MODELING EPIDEMICS ON STRUCTURED POPULATIONS: EFFECTS OF SOCIO-DEMOGRAPHIC CHARACTERISTICS AND IMMUNE RESPONSE QUALITY

Jorge A. Reyes Silveyra M.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2014

APPROVED:

Armin R. Mikler , Major Professor
Arthur Goven, Committee Member
Paul Tarau, Committee Member
Robert Renka, Committee Member
Xiang Gao, Committee Member
Barret Bryant, Chair of the Department of Computer Science and Engineering
Costas Tsatsoulis, Dean of the College of Engineering
Mark Wardell, Dean of the Toulouse Graduate School

Reyes Silveyra, Jorge A. Modeling Epidemics on Structured Populations: Effects of Socio-Demographic Characteristics and Immune Response Quality. Doctor of Philosophy (Computer Science and Engineering), August 2014, 133 pp., 40 tables, 31 illustrations, bibliography, 111 titles.

Epidemiologists engage in the study of the distribution and determinants of health-related states or events in human populations. Eventually, they will apply that study to prevent and control problems and contingencies associated with the health of the population. Due to the spread of new pathogens and the emergence of new bio-terrorism threats, it has become imperative to develop new and expand existing techniques to equip public health providers with robust tools to predict and control health-related crises.

In this dissertation, I explore the effects caused in the disease dynamics by the differences in individuals' physiology and social/behavioral characteristics. Multiple computational and mathematical models were developed to quantify the effect of those factors on spatial and temporal variations of the disease epidemics. I developed statistical methods to measure the effects caused in the outbreak dynamics by the incorporation of heterogeneous demographics and social interactions to the individuals of the population. Specifically, I studied the relationship between demographics and the physiological characteristics of an individual when preparing for an infectious disease epidemic.

Copyright 2014

by

Jorge A. Reyes Silveyra

ACKNOWLEDGMENTS

This research would not have been possible without the tutelage of my supervisor Dr. Armin Mikler. Dr. Mikler once explained to me that in Germany a Ph.D. supervisor is called Doktorvater. This word translates as doctor-father, and I truly believe that his role during this 6 years could not be described with a better word. He became the person that supported me, academically and personally, during this period. I will be forever grateful for all of the time, love, support, and immense amount of knowledge that he has given me.

I would also like to express gratitude to Dr. Goven for teaching me the biological aspects of immunology. To Dr. Gao for introducing me to biostatistics and the language R. To Dr. Renka and Dr. Tarau for their support in the development of this research.

Additionally, I appreciate all those that provided their assistance to me and to this research in other various forms. These include: Sarat for all those hours coding and discussing models with me; Court, Marty, Josh, Angel, Joseph, and all other current and past Cerl and CeCera members for your time and help.

Finally, I would like to thank my family. To my husband Antonio for supporting me with immense love and countless words of encouragement, for always having faith in me, and for his seemingly endless patience. To my mom for her unconditional love and support in every single aspect of my life, I could not have asked for a better mother. To Jos for showing me that passion and perseverance can take you wherever you wish to go in life. To Cuau for inspiring me with your strength, humbleness, and unconditional support. To my dad for teaching me that hard work, honesty, and passion for what you do are a perfect combination for success. To Mamabue, Tito, Abue Oliva and all my cousins, uncles, aunts, and in-laws for their support and love during all these years. To my friends: Teto, Robert, Cass, Luis, Arturo, Eli, Mike, Saul, Conrad, and Fleur for being awesome. And to my amazing girls: Scuppy, Daisha, Sugar, and Cinnamon for rocking my world.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Epidemiology was first established when Aristotle started questioning the magical aspects of illness and noticed that some diseases appeared more frequently in some areas than others [66]. Ever since, epidemiologists have been engaging in the study of the distribution and determinants of health-related states or events in human populations. Eventually, they will apply that study to prevent and control problems and contingencies associated with the health of the population[61]. Because of the spread of new viruses such as H1N1 in 2009 and the emergence of new bioterrorism threats, it has become imperative to develop new and expand existing techniques to equip public health providers with robust tools to predict and control health-related crises.

