2011

Modeling and Surveillance of Pandemic Influenza Outbreaks

Diana Prieto
University of South Florida, dianaprietosanta@gmail.com

Modeling and Surveillance of Pandemic Influenza Outbreaks

by

Diana M. Prieto Santa

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Co-Major Professor: Tapas K. Das, Ph.D.
Co-Major Professor: Alex Savachkin, Ph.D.
Susana Lai-Yuen, Ph.D.
Kandethody Ramachandran, Ph.D.
Yiliang Zhu, Ph.D.

Date of Approval:
July 15, 2011

Keywords: Infectious disease simulation, Real-time disease monitoring, Disease uncertainty management, Concurrent viral strains, Specimen testing strategy

Dedication

Con amor para papa y mama. Todo lo que soy y lo que sere.

Acknowledgements

I would like to thank the economic support for this research work, provided by the Department of Industrial and Management Systems Engineering, and the Graduate School of the University of South Florida. To the professors Alex Savachkin, Ricardo Izurieta and Lillian Stark, I would like to express my sincere gratitude for providing technical support and expert opinion in the background, ideas and methods discussed in this dissertation. To my colleagues, Alfredo Santana and Sharad Malavade, my greatest appreciation for their interest, enthusiasm, and dedication to the challenge grant project. To Diego Martinez, Patricio Rocha, Ioannis Raptis, Felipe Feijoo, Laila Cure, Wilkistar Otieno, Vishnuteja Nanduri, Chaitra Gopalappa, Joshua Candamo, Athina Brintaki and Carlos Paternina for the encouragement, care and support whenever I needed them the most. To Ana Lucia Prieto, for being my angel, my sister, my friend, my co-worker, my roommate and my partner during 29 years. I finally want to thank my mentors Dr. Tapas K. Das and Dr. Marco Sanjuan for their care, support and encouragement, for believing in me and my work, for embracing my ideas and helping me to grow as a human being and as a professional. Not a single step of my doctoral studies would have taken place without their wise, kind, and generous guidance.

Table of Contents

List of Tables

List of Figures

Abstract

Pandemic outbreaks are unpredictable as to their virus strain, transmissibility, and impact on our quality of life. Hence, the decision support models for mitigation of pandemic outbreaks must be user-friendly and operational, and also incorporate valid estimates of disease transmissibility and severity. This dissertation research is aimed at 1) reviewing the existing pandemic simulation models to identify their implementation gaps with regard to usability and operability, and suggesting research remedies, 2) increasing operability of simulation models by calibrating them via an epidemiological model that estimates infection probabilities using viral shedding profiles of concurrent pandemic and seasonal influenza, and 3) developing a testing strategy for the state laboratories, with their limited capacities, to improve their ability to estimate evolving transmissibility parameters. Our review of literature (Aim 1) indicates the need to continue model enhancements in critical areas including updating of epidemiological data during a pandemic, smooth handling of large demographical databases, incorporation of a broader spectrum of social-behavioral aspects, and improvement of computational efficiency and accessibility. As regards the ease of calibration (Aim 2), we demonstrate that the simulation models, when driven by the infection probabilities obtained from our epidemiological model, accurately reproduce the disease transmissibility parameters. Assuming the availability of sufficient disease reporting infrastructure and strong compliance by both infected population and healthcare providers, our testing strategy (Aim 3) adequately supports characterization of real-time epidemiological parameters. Future research on this topic will be aimed at integrating the laboratory testing strategy with our modeling and simulation approach to develop dynamic mitigation strategies for pandemic outbreaks.

Chapter 1: Introduction

Pandemic influenza (PI) has been one of the major causes of illness and death in humans in the 20th century. In 1918, the Spanish influenza inflicted 2-5 new infections per each infected case and approximately 675,000 deaths in the United States [1, 2]. The 1957 and 1968 pandemic outbreaks were milder than the Spanish flu, but more severe than modern seasonal outbreaks (see Table 1.1). Early in the 21^{st} century, with the concern that the H5N1 influenza virus might escalate to pandemic proportions, many countries initiated their preparedness efforts for scenarios of low transmissibility and high severity [3]. In 2009, much of the world experienced a pandemic outbreak, though with a milder H1N1 virus strain comparable to seasonal flu. The above historical perspective suggests that the course and potential impact of an emergent pandemic influenza remain uncertain.

Table 1.1: Impact of pandemic and seasonal influenza outbreaks.

Features/ Outbreaks [References]	**1918-1919** [1, 2]	**1957-1958** [3, 4]	**1968-1969** [5, 6]	**2009-2010** [7, 8]	**Seasonal Outbreaks** [9, 10]
Strain	A/H1N1	A/H2N2	A/H3N2	A/H1N1	e.g., A/H3N2, A/H1N1
Transmissibility†	2-5	1.5-1.7	1.06-2.06	1.4-1.6	1.3
Severity‡	675000	69800	33800	12470	3000-49000

†Transmissibility: number of cases attributable to a single established one. ‡Severity: number of deaths in the U.S. in a year.

To minimize the risk of making pandemic mitigation decisions under uncertainty, decision support tools can be used to evaluate the impact of strategies during an evolving emergency. Recent research on pandemic mitigation have created models that consider the social and viral dynamics in a population and a variety of mitigation strategies to analyze the best course

of action during a pandemic outbreak [11–15]. But these models still have significant limitations. Modeling approaches are built assuming a known virus epidemiology and, as a result, they are not able to adapt to the evolving characteristics of pandemics and mould effective mitigation strategies [3]. Moreover, the existing models do not provide interactive feedback during pandemic outbreaks, and thus lack the ability to support operational decisions that need knowledge of disease progression. Examples of these decisions include optimal allocation of vaccines and antivirals, enforcement of social distancing, and school/workplace closure options.

The Institute of Medicine, in its report on Modeling Community Containment for Pandemic Influenza [3], recommended the research community to take steps on building “decision-aid models that can be readily linked to surveillance data to provide real-time feedback during an epidemic" and to develop, approve, and put in place “research protocols to generate the information needed during an outbreak to inform models, and improve their disease sub-models." A recent report from the Yale New Haven Center for Emergency Preparedness and Disaster Response stated that “resources need to be assigned for the development of models that are specifically intended for use in operational environments and are potentially customizable to meet the decision making needs of the user"[16].

This dissertation is aimed at 1) conducting a major review of the existing pandemic models, to identify the areas of weakness with regard to their usability and operability, and suggest remedial research needs, 2) developing an epidemiological model for calibration of pandemic simulation models, and 3) developing a testing strategy for the state laboratories, with their limited capacity, to improve their ability to assess evolving epidemiological parameters.

The following are the specific deliverables of this dissertation.

1. A systematic review of literature to identify areas of enhancements of pandemic simulation models for higher practical usability. In this review, we conducted a survey of the available

research literature on simulation models for influenza pandemic preparedness. The contributions include a set of areas of enhancement that will enable models to be used in operational environments.

2. An epidemiological model for calibrating simulation models of concurrent pandemic and seasonal influenza outbreaks. The epidemiological model uses the viral shedding profiles of the concurrent viral strains to assess the probabilities of infection that yield the prevailing reproduction behavior. This model allows the user to avoid the time consuming trial-and-error approach commonly used for simulation model calibration.

3. A strategy for statistical sampling and testing of field Influenza-like-illness (ILI) samples at the test laboratories and a Bayesian model to estimate the epidemiological parameters in real-time. The strategy helps to allocate limited testing capacity to the samples received based on ILI information collected by the healthcare providers. The number of confirmed cases from the laboratory testing serves as input to the Bayesian model that infers the actual number of pandemic ILI cases. The strategy and the Bayesian model are evaluated for different reporting rates in a given concurrent outbreak scenario involving one million population and reproduction numbers of 1.8 and 1.3 for pandemic and seasonal, respectively.

Chapter 2: A Systematic Review of Areas of Enhancements of Pandemic Simulation Models

The ability of computer simulation models to “better frame problems and opportunities, integrate data sources, quantify the impact of specific events or outcomes, and improve multi-stakeholder decision making," has motivated their use in public health preparedness (PHP) [17]. In 2006, one such initiative was the creation of the Preparedness Modeling Unit by the Centers for Disease Control and Prevention (CDC) in the U.S. The purpose of this unit is to coordinate, develop, and promote “problem-appropriate and data-centric" computer models that substantiate PHP decision making [18].

Of the existing computer simulation models addressing PHP, those focused on disease spread and mitigation of pandemic influenza (PI) have been recognized by the public health officials as useful decision support tools for preparedness planning [17]. In recent years, computer simulation models were used by the Centers for Disease Control and Prevention (CDC), Department of Health and Human Services (HHS), and other federal agencies to formulate the “U.S. Community Containment Guidance for Pandemic Influenza" [19].

Although the potential of the existing PI models is well acknowledged, it is perceived that the models are not yet usable by the state and local public health practitioners [3, 17, 20]. To identify the challenges associated with the practical implementation of the PI models, the National Network of Public Health Institutes, at the request of CDC, conducted a national

survey of the practitioners [17]. The challenges identified by the survey are summarized in Table 2.1.

Table 2.1: A summary of the survey results on the challenges of practical use of PI models, as perceived by public health practitioners.

Challenge	Description of the challenge
A1. Validity of data support	Model parameters need to be derived from updated demographical and epidemiological data
A2. Credibility and validity of assumptions	Models need to use credible and valid assumptions
A3. Represent human behavior	Models need to incorporate human behavior
A4. Accessibility	Models need to be easily accessible and run on personal computers
A5. Scalability	Models need to be scalable to population specific data from regions of all sizes
A6. Awareness	Available models and best practices need to be disseminated among the practitioners
A7. Action plan	Need to translate models into uniform preparedness and response action plans
A8. Lack of resources	Need to fund staff allocation and specialized training for model implementation
A9. Political implications	Models need to consider second and third tier social implications of containment strategies
A10. Lack of mandates for models	State and federal agencies need to develop mandates for use of model-based strategies

We divided the challenges (labeled A1 through A10 in Table 2.1) into two categories: those (A1 through A5) that are related to model design and implementation and can potentially be addressed by adaptation of the existing models and their supporting databases, and those (A6 through A10) that are related to resource and policy issues, and can only be addressed by changing public health resource management approaches and enforcing new policies. Although it is important to address the challenges A6 through A10, we consider this a prerogative of the public health administrators. Hence, the challenges A6 to A10 will not be discussed in this review.

The challenges A1 through A5 reflect the perspectives of the public health officials, the end users of the PI models, on the practical usability of the existing PI models and databases in supporting decision making. Addressing these challenges would require a broad set of

enhancements to the existing PI models and associated databases, which have not been fully attempted in the literature. In this chapter, we conduct a review of the PI mitigation models available in the published research literature with an objective of answering the question: “how to enhance the pandemic simulation models and the associated databases for higher practical usability?" We believe that our review accomplishes its objective in two steps. First, it exposes the differences between the perspectives of the public health practitioners and the developers of models and databases on the required model capabilities. Second, it derives recommendations for enhancing practical usability of the PI models and the associated databases.

2.1 Design and Implementation Challenges of Pandemic Models and Databases

In this section, we examine each of the design and implementation challenges of the existing PI models (A1-A5) and develop specific recommendations to address them.

2.1.1 A1. Validity of Data Support

Public health policy makers advocate that the model parameters be derived from up to date demographical and epidemiological data during an outbreak [17]. In this paper we examine some of the key aspects of data support, such as data availability, data access, data retrieve, and data translation.

To ensure data availability, a process must be in place for collection and archival of both demographical and epidemiological data during an outbreak. The data must be temporally consistent, i.e., it must represent the actual state of the outbreak. Availability of temporally

consistent demographical data is currently supported by governmental databases including the decennial census and the national household travel survey [21, 22]. To ensure temporal consistency of epidemiological data, the Institute of Medicine (IOM) has recommended enhancing the data collection protocols to support real-time decision making [3]. The interpretation of the term “real-time" in the context of pandemic influenza can be inferred from recent studies for the H1N1, where the information used to estimate real-time epidemiological parameters was obtained daily, and the real-time parameter estimates were calculated day by day [8].

Archival of data must allow expedited *access* for model developers and users. In addition, mechanisms should be available for manual or automatic retrieval of data and its translation into model parameter values in a timely manner.

In our review of the existing PI models, we examined the validity of data that was used in supporting major model parameters including basic reproduction number, illness attack rate, disease natural history, and fractions of symptomatic and asymptomatic individuals. The first row of Table 2.2 summarizes our approach to examine data validity. For each reviewed PI model, and, for each of the major model parameters, we examined the source and the age of data used (A1a, A1b), the type of interface used for data access and retrieval (A1c), and the technique used for translating data into the parameter values (A1d).

2.1.2 A2. Credibility and Validity of Model Assumptions

Public health practitioners have emphasized the need for models with credible and valid assumptions [17]. Credibility and validity of model assumptions generally refer to how closely the assumptions represent reality. However, for modeling purposes, assumptions are often made to balance the analytical tractability and computational feasibility of the models with

Table 2.2: A plan for examination of the challenges for practical use of the existing pandemic models.

Challenges for Practical Use	**Examined Elements of the Models**	
Validity of data support (A1)	Model parameter (demographical or epidemiological; e.g., basic reproduction number)	A1a. Data source for parameter values (actual, simulated, assumed) A1b. Age of data A1c. Type of interface for data access and extraction (manual, automatic) A1d. Technique to translate raw data into model parameter values (e.g., arithmetic conversion, Bayesian estimation)
Credibility and validity of model assumptions (A2)	See Section 2.2 “Credibility and validity of modeling assumptions (A2)". Assumptions analyzed belong to the contact and disease transmission dynamics	
Represent human behavior (A3)	Social-behavioral parameter (e.g., vaccination compliance)	A1a-d
Accessibility and scalability (A4,A5)	- Is the model software available to general public (open source or closed source code - Presence of end user support (user manuals, e-mail/phone technical support) - Information on the number of replicates - Information on the running time - Information on the ways to manage the computational load for implementing large-scale scenarios (e.g., the use of distributed and parallel computing) - Use of replicate minimization techniques - Type of interface for data access, extraction, and translation (manual, automatic). See A1 c,d	

their ability to support timely and correct decisions [20]. Making strong assumptions may produce results that are timely but with limited or no decision support value. On the other hand, relaxing the simplifying assumptions to the point of analytical intractability or computational infeasibility may seriously compromise the fundamental purpose of the models.

As indicated in row 2 of Table 2.2, we review and analyze two most significant areas of model assumptions concerning contact dynamics and disease transmission. We limit our examination of these assumptions to two of the most widely referenced PI models in the pandemic literature [11, 12].

2.1.3 A3. Ability to Represent Human Behavior

It has been observed in [17] that the existing PI models fall short of capturing relevant aspects of human behavior. This observation naturally evokes the following questions. What are the relevant behavioral aspects that must be considered in PI models? Are there scientific evidences that establish the relative importance of these aspects? What temporal consistency is required for data support of the aspects of human behavior? Finally, what are possible reasons that have refrained researchers from incorporating behavioral aspects in PI models?

The third row of Table 2.2 summarizes our plan to examine how the existing models capture human behavior. For each reviewed PI model, we first identify the behavioral aspects that were considered, and then for each aspect we examine the source and the age of data used, the type of interface used for data access and retrieval, and the technique used for translating data into model parameter values (A1 a-d). We also attempt to answer the questions raised above, with a particular focus on the question of why behavioral aspects have not been adequately addressed in the current PI models.

2.1.4 A4 and A5. Accessibility and Scalability

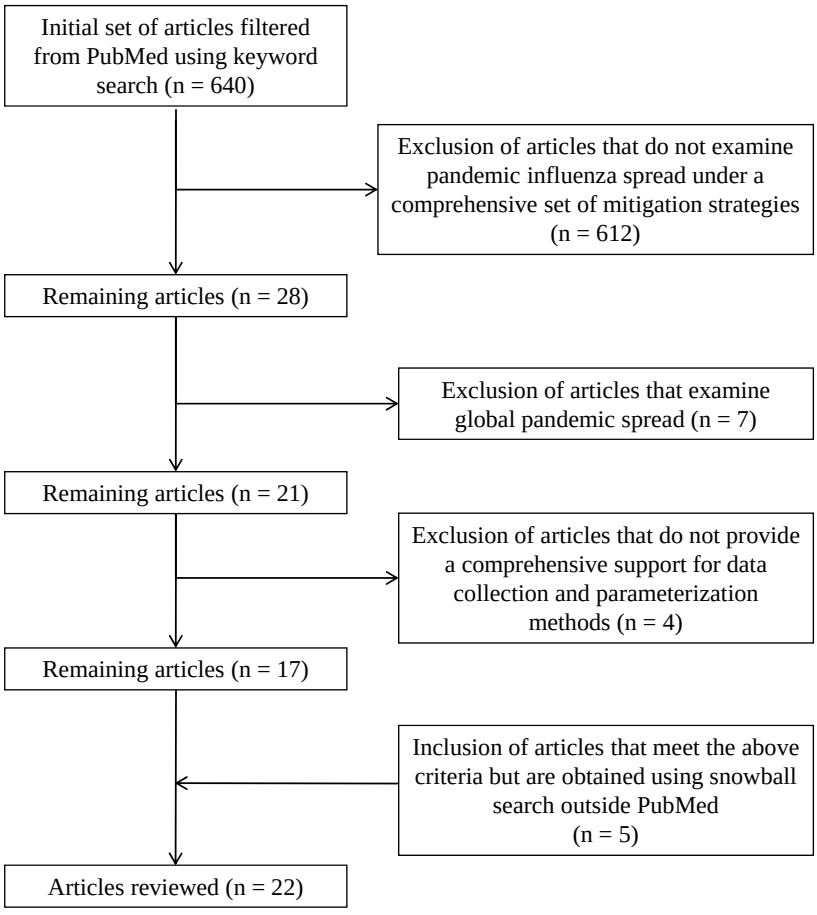
Public health practitioners have indicated the need for openly available models and population specific data that can be downloaded and synthesized using personal computers [17]. While the ability to access the models is essential for end users, executing the PI models on personal computers, in most cases, may not be feasible due to the computational complexities of the models. Some of the existing models feature highly granular description of disease spread dynamics and mitigation via consideration of scenarios involving millions of individuals and refined time scales. While such details might increase credibility and validity of the

models, this can also result in a substantial computational burden, sometimes, beyond the capabilities of personal computers.

There are several factors which contribute to the computational burden of the PI models, the primary of which is the population size. Higher population size of the affected region requires larger datasets to be accessed, retrieved, and downloaded to populate the models. Other critical issues that add to the computational burden are: data interface with a limited bandwidth, the frequency of updating of data during a pandemic progress, pre-processing (filtering and quality assurance) requirement for raw data, and the need for data translation into parameter values using methods, like maximum likelihood estimation and other arithmetic conversions.

The choice of the PI model itself can also have a significant influence on the computational burden. For example, differential equation (DE) models divide population members into compartments, where in each compartment every member makes the same number of contacts (homogeneous mixing) and a contact can be any member in the compartment (perfect mixing). In contrast, agent-based (AB) models track each individual of the population where an individual contacts only the members in his/her relationship network (e.g., neighbors, co-workers, household members, etc.) [23]. The refined traceability of individual members offered by AB models increases the usage of computational resources. Further increases in the computational needs are brought on by the need for running multiple replicates of the models and generating reliable output summaries.

As summarized in the last row of Table 2.2, we examine which models have been made available to general public and whether they are offered as an open or closed source code. We also check for the documentation of model implementation as well as for existence of user support, if any. In addition, we look for the ways that researchers have attempted to address the computational feasibility of their models, including data access, retrieval and translation, model execution, and generation of model outputs.



1.pdf

Figure 2.1: Selection criteria for PI models for systematic review.

2.2 Methods

The initial set of articles for our review was selected using the methodology used by Lee et al. [24]. We used the PubMed search engine with the keyword string “influenza" AND “pandemic" AND “model" in English language. A total of 640 papers were found which were published between 1990 and 2010. We filtered those using the following selection criteria (depicted in Figure 2.1):

- We considered the quantitative and comparative studies that evaluate a comprehensive set of mitigation strategies including social distancing, vaccine, and antiviral application. We selected only those models that evaluate at least three different mitigation options.

- We chose only those articles with single-region simulation models. We excluded the articles that consider of set of single-region simulations linked together to cover an extended geographic area.

- Models that include data sources for most parameter values and,when possible, specify the methods for parameter estimation.

Using the above filtering criteria, an additional snowball search was implemented outside PubMed, which yielded 5 additional papers bringing the total number of papers reviewed to twenty-two. We grouped the twenty-two selected articles in ten different clusters based on their model (see Table 2.3). The clusters are named either by the name used in the literature or by the first author name(s). For example, all three papers in the Imperial-Pitt cluster use the model introduced initially by Ferguson et al. [25]. In each cluster, to review the criteria for the design and implementation challenge (A1), we selected the article with the largest and most detailed testbed (marked in bold in Table 3). As stated earlier, credibility and validity of model assumptions (A2), were examined via two most commonly cited models in the pandemic literature [11, 12]. The challenges A3-A5 were examined separately for each of the selected articles.

