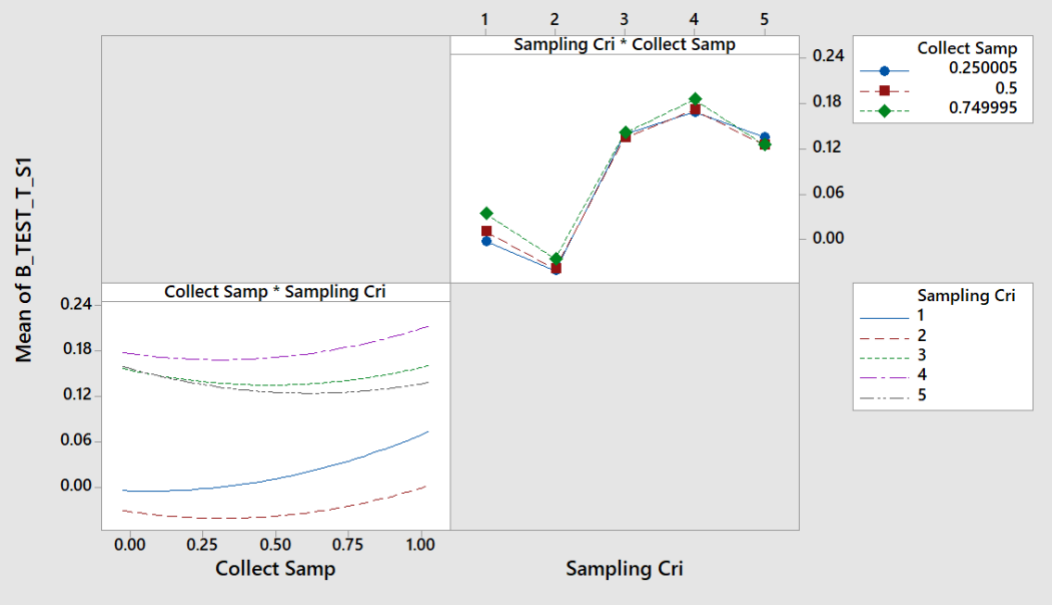
Figure 35. Interaction Plot of Sampling_Criteria and Perceived_Pandemic_Prob on Biases Regarding PHL Overall/Pandemic Test Trends. a) B_Test_T_S1. b) B_Test_T_S2. c) B_Test_P_S1. d) B_Test_P_S2.

Collect_Sample_Prob

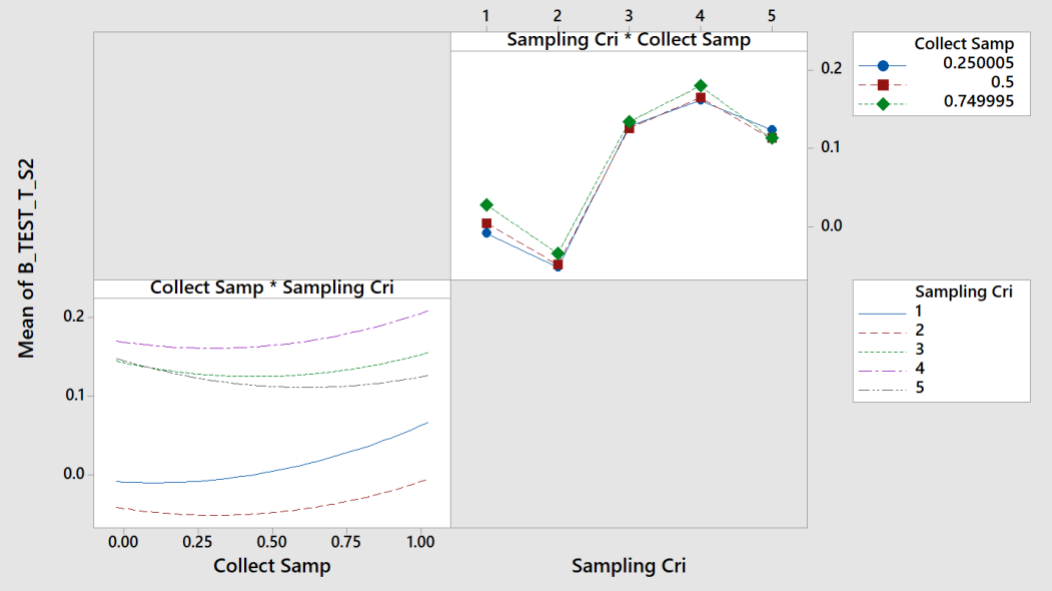
From Figure 36, when Collect_Sample_Prob is at 0.75 or lower, the absolute mean growth rate biases obtained by Sampling_Criteria 1 (FCFS) are smallest. When Collect_Sample_Prob is at 0.75 or higher, the absolute mean growth rate biases obtained by Sampling_Criteria 2 (Restricted FCFS) are smallest, better than Sampling_Criteria 1 (FCFS). When Collect_Sample_Prob is at 0.25 or lower, the absolute mean growth rate biases obtained by Sampling_Criteria 3 (Hybrid) and 5 (1WK Restricted FCFS) are same. When Collect_Sample_Prob is at 0.25 or higher, Sampling_Criteria 5 (1WK Restricted FCFS) obtained

smaller growth biases than Sampling_Criteria 3 (Hybrid). Sampling_Criteria 4 (1WK FCFS) always returned largest growth rate biases.

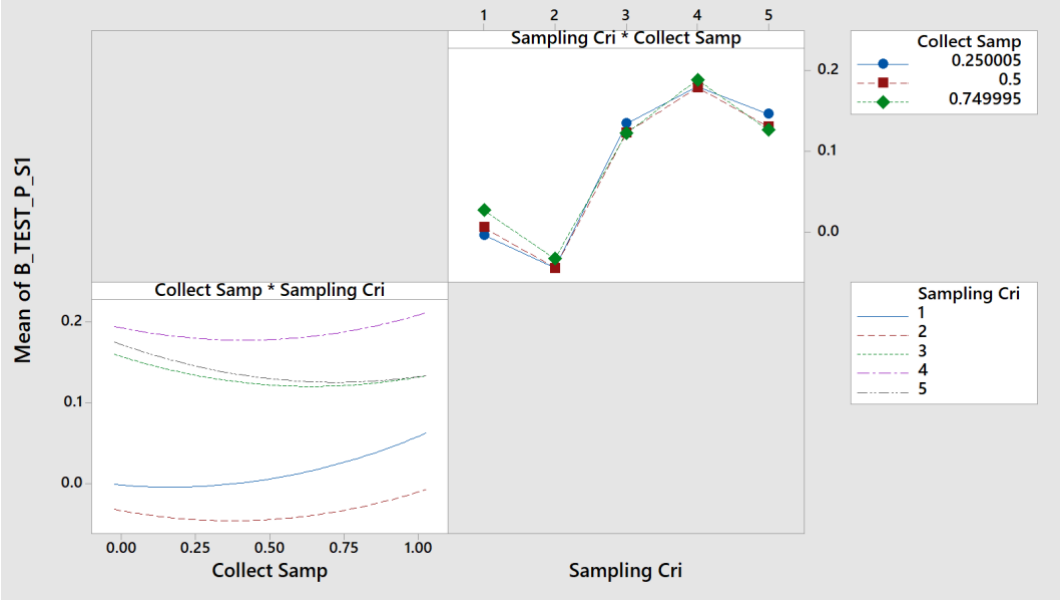
a)



b)



c)



d)

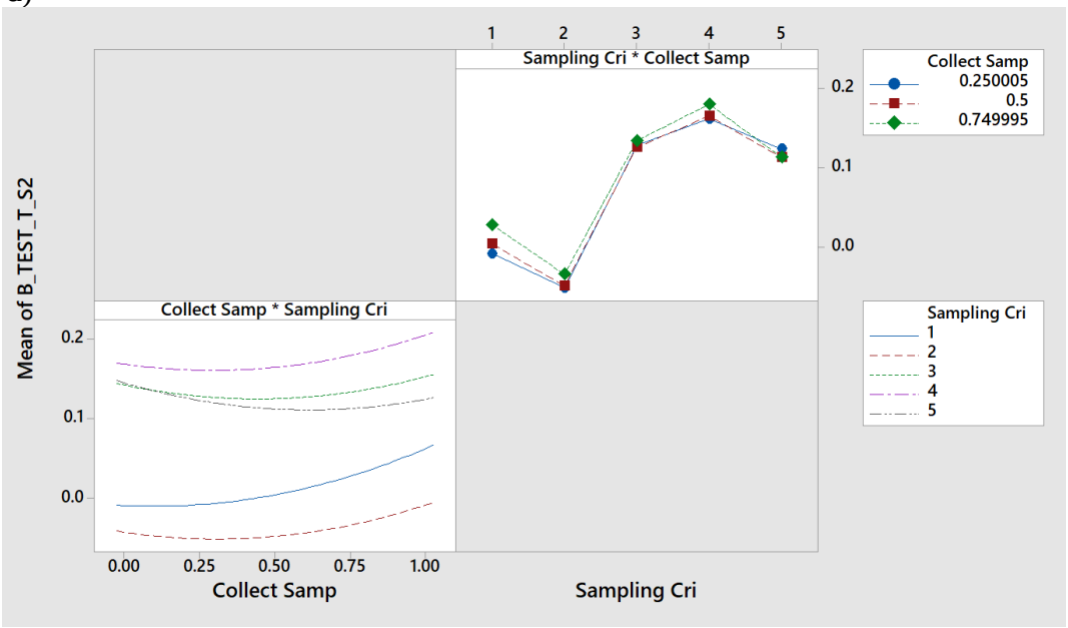
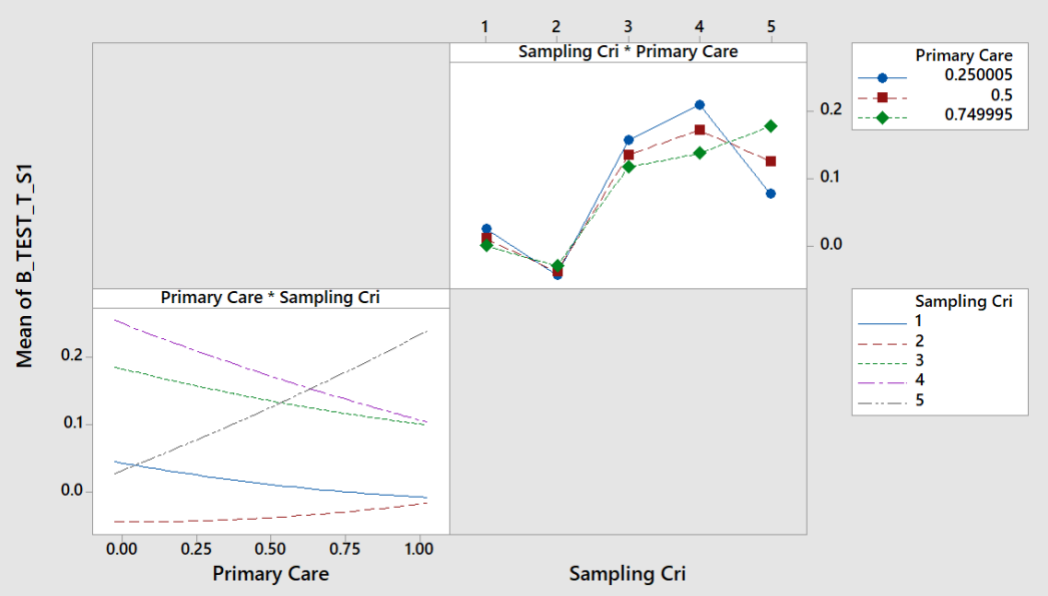