Mathematical and computational models for different types of infectious diseases, including vector-borne and airborne transmitted illnesses, are based on the concept of interactions or contacts. For person-to-person transmitted diseases, a contact can be thought as an abstraction of human interaction, including attendance at the same social event, the same conference or any other situation that could involve a close encounter between two individuals. Such relationships are developed in accordance to the structure of the social networks to which the individual belongs. Chances and frequencies of interactions between individuals will change according to the proximity of their social networks, their association and affinities [34][44].

Early mathematical and computational models consider interactions between elements to be homogeneous, implying that all of the elements have the same likelihood to contact others and to be contacted. Real life behavior, however, indicates that this is not necessarily true; People create social clusters in which the chances of contacting others change according to affinities, geographic location, social position, etc. Creating models that incorporate parameters to create non-homogeneous populations is of uttermost importance to answering questions regarding existing and emerging health concerns.

The most recent approach to infectious disease outbreak modeling incorporates non-homogeneous components to the individuals to be modeled. Studies of the effects of non-homogenous populations on the dynamics of infectious outbreaks have shown the importance of integrating individuals with heterogeneous characteristics [84]. Although the amount of time a person is capable to transmit the disease varies among individuals, many models set that value to be homogeneous for the population. Season and temperature are some of the sources that influence that difference [4] [91]. However, we highlight the capacity of the immune response to diminish the pathogen as a crucial component for that variation as well.

The immune system of an organism provides an extraordinary defense against foreign attacks. Once it recognizes matter as non-self, it activates multiple chemical and physiological processes to control and eliminate the pathogen [47]. These processes are collectively known as immune response. The immune system mounts a response in an attempt to stop the growth of an invading organism in order to retain optimal functionality of the host. Controlling such proliferation is beneficial for the organism, since the quantity of foreign material affects the amount of time during which the organism experiences infection. Further, the duration of the infectious period, during which an organism might transmit the infection to others, is directly related to the quantity of foreign material in the host. Hence, we conjecture that the dynamics of an infectious disease epidemic in a population are driven, among other things, by the organism's immune response. This research establishes this relationship through the integration of immune response into the population at large.

There is an endless number of factors that can contribute to variations in immunucompetence among individuals. However, we concentrate in the elements associated with the host. Age, physical fitness, gender, and nutrition are some of the most commonly studied factors [57]. Further, we examine the role of those demographic characteristics to determine the immune response quality of the different segments of the population. Incorporating the quality of the immune response of a population in epidemiological models could lead to variations in the geographic progression of the epidemic. To analyze those variations, we introduce the concept of epidemic trajectory and define its properties. In order to measure

the properties of the trajectory of an epidemic, we developed a model of the spread of a disease within a population that is located in a realistic geographic space. This model allows the incorporation of individuals with unique demographic characteristics and, consequently, different immune response qualities. To capture the effects of the immune response on the epidemic trajectory, we implemented the model and simulated multiple disease outbreaks in distinct populations. Those populations include important variations on their demographic characteristics. Ultimately, the analysis of the properties of the epidemic trajectory in those simulations will highlight the role of the immune response in the progression of an epidemic. In general, the connection between population immune response and the disease trajectory will reveal any relationships between population demographics and disease prevalence and propagation from an immunological perspective.

The interdisciplinary nature of this research requires multiple areas of science to collaborate. In the following section, we describe the contribution of those areas to this research.

1.1. Areas of Study

Every year new public health policies are being implemented to improve the quality of life of individuals. In recent times, enforcement of the use of car seat belts and the creation of anti-smoking legislations have resulted in multiple positive outcomes for society[1][83]. However, public health professionals still need to address many other issues. To undertake this massive task, public health officials started collaborating with experts from others areas of knowledge. Computer science, biology, mathematics, among others areas, are some of the contributors to the development of new guidelines. The partnership among these disciplines has resulted in a the commencement of a new era of public health.