Out of the ten model clusters presented in Table 2.3, seven are agent-based simulation models, while the rest are differential equation models. Also, while most of the models use purely epidemiological measures (e.g., infection attack rates and reproduction numbers) to assess the effectiveness of mitigation strategies, only a few use economic measures [26–28].

In our review, we examined epidemiological, demographical, and social-behavioral parameters of the pandemic models. We did not examine the parameters of the mitigation strategies as a separate category since those are functions of the epidemiological, demographical, and social-behavioral parameters. For example, the risk groups for vaccine and antiviral (which are mitigation parameters) are functions of epidemiological parameters such as susceptibility to infection and susceptibility to death, respectively. Another example is the compliance to non-pharmaceutical interventions, a mitigation strategy parameter, which can be achieved by altering the social behavioral parameters of the model.

Table 2.3: Clustering of selected review articles based on model type.

Model cluster	**Selected articles for review**
Imperial-Pitt	Ferguson et al. 2005 [25], **Ferguson et al. 2006 [12]**, Halloran et al. 2008 [29]
Wu	**Wu et al. 2006 [30]**
Arino	**Arino et al. 2006 [31]**, Arino et al. 2008 [32]
UW - LANL	Longini et al. 2004 [33], Longini et al. 2005 [34], **Germann et al. 2006 [11]**, Sander et al. 2009 [26], Chao et al. 2010 [35], Halloran et al. 2008 [29]
Gojovic	**Gojovic et al. 2009 [36]**
LOKI - INFECT	**Glass et al. 2006 [13]**, Davey et al. 2008 [37], Davey et al. 2008 [38], Perlroth et al. 2010 [27]
Nuno - Gumel	**Nuno et al. 2007 [39]**, Gumel et al. 2008 [40]
Roberts	**Roberts et al. 2007 [41]**
InfluSim	**Eichner et al. 2007 [42]**
USF	Das et al. 2008 [15], **Uribe et al. 2010 [28]**

2.3 Results and Discussion

In this section, we present the results of our review of the models from the selected articles. We also identify areas of enhancements of the simulation based PI models for higher practical usability.

2.3.1 A1. Validity of Data Support

Our discussion on validity of data support includes both epidemiological and demographic data. Appendix A summarizes the most common epidemiological parameters used in the selected models along with their data sources, interface for data access and retrieval, and techniques used in translating raw data into parameter values. Appendix B presents information similar to above for demographic parameters.

Support for epidemiological data. The most commonly used epidemiological parameters are reproduction number ($\mathscr{R}$), illness attack rate (IAR), disease natural history parameters, and fraction of asymptomatic infected cases. In the models that we have examined, estimates of reproduction numbers have been obtained by fitting case/mortality time series data from the past pandemics into models using differential equations [1], cumulative exponential growth equations [25], and Bayesian likelihood expressions [25]. IARs have been estimated primarily using household sampling studies [33], epidemic surveys [30, 43], and case time series reported for 2009 H1N1 [8, 36]. The parameters of the disease natural history, which is modeled using either a continuous or phase-partitioned time scale (see Appendix A), have been estimated from household random sampling data [25, 33, 44], viral shedding profiles from experimental control studies [35, 42, 45, 46], and case time series reported for 2009 H1N1[8, 36]. Bayesian likelihood estimation methods were used in translating 2009 case time series data [8, 25]. Fraction of asymptomatic infected cases has been estimated using data sources and translation techniques similar to the ones used for natural history.

We note that the parameters $\mathscr{R}$ and IAR, that are used to measure disease spread, are affected by changes in social dynamics during a pandemic. These parameters are critical to support decisions for initiation and scope of the interventions. Hence, a close monitoring of these parameters throughout the pandemic evolution is essential. This need has been acknowledged

in recent publications [47]. Real-time monitoring of parameters describing disease natural history and fraction of asymptomatic cases is generally not necessary since influenza viruses usually evolve between seasons of the year [48]. These parameters are estimated every time a viral evolution is confirmed through laboratory surveillance. Estimation of new values can be done by surveys (e.g., surveys of household members of index cases [49, 50]) or laboratory experiments that inoculate pandemic strains into human volunteers [51].

Current pandemic research literature shows the existence of estimation methodologies for IAR and $\mathscr{R}$ that can be readily used provided that raw data is available [8, 47]. However, as our survey reveals, there is no evidence showing that reliable estimates can be obtained in the critical early pandemic stages to support mitigation decisions. Examples of such decisions include school closures and mass gathering cancellations. Similarly, there is also no evidence that IAR and $\mathscr{R}$ estimates can be properly updated in the course of a pandemic. We believe that the limited ability to obtain prompt and reliable estimates of these pandemic impact parameters constitutes an area of enhancement.

We believe that the above enhancement can be achieved by developing a method for statistical sampling of specimens and conducting confirmatory testing in the public laboratories. In addition, new scheduling protocols will have to be developed for testing the specimens, given limited laboratory testing resources, in order to better assess the epidemiological parameters of an outbreak.

To assure that the above sampling covers the entire population, lab specimens should not only be collected from suspected cases seeking health care services, but also from those who may not seek medical assistance (e.g., uninsured). In addition, cost-effective strategies for dynamic data collection should also consider active surveillance. At present, most of the data collection occurs through passive surveillance.

Our review indicates that currently all of the tasks relating to the access and retrieval of epidemiological data are being done manually. Techniques for translation of data into model parameter values range from relatively simple arithmetic conversions to more time-consuming methods of fitting mathematical and statistical models (Appendix A). In the ideal case when epidemiological data are automatically accessed and retrieved in real-time, complex translation techniques might delay execution of the model. We believe that model developers should consider building (semi)automatic interfaces for epidemiological data access and retrieval and develop translation algorithms that can balance the run time and accuracy.

Support for demographic data. Appendix B shows the most common demographic parameters that are used in the selected models. The parameters are population size/density, distribution of household size, peer-group size, age, commuting travel, and long-distance travel. Estimation of these parameters has traditionally relied on comprehensive public databases, such as U.S. Census, National Household Travel Survey, and Landscan. Our literature review shows that access and retrieval of these data are currently limited to manual procedures. Moreover, there have not been any attempts to create methods for ready translation of such data into model parameters. Hence, the main opportunity for improvement lies on the development of strategies for (semi)automatic data access, retrieval, and translation.

2.3.2 A2. Credibility and Validity of Model Assumptions

As discussed in the background section, the issue of credibility and validity of the model assumptions should be viewed from the perspective of the balance between the analytical tractability and computational complexity of the model and its ability to provide accurate and timely decision support. In this section, we examine this balance for two of the commonly adopted model assumptions - *contact probability* and the *frequency of new infection updates*

(e.g., daily, quarterly, hourly). For this purpose, we selected the Imperial-Pitt [12] and the UW-LANL models [11], since they are among the most cited in the research literature on pandemic influenza mitigation. These models were also used for developing the CDC and HHS “Community Containment Guidance for Pandemic Influenza" [19]. The selection of the Imperial-Pitt and the UW-LANL models was also motivated by the similarities in their mixing groups and the infection transmission processes.

We first examine assumptions that influence the *contact probabilities* within different mixing groups. Within each mixing group, contact probability is a function of proximity and duration. Clearly, considering the impact of small variations in proximity and duration on the contact probability may add significantly to the computational burden. Both the Imperial-Pitt and the UW-LANL models considered a common set of mixing groups (see Table 2.4). However, the models considered different contact patterns for different mixing groups. The contact probabilities within the mixing groups were considered to be either constant, or age dependent [11], or proximity dependent (for community only, [12]). Duration of the contact events was not considered in either of the models. It can be inferred that the mitigation strategies studied in these models, such as vaccination, antiviral application, and contact reduction, did not explicitly require consideration of contact duration. For example, Ferguson et al. studied the effectiveness of applying antiviral prophylaxis to individuals within a ring of a certain radius centered around each detected case. Prophylaxis was administered considering the distance or proximity factor and not the contact duration [25]. It can also be observed from the pandemic preparedness plans offered by the agencies like CDC and HHS that the existing mitigation strategies are generally not aimed at reducing the contact duration [19]. We believe that a limited consideration of contact duration in the above models is a consequence of the lack of data support and the desire to minimize computational needs.

We also examined the assumptions regarding the *frequency of infection updates*. The frequency of update dictates how often the infection status of the contacted individuals is evalu-

ated. In reality, infection transmission occurs (or does not occur) whenever there is a contact event between a susceptible and an infected subject. But the Imperial-Pitt and the UW-LANL models do not evaluate infection after each contact event, since this would require consideration of refined daily schedules to determine the times of the contact events. Instead, the models evaluate infection status every six hours [12] or at the end of the day [11] by aggregating the contact events. While such simplifications do not allow the determination of the exact source and time of infection for each susceptible, they offer a significant computational reduction. Moreover, in a real-life situation, it will be nearly impossible to determine the exact source and time of infection, and hence no practical mitigation strategy should rely on it.

The above analysis reveals how the nature of mitigation strategies drives the the modeling assumptions and the computational burden. We therefore believe that there should be a consensus between policymakers and modelers whereby the degree of model abstractions is dictated by the mitigation needs of the policymakers.

Table 2.4: Factors that influence contact probabilities within mixing groups

Mixing group	**Factors that influence contact probabilities**	
	Imperial-Pitt (Ferguson, 2006)	UW-LANL (Germann, 2007)
Household	Contact probabilities are constant	Contact probabilities vary with age
Neighborhood	Mixing group is not considered	Contact probabilities vary with age
Workplace	75% percent of all workplace contacts occur within a workgroup of close colleagues and the remaining 25% of contacts occur outside the workgroup. Contact probabilities in both cases are constant.	Contact probabilities are constant
School (pre-school, elementary, middle, high, university)	As in workplace	Contact probabilities are constant
Community places, e.g., churches, banks, supermarkets, afterschool	Contact probability between two members varies according to proximity	Contact probabilities vary with age

2.3.3 A3. Represent Human Behavior

Contact rate is the most common social-behavioral aspect considered by the models that we have examined. In these models, except for Eichner et al. [42], the values of the contact rates were assumed due to the unavailability of reliable data required to describe the mobility and connectivity of modern human networks [13, 37, 38]. However, some recent surveys conducted in Europe provide “fresh" estimates of the types, frequency, and duration of human contacts [52].

Other social-behavioral parameters that are considered by the models include reactive withdrawal from work or school due to appearance of symptoms [25], work absenteeism to care for sick relatives or children at home due to school closure [25, 36, 38, 42], and compliance to social distancing, vaccination, and antiviral prophylaxis [29, 37]. Once again, due to the lack of data support, the values of most of these parameters were assumed and their sensitivities were studied to assess the best and worst case scenarios.

A recent survey [53] has explored many additional social-behavioral aspects that were not considered in the models we examined. These include perceived severity, perceived susceptibility, fear, general compliance intentions, compliance to wearing face masks, role of information, wishful thinking, fatalistic thinking, intentions to fly away, stocking, stay indoors, avoid social contact, avoid health care professionals, keep children at home and stay at home, go to work despite being advised to stay at home, and antiviral use [53].

We believe that there is a need for further studies to establish the relative influence of all of the above mentioned social-behavioral factors on pandemic spread and mitigation strategies. Subsequently, the influential factors need to be analyzed to determine how relevant information about those factors should be collected (e.g., in real-time or through surveys before an outbreak), accessed, retrieved, and translated into the final model parameter values.

2.3.4 A4 and A5. Accessibility and Scalability

With regards to accessibility and scalability of the selected models, we first attempted to determine which of the simulation models were made available to general public, either as an open or closed source code. We also checked for available documentation for model implementation and user support, if any. Most importantly, we looked into how the researchers attempted to achieve the computational feasibility of their models. Computational feasibility of a model refers to its ability to provide timely and accurate decision. For the purpose of this review, we identified several computational features affecting feasibility, such as need for replicates, run time per replicate, use of distributed and parallel computing approaches, and statistical means of minimizing the need for replicates (see Appendix C).

Two of the models that make their source codes accessible to general public are InfluSim [42] and FluTE [35]. InfluSim is a closed source differential equation-based model with a graphical user interface (GUI) which allows the evaluation of a variety of mitigation strategies, including school closure, place closure, antiviral application to infected cases, and isolation. FluTE is an open source model, which is an updated version of the UW-LANL [34] agent-based model. The source code for FluTE is also available as a parallelized version that supports simulation of large populations on multiple processors. Among these two softwares, InfluSim's GUI seems to be more user friendly for healthcare policymakers. FluTE, on the other hand, offers more options for mitigation strategies, but requires the knowledge of C/C++ programming language and the communication protocols for parallelization. We note that the policy makers would greatly benefit if the FluTE software can be made available through a cyber-enabled computing infrastructure, such as TeraGrid [54]. This will provide the policy makers access to the program through an web based GUI without having to cope with the issues of software parallelization and equipment availability. This will also eliminate the need to hire and train specialized personnel.

The need for replicates for accurate assessment of the model output measures and the run time per replicate are major scalability issues for pandemic simulation models. Large-scale simulations of the U.S. population reported running times of up to 6 hours per replicate, depending on the number of parallel threads used [35] (see Appendix C for further details). It would then take a run time of one week to execute 28 replicates of only one pandemic scenario. Note that, most of the modeling approaches have reported between 100 to 1000 replicates per scenario [13, 27, 36–40], with the exception of [12, 29] which implemented between 5 to 10 replicates. Clearly, it would take about one month to run 100 replicates for a single scenario involving the entire U.S. population.

While it may not be necessary to simulate the entire U.S. population to address mitigation related questions, the issue of the computation burden is daunting nonetheless. We therefore believe that the modeling community should actively seek to develop innovative methodologies to reduce the computational requirements associated with obtaining reliable outputs. One possible solution is the use of replicate minimization techniques. In principle, these techniques suggest running the replicates, one more at a time, until the confidence intervals for the output variables become acceptable [12, 28]. Use of such techniques would be crucial to provide timely decision support during the evolution of an actual outbreak.

2.4 Conclusions

Though the literature on pandemic models is rich and contains analysis and results that are valuable for public health preparedness, policy makers have raised several questions regarding practical use of these models. The questions are as follows. Is the data support adequate and valid? How credible and valid are the model assumptions? Is human behavior represented appropriately in these models? How accessible and scalable are these models? This

chapter attempts to determine to what extent the current literature addresses the above questions and what the areas of possible enhancements are. The findings with regards to the areas of enhancements are summarized below.

- Enhance the availability of real-time epidemiological data, the access and retrieval of demographical and epidemiological data, and the translation of data into model parameter values. We analyzed the most common epidemiological and demographical parameters that are used in pandemic models, and discussed the need for adequate updating of these parameters during an outbreak. As regards the epidemiological parameters, we have noted the need to obtain prompt and reliable estimates for the IAR and $\mathscr{R}$, which we believe can be obtained by enhancing protocols for expedited and representative specimen collection and testing. During a pandemic, the estimates for IAR and $\mathscr{R}$ should also be obtained as often as possible to update simulation models. For the disease natural history and the fraction of asymptomatic cases, estimation should occur every time viral evolution is confirmed by the public health laboratories. For periodic updating of the simulation models, there is a need to develop interfaces for (semi)automatic data access and retrieval. Algorithms for translating data into model parameters should not delay model execution and decision making. Demographic data are generally available. But most of the models that we examined are not capable of performing (semi)automatic access, retrieval, and translation of demographic data into model parameter values.

- Examine the validity of modeling assumptions from the point of view of the decisions that are supported by the model. By referring to two of the most commonly cited pandemic preparedness models [11, 25], we discussed how simplifying model assumptions are made to reduce computational burden, as long as the assumptions do not interfere with the performance evaluation of the mitigation strategies. Some mitigation strategies require more realistic model assumptions (e.g., location based antiviral prophylaxis would require models that track geographic coordinates of individuals so that those within a radius of an infected

individual can be identified). Whereas other mitigation strategies might be well supported by coarser models (e.g.,'antiviral prophylaxis for household members' would require models that track household membership). Therefore, whenever validity of the modeling assumptions is examined, the criteria chosen for the examination should depend on the decisions supported by the model.

- Incorporate a broader spectrum of social behavioral aspects in modeling disease spread and mitigation. Some of the social behavioral factors that have been considered in the examined models are social distancing and vaccination compliance, natural withdraw from work when symptoms appear, and work absenteeism to care for sick family members. Although some of the examined models attempt to capture social-behavioral issues, it appears that the literature lacks adequate consideration of many of the social behavioral factors and evaluation of their impact on disease spread and intervention effectiveness. Hence, there is a need for research studies or expert opinion analysis to identify which social-behavioral factors are significant for disease spread. It is also essential to determine how the social behavioral data should be collected (in real-time or through surveys), archived for easy access, retrieved, and translated into model parameters.

- Enhance computational efficiency of the solution algorithms. Our review indicates that some of the models have reached a reasonable running time of up to 6 hours per replicate for a large region, such as the entire U.S. [12, 35]. However, reliable decision making requires running much more than one replicate, a task that might be difficult to accomplish in real-time if replicate minimization techniques are not used. We have also discussed the question whether the public health decision makers should be burdened with the task of downloading and running models using local computers (laptops). This task can be far more complex than how it is perceived by the public health decision makers. We believe that models should be housed in a cyber computing environment with an easy user interface for the decision makers.

Chapter 3: A Viral Count Driven Calibration Approach for Simulation Models of Concurrent Pandemic and Seasonal Influenza Outbreaks

Influenza viruses imply major preparedness challenges for public health policymakers (PHP). With the uncertain epidemiology of these viruses, decision making often involves a trade-off between the social costs resulting from the implementation of a mitigation strategy (e.g., school closure) and the harm caused by the uncontrolled spread of an influenza virus [48]. To minimize the trade-off, PHP often need to re-evaluate mitigation strategies as the disease progresses.

Simulations are useful tools to model disease spread and support re-evaluation of strategies. Simulations are usually Differential Equation (DE) or Agent Based (AB) models. Differential equation (DE) models divide population members into compartments, where in each compartment, every member makes the same number of contacts (homogeneous mixing) and a contact can be any member in the compartment (perfect mixing). They are generally fast to run, and, without the presence of stochastic components, they require a single replicate. Simple DE models such as SIR (Susceptible-Infected-Recovered) or SEIR (Susceptible-Exposed-Infected-Recovered) are often used to grasp a basic understanding of the disease spread, and facilitate the analysis of sensitivity with several disease-related parameters. Therefore, simple DE models are the first choice for usage in progressing pandemic outbreaks [31, 42]. However, DE models lack of traceability features to recreate certain individual-based strategies

for the efficient use of limited resources (e.g., quarantine of the household members of an infected case, antiviral prophilaxis for the contacts of an infected case [11]).

Agent Based models are suitable for individual-based strategies, since the population is represented by individual members where contacts occur only with members of the same relationship network (e.g., neighbors, co-workers, household members, etc) [23]. Agent-based models have been previously used to inform preparedness guidelines [11–13], but there are yet challenges to support public health operational response in progressing pandemic outbreaks, since AB models may take long in their replication and calibration processes. As a result, they provide limited support when decisions are to be made by a close deadline. [17].

Replication guarantees statistically confident results. But the time expense in running replicates increases as the observed population grows larger. As an example, as discussed in Chapter 2, running 100 replicates of a simulation of the U.S. might take about one month with the running times reported in the literature. While it may not be necessary to simulate the entire U.S. population to address mitigation related questions, techniques to reduce replication time need to be considered. The issue of replication time is being recently addressed through high performance computing (HPC) and parallelization, by the Models of Infectious Disease Agent Study (MIDAS) network and other groups, as reported by the MIDAS software survey [55].

In the calibration process, internal parameters of a simulation are adjusted until obtaining the desired value for a standard epidemiological parameter. A key epidemiological parameter to characterize the disease transmission potential is the basic reproduction number R_0. Most of the existing simulation models refer to the R_0 since it is widely accepted by the community of theoretical epidemiologists [56], and previous pandemic outbreaks have been characterized through this parameter [1, 8, 25]. Deterministic compartmental models use R_0 as an input parameter and there is no need to calibrate [31, 39, 42]. Several AB simulations calibrate the

R_0 by re-running the model and adjusting internal parameters such as the infection probabilities and contact ratios [13, 33, 34, 57–59]. Some few approaches address calibration by deriving expressions of the R_0 in terms of the internal model parameters, and fitting values for the internal parameters through sampling techniques [12, 25], or nonlinear numerical techniques [30]. In either case, calibration and fitting times might be prohibitively large, and model outputs might become obsolete to support a decision in a progressing outbreak.

The objective of this Chapter is to present an epidemiological model for fast calibration of agent-based simulation of concurrent pandemic and seasonal influenza viruses. This feature increases the operability of simulation models during evolving pandemic outbreaks.

Our approach is unique since it uses a straightforward derivation of the infection probability based on the Reproduction Number and influenza viral shedding profiles in humans. This derivation enables both the calculation of a final R_o, similar to the reproduction number initially introduced in the model, and the inclusion of an alternate virus strain in the population. Both are features that, to our knowledge, have not been previously implemented in agent-based models for pandemic influenza.