Figure 36. Interaction Plot of Sampling_Criteria and Collect_Sample_Prob on Biases Regarding PHL Overall/Pandemic Test Trends. a) B_Test_T_S1. b) B_Test_T_S2. c) B_Test_P_S1. d) B_Test_P_S2,

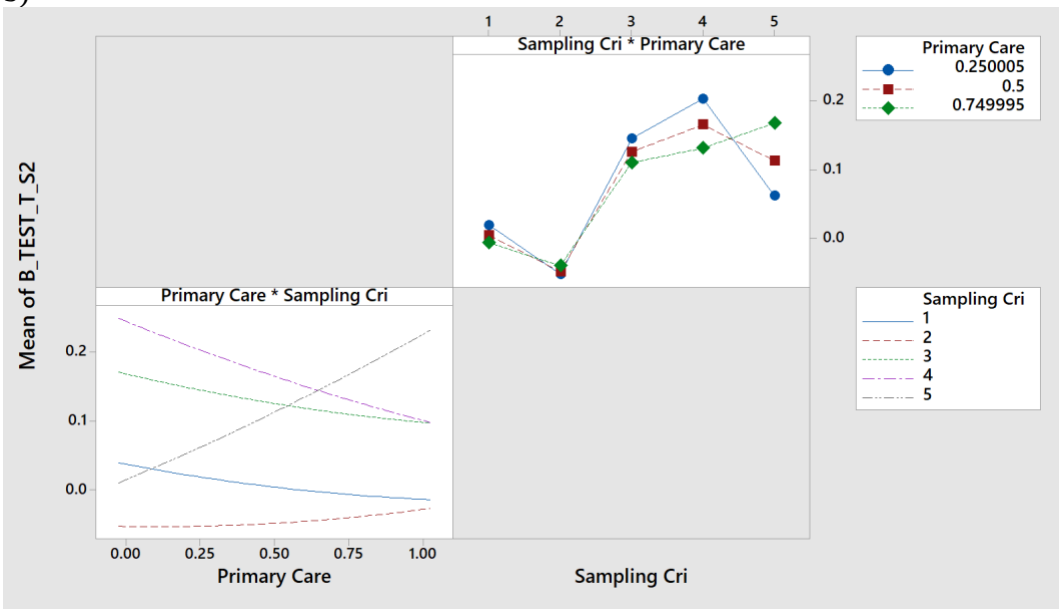
Primary_Care_Accessibility

From Figure 37, when Primary_Care_Accessibility is between 0.00001 and 0.99999, both Sampling_Criteria 1 (FCFS) had smallest absolute mean growth rate biases, Sampling_Criteria 2 (Restricted FCFS) second smallest. But when Primary_Care_Accessibility is as small as 0.00001, Sampling_Criteria 5 (1WK FCFS) had smallest absolute mean growth rate biases. When Primary_Care_Accessibility is between 0.00001 and 0.5, Sampling_Criteria 5 (1WK Restricted FCFS) had smaller absolute mean growth rate biases than Sampling_Criteria 3 and 4.When Primary_Care_Accessibility is between 0.5 and 0.99999, Sampling_Criteria 5 (1WK Restricted FCFS) had larger absolute mean growth rate than Sampling_Criteria 3 and 4.

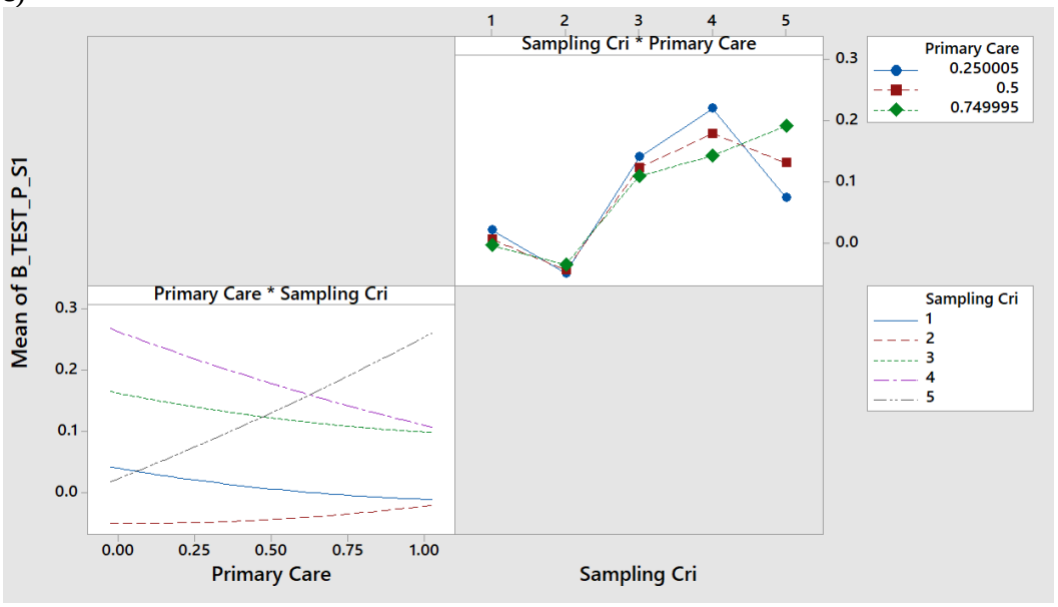
a)



b)


Sampling Cri * Primary Care
Primary Care * Sampling Cri
Mean of B_TEST_T_S2
Primary Care
Sampling Cri
Primary Care
0.250005
0.5
0.749995
Sampling Cri
1
2
3
4
5

c)


Sampling Cri * Primary Care
Primary Care * Sampling Cri
Mean of B_TEST_P_S1
Primary Care
Sampling Cri
Primary Care
0.250005
0.5
0.749995
Sampling Cri
1
2
3
4
5

d)

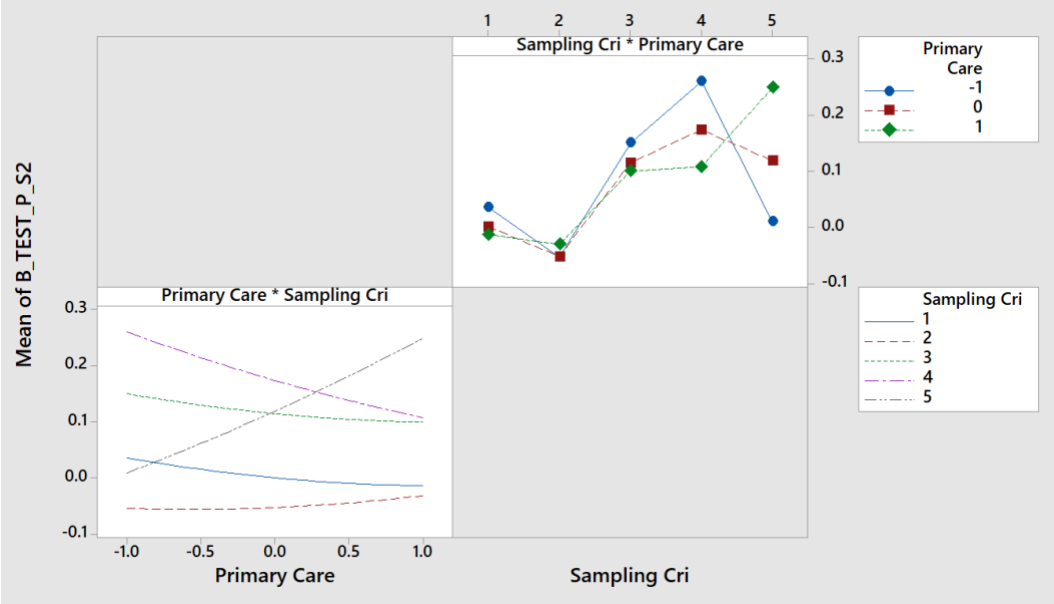
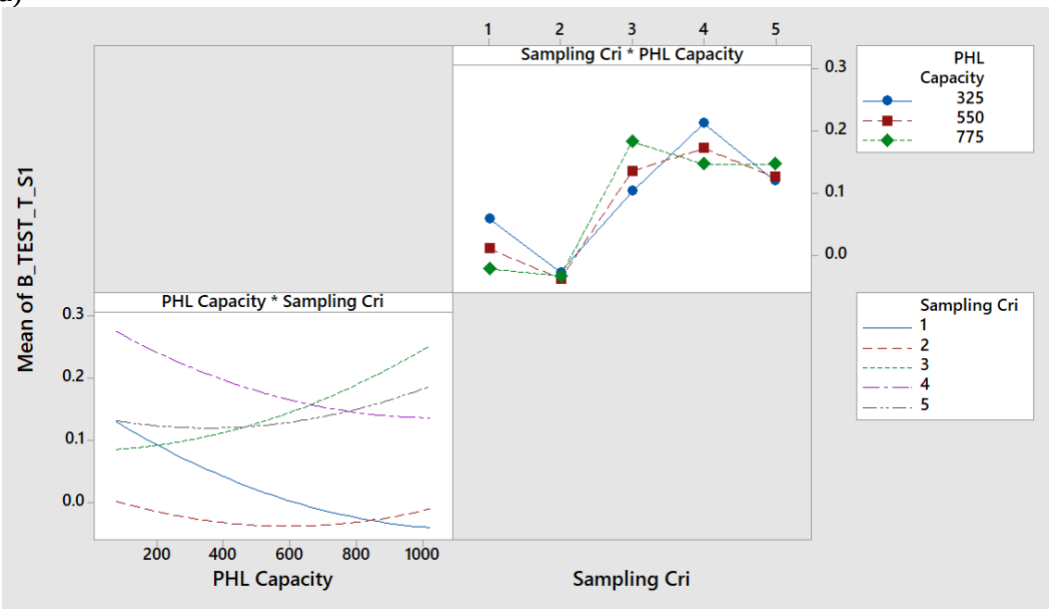


Figure 37. Interaction Plot of Sampling_Criteria and Primary_Care_Accessibility on Biases Regarding PHL Overall/Pandemic Test Trends. a) B_Test_T_S1. b) B_Test_T_S2. c) B_Test_P_S1. d) B_Test_P_S2.