Epidemiology is one of the areas from public health that has obtained great benefits from the collaborations with other areas of knowledge. Experts from mathematics and computer science have supported epidemiologists in the creation of models and simulations of infectious disease outbreaks. The development of these models require comprehension of the biological processes and sociodemographic components associated with an infectious

epidemic. To understand those concepts, epidemiologists obtain assistance from experts from the medical field and biology.

Figure 1.1 depicts the interaction between Computer science, medical field and public health and the contribution of each area to this research. The figure illustrates the augmented role of computer science. computer science provides modeling and simulation techniques that can be implemented to build a system to simulate biological processes. The system is composed of two modules: A model to simulate the immune response of an individual during disease and a model to simulate the transmission of disease among individuals within a population. For this research, the medical field provides biological descriptions of the disease, the immune system and the immune response. Similarly, public health supplies demographic distribution information and epidemiological data.

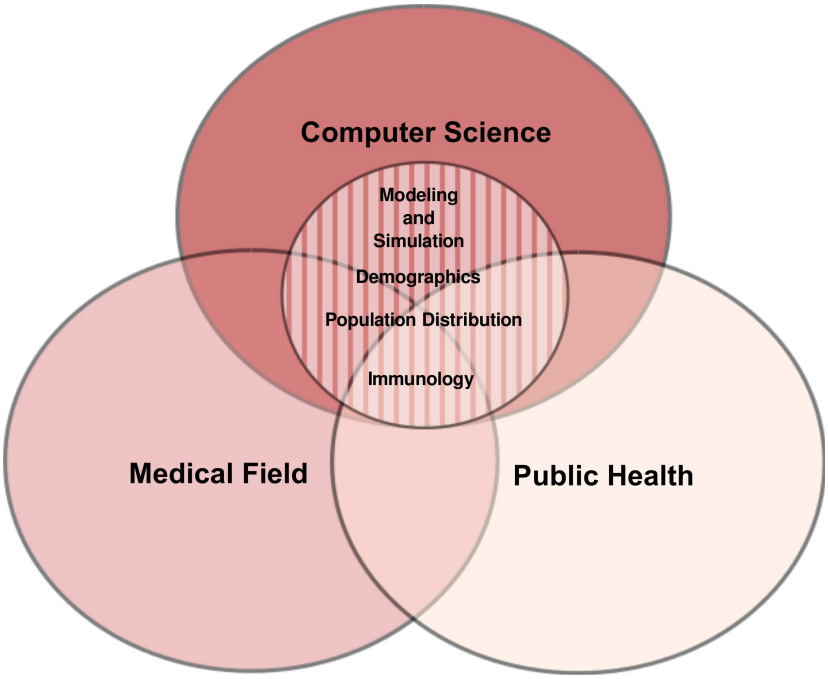

FIGURE 1.1. Areas of study and their contribution to the research

1.2. Contributions

This dissertation is divided in five main sections: Epidemics in Time and Space, Modeling Social/Behavioral Characteristics, Modeling Physiological Differences, and Mod-

eling Outbreaks in Geographic Context. Following, we present a brief description of each section and their contributions.

1.2.1. Epidemics in Time and Space

Epidemiologists build models to simulate the progression of a disease in populations, with special emphasis in infectious diseases. These experts use methods to measure the impact of assigning unique characteristics to the hosts [38], changing the environment [5], among others. A recurrent method to measure the effect caused by those parameters is by assessing the variations in disease incidence caused by their incorporation. However, this methodology completely omits the spatial component of the epidemic. In this section we define the concept of epidemic trajectory and a methodology to measure differences in the properties of the trajectories of different epidemics.

1.2.2. Modeling Social/Behavioral Characteristics

Mathematical and computational models for different types of infectious diseases, including vector-borne and airborne transmitted illnesses, are based on the concept of interactions or contacts. For person-to-person transmitted diseases, a contact can be thought as an abstraction of human interaction, including attendance at the same social event, the same study groups or any other situation that could involve a close encounter between two individuals. Such relationships are developed in accordance to the structure of the social networks to which the individual belongs. Chances and frequencies of interactions between individuals will change according to the proximity of their social networks, their association and affinities [34][44].