This Chapter proceeds as follows. In Section 3.1 we present the epidemiological model used by the simulation. Section 3.2 describes the simulation model. Finally, Section 3.3 presents the results and advantages of the implementation.

3.1 Epidemiological Model

Let N_t^i denote the number of individuals infected by a single infectious individual i between time t and $t+1$. Thus, $R^i = \sum_{t=0}^{\infty} N_t^i$ is the total number of infected cases produced by the i^{th} individual, which is referred to as the reproduction number [56]. We use $\mathscr{N}_t$ to denote the

random variable for the number of new infections created by an individual between time t and t+1 and $\mathscr{R}$ to denote the random variable for the total number of infections by an individual. Note that $\mathscr{N}_t$ and $\mathscr{R}$ are defined over the sample spaces of N_t^i and R^i over all i, respectively, in the population.

Let I^k denote the set of infected cases in the k^{th} generation. Then, I^0 would denote the first set of infected cases that begin an outbreak when the entire population is susceptible. Note that I^1 would include all cases infected by those belonging to I^0. Let $\mathscr{R}^k$ denote the random variable defined over the sample space of R^i, for all $i \in I^k$.

Then $E[\mathscr{R}^0]$, for generation $k = 0$, is called the basic reproduction number and is commonly denoted in the literature by R_o [60]. For the subsequent generations $k = 1, 2, ...$, the susceptible population continues to decrease. If the value of $E[\mathscr{R}^k]$ is higher than one, the outbreak is in the expansion phase, whereas a value of less than one indicates a contracting phase and the outbreak is considered under control.

Let T^i denote the disease generation interval of infected individual i, i.e., the time interval between the infection of individual i and the infection of a secondary case produced by that individual. We use $\mathscr{T}$ to denote the random variable for the disease generation interval. Note that $\mathscr{T}$ is defined over the sample space of T^i over all i in the population. With the notation established above, the probability of creating a secondary infection by the i^{th} infected individual between time t and $t+1$ can be written as $\frac{N_t^i}{R^i}$. Let $w(t)$ denote the probability mass function for the disease generation interval $\mathscr{T}$. Then we have that

$$w(t) = \frac{E[\mathscr{N}_t]}{E[\mathscr{R}]}. \tag{3.1}$$

Also, it has been empirically shown in [25, 51] that $w(t)$ can also be expressed in terms of the absolute (not logarithmic) viral titer count V_t^i for individual i between time t and $t+1$. Let

the viral titer count random variable be $\mathscr{V}_t$ defined over the sample space of V_t^i. Then we can write that

$$w(t) = \frac{E[\mathscr{V}_t]}{E[\sum_{t=0}^{\infty} \mathscr{V}_t]}. \tag{3.2}$$

Calculating $w(t)$ from (3.1) is impractical due to the difficulty of collecting data to estimate expected values. However, since aggregated values for viral shedding profiles of human volunteers are available for influenza [51], (3.2) offers a better method to calculate $w(t)$. From (3.1) and (3.2), it follows that

$$E[\mathscr{N}_t] = E[\mathscr{R}] \frac{E[\mathscr{V}_t]}{E[\sum_{t=0}^{\infty} \mathscr{V}_t]}. \tag{3.3}$$

In what follows, we use (3.3) to obtain an expression for the probability of a contact getting infected.

We assume that $\mathscr{N}_t$ is binomially distributed with parameters c_t and p_t (denoted henceforth as $\mathscr{N}_t \sim bin[c_t, p_t]$), where c_t is the number of contact events that an infected case i makes between time t and $t+1$, and p_t is the probability that an individual becomes infected after being contacted by an infected case. From the binomial assumption, we can write that $E[\mathscr{N}_t] = c_t p_t$, which using (3.3) yields

$$p_t = \frac{E[\mathscr{R}] \frac{E[\mathscr{V}_t]}{E[\sum_{t=0}^{\infty} \mathscr{V}_t]}}{c_t}. \tag{3.4}$$

Note from the expression in (3.4) that the value of p_t can be easily obtained within the simulation model if the reproduction number $E[\mathscr{R}]$ is known. As stated earlier, information regarding the viral shedding profile is available [51], and c_t is a parameter that is generated

by simulation. In what follows, we extend (3.4) to account for the effect of different contact groups and asymptomatic infection on the probability of infection p_t.

In the simulation model, individuals interact within their contact groups such as households, workplaces, schools, and other open places (e.g., stores and restaurants). To account for differences in the closeness of interactions of these contact groups and their relative influence on the infection probability, we consider that $\mathscr{N}_t = \mathscr{N}_{t,h} + \mathscr{N}_{t,w} + \mathscr{N}_{t,o}$, where h, w and o are indices for household, work, school and open places, respectively. We assume that $\mathscr{N}_{t,h} \sim bin[c_{t,h}, p_{t,h}]$, $\mathscr{N}_{t,w} \sim bin[c_{t,w}, p_{t,w}]$, and $\mathscr{N}_{t,o} \sim bin[c_{t,o}, p_{t,o}]$, where $c_{t,h}$ is the number of contact events that an infected case makes in the household between t and $t+1$, and $p_{t,h}$ is the probability that an individual becomes infected after being contacted by an infected case in the household. Expressions $c_{t,w}$, $p_{t,w}$, $c_{t,o}$ and $p_{t,o}$ can be interpreted similarly. Thus we can write from (3.4) that

$$E[\mathscr{N}_{t,h} + \mathscr{N}_{t,w} + \mathscr{N}_{t,o}] = E[\mathscr{R}] \frac{E[\mathscr{Y}_t]}{E[\sum_{t=0}^{\infty} \mathscr{Y}_t]}.$$

Using the binomial assumption, we can rewrite the above as

$$c_{t,h}p_{t,h} + c_{t,w}p_{t,w} + c_{t,o}p_{t,o} = E[\mathscr{R}] \frac{E[\mathscr{Y}_t]}{E[\sum_{t=0}^{\infty} \mathscr{Y}_t]}.$$

We assume that $p_{t,w} = \kappa^w p_{t,h}$ and $p_{t,o} = \kappa^o p_{t,h}$, where $\kappa^w, \kappa^o < 1.0$ and $\kappa^w > \kappa^o$. The use of κ^w and κ^o can be argued from the fact that the duration and closeness of interactions at the households are maximum followed by those for workplaces/schools and open places. We then have that

$$p_{t,h}[c_{t,h} + \kappa^w c_{t,w} + \kappa^o c_{t,o}] = E[\mathscr{R}] \frac{E[\mathscr{V}_t]}{E[\sum_{t=0}^{\infty} \mathscr{V}_t]}.$$

which yields

$$p_{t,h} = \frac{E[\mathscr{R}] \frac{E[\mathscr{V}_t]}{E[\sum_{t=0}^{\infty} \mathscr{V}_t]}}{c_{t,h} + \kappa^w c_{t,w} + \kappa^o c_{t,o}} \tag{3.5}$$

From the above, we get estimates of infection probabilities for an infected individual belonging to contact groups h, w and o, during time interval t and $t+1$.

We can further extend the probability model to account for the fact that some asymptomatic cases could be symptomatic. Presence of asymptomatic cases would naturally lower $E[\mathscr{N}_t]$, since their viral shedding is lower compared to the symptomatic cases [51]. We incorporate the probability $(1-\pi_s)$ of asymptomatic cases as follows.

$$E[\mathscr{R}] = \pi_s * E[\mathscr{R}^s] + (1-\pi_s) * E[\mathscr{R}^a] \tag{3.6}$$

and

$$E[\mathscr{R}^a] = \gamma * E[\mathscr{R}^s]. \tag{3.7}$$

where $\gamma < 1$ and $E[\mathscr{R}^a]$ and $E[\mathscr{R}^s]$ denote the reproduction numbers for asymptomatic and symptomatic cases, respectively.

From (3.6) and (3.7)

$$E[\mathscr{R}^s] = \frac{E[\mathscr{R}]}{(1-\gamma)\pi_s + \gamma}. \tag{3.8}$$

With (3.7) and (3.8) we can extend the expression for $p_{t,h}$ in (3.5) for symptomatic and asymptomatic cases as follows

$$p^s_{t,h} = \frac{\frac{E[\mathscr{R}]}{(1-\gamma)\pi_s+\gamma}\frac{E[\mathscr{Y}^s_t]}{E[\sum_{t=0}^{\infty}\mathscr{Y}^s_t]}}{c_{t,h}+\kappa^w c_{t,w}+\kappa^o c_{t,o}}. \tag{3.9}$$

and

$$p^a_{t,h} = \frac{\frac{\gamma E[\mathscr{R}]}{(1-\gamma)\pi_s+\gamma}\frac{E[\mathscr{Y}^a_t]}{E[\sum_{t=0}^{\infty}\mathscr{Y}^a_t]}}{c_{t,h}+\kappa^w c_{t,w}+\kappa^o c_{t,o}}. \tag{3.10}$$

The expressions for $p^s_{t,w}$, $p^a_{t,w}$, $p^s_{t,o}$ and $p^a_{t,o}$ can be derived and interpreted similarly.

3.1.1 Modeling Co-Circulating Pandemic and Seasonal Influenza Viruses

We now extend our model to account for two co-circulating influenza viruses: pandemic (PI), and seasonal (SI). The PI virus has a high transmission potential, and a high value of the expected reproduction number $E[\mathscr{R}_{PI}]$. Conversely, The SI virus has a low transmission potential and a low value for the $E[\mathscr{R}_{SI}]$. The infection probabilities for the two circulating viruses can then be given from 3.4 as follows.

$$p_{t,PI} = \frac{E[\mathscr{R}_{PI}]\frac{E[\mathscr{Y}_t]}{E[\sum_{t=0}^{\infty}\mathscr{Y}_t]}}{c_t}. \tag{3.11}$$

$$p_{t,SI} = \frac{E[\mathscr{R}_{SI}]\frac{E[\mathscr{Y}_t]}{E[\sum_{t=0}^{\infty}\mathscr{Y}_t]}}{c_t}. \tag{3.12}$$

Considering the variations for contact groups and asymptomatic infection, the infection probabilities can be given from 3.9 and 3.10 as

$$p^s_{t,h,PI} = \frac{\frac{E[\mathscr{R}_{PI}]}{(1-\gamma)\pi_s+\gamma}\frac{E[\mathscr{V}_t^s]}{E[\sum_{t=0}^{\infty}\mathscr{V}_t^s]}}{c_{t,h}+\kappa^w c_{t,w}+\kappa^o c_{t,o}}. \tag{3.13}$$

$$p^a_{t,h,PI} = \frac{\gamma\frac{E[\mathscr{R}_{PI}]}{(1-\gamma)\pi_s+\gamma}\frac{E[\mathscr{V}_t^a]}{E[\sum_{t=0}^{\infty}\mathscr{V}_t^a]}}{c_{t,h}+\kappa^w c_{t,w}+\kappa^o c_{t,o}}. \tag{3.14}$$

$$p^s_{t,h,SI} = \frac{\frac{E[\mathscr{R}_{SI}]}{(1-\gamma)\pi_s+\gamma}\frac{E[\mathscr{V}_t^s]}{E[\sum_{t=0}^{\infty}\mathscr{V}_t^s]}}{c_{t,h}+\kappa^w c_{t,w}+\kappa^o c_{t,o}}. \tag{3.15}$$

$$p^a_{t,h,SI} = \frac{\gamma\frac{E[\mathscr{R}_{SI}]}{(1-\gamma)\pi_s+\gamma}\frac{E[\mathscr{V}_t^a}{E[\sum_{t=0}^{\infty}\mathscr{V}_t^a]}}{c_{t,h}+\kappa^w c_{t,w}+\kappa^o c_{t,o}}. \tag{3.16}$$

Equations for $p^s_{t,w,PI}$, $p^a_{t,w,PI}$, $p^s_{t,w,SI}$, $p^a_{t,w,SI}$, $p^s_{t,o,PI}$, $p^a_{t,o,PI}$, $p^s_{t,o,SI}$ and $p^a_{t,o,SI}$ can be derived similarly. These probabilities are evaluated once an infected case contacts another individual in the population. Table 3.1 shows the six possible outcomes resulting from the contact. If the contacting individual is only infected with PI (column named "Strains in the contacting person", row 1, Table 3.1), the contacted individual can only receive PI (column named “Strains transmitted", row 1, Table 3.1). If the contacting individual is infected with SI, the infected individual can only receive SI (columns named “Strains in the contacting person" and “Strains transmitted", row 2, Table 3.1). But if the contacting individual is infected with both viruses, she can transmit either PI or SI or both (columns named “Strains in the contacting person" and “Strains transmitted", rows 3,4 and 5, Table 3.1).

Table 3.1: Types of infection when a pandemic and a seasonal virus co-circulate in the simulation model.

Strains in the contacting person	**Strains transmitted**	**Strains in the contacted person**	**Possible infection in the contacted person**	**Cross-immunity factor**
PI	PI	None	Infection PI	None
		Recovered from SI	Sequential infection PI	ε_{pr}
		Ongoing SI	Co-infection SI, PI	ε_{p}
SI	SI	None	Infection SI	None
		Recovered from PI	Sequential infection SI	ε_{sr}
		Ongoing PI	Co-infection PI, SI	ε_{s}
PI , SI	PI	None	Infection PI	None
		Recovered from SI	Sequential infection PI	ε_{pr}
		Ongoing SI	Co-infection SI, PI	ε_{p}
PI , SI	SI	None	Infection SI	None
		Recovered from PI	Sequential infection SI	ε_{sr}
		Ongoing PI	Co-infection PI, SI	ε_{s}
PI , SI	PI , SI	None	Simultaneous Co-infection	ε_{ps}
		Recovered from SI	Sequential infection PI	ε_{pr}
		Ongoing SI	Co-infection SI, PI	ε_{p}
		Recovered from PI	Sequential infection SI	ε_{sr}
		Ongoing PI	Co-infection PI, SI	ε_{s}

Once the virus(es) enter(s) the body of the contacted individual, based on the viral strains in the contacting person and the infection probabilities, the internal body infection can occur in many different ways, depending on the viral strains currently infecting the contacted person (column 3 in the Table 3.1). In our model, we consider the three ways that are observed more frequently, and are also shown in Table 3.1:

1. The first way is when only one strain, either PI or SI (in the column “Strains transmitted" in the table), is transmitted to a contacted individual without the PI or the SI strain in her body (denoted as “none" in the column “Strains in the contacted person"). In this case, the possible infection in the contacted person is either SI or PI (denoted as “Infection PI" or “Infection SI" in “Possible infection in the contacted person").

2. The second way is when an individual, once recovered from one virus, gets infected with the other in a process called “sequential infection". Two types of sequential infection are allowed in our model: The first type is sequential infection with PI (“Sequential infection PI", in the column “Possible infection in the contacted person"), which only occurs after the individual recovers from a SI infection (“Recovered from SI" in the column “strains in the contacted person"). The second type is sequential infection with SI (“Sequential infection SI"), which only occurs after the individual recovers from a PI infection (“Recovered from PI"). When an individual i is sequentially infected with PI, we assume that the reproduction number R^i of this individual will be reduced by a cross-immunity factor ε_{pr}, to account for the CD8 T-cell mediated immunity built by the SI virus [61]. When an individual i is sequentially infected with SI, we assume that the R^i of this individual will be reduced by a cross-immunity factor ε_{sr}.

3. The third way is when an individual gets infected with a new virus while still sick from the other virus in a process called "co-infection" [62]. Three types of co-infection may occur in the model: Simultaneous co-infection, which occurs when both viruses infect an individual

at the same time (“Simultaneous Co-infection"); co-infection with PI (“Co-infection SI, PI"), which occurs while an individual is infected with SI (“Ongoing SI"), and co-infection with SI (“Co-infection PI, SI"), which occurs while an individual is infected with PI (“Ongoing PI"). For last the two types of co-infection, we implemented the cross-immunity factors ε_p and ε_s.

We assume that re-infection (Sequential infection with the same virus) or co-infection with the same virus does not occur since these are not typical ways of infection and there is yet the need of research that describes the progression of these two mechanisms in the human body.

3.2 The Simulation

The simulation recreates the daily interactions of urban citizens (individuals) from a single region, and the spread of the influenza viruses through the citizens of the region.

3.2.1 Daily Interactions of Urban Citizens

At the beginning of the simulation, each individual is created and assigned to an age group, a household, a school or work place. In addition, she is assigned to a set of errand places that she will visit during weekdays and weekends. The assignments are based on the distributions shown in Tables 3.2, 3.3 and 3.4 which were calculated based on data obtained from the 2002 Economic Census, the 2001 American Community Survey and the 2001 National Household Travel Survey.

Individuals are set to attend these locations by following an hour by hour schedule. Table 3.5 shows the four types of schedules assumed (weekday-employed-adult, weekday-unemployed-

adult, weekday-children and weekend-everyone). These schedules were adapted from [15] and [63].

Table 3.2: Distribution of population in the Hillsborough County by age.

Age	School type	Cumulative distribution
<5	Pre-school	0.07459
9	Elementary	0.14729
14	Middle-school	0.22422
17	High-school	0.26464
22	University	0.31308
29		0.42222
64		0.88457
>64		1

Data for the age distribution were obtained from [64].

Table 3.3: Distribution of population based on the number of adults and children living in the same household.

Number of adults	Number of children	Cumulative distribution
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

Data for the household distribution in the Unites States were obtained from [65].

The simulation runs day by day. Each day, individuals are hourly placed in a location following their corresponding schedule. Once they are in the location, individuals randomly make a specific number of contacts. The number of contacts in the location was extracted from a recent survey on mixing patterns and contact characteristics during the seven days of the week [52]. Table 3.6 shows in column 4 the number of contacts that each individual makes per location setting. Details on the calculation of these numbers can be found in the footnote of the table.

Table 3.4: Features of household, work, school and errand places in the United States.

Location type	**Number of locations per type (100000 inhabitants)**	**Number of locations per type (1000000 inhabitants)**	**Percentage of individuals in the location**	**Percentage of trips made during weekdays**	**Percentage of trips made during weekends**
Household	100	1	0.06586	†	0
Factory	700	613	0.05802	†	0
Office	240	2266	0.30227	†	0
Pre-school	30	224	0.0048	†	0
Elementary	10	66	0.00997	†	0
Middle	20	134	0.203	†	0
High	10	59	0.09736	†	0
University	10	46	0.10598	†	0
After-school	30	256	0.00681	†	0
Grocery	50	390	0.02599	0.61919	0.51493
Restaurant	30	223	0.08749	0.27812	0.25586
Entertainment	40	360	0.03181	0.06601	0.1162
Church	10	86	0.00064	0.03668	0.113

Workplace features. The type of location and the number of locations per type (columns 2, 3 and 4) were obtained from [66]. The percentage of trips made during weekdays (column 5) and percentage of trips made during weekends (column 6) are based on the total number of trips made to all the location types during a weekday/weekend. Data for these percentages were obtained from [22]. †Indicates that all this location are for mandatory attendance if they are included in the daily weekday schedule of the individual.

3.2.2 Spread of the Influenza Viruses

At the beginning of the simulation, a set of $M = M_{PI} + M_{SI}$ randomly selected cases are introduced in the simulation. For these M initial cases and for any other subsequent case, the simulation randomly determines whether a case is symptomatic or asymptomatic. Of the cases in the simulation, 66.9% present symptomatic infection and 33.1% present asymptomatic infection, consistent with recent review studies on controlled groups of influenza volunteers [51]. To each case i, the simulation randomly assigns a probability mass function for the viral shedding profile $\frac{E[\mathscr{V}_t^s]}{E[\sum_{t=0}^{\infty}\mathscr{V}_t^s]}$, when the case is symptomatic, and $\frac{E[\mathscr{V}_t^a]}{E[\sum_{t=0}^{\infty}\mathscr{V}_t^a]}$ when the case is asymptomatic. For the symptomatic group, we considered three probability mass functions: 1) Lognormal with shape parameter, $a = 3.98098$ and scale parameter $b = 0.286479$, 2) Gamma with $a = 11.7534$ and $b = 4.75874$, and 3) Weibull with $a = 62.1897$ and $b =$

Table 3.5: Schedules assumed in the simulation model.

Hour	Weekday employed	Weekday unemployed	Weekday children	Weekend everyone
1	household	household	household	household
2	household	household	household	household
3	household	household	household	household
4	household	household	household	household
5	household	household	household	household
6	household	household	household	household
7	household	household	household	household
8	workplace	errands	workplace	household
9	workplace	household	workplace	household
10	workplace	household	workplace	errands
11	workplace	household	workplace	errands
12	workplace	errands	workplace	household
13	workplace	household	workplace	household
14	workplace	household	workplace	errands
15	workplace	household	workplace	household
16	workplace	errands	after-school	household
17	workplace	household	after-school	household
18	errands	household	after-school	household
19	errands	household	household	household
20	household	household	household	household
21	household	household	household	household
22	household	household	household	household
23	household	household	household	household
24	household	household	household	household

The four types of schedules assumed. Employed adults and children follow a fixed routine during weekdays. Unemployed adults are allowed to go to errand places for three randomly selected hours/day, between the 8th and the 19th hour. The table presents an example of a weekday schedule for an unemployed adult. During weekends, all individuals are allowed three randomly selected hours of errand, between the 10th and the 20th hour. The table presents an example of a weekend schedule.