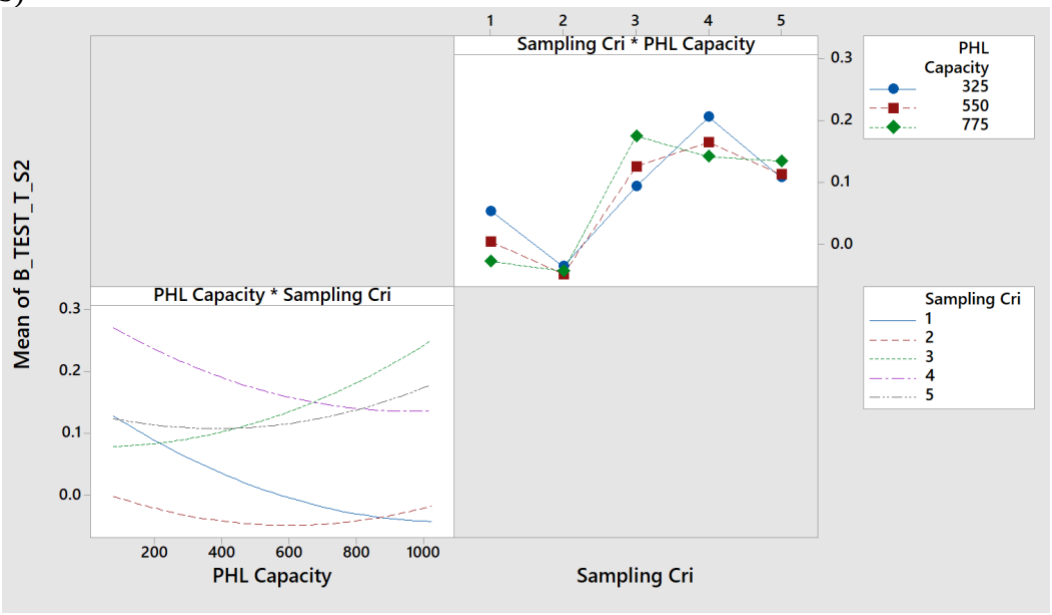
PHL_Capacity

From Figure 38, when PHL_Capacity is between 550 and 1000, Sampling_Criteria 1 (FCFS) obtained smallest absolute mean growth rate biases. When PHL_Capacity is between 100 and 550, Sampling_Criteria 2 (Restricted FCFS) had smallest absolute mean growth rate biases. When PHL_Capacity is between 775 and 1000, Sampling_Criteria 3 (Hybrid) obtained larger absolute mean growth rate biases than Sampling_Criteria 4 (1WK FCFS) and 5 (1WK Restricted FCFS).

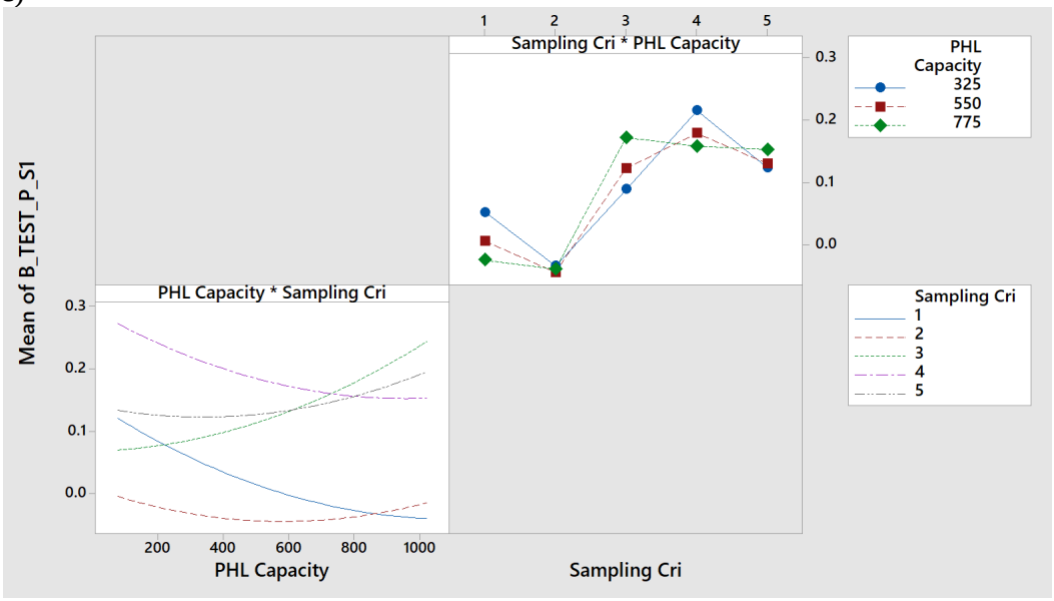
a)


Mean of B_TEST_T_S1
Sampling Cri * PHL Capacity
PHL Capacity * Sampling Cri
PHL Capacity
325
550
775
Sampling Cri
1
2
3
4
5
0.0
0.1
0.2
0.3
200
400
600
800
1000
PHL Capacity
Sampling Cri

b)


Mean of B_TEST_T_S2
Sampling Cri * PHL Capacity
PHL Capacity * Sampling Cri
PHL Capacity
325
550
775
Sampling Cri
1
2
3
4
5
0.0
0.1
0.2
0.3
200
400
600
800
1000
PHL Capacity
Sampling Cri

c)



d)

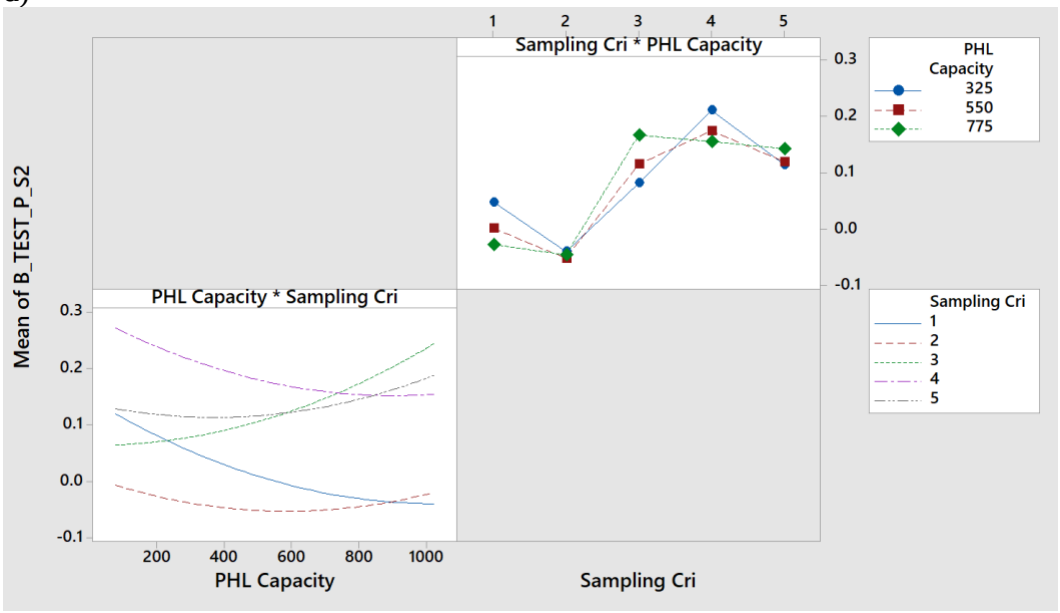
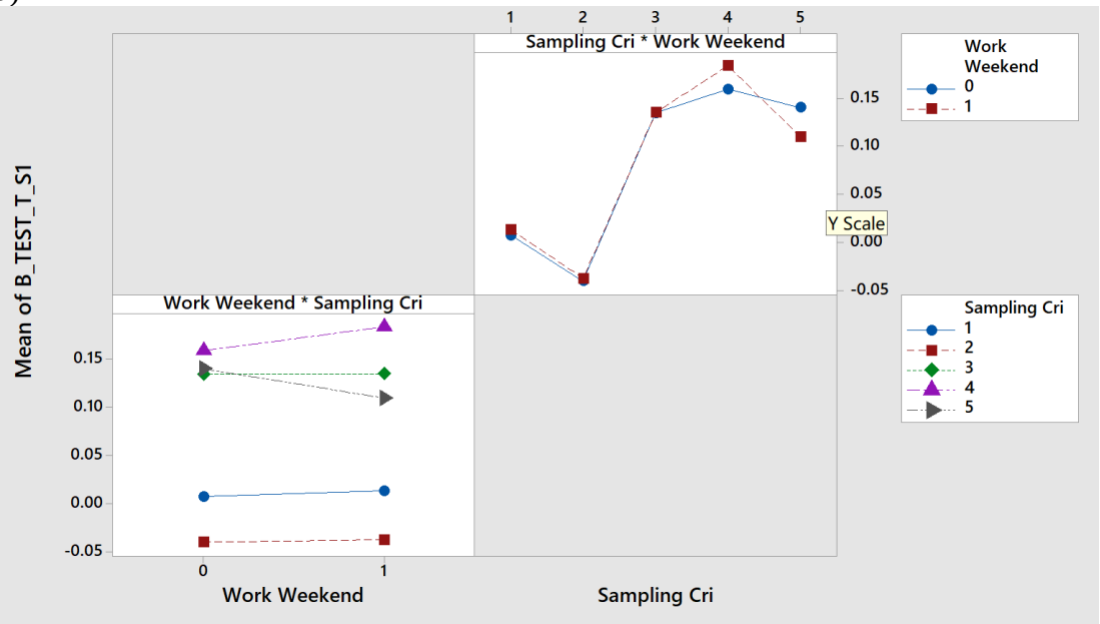


Figure 38. Interaction Plot of Sampling_Criteria and PHL_Capacity on Biases Regarding PHL Overall/Pandemic Test Trends. a) B_Test_T_S1. b) B_Test_T_S2. c) B_Test_P_S1. d) B_Test_P_S2.