The majority of the mathematical and computational models in literature consider interactions between elements to be homogeneous, implying that all of the elements have the same likelihood to contact others and to be contacted. Real life behavior, however, indicates that this is not necessarily true; People create social clusters in which the chances of contacting others change according to affinities, geographic location, social status, etc. In this section, we introduce important epidemiology concepts and highlight the importance

of incorporating social/behavioral characteristics in epidemiological models. Additionally, we present a model that incorporates social/behavioral characteristics, and a quantitative analysis of the importance of the incorporation of those characteristics.

1.2.3. Modeling Physiological Differences

The infective process in the host is driven by the competition of the pathogen and the immune system. Once a pathogen starts replicating in the host, the immune response attempts to stop its proliferation. During this process the amount of pathogen load in the blood can be quantified. This value is used to measure the disease severity and the capacity of the host to transmit the disease. Additionally, in this research we analyze its role to determine the duration of the infectious periods.

Every individual is different. Due to this distinctness, their immune system is different as well. The quality of the immune response of an individual is determined by multiple demographic characteristics, among other factors. Age, physical fitness, gender, and nutrition are some of the most commonly studied factors. This research explores and discusses the importance of considering the variation in the immune response quality caused by the characteristics of the population and their impact in the epidemics.

The concept of population immune response (PIR) is introduced and its contributions to the infectious outbreak modeling are highlighted. Population immune response (PIR) is a new concept that captures the collective immune response (IR) of individuals in a population represented by the superposition of individual immune responses. To represent this concept, a computational model is proposed in which three modules are required: population and disease database, immune competence and infectious disease outbreak simulation. The computational model is implemented to highlight the importance of PIR.

1.2.4. Modeling Outbreaks in Geographic Context

This research highlights the crucial role of the demographic characteristics of the population during the progression of an epidemic in a geographic space. In this section, we present a model that simulates the interactions of a population at the census block level. This

model is used to measure the properties of the spatial progression of an epidemic in a specific geographic location. The properties of the trajectory are captured in the force of infection and the transmission trend of the census blocks of the geographic space. Further, the properties of the trajectory are analyzed to measure variations in the outbreak dynamics caused by the effects produced by PIR. Ultimately, this study compares the effects that multiple geographic, immunological, and socio/demographic characteristics have on the disease dynamics.

1.3. Overview

This dissertation is divided in six chapters. Chapter 2 presents a background review of concepts and literature relevant to this research. The chapter is divided in sections that specifically address each chapter of this document. Chapter 3 describes a method that captures the underlying process of the trajectory of an epidemic in a population with the introduction of the concept of epidemic trajectory. Further, that concept is employed to define a methodology to measure differences in the properties of the trajectories of different epidemics. In chapter 4, the importance of social/behavioral characteristics in epidemiological models are studied. In this chapter a computational model to simulate the spread of infectious epidemics in a heterogeneous population is presented. This model is used to highlight the importance of incorporating heterogeneous contact rates and affinities between individuals during simulations of infectious outbreaks. In chapter 5, we describe several physiological characteristics of an individual and investigate their role during infection. Further, we highlight the importance of understanding and integrating the immune response effects to the study of an infective process and its relationship to an infectious outbreak. In this chapter we introduce the concept of population immune response and present a framework to build a computational implementation. Finally, in chapter 6 we highlight the crucial role of the demographic characteristics of the population during the progression of an epidemic in a geographic space. A computational model is implemented to compare the effects that multiple geographic, immunological, and socio/demographic characteristics have in the disease dynamics. Ultimately, this model is used to compare the effects of PIR in two populations with distinct demographic distributions.

CHAPTER 2

BACKGROUND

The multidisciplinary nature of this research requires the study of different concepts from multiple academic disciplines. In the following sections, we present a review of the previous work and background concepts of importance for this research.