2.93571. For the asymptomatic group we also considered three probability mass functions: 1) Lognormal with shape parameter, $a = 4.11187$ and scale parameter $b = 0.491289$, 2) Gamma with $a = 4.23216$ and $b = 16.3124$, and 3) Weibull with $a = 78.3308$ and $b = 2.01736$. These functions were derived from the data in [51]. Table 3.7, columns 3 and 4 show the data used to fit the probability mass functions.

Infected cases contact other individuals in the region as described in Subsection 3.2.1. The contacts an infected case i makes per location are recorded and aggregated to compute the values of $c^i_{t,h}$, $c^i_{t,w}$ and $c^i_{t,o}$, in the interval between the beginning of day t and the beginning of day $t+1$ (see Section 3.1). Before the beginning of day $t+1$, a new set of infected cases is

Table 3.6: Number of contacts an individual makes per day and per location setting.

Location type	**Number of contacts in the location per day**	**Proportion of contacts occurring in the location**	**Number of contacts in the location per day**
Home	13.4	0.23	3
Factory	14	0.21	3
Office	14	0.21	3
Pre-school	10.21	0.14	2
Elementary	14.81	0.14	2
Middle school	18.22	0.14	3
High school	17.6	0.14	3
University	15.98	0.14	2
Afterschool	13.4	0.14	2
Grocery	13.4	0.16	2
Restaurant	13.4	0.16	2
Entertainment	13.4	0.16	2
Church	13.4	0.16	2

Data for the contacts was extracted from [52], which is a survey on contact patterns in eight European countries. To our knowledge, no such studies exist for the American society and we consider the European social behavior as the best existing proxy for a developed country, since the contact features are remarkably similar across the eight countries included in the sample. In the table, column 2 "Number of contacts in the location per day" and column 3 "Proportion of contacts occurring in the location" show the information provided by the survey. Column 4 (Number of contacts in the location per day) results from multiplying the number of contacts in a location per day and the proportion of contacts occurring in the location.

Table 3.7: Daily viral titer count in symptomatic and asymptomatic volunteers inoculated with influenza A/H1N1.

Days	**Hours**	**Viral titer in a symptomatic individual (log scale)**	**Viral titer in an asymptomatic individual (log scale)**
0-1	0-24	0.1	0.05
1-2	24-48	1.75	0.875
2-3	48-72	3	1.5
3-4	72-96	2.5	1.25
4-5	96-120	1.8	0.9
5-6	120-144	1.2	0.6
6-7	144-168	0.7	0.35
7-8	168-192	0.5	0.25
8-9	192-216	0.2	0.1

Column3. Daily viral titer count in symptomatic volunteers inoculated with influenza A/H1N1. Column 4. Calculated daily viral titer count in asymptomatic volunteers inoculated with influenza A/H1N1. These values were obtained by assuming that the daily quantity of virus from symptomatic volunteers who shed virus and develop illness was two log_{10} times higher than individuals who did not develop illness. This assumption is supported by studies of influenza A/H3N2 on volunteers with and without clinical illness. In those studies, the mean quantity of virus from volunteers who shed virus and develop illness was from two log_{10} to three log_{10} times higher than individuals who did not develop illness.

selected from the individuals contacted by case i during day t. If the i is a case of pandemic flu, the equation 3.11 is used to calculate the infection probability $p^i_{t,PI}$. If i is a case of seasonal flu, the equation 3.12 is used for $p^i_{t,SI}$

When using the extended version of the model to account for contact groups and asymptomatic infection, the following set of rules are used to determine the new set of infected cases from the individuals contacted by case i. If the individuals are contacted in the household, and i is a symptomatic case of pandemic influenza, the equation 3.13 is used to calculate $p^{i,s}_{t,h,PI}$, the probability of infecting the household contacts of case i. If i is an asymptomatic case of seasonal influenza, the equation 3.15 is used to calculate $p^{i,s}_{t,h,SI}$. Subsection 3.1 enumerates the remaining probabilities that are used in the determination of the new infected cases. Table 3.8 shows the parameter values used to calculate the probabilities.

Table 3.8: Parameter values used to calculate the infection probabilities

Parameter	**Value**	**Reference**
$E[R^i_{PI}]$	3,2.5,2.0,1.8	[11, 25, 25]
$E[R^i_{SI}]$	1.3	[10]
$\frac{E[V_t^{i,s}]}{E[\sum_{t=0}^{\infty} V_t^{i,s}]}$	Refer to this subsection and table 3.7	[51]
γ	0.22†	[51]
π_s	0.669‡	[51]
κ^w	0.67	[52]
κ^o	0.44	[52]

† This value represents ratio between the average viral shedding of a symptomatic individual (average of the values in column 3, Table 3.7) and the average viral shedding of an asymptomatic individual (average of the values in column 4, Table 3.7). The ratio is then converted to absolute scale. ‡ This value is derived from the proportion of symptomatic individuals produced by an influenza virus. For this parameter, we interpreted the proportion as a probability.

3.3 Results

In this section, we illustrate how the epidemiological model derived in Section 3.1, and embedded in the simulation described in Section 3.2 attains the objective of avoiding calibration of R_o, the basic reproduction number.

The parameter R_o is mathematically defined in Section 3.1 as the number of infected cases per each case in generation zero, just before it begins the depletion of susceptible population. However, observing the value of this parameter in generation zero is not the case for all the simulation replicates, since the initial number of infected cases is generally small to provide enough information to determine the true value of R_o. When the number of infected cases in a generation $k \geq 0$ is significant, an estimation of the R_o can also be observed from an $E[\mathscr{R}^k]$ that peaks and then decreases in the subsequent generation [13].

For a population of around 100000 inhabitants and 10 cases in generation zero, Table 3.9, shows the average $\widehat{R}_o$ obtained when different expected reproduction numbers $E[\mathscr{R}]$ are introduced in the simulation and (3.11) is used to model the infection probability. For an $E[\mathscr{R}] = 2.0$ (first sub-table), 48 replicates expressed a peak of 2.17292 in generation $k = 0$, and 35, 24, 17, 11 and 8 replicates peaked on $k = 1$, $k = 2$, $k = 3$, $k = 4$ and $k = 5$ with a value of 2.07876, 2.07226, 2.00232, 1.94093, 1.96075 and 2.08306, respectively. Averaging the peaks of the 143 replicates, an average $\widehat{R}_o$ of 2.08306 is obtained, with a lower confidence interval (CI) of 2.04431, just slightly above the value of two. For an $E[\mathscr{R}] = 2.5$, (first sub-table), an average of the peaks of 135 replicates provides an estimation of 2.50199, with a lower CI of 2.45367 and an upper CI of 2.55032. For an $E[\mathscr{R}] = 3.0$, (third sub-table), an average of the peaks of 150 replicates provides an estimation of 2.92014, with an upper confidence interval (CI) of 2.96379, just slightly below the value of three. For a population of around 1000000 inhabitants and 10 cases in generation zero, Table 3.9 presents the results for the average $\widehat{R}_o$ obtained when $E[\mathscr{R}] = 2.0$, $E[\mathscr{R}] = 2.5$ and $E[\mathscr{R}] = 3.0$ are introduced in the simulation.

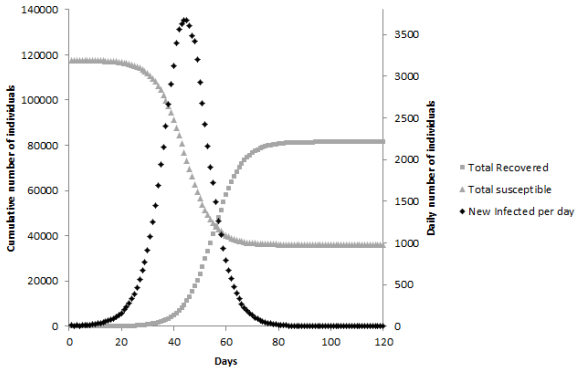
These results indicate that it is possible to achieve an average $\widehat{R}_o$ similar to the expected reproduction number $E[\mathscr{R}]$ introduced in the simulation. This conclusion holds for different population levels. The slight variations between the initial $E[\mathscr{R}]$ and the average $\widehat{R}_o$ can be attributed to the stochastic disease spread process embedded in the simulation and possible instabilities that simulations normally present at their initial stages.

Graphs 3.1(a) and 3.1(b) show the features of an outbreak with an average $\widehat{R}_o = 2$ expressed in generation zero. For a population of around 100000 inhabitants, the outbreaks peak around day 44 with 3674 cases (read using the right scale "Daily number of individuals", Figure 3.1(a)) and last around 100 days, leaving a total of 36147 susceptible and 81479 recovered cases (cumulative susceptible and cumulative recovered cases are read using the left scale "cumulative number of individuals" in graph 3.1(a)). Graph 3.1(b) shows the average $\widehat{R}_0$ in generation zero and the decreasing values for the average $E[\mathscr{R}^k]$ for generations $k > 0$. For a population of around 1000000 inhabitants (Graph 3.1(c)), the outbreaks peaks around day 55 with 34473 cases and last around 120 days, with 315618 susceptible individuals and 719914 recovered cases. The average $\widehat{R}_o = 2.2$ is observed in generation zero in the graph 3.1(d).

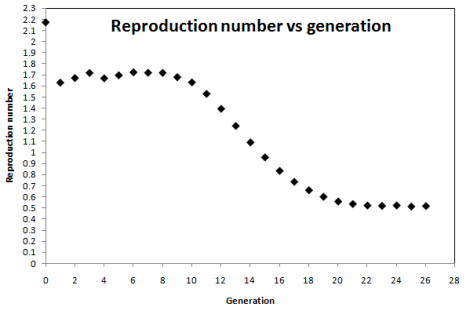
Table 3.9: Values for the $\widehat{R}_0$ obtained for an outbreak region with population 100000 inhabitants.

$E[\mathscr{R}] = 2.0$	$E[\mathscr{R}^0]$	$E[\mathscr{R}^1]$	$E[\mathscr{R}^2]$	$E[\mathscr{R}^3]$	$E[\mathscr{R}^4]$	$E[\mathscr{R}^5]$	**Average**
Number of replicates with the peak on the generation	48	35	24	17	11	8	143 (replicates)
Average $\widehat{R}_o$ over the replicates	2.17292	2.07876	2.07226	2.00232	1.94093	1.96075	2.08306
Standard deviation	0.29589	0.19610	0.25352	0.12903	0.14155	0.15082	0.24017
Lower 95% CI	2.08700	2.0114	1.96520	1.93598	1.84583	1.83466	2.04431
Upper 95% CI	2.25883	2.14612	2.17931	2.06866	2.03602	2.08684	2.12181
$E[\mathscr{R}] = 2.5$	$E[\mathscr{R}^0]$	$E[\mathscr{R}^1]$	$E[\mathscr{R}^2]$	$E[\mathscr{R}^3]$	$E[\mathscr{R}^4]$	$E[\mathscr{R}^5]$	**Average**
Number of replicates with the peak on the generation	48	35	18	16	11	7	135 (replicates)
Average $\widehat{R}_o$ over the replicates	2.73958	2.50242	2.37344	2.37237	2.317093	2.33657	2.50199
Standard deviation	0.370589	0.18789	0.15814	0.09495	0.08997	0.060369	0.29951
Lower 95% CI	2.631976	2.43788	2.29480	2.32177	2.25665	2.280741	2.45367
Upper 95% CI	2.847191	2.56696	2.45208	2.42297	2.37754	2.392404	2.55032
$E[\mathscr{R}] = 3.0$	$E[\mathscr{R}^0]$	$E[\mathscr{R}^1]$	$E[\mathscr{R}^2]$	$E[\mathscr{R}^3]$	$E[\mathscr{R}^4]$	$E[\mathscr{R}^5]$	**Average**
Number of replicates with the peak on the generation	49	31	15	19	19	17	150 (replicates)
Average $\widehat{R}_o$ over the replicates	3.14898	2.93232	2.85789	2.79178	2.71286	2.66838	2.92014
Standard deviation	0.30285	0.14774	0.21743	0.10441	0.08294	0.05809	0.27056
Lower 95% CI	3.06199	2.87813	2.73748	2.74145	2.67289	2.63851	2.87649
Upper 95% CI	3.23597	2.98651	2.97829	2.8421	2.75284	2.69824	2.96379

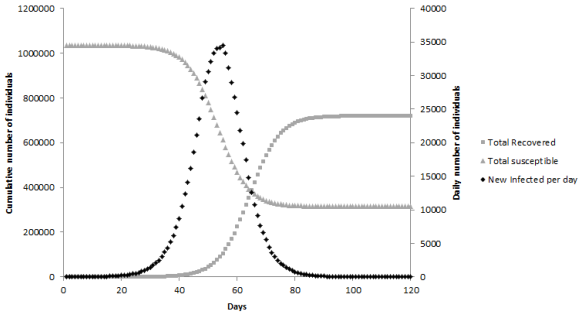
Besides the average R_o, another measure of disease transmissibility is the Infection Attack Rate (IAR). The infection attack rate is defined as the ratio between the final number of infected cases and the initial number susceptible cases. With both the average R_o and the IAR,



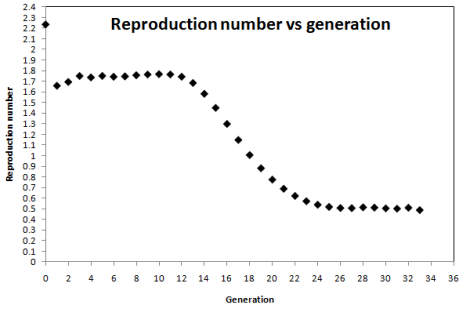
(a) Population size = 100000 inhabitants



(b) Population size = 100000 inhabitants



(c) Population size = 1000000 inhabitants



(d) Population size = 1000000 inhabitants

Figure 3.1: Disease spread behavior and expected reproduction number per generation for a PI outbreak with an $\widehat{R}_o = 2$ expressed in generation zero.

Table 3.10: Values for the $\widehat{R}_o$ obtained for an outbreak region with population 1000000 inhabitants.

$E[\mathscr{R}] = 2.0$	$E[\mathscr{R}^0]$	$E[\mathscr{R}^1]$	$E[\mathscr{R}^2]$	$E[\mathscr{R}^3]$	$E[\mathscr{R}^4]$	$E[\mathscr{R}^5]$	**Average**
Number of replicates with the peak on the generation	41	32	18	24	12	8	135 (replicates)
Average $\widehat{R}_o$ over the replicates	2.22927	2.16272	2.11159	1.98649	1.92760	1.94942	2.11124
Standard deviation	0.25124	0.24888	0.26396	0.12306	0.11900	0.13722	0.24295
Lower 95% CI	2.14997	2.07298	1.98032	1.93453	1.85199	1.83470	2.06993
Upper 95% CI	2.30857	2.25245	2.24285	2.03846	2.00321	2.06414	2.15256
$E[\mathscr{R}] = 2.5$	$E[\mathscr{R}^0]$	$E[\mathscr{R}^1]$	$E[\mathscr{R}^2]$	$E[\mathscr{R}^3]$	$E[\mathscr{R}^4]$	$E[\mathscr{R}^5]$	**Average**
Number of replicates with the peak on the generation	46	37	24	14	10	6	137 (replicates)
Average $\widehat{R}_o$ over the replicates	2.64130	2.58423	2.47563	2.36186	2.34260	2.34506	2.53353
Standard deviation	0.24274	0.22850	0.14634	0.10948	0.11887	0.06663	0.22734
Lower 95% CI	2.56922	2.50805	2.41384	2.29865	2.25757	2.27513	2.49515
Upper 95% CI	2.71339	2.66042	2.53743	2.42507	2.42764	2.41499	2.57191
$E[\mathscr{R}] = 3.0$	$E[\mathscr{R}^0]$	$E[\mathscr{R}^1]$	$E[\mathscr{R}^2]$	$E[\mathscr{R}^3]$	$E[\mathscr{R}^4]$	$E[\mathscr{R}^5]$	**Average**
Number of replicates with the peak on the generation	43	29	19	10	6	7	114 (replicates)
Average $\widehat{R}_o$ over the replicates	3.12791	2.94846	2.82230	2.73569	2.71901	2.71981	2.95034
Standard deviation	0.28894	0.22926	0.08772	0.08461	0.04867	0.08942	0.26711
Lower 95% CI	3.03898	2.86125	2.78002	2.67516	2.66793	2.63711	2.90090
Upper 95% CI	3.21683	3.03566	2.86458	2.79622	2.77009	2.80250	2.99977

a simulated outbreak can be compared to other modeling approaches to guarantee the validity of the results. But the average R_o and the IAR of most modeling approaches are highly dependent on the demographical, epidemiological and behavioral assumptions of each approach, and in many cases a comparison might not provide a good evidence of quality. In our case, we followed the methodology proposed in [67], where the comparisons are made with the Kermack-McKendrick Susceptible-Infected-Recovered (SIR) model. The SIR model is a a set of three differential equations that assume homogeneous and perfect mixing of the population modeled (refer to Chapter 2, for a definition of both terms). It is analytically tractable, and the relationship between average R_o and the IAR can be easily obtained through the final size equation

$$ln(IAR) = R_o(IAR - 1). \tag{3.17}$$

Table 3.11 shows the IARs obtained for all the scenarios presented in Tables 3.9 and 3.10. These IARs are always lower to the ones obtained through (3.17), a logical result when the assumptions of homogeneous and perfect mixing are relaxed, as is the case of our model.

Table 3.11: Infection attack rates for different simulation scenarios.

Population	$\widehat{R}_o$	**Mean IAR from simulation**	**Upper 95% CI**	**Lower 95% CI**	**IAR from Kermack-McKendrick model**
100000	2	0.69338	0.69056	0.69620	0.84
100000	2.5	0.82577	0.82386	0.82768	0.895
100000	3	0.89529	0.89402	0.89656	0.935
1000000	2	0.68823	0.68538	0.69108	0.84
1000000	2.5	0.81852	0.81657	0.82048	0.895
1000000	3	0.88730	0.88599	0.88862	0.935

For co-circulating outbreaks, we simulated two scenarios assuming there is no cross-immunity (all the ε factors in Table 3.1 and column 5 are equal to 1). The first scenario considered both outbreaks with $E[\mathscr{R}] = 1.8$, and infection probabilities (3.11) and (3.12). For a population of 100000 inhabitants and 30 replicates, we obtained an average $\widehat{R}_o = 1.7725$ with CI (1.7281, 1.8169) for the first outbreak, and a $\widehat{R}_o = 1.7724$ with CI (1.7237, 1.8212) for the second outbreak. The corresponding IAR values were 0.6090 with CI (0.5928, 0.6253), and 0.608 with CI (0.5917, 0.6242), respectively. Figure 3.2 shows the daily number of infected cases for a typical replicate, which peaks at day 49 and finishes around day 120. Figure 3.3 shows the inmediate depletion of the $E[\mathscr{R}^k]$ after the second generation. The instabilities occuring after the 27th generation are due to the small number of infected cases used to calculate the reproduction number, creating inconsistent values that disappear when the total extinction of the outbreaks occurs at generations 36 and 42.

The second scenario is set with a pandemic outbreak with $E[\mathscr{R}_{PI}] = 1.8$, a seasonal outbreak with $E[\mathscr{R}_{SI}] = 1.3$, and infection probabilities (3.14), (3.13), (3.16) and (3.15). For 100000 inhabitants and 146 replicates, The average $\widehat{R}_o$ for PI was equal to 1.67 with CI(1.6432,1.6948), and the average $\widehat{R}_o$ for SI was equal to 1.39 with CI(1.3633,1.4251). The corresponding IAR values were 0.5373 with CI (0.534,0.541), and 0.0672 with CI (0.0663, 0.068), respectively.