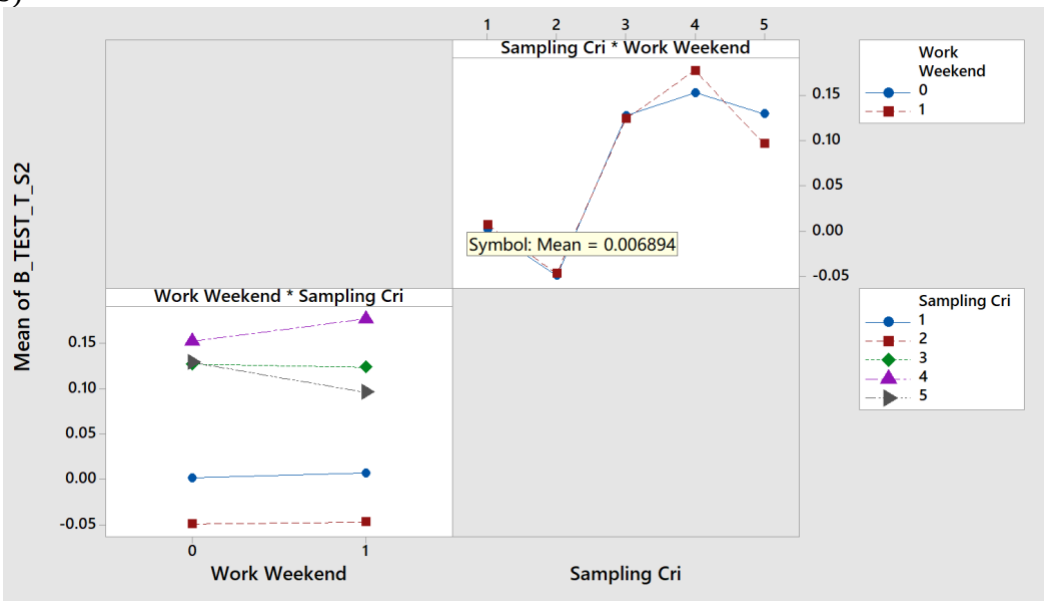
PHL_Work_Weekend

From Figure 39, sampling_Criteria 1 (FCFS) had the smallest absolute mean growth rate biases. There is no statistically significant difference between groups of PHL_Work_Weekend (0 and 1) for Sampling_Criteria 1 (FCFS), 2 (Restricted FCFS) or 3 (Hybrid). When PHL_Work_Weekend is 1 (work), Sampling_Criteria 4 (1WK FCFS) had largest absolute mean growth rate biases, and Sampling_Criteria 5 (1WK Restricted FCFS) had smaller absolute mean growth rate biases than Sampling_Criteria3 (Hybrid) and 4 (1WK FCFS). When PHL_Work_Weekend is 0 (not work), Sampling_Criteria 3 (Hybrid) and 5 (1WK Restricted FCFS) had no statistically significant difference, and they both had smaller absolute mean growth rate biases than Sampling Criteria 4 (1WK FCFS).

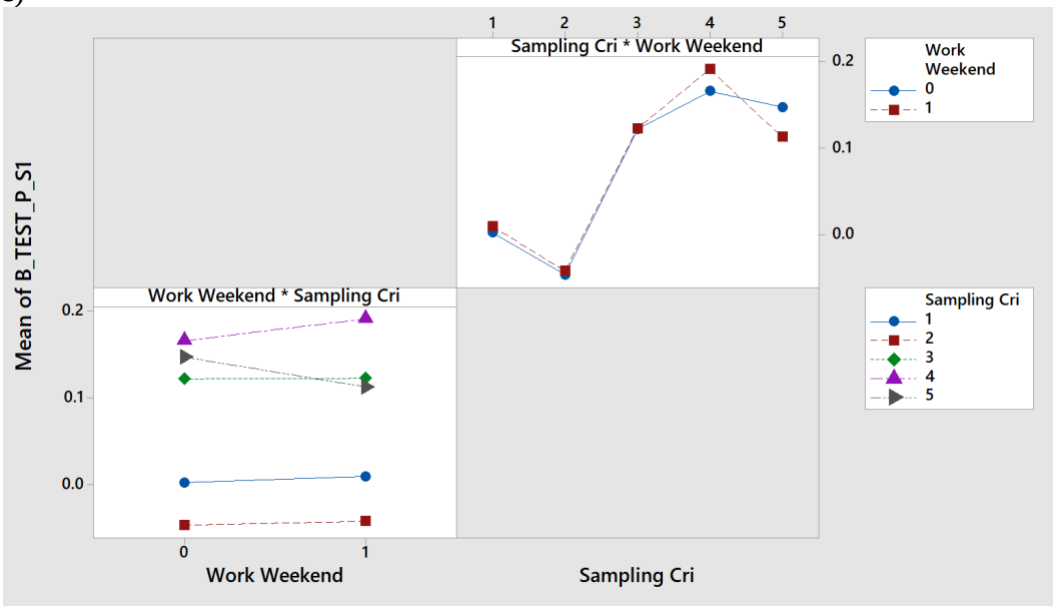
a)



b)


Mean of B_TEST_T_S2
Sampling Cri * Work Weekend
Work Weekend * Sampling Cri
Symbol: Mean = 0.006894
Work Weekend
Sampling Cri
Work
Weekend
0
1
Sampling Cri
1
2
3
4
5

c)


Mean of B_TEST_P_S1
Sampling Cri * Work Weekend
Work Weekend * Sampling Cri
Work Weekend
Sampling Cri
Work
Weekend
0
1
Sampling Cri
1
2
3
4
5

d)

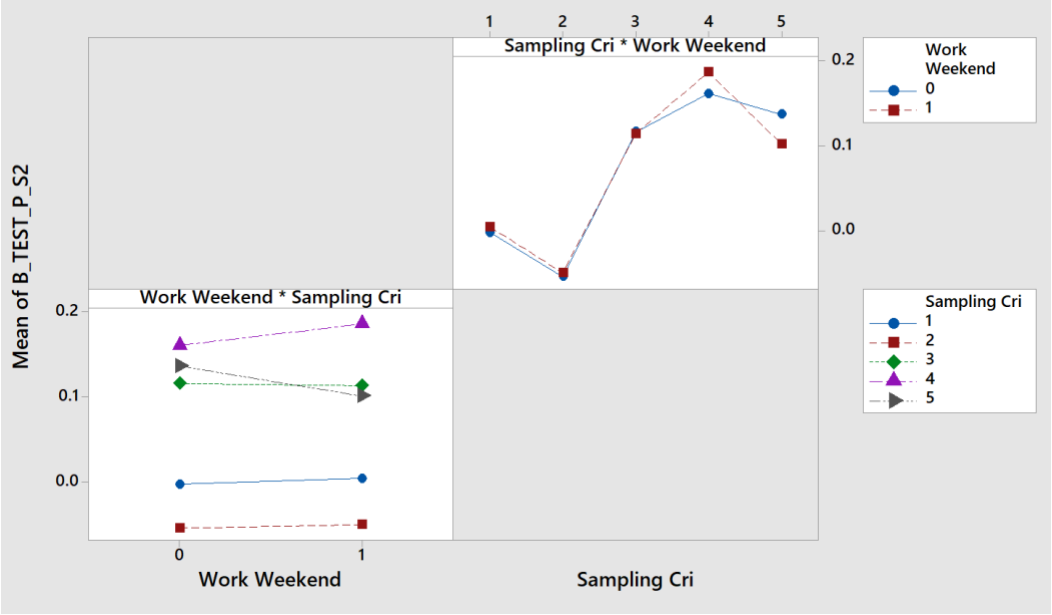


Figure 39. Interaction Plot of Sampling_Criteria and PHL_Work_Weekend on B_Test_P_S2

Results of Research Question 4

Explanatory Power of Influenza Surveillance Trends

The CCI experimental design in section methods generated 3200 replicates, consisting of 320 combinations with 10 replicates. Each replicate simulated the observed MSSS trends, the observed Submit trends, the observed Test trends, the real overall influenza incidence trends and the real pandemic influenza incidence trends. The growth rates were then estimated from these simulated trends using Exponential Growth Model. To find out which observed growth rates can best predict the real growth rates of overall influenza and pandemic influenza trends, respectively, a simple linear regression model was employed the test the coefficients of determination between the observed growth rates and the real growth rates under each of the 320

combinations. The trends yielding highest coefficient of determination was chosen for a combination.

Under two influenza outbreak scenarios, the number of combinations that are explained by each type of the observed trends were summarized in Table 33 and Table 34. It also shows that the Submit trends explain the majority of the combinations in both scenarios. In scenario 1, the growth rate of the Submit trends better explain the growth rates of the real overall influenza incidence trends in 96.6% of the combinations with a median explanatory power of 86.36%. Similarly, the growth rates of the Submit trends also better explain the growth rates of the pandemic influenza incidence trends in 95% of the combinations with a median explanatory power of 81.30%. In scenario 2, the growth rate of the Submit trends better explain the growth rates of the real overall influenza incidence trends in 94.7% of the combinations with a median explanatory power of 84.5%. Similarly, the growth rates of the Submit trends also better explain the growth rates of the pandemic influenza incidence trends in 92.8% of the combinations with a median explanatory power of 81.50%. Figure 40 shows the coefficients of determination Box-plot of the Submit trends and the real trends.