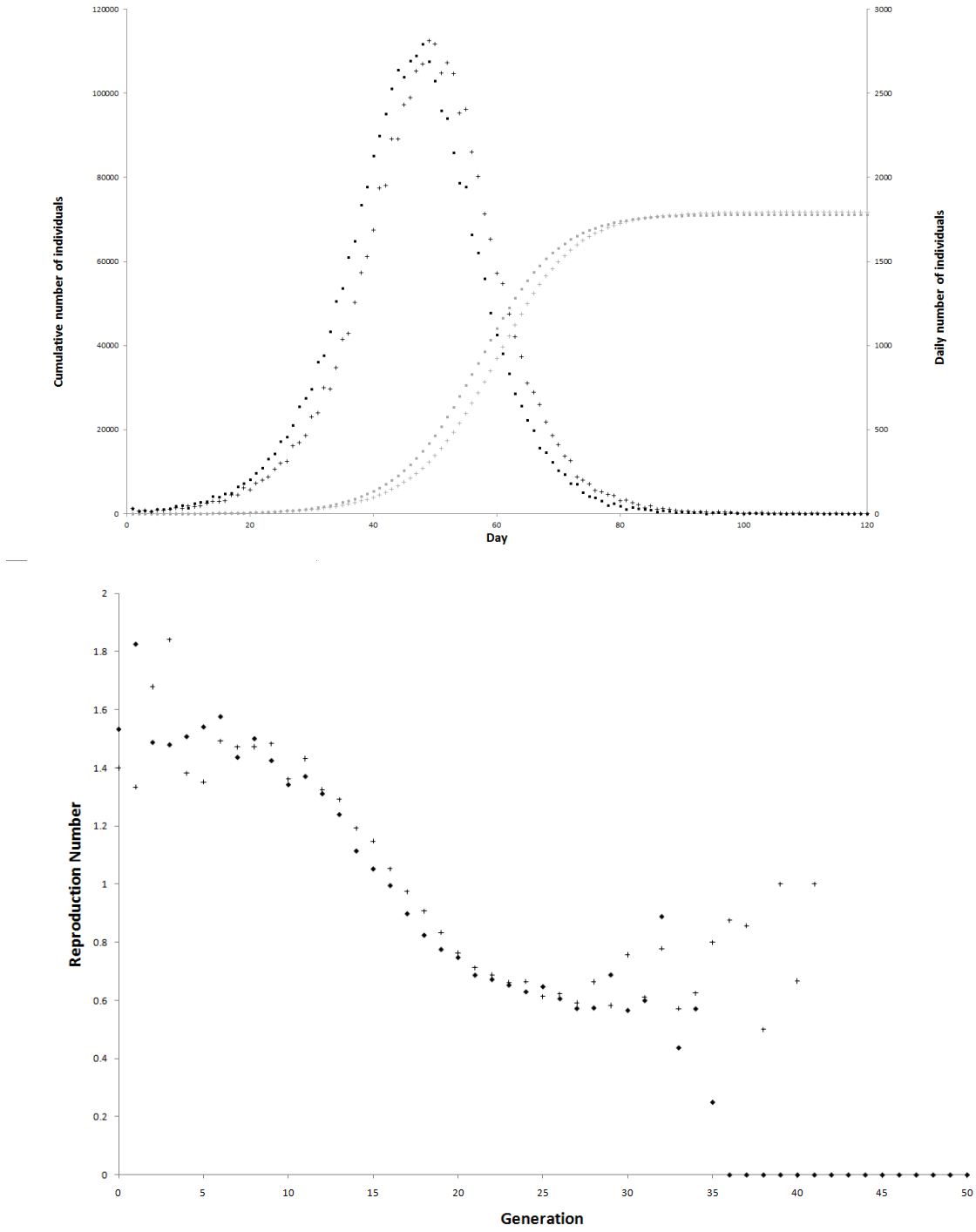


Figure 3.2: Disease spread behavior and expected reproduction number per generation for two outbreaks with $\widehat{R}_o = 1.8$ expressed in generation two.

Note that the IAR obtained with $E[\mathscr{R}_{PI}] = 1.8$ is lower than the IAR obtained while using infection probabilities (3.11) and (3.12). This reduction results from the assumption that interaction within households are higher than those in workplaces/schools and other places ($\kappa^w, \kappa^o < 1.0$ in equations 3.14, 3.13,3.16 and 3.15). The households are the contact groups with less members, and once all the members are infected and trying to infect each other, many opportunities of infection are wasted, reducing the IAR. Figure 3.3 shows the disease spread and reproduction number of both the pandemic and the seasonal outbreaks. Note in Figure 3.3(a) that the pandemic outbreak completely shadows the seasonal outbreak, which takes 350 days more in disappearing than the pandemic outbreak. This is the natural behavior of seasonal outbreaks since they remain infecting the population for extended time periods. It is also interesting to observe in Figure 3.3(a) the fluctuations of the $E[\mathscr{R}^k]$ over many generations until the outbreak disappear after 170 generations, which are attributed to the small sample sizes that are available for the calculation of the expected reproduction number.

With these results, we have demonstrated how the proposed epidemiological model reduces the time expense in calibrating the simulation of concurrent pandemic and seasonal influenza outbreaks. Therefore, the model can be useful for simulations supporting decision making in real-time. We have proposed a basic probability model with (3.11) and (3.12), which can be used to understand general disease spread patterns and test novel disease surveillance surveillance strategies like the one explored in Chapter 4. The extended probability model described by (3.14), (3.13), (3.16) and (3.15) can be useful to test pharmaceutical and non-pharmaceutical interventions, aiming at reducing the percentage of symptomatic infection or reducing contact with the closure of schools, workplaces and public places. Our model can also useful to understand the effect of cross-immunity in the final size of pandemic and seasonal outbreaks, a future research area that necessitates a better understanding of the internal interaction of both viruses in the human body.

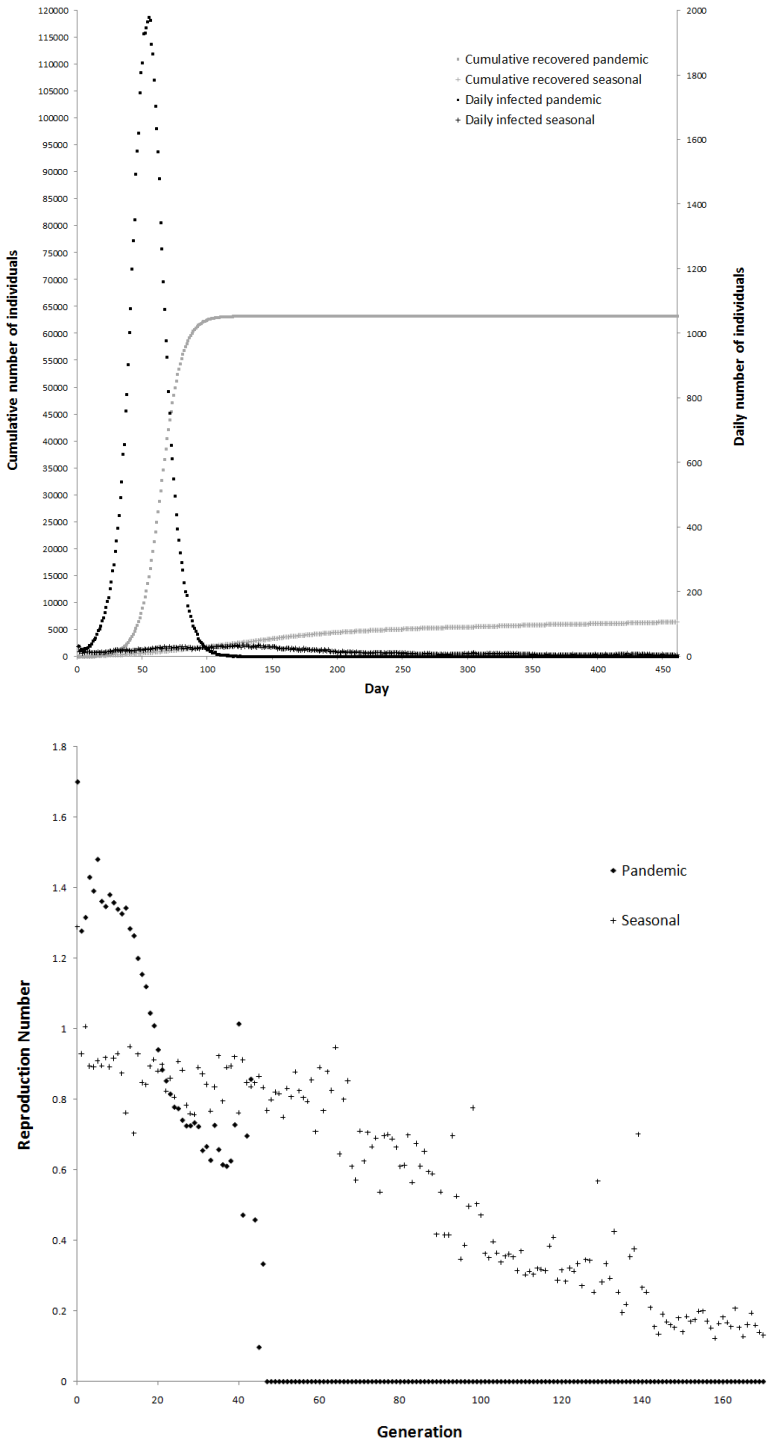


Figure 3.3: Disease spread behavior and expected reproduction number per generation for a pandemic outbreak with $\widehat{R}_o = 1.8$ and a seasonal outbreak with $\widehat{R}_o = 1.3$, both expressed in generation zero.

Chapter 4: A Testing Strategy of Influenza-Like-Illness Specimens for Real-Time Characterization of Pandemic Outbreaks

As discussed in Chapter 2, one of the main requirements for the implementation of a model-based, decision making process for real-time pandemic mitigation, is the availability of data that can be used to characterize population epidemiological parameters.

Currently, characterizing pandemic outbreaks in real-time constitutes a challenge given the status of the surveillance networks. The Florida Department of Health (FDOH) monitors influenza activity through multiple surveillance systems: 1) Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE), 2) data from the Florida Bureau of Laboratories (BOL), 3) county influenza activity levels, 4) the Florida Pneumonia and Influenza Mortality Surveillance System (FPIMSS), 5) Notifiable disease reports (Merlin), 6) Florida Outpatient Influenza-like Illness Surveillance Network (ILINet), 7) pediatric influenza-associated mortality, and 8) clusters of influenza-like illness (EpiCom) [68]. The reported information helps to determine the location, time and type of influenza viruses that are circulating, and also whether pandemic influenza activity is increasing or decreasing on a daily basis [69]. However, the reported information cannot be used to directly determine population estimates for the daily number of pandemic cases with symptoms onset, which is key to determine transmissibility.

In this Chapter, we address the question of how to estimate the number of pandemic cases with symptoms onset in real-time. We initially consulted members of the current surveillance

system in Hillsborough, to understand and evaluate the system. With this information, we developed a strategy to implement real-time testing and parameter estimation. The strategy proposes to infer the number of pandemic cases with symptoms onset from the daily cases confirmed by the BOL, and the daily number of symptomatic cases reported by the healthcare providers. We evaluate the strategy via simulation, and determine its conditions for implementation by means of a design of experiments, where the factors include the laboratory capacity and the reporting rates.

Section 4.1 discusses our findings from the field investigation. The strategy proposed is detailed in Section 4.2 and the details for the analysis of the strategy are discussed in Section 4.3.

4.1 Hillsborough County Surveillance System

4.1.1 Description

We investigated the features of the Hillsborough surveillance system during the 2009 pandemic. We initially looked for documentation in institutional websites, with descriptions of the surveillance programs in place. We searched into the websites of the World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), the Florida Department of Health (FDOH)and the Hillsborough County Health Department (HCHD). In addition, we interviewed several representatives from the Hillsborough surveillance system including members from HCHD, the Bureau of Laboratories (BOL) and Hillsborough medical providers. A list of our collaborators is provided in Appendix D.

With this information, we created a diagram showing the flow of an infected individual through the surveillance system. The complete flow diagram is shown in Appendix E and we describe the diagram as follows. During the 2009 pandemic outbreak, the surveillance system was passive, meaning that it only covered a symptomatic individual who decided to look for medical attention (Appendix E, page 1). If the individual goes to the Emergency Room (ER) and presents Influenza Like Illness (ILI) symptoms, she is reported in through the ESSENCE software. The ESSENCE software works almost in real-time, since chief complaints can be observed by the HCHD one day after they are received in an ER.

If the individual is assisted by a sentinel physician or nurse practitioner, she will be reported to CDC through the ILINet, and diagnosis will proceed according with the symptoms (Appendix E page 2). An individual with mild symptoms and with high risk of complications might be selected for further analysis by the Florida Bureau of Laboratories (BOL), as part of the 5 (or less) specimens per week that sentinels are recommended to submit to the lab. Once in the BOL lab, the specimen is tested through a Polymerase Chain Reaction (PCR) test. In case the PCR is positive, it is reported through the Merlin system at the state level. A death is always a possibility for any type of symptomatic patient (Appendix E page 4). If a death occurs, it is reported to the HCHD and the Florida Pneumonia and Influenza Mortality System (FPIMSS). Symptomatic patients with life threatening illnesses are always reported to the HCHD.

Individuals with severe symptoms are commonly hospitalized (Appendix E pages 2 and 3). Some individual's specimens are PCR tested (at the clinician's discretion). Deaths, pregnant hospitalized women and patients with life threatening illnesses are reported to HCHD.

If the physician or nurse practitioner is not a sentinel, the system works as in the sentinel case. The only difference is that non-sentinels do not receive any recommendation on the number of samples to send to BOL for PCR diagnosis.

4.2 Data Collection and Sampling Strategy

As described before, the BOL determines the strain of circulating viruses. Specimens are generally collected from the sentinel healthcare providers for routine surveillance. However, during the recent H1N1 pandemic, the BOL also received specimens from non-sentinel healthcare providers. Since there is no definite sampling policy to prioritize the specimen to test, the BOL tested all the specimens received, extending the desired testing timeline. These delays resulted in a slower pace of publication of confirmed cases, affecting the real-time assessment of the number of pandemic cases with symptoms onset.

To maximize the use of the lab capacity while achieving the goal of real-time characterization of symptomatic pandemic cases, we propose a testing strategy in which the lab is recommended to distribute its daily testing capacity considering the arrival behavior of ILI cases to healthcare providers everyday.

Figure 4.1 presents a schematic of the flow of specimens through the surveillance system on a daily basis. The solid arrows represent individuals with ILI symptoms onset on day t, who are assumed to seek healthcare either on day t, $t+1$ or $t+2$. The dotted arrows represent symptomatic individuals on day $t+1$ having $t+1$, $t+2$ and $t+3$ as possible days to seek healthcare. The dotted-and-striped arrows represent symptomatic individuals with $t+2$, $t+3$ and $t+4$ as the days to seek for healthcare.

Once ILI symptomatic cases seek for healthcare with probability r_h, their specimens can be submitted to the BOL laboratory with probability r_l. This is represented in the diagram with the box “Healthcare providers sending specimens to the lab".

Finally, the box “The BOL lab processing specimens" represents the final part of the flow where specimens are tested to confirm the presence of the pandemic virus.

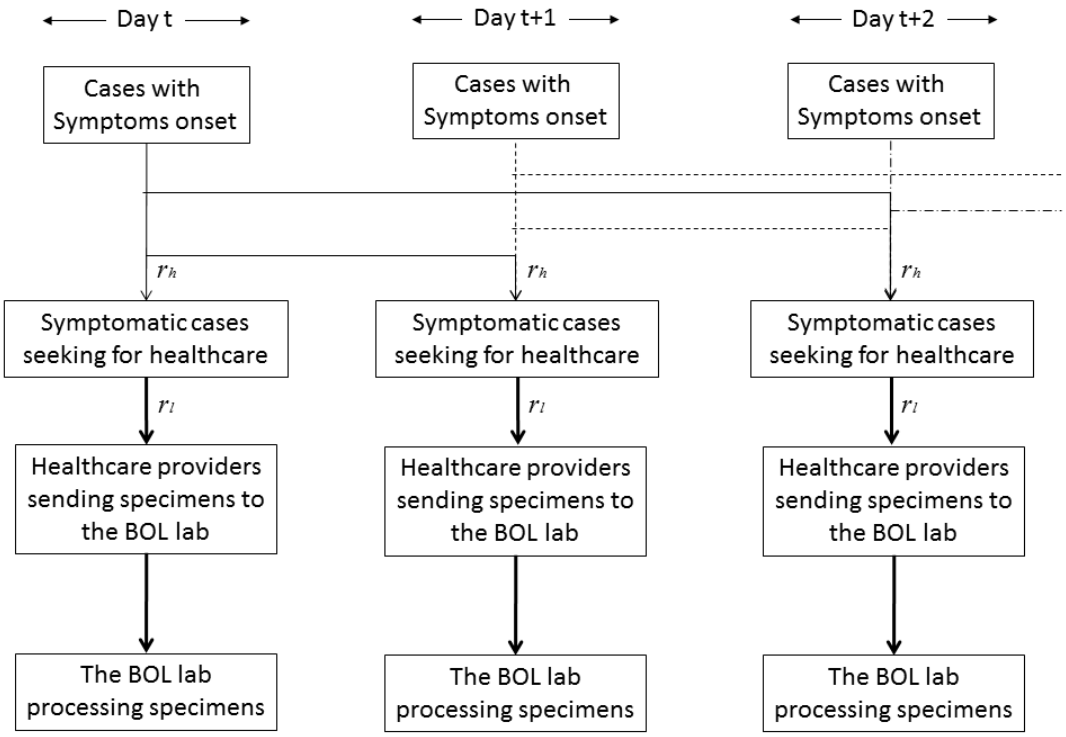


Figure 4.1: Flow of specimens through the surveillance system on a daily basis

Let X_t be the number of symptomatic cases seeking healthcare on day t and $X_{a,b}$ be the number of symptomatic cases seeking healthcare on day a with symptoms onset on day b. Then we assume that

$$X_t = X_{t,t} + X_{t,t-1} + X_{t,t-2}. \tag{4.1}$$

We also assume that the terms on the right hand side of (4.1) for any t are known to the laboratory. With this information, the BOL can determine the daily number of specimens to sample as follows. Let $Y_{a,b}$ denote the number of specimens to test on day a with symptoms onset on day b. Then we have that

$$Y_{t,k} = L\frac{X_{t,k}}{X_t}, \quad for\, k = t,\, t-1,\, t-2, \tag{4.2}$$

where L denotes the daily testing capacity of the laboratory. The quantities $Y_{t,t}$, $Y_{t,t-1}$ and $Y_{t,t-2}$ are sampled from $X_{t,t}$, $X_{t,t-1}$ and $X_{t,t-2}$, respectively. From $Y_{t,t}$, $Y_{t,t-1}$ and $Y_{t,t-2}$, the laboratory determines $B_{t,t}$, $B_{t,t-1}$, and $B_{t,t-2}$, which denote the number of confirmed pandemic cases on day t with symptoms onset on day t, $t-1$, and $t-2$, respectively.

4.2.1 Bayesian Inference Engine

Let $A_{a,b}$ denote the number of pandemic cases seeking healthcare on day a and with symptoms onset on day b. Let $\mathscr{A}_0$, $\mathscr{A}_1$ and $\mathscr{A}_2$ denote the random variables defined over outcomes $A_{t,t}$, $A_{t,t-1}$ and $A_{t,t-2}$, respectively, for all t. Defined similarly are $\mathscr{B}_0$, $\mathscr{B}_1$, and $\mathscr{B}_2$. We can write that

$$P[\mathscr{A}_k \mid \mathscr{B}_k] \propto P[\mathscr{B}_k \mid \mathscr{A}_k]P[\mathscr{A}_k], \quad for\, k = 0,\ 1,\ 2. \tag{4.3}$$

Note that $P[\mathscr{B}_k \mid \mathscr{A}_k]$ can be derived from a hypergeometric distribution with parameters $X_{t,t-k}$, $A_{t,t-k}$ and $Y_{t,t-k}$. For example, $P[\mathscr{B}_1 \mid \mathscr{A}_1]$ can be given as

$$P[\mathscr{B}_1 \mid \mathscr{A}_1] = \frac{\binom{A_{t,t-1}}{B_{t,t-1}}\binom{X_{t,t-1}-A_{t,t-1}}{Y_{t,t-1}-B_{t,t-1}}}{\binom{X_{t,t-1}}{Y_{t,t-1}}}$$

The prior for $\mathscr{A}_k$ is set to be a uniform distribution with parameters $B_{t,t-k}$, $X_{t,t-k}$.

4.2.2 Symptoms Onset Time Series and Instantaneous Reproduction Number

Let I_t be the set of pandemic cases with symptoms onset between time interval t and $t+1$. At every time t, it is possible to obtain the following estimations.

The first level estimation is

$$\bar{I}_t = A_{t,t}. \tag{4.4}$$

The second level estimation is

$$\bar{I}_{t-1} = A_{t-1,t-1} + A_{t,t-1}. \tag{4.5}$$

And the third level estimation is

$$\bar{I}_{t-2} = A_{t-2,t-2} + A_{t-1,t-2} + A_{t,t-2}. \tag{4.6}$$

The time series created by each of the estimation level can be converted into a time series for the instantaneous reproduction number $R(t)$ [70] by means of the expression

$$R(t) = \frac{I_t}{\sum_{j=0}^{\infty} w(j) I_{t-j}} \tag{4.7}$$

where $\sum_{j=0}^{\infty} w(j) = 1$. It is realistic to truncate the function $w(j)$ at a time m such that $w(j) = 0$ for all $j \geq m$. In practice, $R(t)$ represents the number of cases during time interval t and $t+1$ that were created by a case infected before and during time interval t and $t+1$.

4.3 Performance of the Sampling and Testing Strategy

We tested the performance of the strategy by implementing the scenario depicted in Figure 4.1, in the baseline simulation described in Chapter 3. A population of around 1000000 inhabitants was simulated with all the features of the Hillsborough county, described previously in Section 3.2.

Individuals were assumed to present symptoms after two or three days of infection with either pandemic or seasonal influenza [51]. It is assumed that all symptomatic individuals seek healthcare with probabilty r_h^1 before the pandemic is regionally declared, and r_h^2 after the outbreak is declared. It is also assumed that the healthcare providers submit a specimen to the BOL with probability r_l^1 and r_l^2.

We evaluated the strategy with $L = 1000$ specimens per day and under a perfect reporting scenario, where r_h^1, r_h^2, r_l^1, and r_l^2 are equal to 1. Figure 4.2 plot the time series for two instantaneous reproduction numbers: 1) the reproduction number for the daily cases with symptoms onset, denoted as R(t), and 3) the reproduction number for the first level estimation of the

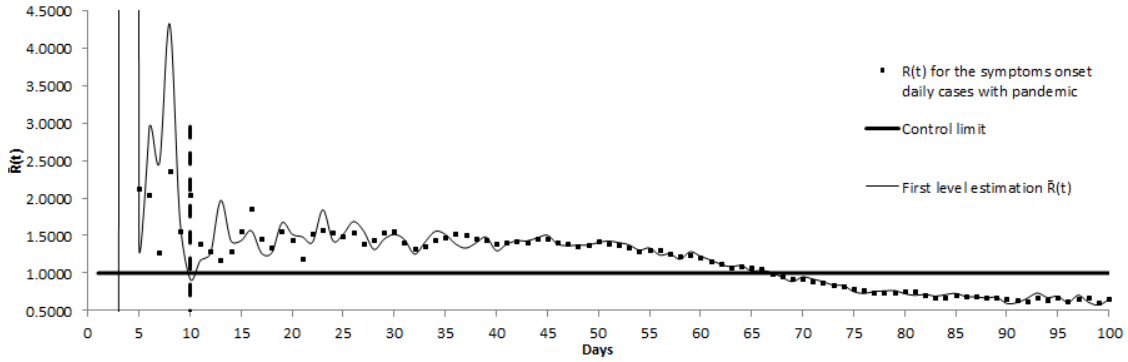
daily cases with symptoms onset, denoted as $\bar{R}(t)$. The time series should be observed after day ten (vertical striped line) since the instantaneous reproduction numbers need a burning period to account for enough values of $w(t)$, the disease generation interval probability mass function (Equation 3.2 provides the formula for $w(t)$).

The first level $\bar{R}(t)$ time series is meant to follow the R(t) for the symptoms onset series, but this is poorly achieved during the first 30 days of estimation. The second and third level estimations fall closer to R(t). This behavior can be measured in terms of the Sums of Squared errors $SSq = \sum_{t=k}^{l}(R(t) - \bar{R}(t))^2$ where, in this scenario, $k = 10$ and $l = 100$. The SSq values for the first, second and third level estimations are, 2.539195, 0.305079 and 0.000555, respectively.

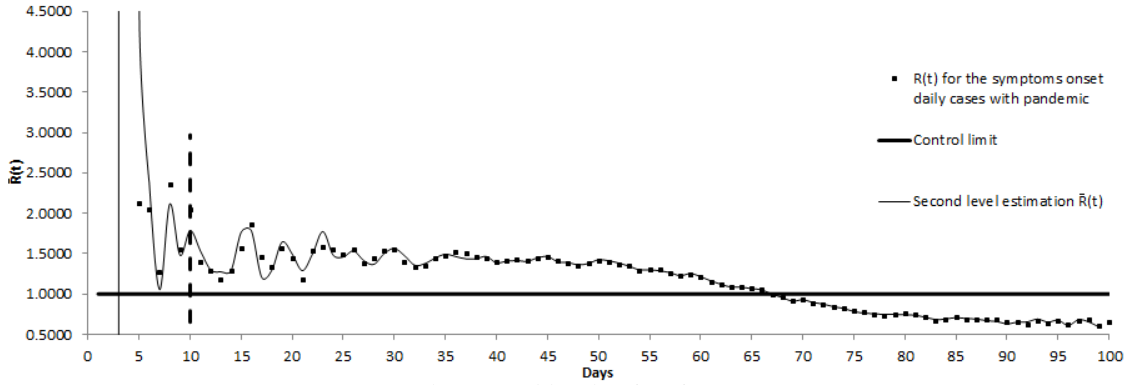
The black horizontal line indicates the control limit for R(t) [8]. The second and third level $\bar{R}(t)$ provide the same information about the time when the outbreak is control($R(t) \leq 0$), which occurs in day 66 . Note that information for day 66 will be known by days 67 and 68, when the second and third level $\bar{R}(t)$ can be estimated.

Assuming that the pandemic is declared 10 days after the first infected case, we set $r_h^1 = r_l^1 = 0.2$ and $r_h^1 = r_l^1 = 0.8$, to recreate the increase in reporting due to the fear for the pandemic. Figure 4.3 shows the effect of the sudden shift in the reporting probabilities. During the first periods following day 10, the discrepancies between the $R(t)$ and $\bar{R}(t)$ are huge for the first, second and third level estimations, indicating the negative effect of high variations in the reporting behavior.

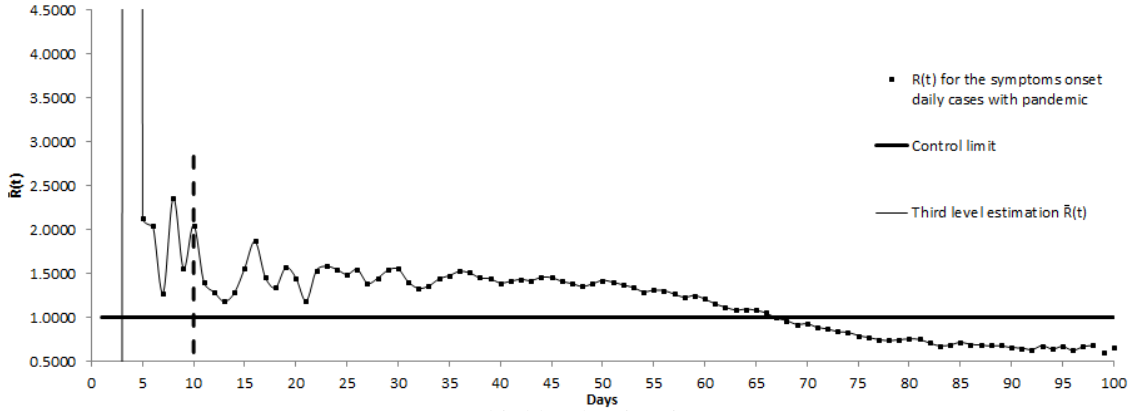
To eliminate the effect of the high variations in the reporting probabilities, we discarded $\bar{R}(10)$, $\bar{R}(11)$, and $\bar{R}(12)$ from the first level estimation. Note that these values correspond to $\bar{R}(11)$, $\bar{R}(12)$, and $\bar{R}(13)$ for the second level estimation, and $\bar{R}(12)$, $\bar{R}(13)$, and $\bar{R}(14)$ for the third level estimation. These value are chosen since they are the most heavily driven by daily counts for symptoms onset occurring before day 10.



(a) First level estimation

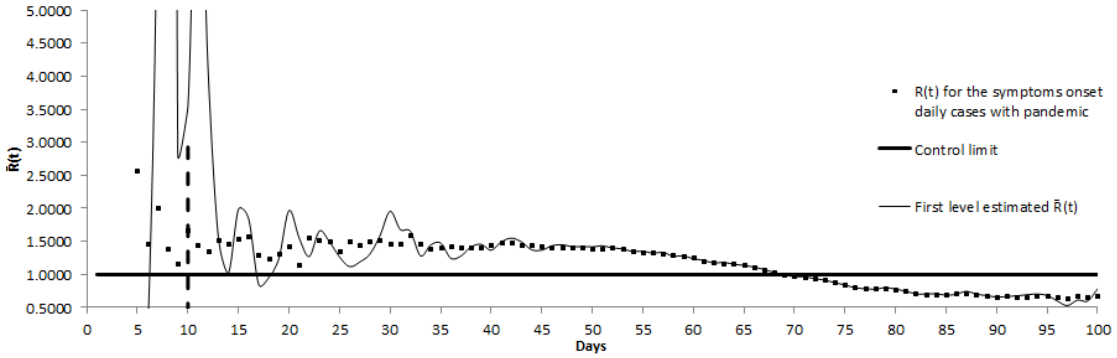


(b) Second level estimation

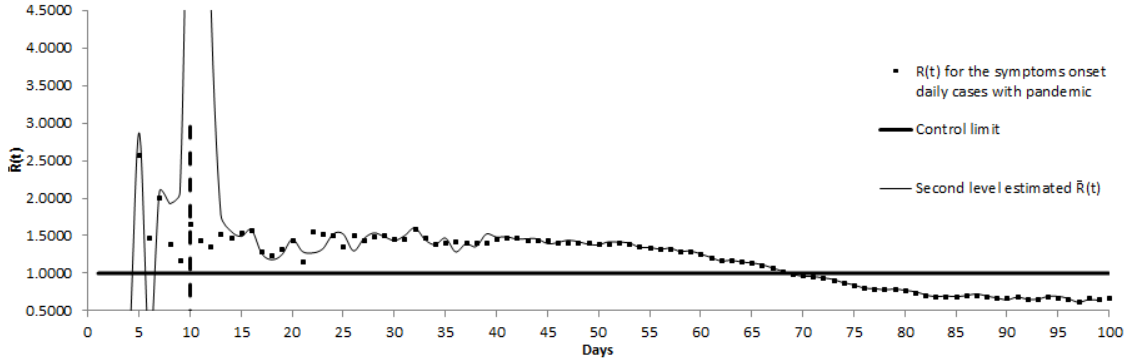


(c) Third level estimation

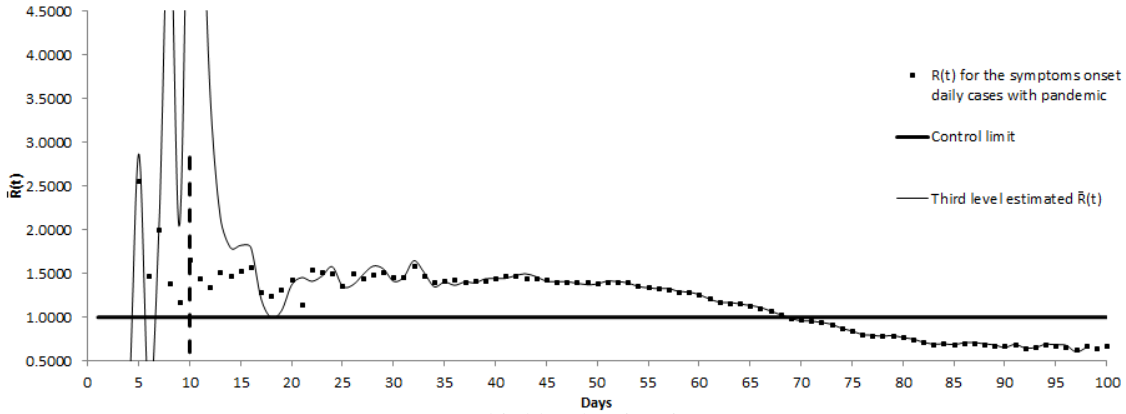
Figure 4.2: Results for the $R(t)$ assuming perfect reporting.



(a) First level estimation



(b) Second level estimation



(c) Third level estimation

Figure 4.3: Results for the $R(t)$ assuming the pandemic is declared 10 days after the beginning of the outbreak.

Once the effect of high variations is discarded, we analyzed the effect of the reporting rates and the laboratory capacity L in the SSq. We developed a 3^3 factorial experiment for the first, second and third estimation level. Table 4.1 presents the description of the experiment.

Table 4.1: Design of experiments for the sampling and testing strategy.

Factor	**Low level**	**Mid level**	**High level**
r_h^2	0.2	0.5	0.8
r_l^2	0.2	0.5	0.8
L	1000	5500	10000

The analysis of variance for the experiment (Appendix E) indicates that the capacity is not significant for any of the estimation levels. This confirms our results that the sampling and testing strategy effectively elevates the restriction of capacity in the laboratory. The contour plots in Figure 4.4 illustrate the effect of the reporting rates in the SSq for the first, second and third estimation level (Tables 4.4(a), 4.4(b) and 4.4(c), respectively). From the contour plots it is observed the proportional increase of the SSq as reporting probabilities increase. In addition, the second and third level are more robust to lower attack rates than the first estimation level.

With these results we have demonstrated that it is possible to achieve real-time characterization of the transmissibility of pandemic viruses via the three levels of estimation proposed. The first level estimation is poor, but it rapidly improves as more data is collected in the subsequent days and the second and third level estimations take place. Since these results are obtained under good reporting conditions, there is the need of research that determines actions to either increase the reporting probabilities, or eliminate their effect through the implementation of active surveillance (e.g., phone surveys to determine the number of ILI cases in the poputation).

The strategy assumes that the information about the number of symptomatic cases and their date of symptoms onset are known by the laboratories. We believe that this assumption is implementable since there are surveillance networks (e.g., ILINet and ESSENCE) tracking

the number of symptomatic cases in the population, and these networks that can be integrated to the laboratories for the sampling of specimens in real-time. A future research need is to investigate the requirements and associated costs of this integration.

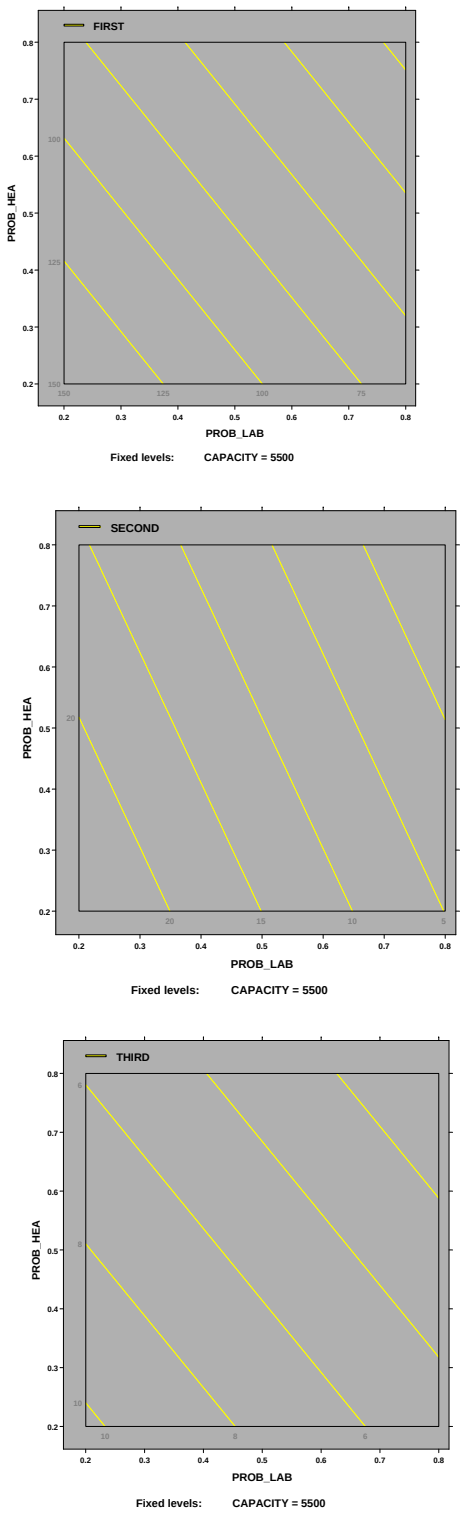


Figure 4.4: Contour plots for the effect of the reporting rates in the sum of squares.

Chapter 5: Conclusions and Future Work

Given the uncertain nature of pandemic viruses, public health officials face multiple decision making challenges whenever a pandemic emerges. Simulation models have provided valid support for these decisions, and model-driven recommendations have been used to create guidelines for implementation of strategies to mitigate pandemic outbreaks.

The two common modeling approaches are the differential equation (DE) and the agent-based (AB). Differential Equation (DE) models can be calibrated and replicated quickly, but they are only suitable for evaluation of mass-action mitigation strategies (e.g., mass antiviral prophilaxis for children). AB models provide a more comprehensive decision support since they possess the granularity and traceability features to recreate both mass-action and individual-based mitigation strategies (e.g., antiviral prophilaxis for the contacts of an infected case).

Ideally, modeling frameworks should be useful not only for preparedness but also for emergency management during the pandemic. However, there exists significant gaps between the needs of the strategic and tactical decision makers and the support that the current pandemic AB models provide. There is a current lack of means for access, retrieval and translation of epidemiological, demographic and social-behavioral data into model parameter values. Epidemiological data can be obtained through surveillance networks like ESSENCE, ILINet and Bureau of Laboratories, but their information is not integrated to provide real-time characterization of epidemiological parameters at the population level. Demographic data can

be obtained from databases such as the U.S. Census Bureau and the National Household Travel Survey, but their access, retrieval and translation is not automatic. Social-behavioral information is scarce and there is the need of research to determine data collection priorities and methods. Operability and usability of the models is poor since most of the models are academic, and they are not adapted for public health policymakers.

We focused on the challenges of epidemiological data availability and model operability. With regards to data availability, we proposed a data collection strategy that suggests the integration of the ILI case counts from the networks, such as ILINet or ESSENCE, with the specimen data from the Bureau of Laboratories. This integration would provide estimates for the daily number of symptoms onset, useful for assessing transmissibility. With regards to operability, we proposed an epidemiological model for calibrating simulation models of concurrent pandemic and seasonal influenza outbreaks. The epidemiological model automatically translates basic reproduction numbers and viral shedding profiles into the infection probabilities of the simulation model, avoiding the need for trial-and-error calibration.

Our future work consists of integrating the testing strategy and the modeling framework to create a model-based, decision support system for mitigation of pandemic outbreaks in real-time. Mitigation actions recommended by this system would produce an effect in the real environment, and this effect can be re-assessed by means of the testing strategy, creating a feedback loop for continuous observation and mitigation of the outbreak. This system would require careful study of the implementation gaps of pandemic models, and a strong partnership between modelers and public health decision makers.

List of References

[1] C. Mills, J. Robins, and M. Lipsitch, "Transmissibility of 1918 pandemic influenza," *Nature*, vol. 432, pp. 904–906, 2004.

[2] "Researchers Reconstruct 1918 Pandemic Influenza Virus; Effort Designed to Advance Preparedness." [Online]. Available: http://www.cdc.gov/media/pressrel/r051005.htm

[3] Committee on Modeling Community Containment for Pandemic Influenza, "Modeling community containment for pandemic influenza: A letter report," http://www.nap.edu/catalog/11800.html, 2006, last accessed on 12/09/2008.

[4] L. A. Rvachev and I. M. Longini, "A mathematical model for the global spread of influenza," *Math Biosci*, vol. 75, pp. 3–22, 1985.

[5] "History of Flu Pandemics," http://www.flu.gov/individualfamily/about/pandemic/history.html, 2011, last accessed on 07/05/2011.

[6] P. M. C. Jackson, E. Vynnycky, "Estimates of the Transmissibility of the 1968 (Hong Kong) Influenza Pandemic: Evidence of Increased Transmissibility Between Successive Waves," *American Journal of Epidemiology*, vol. 171, no. 4, pp. 465–478, 2010.

[7] Centers for Disease Control and Prevention (CDC), "Updated CDC Estimates of 2009 H1N1 Influenza Cases, Hospitalizations and Deaths in the United States, April 2009 - April 2010." [Online]. Available: http://www.cdc.gov/h1n1flu/estimates_2009_h1n1.htm

[8] C. Fraser and C. Donnelly, "Pandemic potential of a strain of influenza A (H1N1): Early Findings," *Science*, vol. 324, pp. 1557–1561, 2009.

[9] Centers for Disease Control and Prevention (CDC), "Estimating Seasonal Influenza-Associated Deaths in the United States: CDC Study Confirms Variability of Flu." [Online]. Available: http://www.cdc.gov/flu/about/disease/us_flu-related_deaths.htm

[10] G. Chowell, M. A. Miller, and C. Viboud, "Seasonal influenza in the United States, France, and Australia: Transmission and prospects for control," *J R Soc interface*, vol. 4, pp. 155–156, 2008.

[11] T. Germann, K. Kadau, I. M. Longini, and C. Macken, "Mitigation strategies for pandemic influenza in the United States," *PNAS*, vol. 103, pp. 5935–5940, 2006.

[12] N. Ferguson, D. A. Cummings, C. Fraser, J. Cajka, P. C. C., and D. S. Burke, "Strategies for mitigating an influenza pandemic," *Nature*, vol. 442, no. 27, pp. 448–452, 2006.

[13] R. Glass, W. Beyeler, and H. Min, "Targeted social distancing design for pandemic influenza," *Emerging Infectious Diseases*, vol. 12, no. 11, pp. 1671–1681, 2006.