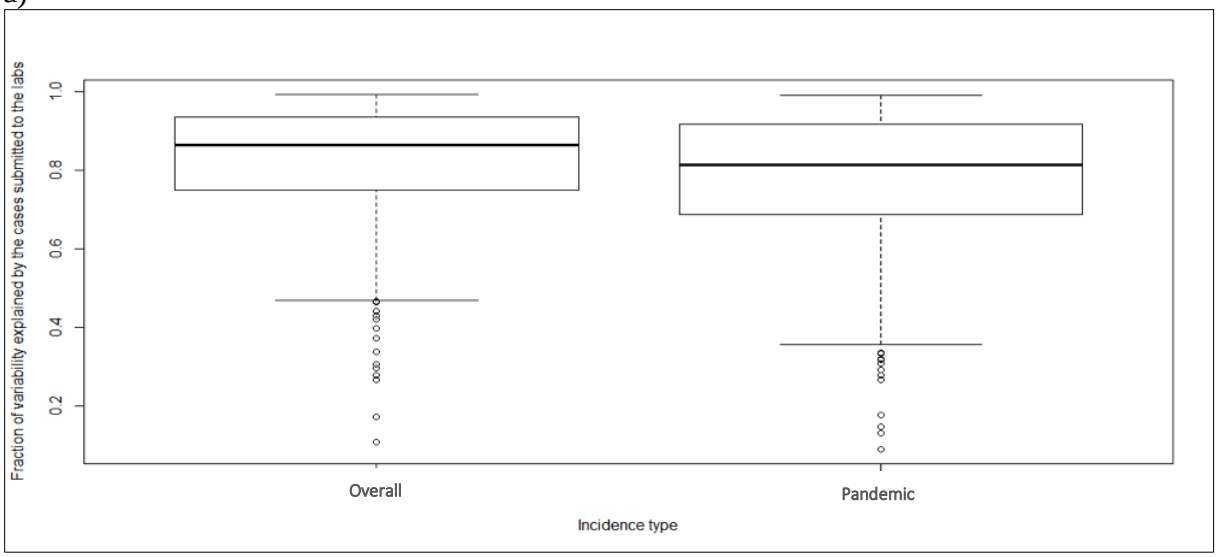
Table 33 *Number of Combinations Explained By the Observed Trends Under Scenario 1*

First scenario	Incidence trend	Number of combinations that are explained		
		MSSS	SUBMIT	TEST
	Overall	7	309	4
	Pandemic	13	304	3

Table 34 *Number of Combinations Explained By the Observed Trends Under Scenario 2*

Second scenario	Incidence trend	Number of combinations that are explained		
		MSSS	SUBMIT	TEST
	Overall	14	303	3
	Pandemic	17	297	6

a)



b)

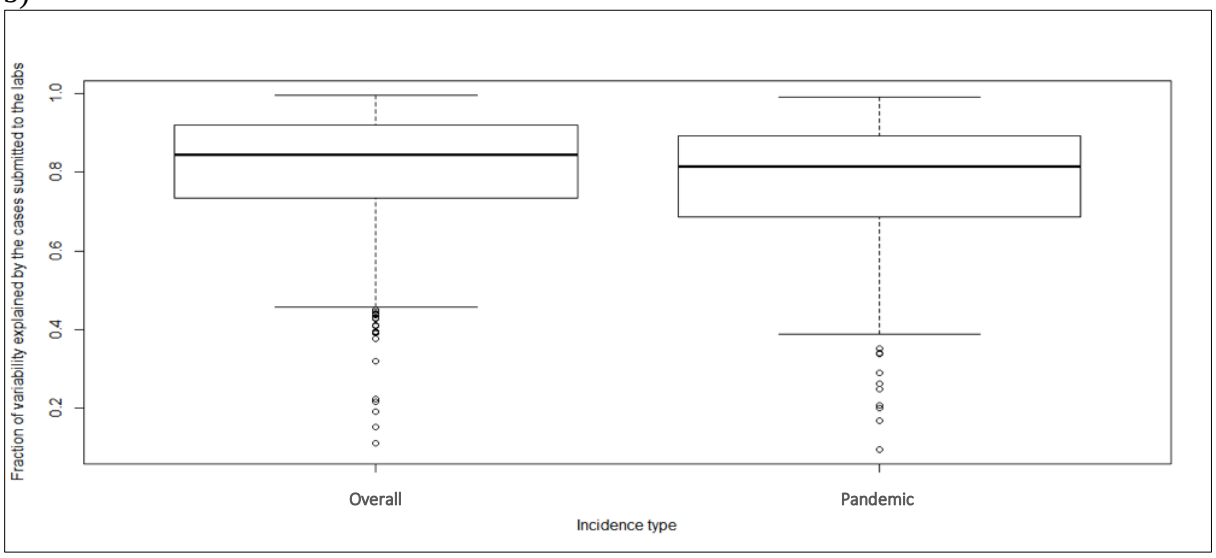


Figure 40. Boxplot of the Coefficients of Determination for the Overall and Pandemic Influenza Incidence Trends. a) Under Scenario 1. b) Under Scenario 2.

Discussion

This chapter investigates the impact of data collection operations on the biasedness in the syndromic surveillance (MSSS trends), the virological surveillance (PHL Overall/Pandemic Test trends), and the alternative syndromic surveillance (Submit trends) in the state of Michigan.

Perceived_Pandemic_Prob

Perceived_Pandemic_Prob is found to consistently have major effects in reducing the biasedness in syndromic surveillance. Influenza surveillance relies on symptomatic cases seeking healthcare. Higher perception of pandemic illness can increase the number of ERs visits and primary/urgent care consults, which reduces the growth rate biases in MSSS trend and Submit trend. Perceived_Pandemic_Prob also affects the effect of Primary_Care_Accessibility and primary/urgent care Collecting Sample Probability. With higher perception of pandemic symptoms, Primary_Care_Accessibility and primary/urgent care Collect_Sample_Prob yield smaller growth rate biases in MSSS trend and Submit trend.

Perceive_Pandemic_Probability also has major effect on virological surveillance biasedness. In average, mid-to-low level pandemic perception yields the smallest growth rate biases in PHL Overall/Pandemic trends. However, the effect also depends on the sampling methods that the PHL uses during the outbreak. When the PHL tests the submitted specimens using FCFS and Restricted FCFS, mid-to-high level of pandemic perception appears to minimize the growth rate biases in PHL Overall/Pandemic trends. When other sampling criteria are used (Hybrid, 1 WK FCFS or 1WK Restricted FCFS), opposite effect is observed: More people with pandemic perception instead worsen the biasedness in PHL Overall/Pandemic Test trends. This

result shows that when specimen submission to the PHL grows, these sampling criteria seem to intensify the underestimation of growth rate.

These findings suggest that high perception of pandemic illness is important to obtain unbiased MSSS trend and Submit trend, as it increases the submission proportion of the infected cases. The same suggestion is also applied to virological surveillance that collects PHL Overall/Pandemic trends, when the PHL uses FCFS or Restricted FCFS.

Sampling_Criteria

Sampling_Criteria is another factor that has major effect on the biasedness in virological surveillance that collects PHL Overall/Pandemic trends. Overall, FCFS sampling method turns out to have the smallest biasedness.

However, the performance of FCFS is affected by the amount of submitted specimens and the PHL_Capacity. It is already explained in the last part (Perceive_Pandemic_Probability) that large specimen submission generally reduces the biasedness in both syndromic surveillance and virological surveillance. When PHL_Capacity is constrained and cannot meet the high volume of specimens, Restricted FCFS is suggested to be used. Note that Restricted FCFS should be used the whole time after pandemic influenza declaration, especially in the PHL with constrained capacity (which is a common phenomenon).

Nevertheless, FCFS is still the best sampling method when PHL_Capacity is not a problem. Besides, to achieve that a larger proportion of infected cases submit the specimens to the PHL, it is recommended to increase the pandemic perception level and primary/urgent care sample collection level.

Other three sampling criteria (Hybrid, 1WK FCFS, and 1WK Restricted FCFS) are found to always underestimate the growth rate. Changing the sampling procedure from one to another, or using a truncated trends, would alter the disease portrayal to a large extent.

Other Factors

PHL_Work_Weekend is found to not improve the data collection unbiasedness when FCFS or Restricted FCFS sampling methods are used for specimen testing. It is worth mentioning that the specimens are supposed to expire 3 days. When specimens can survive longer, the effect of PHL Work Weekend might be changed.

Poverty_Level affects the healthcare seeking decision that a case makes in choosing ED or urgent care providers. The effect of Poverty_Level seems to work in different way among MSSS, Submit and PHL trends. Either low or high Poverty_Level reduces a the growth rate biases in MSSS trend. While lower Poverty_Level reduces the growth rate biases in Submit trend, but higher Poverty_Level reduce the biases in PHL Overall/Pandemic trends. Poverty_Level is not a controllable factor, but it might be useful as a predictor of the biases in the surveillance trends.

Submit Trend

Among the surveillance trends (MSSS, Submit, PHL Overall and Pandemic), Submit trend appears to have the most descriptive power for Real Overall and Real Pandemic trends. Submit trend is the daily number of specimens submitted to the PHL for testing and confirming, and has been collected in the current virological surveillance. This result suggests that Submit

trend is very useful to predict the growth rate of PHL Overall/pandemic influenza trends, and improve the predictive performance of many influenza outbreak models.

CHAPTER VI

CONCLUSION

Summary of the Study

This study accessed the descriptive power of the current data collection systems for influenza surveillance during an emergency in the state of Michigan with two steps: 1) preparing a simulation test bed for the state of Michigan, 2) evaluating the data collection processes with the simulation test bed.

A simulation testbed was first developed on top of a baseline agent-based influenza co-circulating model with the state of Michigan demographic, geographic, travel, and mitigation features. The testbed was accelerated by about 300% with OpenMP parallelization on a high performance cluster system equipped with 8 CPU cores. The simulation testbed was calibrated to model the initial growth phase of the pandemic outbreak in 2009 Michigan using a simulation-based optimization approach where Central Composite Design and Multi-Response Surface Methodology were employed.

The simulation testbed was further expanded with the data collection processes for influenza surveillance in Michigan, considering factors related with healthcare seeking behaviors (Poverty_Level, Perceived_Pandemic_Prob), specimen collection (Primary_Care_Accessibility, Collect_Sample_Prob), specimen sampling and testing (PHL_Capacity, PHL_Work_Weekend, Sampling_Criteria). A Central Composite Inscribed Design and General Linear Model were used

to perform simulation based experiments and investigate the influence of these data collection factors on the surveillance biasedness in the observed trends (MSSS trend, Submit trend, PHL Overall trend, PHL Pandemic trend).

Summary of the Results

Research objective 1. To prepare a high performance simulation testbed embedded with the realistic demographic, geographic, travel patterns in the state of Michigan in presence of mitigation and surveillance policies.

Research objective 2. To estimate the simulation model parameter space by calibrating the model with the MSSS influenza incidence trends during the initial phase of the 2009 pandemic influenza outbreak in state of Michigan.

Research question 1. Does the observed real-time trend of influenza incidence exclusively capture the disease behavior in presence of mitigation or is it also affected by other factors (e.g., public communication about the outbreak status)?