[14] L. M. Wein and M. P. Atkinson, "Assessing infection control measures for pandemic influenza," *Risk Analysis*, vol. 29, no. 7, pp. 949–962, 2009.

[15] T. Das, and A. Savachkin, "A large scale simulation model for assessment of societal risk and development of dynamic mitigation strategies," *IIE Transactions*, vol. 40, no. 9, pp. 893–905, 2008.

[16] Yale New Haven Center for Emergency Preparedness and Disaster Response, and US Northern Command, "Study to Determine the Requirements for an Operational Epidemiological Modeling Process in Support of Decision Making During Disaster Medical and Public Health Response Operations," Draft, 2011, updated on 24/05/2011.

[17] L. A. Rosenfeld, C. E. Fox, D. Kerr, E. Marziale, C. A, K. Lota, J. Stewart, and M. Thompson, "Use of computer modeling for emergency preparedness functions by local and state health official: A needs assessment," *J Public Health Management Practice*, vol. 15, no. 2, pp. 96–104, 2009.

[18] CDC's new preparedness modeling initiative: Beyond (and before) crisis response. [Online]. Available: http://www2.isye.gatech.edu/humlog09/program/longtermCDCModelingPresentationHupert20090219.pdf

[19] Centers for Disease Control and Prevention (CDC), "Interim pre-pandemic planning guidance: Community strategy for pandemic influenza mitigation in the United States," http://www.pandemicflu.gov/plan/community/community_mitigation.pdf, 2007, last accessed on 04/01/2009.

[20] M. Brandeau, J. H. McCoy, N. Hupert, J. E. Holty, and B. D. M, "Recommendations for modeling disaster responses in public health and medicine: A position paper of the society for medical decision making," *Medical Decision Making*, vol. 29, no. 4, pp. 438–460, 2009.

[21] United States Census Bureau, "Census 2000," http://www.census.gov, 2000.

[22] Bureau of transportation statistics, "2001 National household travel survey (NTHS)," http://www.bts.gov/programs/national_household_travel_survey/, 2002, last accessed on 12/09/2008.

[23] H. Rahmandad and J. Sterman, "Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models," *Management Science*, vol. 54, no. 5, pp. 998–1014, 2008.

[24] V. J. Lee, D. C. Lye, and W.-S. A, "Combination strategies for pandemic influenza response: a systematic review of mathematical modeling studies," *BMC Medicine*, vol. 7, no. 76, 2009.

[25] N. M. Ferguson, D. Cummings, S. Cauchemez, C. Fraser, S. Riley, M. Aronrag, S. Lamsirithaworn, and D. Burke, "Strategies for containing an emerging influenza pandemic in Southeast Asia," *Nature*, vol. 437, pp. 209–214, 2005.

[26] B. I. Sander, A. Nizam, L. P. Garrison, M. J. Postma, M. E. Halloran, and I. M. Longini, "Economic evaluation of influenza pandemic mitigation strategies in the united states using a stochastic microsimulation transmission model," *Value in Health*, vol. 12, no. 2, pp. 226–233, 2009.

[27] D. Perlroth, R. J. Glass, V. J. Davey, D. Cannon, A. M. Garber, and D. K. Owens, "Health outcomes and costs of community mitigation strategies for an influenza pandemic in the United States," *Clinical Infectious Diseases*, vol. 50, pp. 000–000, 2010.

[28] A. Uribe, A. Savachkin, A. Santana, T. K. Das, and D. Prieto, "A predictive decision aid methodology for dynamic mitigation of influenza pandemics," *Special Issue on Optimization in Disaster Relief, OR Spectrum (6 May 2011)*, pp. 1–36, 2010.

[29] M. E. Halloran, N. M. Ferguson, S. Eubank, and I. Longini, "Modeling targeted layered containment of an influenza pandemic in the United States," *PNAS*, vol. 105, no. 12, pp. 4639–4644, 2008.

[30] J. T. Wu, S. Riley, C. Fraser, and G. Leung, "Reducing the impact of the next influenza pandemic using household-based public health interventions," *PLoS Med*, vol. 3, no. 9, pp. 1532–1540, 2006.

[31] J. Arino, F. Brauer, P. den Driessche, J. Watmough, and J. Wu, "Simple models for containment of a pandemic," *J. R. Soc. Interface*, vol. 3, pp. 453–457, 2006.

[32] J. Arino, F. Brauer, P. Van den Driessche, J. Watmough, and J. Wu, "A model for influenza with vaccination and antiviral treatment," *J. R. Soc. Interface*, vol. 253, pp. 118–130, 2008.

[33] I. M. Longini, M. E. Halloran, A. Nizam, and Y. Yang, "Containing pandemic influenza with antiviral agents," *Am J Epidemiol*, vol. 159, no. 7, pp. 623–633, 2004.

[34] I. M. Longini, A. Nizam, X. Shufu, K. Ungchusak, W. Hanshaoworakul, D. Cummings, and M. E. Halloran, "Containing pandemic influenza at the source," *Science*, vol. 309, pp. 1083–1087, 2005.

[35] D. L. Chao, M. E. Halloran, V. J. Obenchain, and I. M. Longini, "FluTE, a publicly available stochastic influenza epidemic simulation model," *Plos Computational Biology*, vol. 6, no. 1, pp. 1–8, 2010.

[36] M. Z. Gojovic, B. Sander, D. Fisman, K. M. D, and B. C. T, "Modelling mitigation strategies for pandemic (H1N1) 2009," *CMAJ*, vol. 181, no. 10, pp. 673–680, 2009.

[37] V. J. Davey, R. J. Glass, H. J. Min, W. E. Beyeler, and M. L. Glass, "Effective, Robust Design of Community Mitigation for Pandemic Influenza: A Systematic Examination of Proposed US Guidance," *Plos one*, vol. 3, no. 7, pp. 1–14, 2008.

[38] V. J. Davey and R. J. Glass, "Rescinding community mitigation strategies in an influenza pandemic," *Emerging Infectious Diseases*, vol. 14, no. 3, pp. 365–372, 2008.

[39] M. Nuno, G. Chowell, and A. B. Gumel, "Assessing the role of basic control measures, antivirals and vaccine in curtailing pandemic influenza: scenarios for the US,UK and the Netherlands," *J. R. Soc. Interface*, vol. 4, pp. 505–521, 2007.

[40] A. B. Gumel, M. Nuno, and G. Chowell, "Mathematical assessment of Canada's pandemic influenza preparedness plan," *Can J Infect Dis Med Microbiol*, vol. 19, no. 2, pp. 185–192, 2008.

[41] M. G. Roberts, M. Baker, L. C. Jennings, G. Sertsou, and N. Wilson, "A model for the spread and control of pandemic influenza in an isolated geographical region," *J. R. Soc. Interface*, vol. 4, pp. 325–330, 2007.

[42] M. Eichner, M. Schwehm, H. Duerr, and S. O. Brockmann, "The influenza pandemic preparedness planning tool InfluSim," *BMC Infectious Diseases*, vol. 7, no. 17, 2007.

[43] E. O. Jordan, *Epidemic influenza; A survey*. Chicago: American Medical Association, 1927.

[44] L. R. Elveback, J. P. Fox, E. Ackerman, A. Langworthy, M. Boyd, and L. Gatewood, "An influenza simulation model for immunization studies," *Am J Epidemiol*, vol. 103, no. 2, pp. 152–155, 1976.

[45] D. M. Bell, "Non-pharmaceutical interventions for pandemic influenza, national and community measures," *Emerging infectious diseases*, vol. 12, no. 1, p. 88, 2006.

[46] F. G. Hayden, R. Scott Fritz, M. Lobo, G. Alvord, W. Strober, and S. S. E, "Local and systemic cytokine responses during experimental human influenza a virus infection," *The journal of clinical investigation*, vol. 101, no. 3, pp. 643–649, 1998.

[47] S. Cauchemez, P. Boelle, G. Thomas, and A. Valleron, "Estimating in real time the efficacy of measures to control emerging communicable diseases," *Am J Epidemiol*, vol. 164, no. 6, pp. 591–597, 2006.

[48] M. Lipsitch, S. Riley, S. Cauchemez, A. Ghani, and N. Ferguson, "Managing and reducing uncertainty in an emerging influenza pandemic," *New. Eng. J. Med*, vol. 361, no. 2, pp. 112–115, 2009.

[49] F. Carrat, C. Sahler, S. Rogez, M. Leruez-Ville, F. Freymuth, C. Le Gales, M. Bungener, B. Housset, M. Nicolas, and C. Rouzioux, "Estimates from a national prospective survey of household contacts in france," *Arch Intern Med*, vol. 162, pp. 1842–1848, 2002.

[50] S. Cauchemez, C. Donelly, C. Reed, A. Ghani, C. Fraser, C. Kent, L. Finelly, and N. Ferguson, "Household transmission of 2009 pandemic influenza a (H1N1) virus in the United States," *N Engl J Med*, vol. 361, pp. 2619–2627, 2009.

[51] F. Carrat, E. Vergu, N. Ferguson, M. Lemaitre, S. Cauchemez, and S. Leach, "Timelines of infection and disease in human influenza: A review of volunteer challenge studies," *Am J Epidemiol*, vol. 167, pp. 775–785, 2008.

[52] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mickolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and M. J. Edmunds, "Social contacts and mixing patterns relevant to the spread of infectious diseases," *PLoS Med*, vol. 5, no. 3, pp. 381–391, 2008.

[53] G. Kok, R. Jonkers, R. Gelissen, R. Meertens, H. Schaalma, and O. de Zwart, "Behavioural intentions in response to an influenza pandemic," *BMC Public Health*, vol. 10, no. 174, 2010.

[54] TeraGrid. [Online]. Available: https://www.teragrid.org/web/about/

[55] "MIDAS Software Survey Results." [Online]. Available: https://mission.midas.psc.edu/index.php?option=com_content&view=article&id=84&Itemid=114

[56] J. Wallinga and M. Lipsitch, "How generation intervals shape the relationship between growth rates and reproduction numbers," *Proceedings of the Royal Society B*, vol. 274, pp. 599–604, 2007.

[57] F. Carrat, J. Luong, H. Lao, A. Salle, C. Lajaunie, and H. Wackernagel, "A 'small-world-like' model for comparing interventions aimed at preventing and controlling influenza pandemics," *BMC Medicine*, vol. 4, no. 26, 2006.

[58] G. J. M. N. Halder, J. Kelso, "Analysis of the effectiveness of interventions used during the 2009 a/h1n1 influenza pandemic," *BMC Public Health*, vol. 10, no. 168, 2010.

[59] A. H. V. Sypsal, "School closure is currently the main strategy to mitigate influeza a(h1n1): A modeling study," *Eurosurveillance*, vol. 14, no. 24, 2009.

[60] O. Diekmann and J. Heesterbeek, *Mathematical epidemiology of infectious diseases: Model building, analysis and interpretation*. Wiley, 2000.

[61] H. Guo, F. Santiago, K. Lambert, T. Takimoto, and D. Topham, "T cell-mediated protection against lethal 2009 pandemic H1N1 influenza virus infection in a mouse model," *Journal of Virology*, vol. 85, no. 1, pp. 448–455, 2011.

[62] W. Liu, Z. Li, F. Tang, M. Wei, Y. Tong, L. Zhang, Z. Xin, M. Ma, X. Zhang, L. Liu, L. Zhan, C. He, H. Yang, C. Boucher, J. Hendrik, and W. Cao, "Mixed infections of pandemic H1N1 and seasonal H3N2 viruses in one outbreak," *Clinical Infectious Diseases*, vol. 50, no. 10, pp. 1359–1365, 2010.

[63] A. Savachkin, D. Prieto, A. Uribe, and T. Das, "Developing dynamic predictive strategies for mitigation of cross-regional pandemic outbreaks," *In review with Manufacturing and Service Operations Management*, 2011.

[64] U.S. Census Bureau, "2001 American Community Survey," http://factfinder.census.gov/home/saff, last accessed on 04/27/2009.

[65] ——, "2001 American Community Survey," http://www.census.gov/prod/2001pubs/statab/sec01.pdf, last accessed on 03/27/2009.

[66] ——, "2002 Economic Census," http://factfinder.census.gov/home/saff.

[67] J. T. Wu, S. Riley, C. Fraser, and G. Leung, "Spatial considerations for the allocation of pre-pandemic influenza vaccination in the United States," *Proc. R. Soc. Bio. Sci.*, vol. 274, no. 1627, pp. 2811–2817, 2006.

[68] Florida Department of Health, "Florida Influenza Weekly Surveillance Reports - 2008 / 2009," http://www.doh.state.fl.us/disease_ctrl/epi/htopics/flu/2009/index.html, 2009, last accessed on 10/22/2009.

[69] Centers for Disease Control and Prevention (CDC), "Overview of Influenza Surveillance in the United States," http://www.cdc.gov/flu/weekly/pdf/overview.pdf, 2009, last accessed on 10/04/2009.

[70] C. Fraser, "Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic," *PLoS ONE*, vol. 2, no. 8, p. e758, 2007.

Appendices

Appendix A Epidemiological Parameters in Models for Pandemic Influenza Preparedness

Table A1: Epidemiological parameters explored in the models

Epidemiological model parameter (# models using the parameter)	Sources used by the models to obtain the parameter value Author/year (type of study, e.g., research paper, survey paper)	Data sources supporting the parameter value determination	Type of interface for data access & retrieval	Techniques to translate raw data into model parameter values
Basic reproduction number, *R* (10)	Mills et al. 2004 (research paper) Ferguson et al. 2005 (research paper) Ferguson et al. 2006 (research paper) Chowell et al. 2006 (research paper)	Excess weekly mortality in UK (1918-1919). Weekly mortality attributed to influenza in England and Wales (1957-1958). Daily Case notification data in San Francisco, California (1918 – 1919). Statistical estimates from England and Wales (1958-1973) . Influenza case incidence data for the first wave of pandemic influenza A starting in July 1968 in Hong Kong.	Manual	Examples: Fitting a SEIR (Susceptible Exposed Infectious Recovered) differential equation model. Fitting an equation for cumulative exponential growth; Bayesian likelihood analysis

Appendix A Continued

Table A1: Continued

Epidemiological model parameter (# models using the parameter)	Sources used by the models to obtain the parameter value Author/year (type of study, e.g., research paper, survey paper)	Data sources supporting the parameter value determination	Type of interface for data access & retrieval	Techniques to translate raw data into model parameter values
Illness attack rates, IAR (8)	Jordan et al. 1927 (survey paper), Elveback et al. 1976 (research paper) Longini et al. 1988 (research paper) Fraser et al. 2009 (research paper)	Survey on epidemic influenza, Chicago, 1927 Attack rates and estimates from Long-term data collection from a random sample of households, Tecumseh study of respiratory Illness, 1965-1971, 1976-1981. Confirmed and suspected deaths in Mexico, time series of the Mexican reported case numbers, accounting for early under-reporting	Manual	Examples: Arithmetic conversion (ratio of the number of infected over total susceptible population)

Appendix A Continued

Table A1: Continued

Epidemiological model parameter (# models using the parameter)	**Sources used by the models to obtain the parameter value Author/year (type of study, e.g., research paper, survey paper)**	**Data sources supporting the parameter value determination**	**Type of interface for data access & retrieval**	**Techniques to translate raw data into model parameter values**
Disease Natural History: Continuous time scale (e.g., Infectiousness profile and its parameter, the disease generation time), (8)	Carrat et al. 2002 (research paper) Cauchemez et al. 2004 (research paper) Murphy et al. 1980 (research paper) Baccam et al. 2006 (research paper)	Data from a follow up study of influenza infections in households, Epigrippe study, France, 2000 Viral titers per day from experimentally infected patients, 1980, 2006	Manual	Examples: Fitting a likelihood function to find the joint posterior distribution of augmented data and parameters; Fitting a distribution directly from the viral data

Appendix A Continued

Table A1: Continued

Epidemiological model parameter (# models using the parameter)	Sources used by the models to obtain the parameter value Author/year (type of study, e.g., research paper, survey paper)	Data sources supporting the parameter value determination	Type of interface for data access & retrieval	Techniques to translate raw data into model parameter values
Disease Natural History: Phase-partitioned time scale (e.g., non-infectious and infectious periods and relative infectivity of the disease states), (2)	Elveback et al. 1976 (research paper) Longini et al. 1988 (research paper) Ferguson et al. 2005 (research paper) Fraser et al. 2009 (research paper) Bell et al. 2006 (research paper) Hayden et al. 1998 (research paper)	Attack rates and estimates from Long-term data collection from a random sample of households, Tecumseh study of respiratory Illness, 1966-1971, 1976-1981. Non-infectious period: Human influenza data on multiple exposure event occurring on an aeroplane, 1979 Data from a follow up study of influenza infections in households, Epigrippe study, France, 2000 Time series of the Mexican reported case numbers, 2009 Experimental Laboratory study - Viral shedding profile of 20 volunteers, 1979 (From Bell et al. 2006) Experimental Laboratory study - Viral shedding profile of 20 volunteers (Hayden 1998)	Manual	Examples: Distribution fitting from the frequency data; Fitting a likelihood function to find the posterior distribution of parameters

Appendix A Continued

Table A1: Continued

Epidemiological model parameter (# models using the parameter)	Sources used by the models to obtain the parameter value Author/year (type of study, e.g., research paper, survey paper)	Data sources supporting the parameter value determination	Type of interface for data access & retrieval	Techniques to translate raw data into model parameter values
Fraction of infectious population that become symptomatic/ asymptomatic, (10)	Carrat et al. 2002 (research paper) Cauchemez et al. 2004 (research paper) Fraser et al. 2009 (research paper) Gani et al. 2005 (research paper)	Data from a follow up study of influenza infections in households, Epigrippe study, France, 2000 Localized outbreak data in La Gloria, Mexico, May 30 2009 Five year study of influenza in families, London, 1981	Manual	Examples: Fitting an age stratified mathematical model to outbreak data

Appendix B Demographic Parameters in Models for Influenza Pandemic Preparedness

Table B1: Demographic parameters explored in the models

Demographic model parameter (# models using the parameter)	**Data sources**	**Nature of data sources: database (D), survey paper (SP), official document (OD)**	**Type of interface for data access & retrieval**	**Technique to translate raw data into model parameter values**
Population size or density (9)	Landscan dataset (2003) US Census data and number of USA census tracts (2000) Census of Canada, Ottawa (2006)	D	Manual	Direct translation
Household size distribution (6)	US Census data (2000) Census of Canada, Ottawa (2006) Hong Kong survey data (2005)	D, SP	Manual	Direct translation; distribution fitting
Peer-group size distribution (6)	Establishment size data (Categorical distribution of the number of workers versus size) (2003) Census tract to tract worker flow data (2000) Census of Canada, Ottawa (2006) Enrollment by grade, London, Ontario (2008)	D	Manual	Data converted into a power law distribution with maximum likelihood methods; Direct translation

Appendix B Continued

Table B1: Continued

Demographic model parameter (# models using the parameter)	Data sources	Nature of data sources: database (D), survey paper (SP), official document (OD)	Type of interface for data access & retrieval	Technique to translate raw data into model parameter values
Age distribution (5)	US Census data (2000) [1] US school geo-referenced database (2004) [7] German National Pandemic Preparedness Plan (2005) [8]	D, OD	Manual	Distribution fitting; direct translation
Distribution of home to work commuting distances (Number of individuals traveling a certain distance) (2)	Census track of work by census track of residence (STP64) (2004) [9] Census tract to tract worker flow data (2000) [1]	D	Manual	Distribution fitting

Appendix B Continued

Table B1: Continued

Demographic model parameter (# models using the parameter)	Data sources	Nature of data sources: database (D), survey paper (SP), official document (OD)	Type of interface for data access & retrieval	Technique to translate raw data into model parameter values
Long distance traveling parameters (e.g., probability of flight from an origin to a destination, distribution of nights spent away, number of trips per person, average trip duration) (3)	Origin and destination of each flight included in an itinerary (Origin-destination survey of airline passenger dataset, BTS) (2005) Nights spend away for long distance air and car travel, National Household Travel Survey (NHTS) (1991) Person trips along the USA whole population. Long distance travel data (NHTS) (1995) Average trip duration. Long distance travel data (NHTS), (1995)	D	Manual	Distribution fitting; direct translation

Appendix C Accessibility and Scalability Features Investigated in the Models

Table C1: Accessibility and scalability features investigated in the models

Model name	Supporting papers	Is the model software available to general public? (open source or closed source code)	Presence of end user support (user manuals, e-mail/phone technical support)	Information on the number of replicates per implemen-tation scenario	Information on the running time per replicate (population size)	Information on the ways to manage the computational load for implementing large-scale scenarios (e.g., the use of distributed and parallel computing)	Use of replicate minimization techniques?
Imperial-Pitt	Ferguson, 2005	No	N/A	At most 1000	None stated	None	Stop if 50 replicates resulted in a failed mitigation policy.
Imperial-Pitt	Ferguson, 2006	No	N/A	At least 5	8 CPUs - 1-2 hour, 16 CPUs - 2-5 hour (for U.S. testbed)	Implemented on a distributed environment with either 8 or 16 servers.	Yes, until the average values of the variables with higher dispersion were "reasonably" determined.

Appendix C Continued

Table C1: Continued

Model name	Supporting papers	Is the model software available to general public? (open source or closed source code)	Presence of end user support (user manuals, e-mail/phone technical support)	Information on the number of replicates per implemen-tation scenario	Information on the running time per replicate (population size)	Information on the ways to manage the computational load for implementing large-scale scenarios (e.g., the use of distributed and parallel computing)	Use of replicate minimization techniques?
UW_LANL	Longini 2005	No	N/A	At least 1000	None stated	N/A	None stated
UW_LANL	Germann 2006	No	N/A	200	8-12 hr. (for U.S. testbed)	Yes (no details reported)	None stated
UW_LANL	Sander 2009	No	N/A	100	None stated	N/A	None stated
UW_LANL	Chao 2010	Yes (open source)	Instructions on the model inputs and outputs	None	32 CPUs - 6 hours (for U.S. testbed)	A parallelized version assigns population to different processors and open MPI to update the status of the variables on different CPUs	None stated

Appendix C Continued

Table C1: Continued

Model name	Supporting papers	Is the model software available to general public? (open source or closed source code)	Presence of end user support (user manuals, e-mail/phone technical support)	Information on the number of replicates per implemen-tation scenario	Information on the running time per replicate (population size)	Information on the ways to manage the computational load for implementing large-scale scenarios (e.g., the use of distributed and parallel computing)	Use of replicate minimization techniques?
LOKI-INFECT	Glass 2006	No	N/A	6 scenarios w/ 100 replicates each; 1 scenario w/ 1,000 replicates	None stated	N/A	None stated
LOKI-INFECT	Davey 2008a	No	N/A	100	None stated	N/A	None stated
LOKI-INFECT	Davey 2008b	No	N/A	100	None stated	None	None stated
LOKI-INFECT	Pelroth 2010	No	N/A	100	None stated	N/A	None stated

Appendix C Continued

Table C1: Continued

Model name	Supporting papers	Is the model software available to general public? (open source or closed source code)	Presence of end user support (user manuals, email/phone technical support)	Information on the number of replicates per implemen-tation scenario	Information on the running time per replicate (population size)	Information on the ways to manage the computational load for implementing large-scale scenarios (e.g., the use of distributed and parallel computing)	Use of replicate minimization techniques?
Gojovic	Gojovic, 2009	No	N/A	100	None stated	N/A	None stated
Nuno	Nuno, 2007	No	N/A	100	None stated	None	None stated
Gumel	Gumel, 2008	No	N/A	100	None stated	None	None stated
InfluSim	Eichner, 2007	Yes (closed source)	Technical paper	Not reported	None stated	None	None stated

Appendix C Continued

Table C1: Continued

Model name	**Supporting papers**	**Is the model software available to general public? (open source or closed source code)**	**Presence of end user support (user manuals, e-mail/phone technical support)**	**Information on the number of replicates per implemen-tation scenario**	**Information on the running time per replicate (population size)**	**Information on the ways to manage the computational load for implementing large-scale scenarios (e.g., the use of distributed and parallel computing)**	**Use of replicate minimization techniques?**
Imperial-Pitt UW_LANL	Halloran, 2008	No	N/A	UW_LANL - 5, Imperial Pitt - 10	None stated	None	See Imperial Pitt
USF	Das08, Uribe10	Yes (closed source)	Manual, e-mail	15	1 CPU - 20 mins (for 1 million people)	None	Yes, until confidence intervals of selected parameters reached prespecified width

Appendix D List of Collaborators

- Tapas Das, Professor, Department of Industrial and Management Systems Engineering, University of South Florida.

- Alex Savachkin, Assistant Professor, Department of Industrial and Management Systems Engineering, University of South Florida.

- Ricardo Izurieta, Assistant Professor, College of Public Health, Global Health, Communicable Diseases Track, University of South Florida.

- Karlette Peck, Epidemiologist, St. Lucie County Health Department.

- Warren McDougle, Epidemiology Program Manager, Hillsborough County Health Department.

- David Atrubin, Epidemiologist, Hillsborough County Health Department.

- Lillian Stark, Virology Administrator, Florida Department of Health, Bureau of Laboratories.

- Jackie Whitaker, Infection Control Director/NPSG Administrator, University Community Hospital, Tampa.

- Egilda Terenzi, Director, Student Health Services, University of South Florida.

Appendix E Current Surveillance System at the State Level

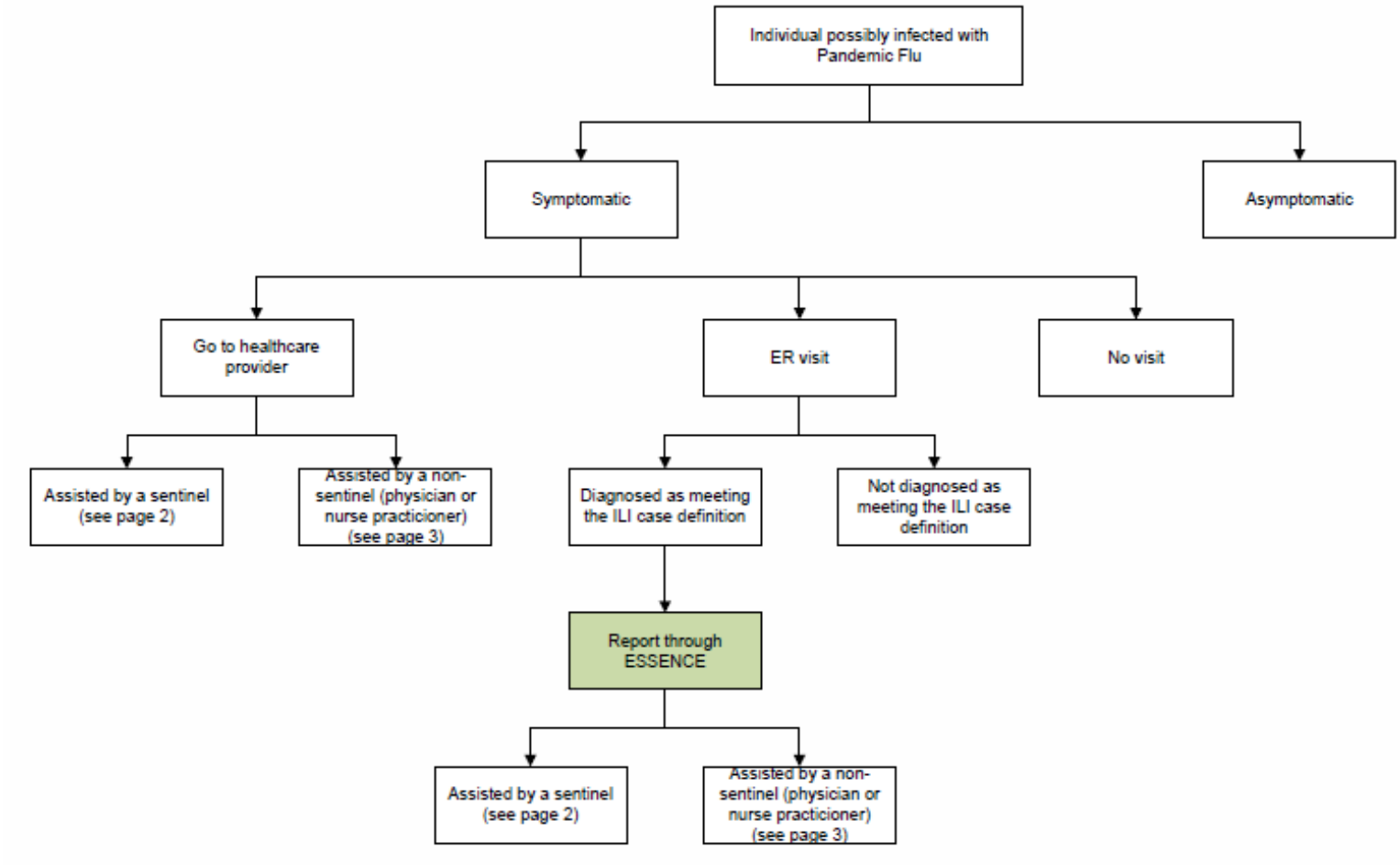


Figure E1: Flow of infected individuals through the surveillance system

Appendix E Continued

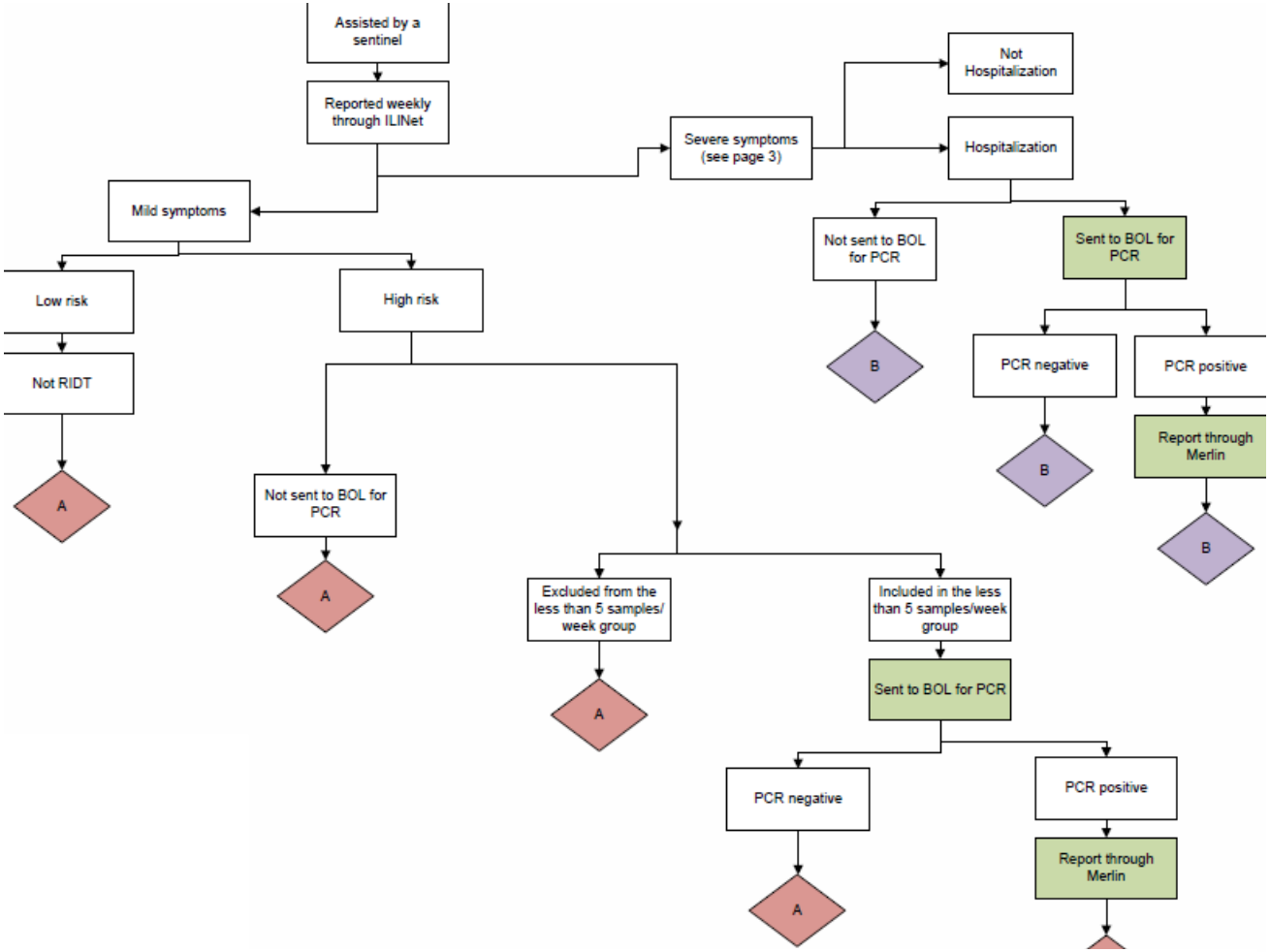


Figure E1: Continued

Appendix E Continued

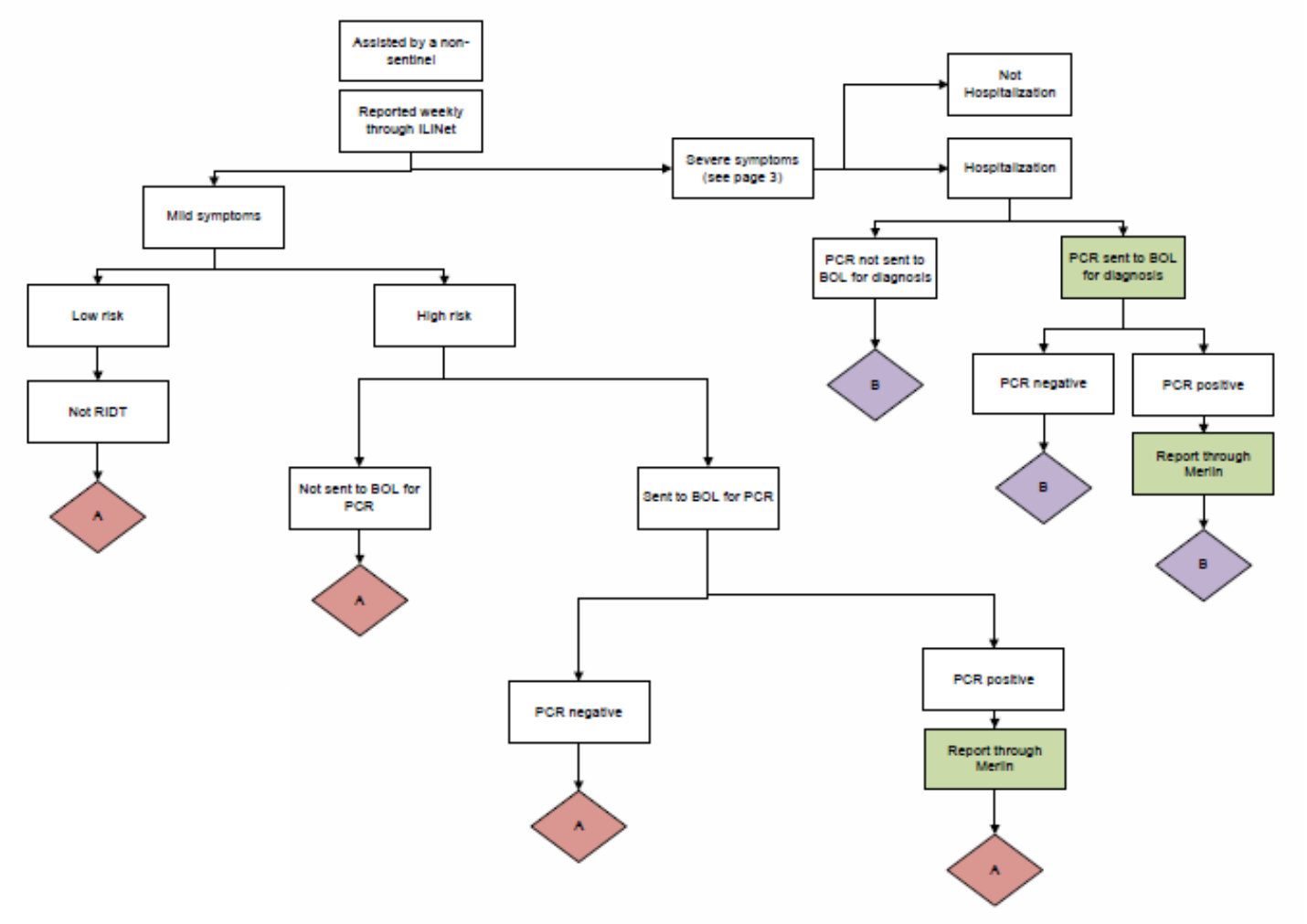


Figure E1: Continued

Appendix E Continued

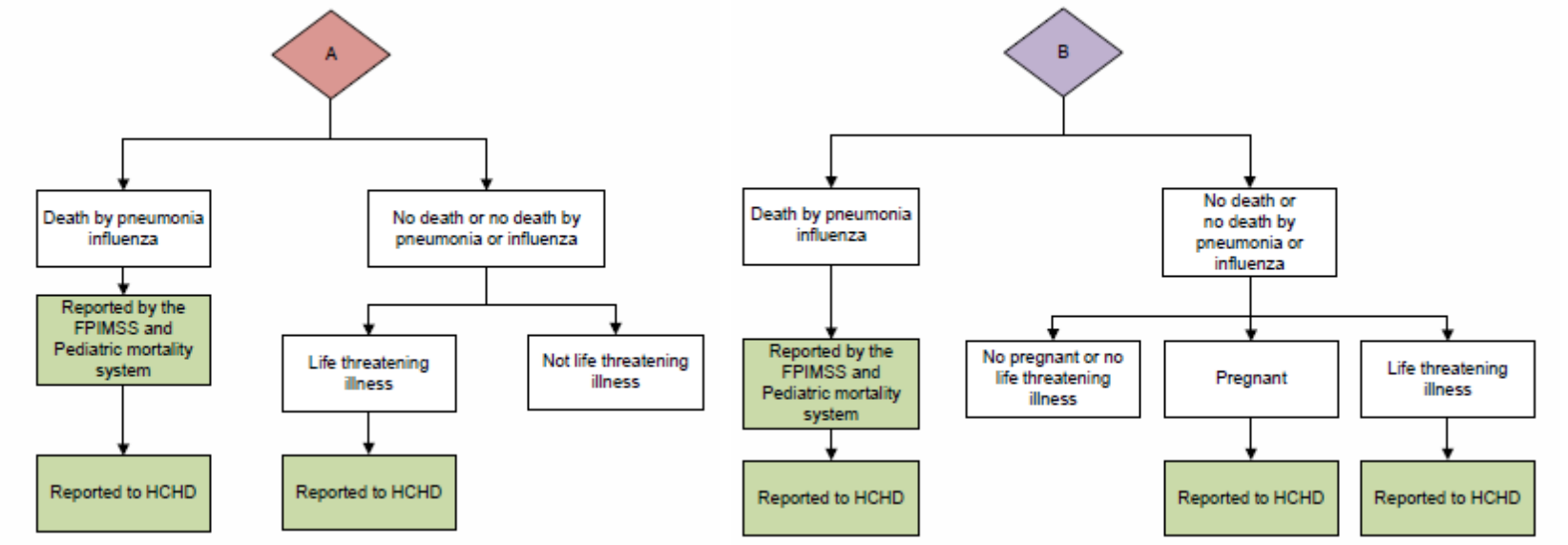


Figure E1: Continued

Appendix F Effect of the Reporting Rates in Sum of Squares of the Instantaneous Reproduction Number *R(t)*

Table F1: Factors of the experiment

Factor	First	Second	Third
Probability that a symptomatic individual seek for healthcare (r_h^1)	0.2	0.5	0.8
Probability that a healthcare provider submit a specimen to the BOL (r_l^1)	0.2	0.5	0.8
Lab capacity (L)	1000	5500	10000

The experiment included three responses, namely, the Sum of Squares for the first estimation level, the Sum of Squares for the second estimation level, and the Sum of squares for the third estimation level

Table F2: ANOVA for the first estimation level

Source	DF	SS	MS	F	Pr > F
r_h^1	1	33409.14	33409.14	6.765992	0.0160
r_l^1	1	21742.99	21742.99	4.403372	0.0470
L	1	0.07241	0.07241	0.000015	0.9970
Model	3	55152.2	18384.07	3.723126	0.0256
Error	23	113569.5	4937.804		
Total	26	168721.7			

The uncoded predictive model for the first estimation level is as follows
Sum of Squares for the first estimation level $= 201.8986 - 143.6069 r_h^1 - 115.8516 r_l^1$

Table F3: ANOVA for the second estimation level

Source	DF	SS	MS	F	Pr > F
r_h^1	1	1812.421	1812.421	16.73723	0.0004
r_l^1	1	398.7138	398.7138	3.682018	0.0675
L	1	0.004667	0.004667	0.000043	0.9948
Model	3	2211.139	737.0464	6.806432	0.0019
Error	23	2490.595	108.2868		
Total	26	4701.735			

The uncoded predictive model for the second estimation level is as follows

Appendix F Continued

Sum of Squares for the second estimation level $= 34.81232 - 33.44814r_h^1 - 15.6882r_l^1$

Table F4: ANOVA for the third estimation level

Source	DF	SS	MS	F	Pr > F
r_h^1	1	132.5171	132.5171	17.15874	0.0004
r_l^1	1	88.75272	88.75272	11.49199	0.0025
L	1	0.005416	0.005416	0.000701	0.9791
Model	3	221.2752	73.7584	9.550478	0.0003
Error	23	177.6291	7.723006		
Total	26	398.9043			

The uncoded predictive model for the second estimation level is as follows
Sum of Squares for the second estimation level $= 13.58697 - 9.044371r_h^1 - 7.401731r_l^1$