To answer the research question 1, the effect of epidemiological factors, mitigation and containment factors on the growth rate of the simulated overall influenza incidence trends were investigated. All these experimental factors have significant impacts on the simulated growth rate. Two optimal solutions were found that yielded the same growth rate of the overall influenza trend using the 2009 ED constitutional data after the 2009 pandemic declaration. Both scenarios portrayed high pandemic reproduction number, moderate seasonal reproduction number, high withdrawal percentage, and low antiviral usage coverage, high initial pandemic cases, and high initial seasonal cases. These results indicate that disease progression might have started before

the pandemic declaration. Presumably the declaration influenced the reported number of ED constitutional cases as a behavioral response due to fear of severe consequences from the disease. The surge of ED visit counts seems to be not only related to the evolution of the disease itself, but also related with the timing of first media declaration.

In the following, research objectives 3-4 were achieved by answering research questions 3-4.

Research objective 3. To develop the surveillance operations with the data collection processes in the Michigan simulation testbed;

Research objective 4. To explore the effect of the data collection operational factors in the exponential growth rate biases between the real and observed incidence trends with the simulation based experiments.

Research question 2. Which factors of the data collection operations have significant impacts on biasedness between the real and observed incidence trends?

To answer 2, the effect of behavior factors and data collection operational factors on the biases between each of the surveillance trend in the three stages and the real incidence trend was investigated. In stage 1, Perceived_Pandemic_Prob, Poverty_Level, and Primary_Care_Accessibility have significant effect on the biases between MSSS trend and the real incidence trend. In stage 2, Perceived_Pandemic_Prob, Poverty_Level, Primary_Care_Accessibility, together with Collect_Sample_Prob have significant impact on the biases between Submit trend and the real incidence trend. In stage 3, Perceived_Pandemic_Prob, Poverty_Level, Collect_Sample_Prob, Sampling_Criteria and PHL_Capacity have significant impact on the biases between PHL Test trend and the real incidence trend.

Primary_Care_Accessibility and PHL schedule has insignificant main effect on the biases regarding PHL Test trend, but the interaction between Primary_Care_Accessibility and Sampling_Criteria are significant.

Research question 3. How do Sampling_Criteria affect the biasedness between the real and observed incidence trends?

To answer research question 3, the main effect of Sampling_Criteria and its interaction effect with all the experimental factors on the biases between PHL Overall/Pandemic Test trend and the real incidence trend. FCFS and FCFS Restricted Sampling_Criteria outperform other Sampling_Criteria. Furthermore, FCFS is the best method that minimized the biases between PHL Overall/Pandemic Test trend and the real incidence trend. The interaction effect between Sampling_Criteria and Perceived_Pandemic_Prob, Poverty_Level, Collect_Sample_Prob have significant effect on the biases regarding PHL Overall/Pandemic Test trend. PHL schedule doesn't have significant impact on the biases created by either FCFS and FCFS Restricted sampling method. PHL_Capacity has significant impact on FCFS and FCFS Restricted performance. FCFS works the best when PHL_Capacity is less than 800, but FCFS Restricted works the best when PHL_Capacity is greater than 800.

Research question 4. To what extent do the observed influenza incidence trends explain the real influenza incidence trends?

The power of surveillance trends to explain the variability of growth rate of the real influenza trend was analyzed. Submit trend outperforms MSSS trend and PHL Test trend in most of the experimental scenarios and explained over a median of 80% of the total variances of the real influenza trend.

Findings and Recommendations

The surveillance system can be seen as a set of connected pipelines through which cases are constantly routed. The parameters Perceived_Pandemic_Prob, Primary_Care_Accessiblity, Powerty_Level, Collect_Sample_Prob, and PHL_Capacity act as valves that divert or route the flow of specimens to the data collection points. The effect of increasing or decreasing the value of each parameter in the growth rate biases in each reported trends has been previously discussed. In general, opening a valve that leads to a data collection point contributes to cases that reduce the bias, and closing a valve has the opposite effect.

Perceived_Pandemic_Prob is one of the first valves of the simulated surveillance system and can route or divert a high volume of cases. Hence, it has a highly significant impact in all the biases under study. The results suggest that high risk perception due to the influenza emergency increases the reporting, which leads to the reduction of biases in the growth rates. As the public interest decreases, so it will the reporting pattern (Birrell, Pebody, Charlett, Zhang, & De Angelis, 2017). Therefore, a recently declared emergency should be seen as an opportunity to support the collection of larger sample sizes, and public communication should provide a realistic sense of the risk posed by the influenza emergency.

Poverty_Level is a key valve in routing or diverting cases reporting over the weekend, and hence it is important to encourage **Urgent Care** and EDs to report their weekend cases. This practice will likely produce more representative samples in the incidence trends for each of the days reported during the weekend. The PHL should also consider that misrepresentation may occur for days with small sample sizes.

To the sampling decisions made in Michigan during the H1N1 Pandemic, this study designed and tested five types of **Sampling_Criteria**. Two of these criteria implied that the tested sampling method was maintained over the course of the growth phase, while three of these criteria implied a switch in the sampling method. The results show that switching from one of the Sampling_Criteria to another contributes to higher biases than maintaining the same sampling criteria. When the transition happens from FCFS or FCFS restricted to no sampling, the growth rate is estimated only with one week of data, which increases the bias. When the transition happens from FCFS to FCFS restricted, the pool of samples is reduced, and therefore the samples that tested positive out of this pool do not represent incidence at the same scale that the FCFS samples. Estimation methods should ensure that proper adjustment strategies are used to account for abrupt changes in the pool of samples.

FCFS restricted performs better when the **PHL_Capacity** for daily testing is less than 800 specimens while FCFS performs better as the capacity increases. Also, whether the **PHL_Work_Weekend** or not does not affect the biases under these two criteria. Sample size restrictions can help manage, but not eliminate the biases during high demand for testing services when the capacity is constrained. The restricted sample should represent all the features of the population that estimates should consider. For example, if the objective is to estimate the growth pattern per age groups, testing data only for those in high risk will not allow the estimation for other age groups.

In this study, FCFS and FCFS restricted outperformed other 3 sampling criteria when the objective was to minimize the growth rate bias for the overall incidence trend. Since the growth rates were not estimated considering additional features of the sample (e.g., growth rate estimates

by age or geography), samples collected in FCFS order would provide similar estimates than samples collected in FCFS order restricted to patients with ages of higher mortality risk (in the simulation, having an age of high mortality risk did not imply an additional chance of infection). However, if growth rate per age or per geography are to be estimated, FCFS restricted to high risk groups will not allow sampling across ages, and FCFS will not be suitable for geography-based estimation, because samples from closer areas will arrive before than samples from more distant areas. Samples from distant areas will also have a higher likelihood to be discarded as they will have to wait longer.

The results also show that the growth rate of ILI cases submitted to the PHL has a good **explanatory power** over the growth rate of overall ILI cases. Similarly, the growth rate of pandemic cases submitted to the PHL has a good explanatory power over the growth rate of overall pandemic cases. This occurs as a policy was simulated where the public health system encourages every healthcare provider to submit to the PHL, and the PHL was recording all the submissions. In reality, data collection did not occur as exactly as proposed in the simulation. After the pandemic was declared in Michigan, everyone was encouraged to submit their specimens, but some healthcare providers called the PHL ahead to determine whether they should submit their specimens or not. As a result, the PHL only documented the specimens that arrived to the lab but not all that were reported (either by phone or arriving). An online system managing the submission of specimens in real-time would potentially enhance the understanding of incidence progression in the overall population.

The growth rate of ILI cases submitted to the PHL outperformed the MSSS growth rate in its explanatory power of the overall growth rate. This result seems to demonstrate that the

node connected to the most and the widest variety of healthcare providers in the surveillance network also collects the most representative data, leading to lower biases in the growth rates. In countries like England, Public Health authorities are in search of better predictors of ILI activity, to complement or replace the existing PCP consultation data (Birrell, Pebody, Charlett, Zhang, & De Angelis, 2017). Convergent data sources that receive reporting from different types of providers (i.e., urgent care, PCP, or hospital) might present as enhanced alternatives than PCP consultation for the estimation of ILI incidence.

Limitations

The Restricted FCFS sampling method in the study only tests children and elders. But other high risk groups such as pregnant women and people with chronic diseases might affect the efficiency of this sampling method.

Operational Factors are not analyzed with the regional effect with geographic heterogeneity among the four simulated regions. But the agent-based model still considers spatial effect on influenza co-circulation.

Exponential growth rate is used in measuring the biasedness. However, other metrics have not been compared with the results using exponential growth rate, such as instantaneous reproduction number. It might be useful to catch the daily change rate but would increase the complexity in model calibration.

Future Research Directions

To improve influenza surveillance, a couple of future research directions can be proposed:

Better distribution of the PHL testing resources: Novel sampling schemes that aim to better represent incidence across day, age, and geography can minimize the biases while constrained by the PHL testing capacity. Development and testing of these schemes could provide implementable recommendations.

Simulation-based learning for operational decision making: Influenza simulation models can be used in table-top exercises by allowing decision makers to play and learn about disease progression from different simulated scenarios. It would be valuable to understand whether emergency planning is enhanced by such learning strategies.

APPENDIX

Table A1. Complete Design Matrix of Half Fractional CCI for Evaluation

https://www.dropbox.com/s/mjoqcpk8tpln1qc/Chapter_5_DOE.xls?dl=0

REFERENCES

2009 Comprehensive Household Travel Data Collection Program MI Travel Counts II Comparison Report [Pdf]. (2010, July 30). Michigan Department of Transportation. Retrieved from https://www.michigan.gov/documents/mdot/mdot_MTC_II_comparison_report_333906_7.pdf

2009 flu pandemic. (2019, July 02). Retrieved from https://en.wikipedia.org/wiki/2009_flu_pandemic#Mexico

2009 flu pandemic timeline. (2019, July 04). Retrieved from https://en.wikipedia.org/wiki/2009_flu_pandemic_timeline

Abar, S., Theodoropoulos, G. K., Lemarinier, P., & O'Hare, G. M. (2017). Agent based modelling and simulation tools: A review of the state-of-art software. *Computer Science Review, 24*, 13-33. doi:10.1016/j.cosrev.2017.03.001

Ahmed, A., & Fincham, J. E. (2010). Physician office vs retail clinic: patient preferences in care seeking for minor illnesses. *Annals of Family Medicine, 8*(2), 117–123. doi:10.1370/afm.1052

Antunes, A. C., Jensen, D., Halasa, T., & Toft, N. (2017). A simulation study to evaluate the performance of five statistical monitoring methods when applied to different time-series components in the context of control programs for endemic diseases. *Plos One, 12*(3). doi:10.1371/journal.pone.0173099

Asian Lineage Avian Influenza A(H7N9) Virus | Avian Influenza (Flu). (2018). Retrieved from https://www.cdc.gov/flu/avianflu/h7n9-virus.htm

Basta, N. E., Halloran, M. E., Matrajt, L., & Longini, I. M. (2008). Estimating influenza vaccine efficacy From challenge and community-based study data. *American Journal of Epidemiology, 168*(12), 1343-1352. doi:10.1093/aje/kwn259

Berry, D. (2004). *Risk, communication and health psychology*. McGraw-Hill Education (UK).

Birrell, P. J., Ketsetzis, G., Gay, N. J., Cooper, B. S., Presanis, A. M., Harris, R. J., . . . Angelis, D. D. (2011). Bayesian modeling to unmask and predict influenza A/H1N1pdm dynamics in London. *Proceedings of the National Academy of Sciences, 108*(45), 18238-18243. doi:10.1073/pnas.1103002108

Birrell, P. J., Pebody, R. G., Charlett, A., Zhang, X. S., & De Angelis, D. (2017). Real-time modelling of a pandemic influenza outbreak. *Health Technology Assessment, 21*(58), 1–118. doi:10.3310/hta21580

Bish, A., & Michie, S. (2010). Demographic and attitudinal determinants of protective behaviours during a pandemic: A review. *British Journal of Health Psychology, 15*(4), 797–824. doi:10.1348/135910710X485826

Bjørkdahl, K., & Carlsen, B. (2017). Fear of the fear of the flu: Assumptions about media effects in the 2009 Pandemic. *Science Communication, 39*(3), 358–381. doi.10.1177/1075547017709792

Blendon, R. J., Benson, J. M., DesRoches, C. M., Raleigh, E., & Taylor-Clark, K. (2004). The public's response to severe acute respiratory syndrome in Toronto and the United States. *Clinical Infectious Diseases*, *38*(7), 925–931. doi:10.1086/382355

Briand, S., Mounts, A., & Chamberland, M. (2011). Challenges of global surveillance during an influenza pandemic. *Public Health, 125*(5), 247-256. doi:10.1016/j.puhe.2010.12.007

Burman, M. E. (1996). Daily symptoms and responses in adults: A review. *Public Health Nursing,13*(4), 294-301. doi:10.1111/j.1525-1446.1996.tb00253.x

Cao, P. H., Wang, X., Fang, S. S., Cheng, X. W., Chan, K. P., Wang, X. L., … Yang, L. (2014). Forecasting influenza epidemics from multi-stream surveillance data in a subtropical city of China. *PLoS ONE*, *9*(3), 1–8. doi:10.1371/journal.pone.0092945

Carrat, F., & Flahault, A. (2007). Influenza vaccine: The challenge of antigenic drift. *Vaccine, 25*(39-40), 6852-6862. doi:10.1016/j.vaccine.2007.07.027

Carrat, F., Vergu, E., Ferguson, N. M., Lemaitre, M., Cauchemez, S., Leach, S., & Valleron, A. (2008). Time lines of infection and disease in human influenza: A review of volunteer challenge studies. *American Journal of Epidemiology*, *167*(7), 775-785. doi:10.1093/aje/kwm375

CDC's World Health Organization (WHO) Collaborating Center for Surveillance, Epidemiology and Control of Influenza | CDC. (2017). Retrieved from https://www.cdc.gov/flu/weekly/who-collaboration.htm

Chretien, J., George, D., Shaman, J., Chitale, R. A., & Mckenzie, F. E. (2014). Influenza forecasting in human populations: A scoping review. *PLoS ONE, 9*(4). doi:10.1371/journal.pone.0094130

Clinical Signs and Symptoms of Influenza | CDC. (2019). Retrieved from https://www.cdc.gov/flu/professionals/acip/clinical.htm

Deangelis, D. L., & Grimm, V. (2014). Individual-based models in ecology after four decades. *F1000Prime Reports, 6*. doi:10.12703/p6-39

Derringer, G., & Suich, R. (1980). Simultaneous optimization of several response variables. *Journal of Quality Technology, 12*(4), 214-219. doi:10.1080/00224065.1980.11980968

Dórea, F. C., Mcewen, B. J., Mcnab, W. B., Sanchez, J., & Revie, C. W. (2013). Syndromic surveillance using veterinary laboratory data: Algorithm combination and customization of alerts. *PLoS ONE, 8*(12). doi:10.1371/journal.pone.0082183

Dupuy, C., Morignat, E., Dorea, F., Ducrot, C., Calavas, D., & Gay, E. (2015). Pilot simulation study using meat inspection data for syndromic surveillance: Use of whole carcass condemnation of adult cattle to assess the performance of several algorithms for outbreak detection. *Epidemiology and Infection*, *143*(12), 2559–2569. doi:10.1017/S0950268814003495

Emergency Department Syndromic Surveillance in Michigan [PDF]. (2017). Retrieved from http://www.cec.org/sites/default/files/documents/agenda/jeu-7-sept.-2017/jay_fiedler.pdf

Fairchild, G., Polgreen, P. M., Foster, E., Rushton, G., & Segre, A. M. (2013). How many suffice? A computational framework for sizing sentinel surveillance networks. *International Journal of Health Geographics*, *12*, 1–11. doi:10.1186/1476-072X-12-56

Fauci, A. S. (2006). Emerging and re-emerging infectious diseases: Influenza as a prototype of the host-pathogen balancing act. *Cell, 124*(4), 665–670. doi:10.1016/j.cell.2006.02.010

Flannery, B., Chung, J. R., Belongia, E. A., Mclean, H. Q., Gaglani, M., Murthy, K., . . . Fry, A. M. (2018). Interim estimates of 2017–18 seasonal influenza vaccine effectiveness — United States, February 2018. *MMWR. Morbidity and Mortality Weekly Report, 67*(6), 180-185. doi:10.15585/mmwr.mm6706a2

Flu Symptoms & Complications. (2017, July 28). Retrieved April 10, 2018, from https://www.cdc.gov/flu/consumer/symptoms.htm

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012–1014. doi:10.1038/nature07634

Glik, D. C. (2007). Risk communication for public health emergencies. *Annual Review of Public Health*, *28*(1), 33–54. doi:/10.1146/annurev.publhealth.28.021406.144123

Global Influenza Surveillance and Response System (GISRS). (2019, June 21). Retrieved from https://www.who.int/influenza/gisrs_laboratory/en/

Gunst, R. F., Myers, R. H., & Montgomery, D. C. (1996). Response surface methodology: Process and product optimization using designed experiments. *Technometrics, 38*(3), 285. doi:10.2307/1270613

Holvenstot, P., Prieto, D., & Doncker, E. D. (2014). GPGPU parallelization of self-calibrating agent-based influenza outbreak simulation. *2014 IEEE High Performance Extreme Computing Conference (HPEC)*. doi:10.1109/hpec.2014.7041000

Influenza A virus subtype H2N2. (2019, January 26). Retrieved from https://en.wikipedia.org/wiki/Influenza_A_virus_subtype_H2N2#Asian_flu

Influenza pandemic. (2019, April 25). Retrieved from https://en.wikipedia.org/wiki/Influenza_pandemic

Influenza (Seasonal). (2018). Retrieved from https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)

Influenza Virologic Surveillance Right Size Roadmap [PDF]. (2012). Association of Public Health Laboratories. Retrieved from http://www.aphl.org/AboutAPHL/publications/Documents/ID_July2013_Influenza-Virologic-Surveillance-Right-Size-Roadmap.pdf

Influenza virus infections in humans [PDF]. (2014). World Health Organization. Retrieved from https://www.who.int/influenza/human_animal_interface/virology_laboratories_and_vaccines/influenza_virus_infections_humans_feb14.pdf

Iuliano, A. D., Roguski, K. M., Chang, H. H., Muscatello, D. J., Palekar, R., Tempia, S., … Mustaquim, D. (2018). Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *The Lancet, 391*(10127), 1285–1300. doi:10.1016/S0140-6736(17)33293

Jeong, I., & Kim, K. (2009). An interactive desirability function method to multiresponse optimization. *European Journal of Operational Research, 195*(2), 412-426. doi:10.1016/j.ejor.2008.02.018

Jit, M., & Brisson, M. (2011). Modelling the epidemiology of infectious diseases for decision analysis. *PharmacoEconomics, 29*(5), 371-386. doi:10.2165/11539960-000000000-00000

Kasaie, P., & Kelton, W. D. (2013). Simulation optimization for allocation of epidemic-control resources. *IIE Transactions on Healthcare Systems Engineering, 3*(2), 78-93. doi:10.1080/19488300.2013.788102

Kegeles, S. S. (1963). Why people seek dental care: A test of a conceptual formulation. *Journal of Health and Human Behavior, 4*(3), 166. doi:10.2307/2948658

Kleef, E. V., Robotham, J. V., Jit, M., Deeny, S. R., & Edmunds, W. J. (2013). Modelling the transmission of healthcare associated infections: A systematic review. *BMC Infectious Diseases,13*(1). doi:10.1186/1471-2334-13-294

Laboratory-Confirmed Influenza Hospitalizations. (n.d.). Retrieved from https://gis.cdc.gov/GRASP/Fluview/FluHospRates.html

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science, 343*(6176), 1203-1205. doi:10.1126/science.1248506

Lee, E. C., Arab, A., Goldlust, S. M., Viboud, C., Grenfell, B. T., & Bansal, S. (2018). Deploying digital health data to optimize influenza surveillance at national and local scales. *PLoS Computational Biology*, *14*(3), 1–23. doi:10.1371/journal.pcbi.1006020

*Lessons from a virus: Public health laboratories respond to the H1N1 pandemic [*PDF]. (2011). Association of Public Health Laboratories. Retrieved from https://www.aphl.org/AboutAPHL/publications/Documents/ID_2011Sept_Lessons-from-a-Virus-PHLs-Respond-to-H1N1-Pandemic.pdf

Longini Jr, I. M., Halloran, M. E., Nizam, A., & Yang, Y. (2004). Containing pandemic influenza with antiviral agents. *American Journal of Epidemiology, 159*(7), 623-633. doi:10.1093/aje/kwh092

Magid, A., Gesser-Edelsburg, A., & Green, M. S. (2018). The role of informal digital surveillance systems before, during and after infectious disease outbreaks: a critical analysis. *Defence Against Bioterrorism NATO Science for Peace and Security Series A: Chemistry and Biology,* 189-201. doi:10.1007/978-94-024-1263-5_14

McIver, D. J., & Brownstein, J. S. (2014). Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. *PLoS Computational Biology, 10*(4). doi.10.1371/journal.pcbi.1003581

Michigan Department of Health, Human Services (2016) Influenza update. MACI Meeting p.18

Michigan Syndromic Surveillance System Emergent/Urgent Data Submission Guide [PDF]. (2017, May 19). Michigan Department of Health and Human Services. Retrieved from http://www. michigan.gov/mdch/0,1607,7-132-2945_5104_31274-107091--,00.htm

Mittal, N., & Medhi, B. (2007). The Bird Flu: A New Emerging Pandemic Threat And Its Pharmacological Intervention. *International journal of health sciences*, *1*(2), 277-283.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., . . . Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Medicine, 5(3). doi:10.1371/journal.pmed.0050074

Morse, S. S. (2007). Global infectious disease surveillance and health intelligence. *Health Affairs,26*(4), 1069-1077. doi:10.1377/hlthaff.26.4.1069

Myers, R. H., Vining, G. G., Giovannitti-Jensen, A., & Myers, S. L. (1992). Variance dispersion properties of second-order response surface designs. *Journal of Quality Technology, 24*(1), 1-11. doi:10.1080/00224065.1992.11979368

Neill, S., Roland, D., Jones, C. H. D., Thompson, M., & Lakhanpaul, M. (2015). Information resources to aid parental decision-making on when to seek medical care for their acutely sick child: A narrative systematic review. *BMJ Open*, *5*(12), e008280. doi:10.1136/bmjopen-2015-008280

Neumann, G., Noda, T., & Kawaoka, Y. (2009). Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature, 459*(7), 931–939. doi:10.1038/nature08157

Nsoesie, E., Mararthe, M., & Brownstein, J. (2013). Forecasting peaks of seasonal influenza epidemics. *PLoS Currents*. doi:10.1371/currents.outbreaks.bb1e879a23137022ea79a8c508b030bc

Nsoesie, E. O., Buckeridge, D. L., & Brownstein, J. S. (2014). Guess who's not coming to dinner? Evaluating online restaurant reservations for disease surveillance. *Journal of Medical Internet Research*, *16*(1), e22. doi:10.2196/jmir.2998

Olshaker, J. S., & Rathlev, N. K. (2006). Emergency Department overcrowding and ambulance diversion: The impact and potential solutions of extended boarding of admitted patients in the Emergency Department. *The Journal of Emergency Medicine*, *30*(3), 351-356. doi:10.1016/j.jemermed.2005.05.023

O'Meara, W. P., Karuru, S., Fazen, L. E., Koech, J., Kizito, B., Tarus, C., & Menya, D. (2014). Heterogeneity in health seeking behaviour for treatment, prevention and urgent care in four districts in western Kenya. *Public Health, 128*(11), 993–1008. doi:10.1016/j.puhe.2014.08.010

Ostroy, G., Prieto, D., Gu, Y., Dedoncker, E., & Paul, R. (2017). Flu MODELO 1.0: A simulation model and graphic interface for training and decision support for influenza management (pp. 2287–2289). Presented at the *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE. http://doi.org/10.1109/BIBM.2017.8218029

Outbreak Communication: Best practices for communicating with the public during an outbreak [Pdf]. (2004.). World Health Organization.

Overview of Influenza Surveillance in the United States | CDC. (2017). Retrieved from https://www.cdc.gov/flu/weekly/overview.htm

Paleshi, A., Bae, K. H., Evans, G., & Heragu, S. (2017). A simulation-based optimization approach for mitigation of pandemic influenza. *IISE Transactions on Healthcare Systems Engineering*, *7*(2), 107–120. doi:10.1080/24725579.2017.1302525

Past Seasons Estimated Influenza Disease Burden | CDC. (2018). Retrieved from https://www.cdc.gov/flu/about/burden/past-seasons.html

People at High Risk For Flu Complications | CDC. (2018). Retrieved from https://www.cdc.gov/flu/highrisk/index.htm?CDC_AA_refVal=https://www.cdc.gov/flu/about/disease/high_risk.htm

Polgreen, P. M., Chen, Y., Pennock, D. M., & Nelson, F. D. (2008). Using internet searches for influenza surveillance. *Clinical Infectious Diseases*, *47*(11), 1443–1448. doi:10.1086/593098

Polgreen, P. M., Chen, Z., Segre, A. M., Harris, M. L., Pentella, M. A., & Rushton, G. (2009). Optimizing influenza sentinel surveillance at the state level. *American Journal of Epidemiology*, *170*(10), 1300–1306. doi:10.1093/aje/kwp270

Prieto, D. M., Das, T. K., Savachkin, A. A., Uribe, A., Izurieta, R., & Malavade, S. (2012). A systematic review to identify areas of enhancements of pandemic simulation models for operational use at provincial and local levels. *BMC Public Health, 12*(1). doi:10.1186/1471-2458-12-251

Prieto, D., & Das, T. K. (2016). An operational epidemiological model for calibrating agent-based simulations of pandemic influenza outbreaks. *Health Care Management Science, 19*(1). doi:10.1007/s10729-014-9273-3

Safer, M. A., Tharps, Q. J., Jackson, T. C., & Leventhal, H. (1979). Determinants of three stages of delay in seeking care at a medical clinic. *Medical Care, 17*(1), 11–29. doi:10.1097/00005650-197901000-00002

Santillana, M., Nsoesie, E. O., Mekaru, S. R., Scales, D., & Brownstein, J. S. (2014). Using clinicians' search query data to monitor influenza epidemics. *Clinical Infectious Diseases, 59*(10), 1446–1450. doi:10.1093/cid/ciu647

Santillana, M., Nguyen, A. T., Dredze, M., Paul, M. J., Nsoesie, E. O., & Brownstein, J. S. (2015). Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. *PLOS Computational Biology, 11*(10). doi:10.1371/journal.pcbi.1004513

Santillana, M., Nguyen, A. T., Louie, T., Zink, A., Gray, J., Sung, I., & Brownstein, J. S. (2016). Cloud-based electronic health records for real-time, region-specific influenza surveillance. *Scientific Reports, 6*(1). doi:10.1038/srep25732

Scarpino, S. V., Dimitrov, N. B., & Meyers, L. A. (2012). Optimizing provider recruitment for influenza surveillance networks. *PLoS Computational Biology, 8*(4). doi:10.1371/journal.pcbi.1002472

Scheduled Passenger Deplanements 2009 Airport Facility Year to Date Record [PDF]. (2009). Michigan Department of Transportation. Retrieved from https://mdotjboss.state.mi.us/AIRSTATS/searchByReportType.htm?year=2009&reportType=2&formatType=pdf

Schlegelmilch, J., Lenart, B., King, L. B., Schnelle, D. D., DiFato, T. L., Smith, S. D., … Wireman, J. R. (2012). Summary of the current operational epidemiological modelling landscape. *Journal of Business Continuity & Emergency Planning*, *5*(4), 338–351. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22576138

Shaman, J., & Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences, 109*(50), 20425-20430. doi:10.1073/pnas.1208772109

Shaman, J., Karspeck, A., Yang, W., Tamerius, J., & Lipsitch, M. (2013). Real-time influenza forecasts during the 2012–2013 season. *Nature Communications, 4*(1). doi:10.1038/ncomms3837

Shekh, B., Doncker, E. D., & Prieto, D. (2015). Hybrid multi-threaded simulation of agent-based pandemic modeling using multiple GPUs. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). doi:10.1109/bibm.2015.7359894

Signorini, A., Segre, A. M., & Polgreen, P. M. (2011). The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PLoS ONE, 6*(5). doi.10.1371/journal.pone.0019467

Soto-Ferrari, M., Holvenstot, P., Prieto, D., Doncker, E. D., & Kapenga, J. (2013). Parallel programming approaches for an agent-based simulation of concurrent pandemic and seasonal influenza outbreaks. *Procedia Computer Science, 18*, 2187-2192. doi:10.1016/j.procs.2013.05.389

Souty, C., & Boëlle, P. Y. (2016). Improving incidence estimation in practice-based sentinel surveillance networks using spatial variation in general practitioner density. *BMC Medical Research Methodology*, *16*(1), 1–8. doi:10.1186/s12874-016-0260-x

Spanish flu. (2019, June 19). Retrieved from https://en.wikipedia.org/wiki/Spanish_flu#cite_note-3

State Long Range Transportation Plan 2005-2030: Travel Characteristics Technical Report [PDF]. (2006, August 8). Michigan Department of Transportation. Retrieved from https://www.michigan.gov/documents/MDOT_TravCharTR_Final20060804_167340_7.pdf

Summary of the 2017-2018 Influenza Season | CDC. (2018). Retrieved from https://www.cdc.gov/flu/about/season/flu-season-2017-2018.htm

Types of Influenza Viruses | CDC. (2017). Retrieved from https://www.cdc.gov/flu/about/viruses/types.htm

US Census Bureau. (2018, May 04). 2010 Census: Population Density Data (Text Version). Retrieved from https://www.census.gov/content/census/en/data/tables/2010/dec/density-data-text.html

Venkatraman, S., & Selvagopal, D. (2018). An optimized multiobjective CPU job scheduling using evolutionary algorithms. *Turkish Journal Of Electrical Engineering & Computer Sciences, 26*, 101-114. doi:10.3906/elk-1701-22

Wallinga, J., & Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences,274*(1609), 599-604. doi:10.1098/rspb.2006.3754

WHO global technical consultation: Global standards and tools for influenza surveillance [PDF]. (2011). Geneva: World Health Organization.

WHO Interim Global Epidemiological Surveillance Standards for Influenza [PDF]. (2012, July). World Health Organization.

Willem, L., Verelst, F., Bilcke, J., Hens, N., & Beutels, P. (2017). Lessons from a decade of individual-based models for infectious disease transmission: A systematic review (2006-2015). *BMC Infectious Diseases, 17*(1). doi:10.1186/s12879-017-2699-8

Wouterse, F., & Tankari, M. (2015). Household out-of-pocket expenses on health: Does disease type matter? *Journal of African Economies*, *24*(2), 254–276. doi:10.1093/jae/ejv001

Yax, L. K. (2010). America's Families and Living Arrangements: 2010. Retrieved from https://www.census.gov/population/www/socdemo/hh-fam/cps2010.html

Yuan, Q., Nsoesie, E. O., Lv, B., Peng, G., Chunara, R., & Brownstein, J. S. (2013). Monitoring influenza epidemics in China with search query from baidu. *PLoS ONE, 8*(5). doi.10.1371/journal.pone.0064323

Zhang, Y., & Brown, D. (2014). Simulation optimization of police patrol districting plans using response surfaces. *Simulation, 90*(6), 687-705. doi:10.1177/0037549714533159