2.1. Epidemics in Time and Space

An epidemic is the occurrence in a community or region of cases of an illness, specific health-related behavior, or other health-related events clearly in excess of normal expectancy [90]. Although the majority of epidemics are caused by non-infectious agents, the impact of infectious epidemics in history has been of colossal importance. McNeill [58] lists historic events in which the presence of infectious pathogens have affected the course of human history. The presence of an epidemic results in great expenses in different types of resources, including time, wealth, and loss of life [22][81]. Implementing measures to limit and stop epidemics is part of the role of epidemiology and epidemiologists.

Understanding the intrinsic characteristics of an epidemic is crucial to implement strategies to attempt to control or stop it. Determining the nature of the source of transmission for the disease causing the epidemic is one of those characteristics. Epidemiologists consider four types of transmission sources: fomites, vector, reservoir, and carrier [90]. A *fomite* is an object that can harbor a disease and is also capable of transmitting it. A *vector* is an invertebrate animal capable of transmitting the disease. A *reservoir* is the habitat in or on which the infectious agent lives, grows, and multiplies. As infectious organisms reproduce in this reservoir, they do so in a manner that allows disease to be transmitted. Finally, a *carrier* contains, spreads, or harbors an infectious organism. The carrier is later capable to transmit the disease.

In this research, we consider epidemics in which transmissions occur through contacts between carriers and susceptible individuals (an individual that does not carry the disease and could be infected). However, independent of the method of transmission, for an epidemic

to flourish it is necessary the presence and interaction of the four epidemiological factors: environment, time, host, and agent [90]. Figure 2.1 depicts the interaction and interdependence of those factors. The *environment* is the cluster of conditions and surroundings external to the human or animal that cause or allow disease transmission. It can include biological aspects as well as the social, cultural and physical aspects of the environment. The *host* is the organism that harbors the disease, e.g. human, animal, etc. The *agent* is the cause of the disease, e.g. virus, bacteria. Finally, the *time* accounts for incubation periods, life expectancy of the host or pathogen, and duration of the course of the illness or condition. The primary goal of epidemiology is to perturb the triangle of epidemiology to stop the spreading of the disease.

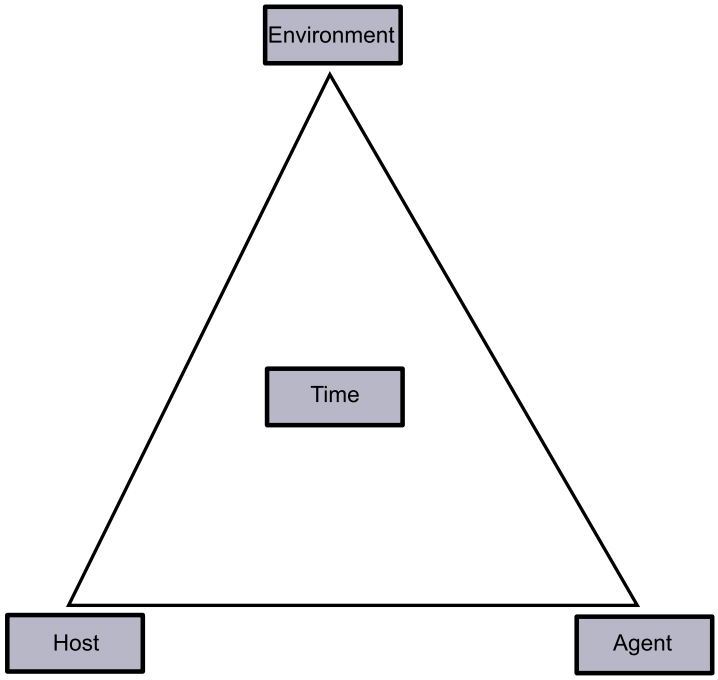

FIGURE 2.1. Epidemiological triangle

2.2. Heterogeneous Populations Modeling

Most epidemiological mathematical models are based on the interaction principles between groups of susceptible(S), latent(L), infectious(I), and recovered/removed(R) individuals, i.e., the SIR/SLIR model. Susceptibles are members of a population who can be

infected by the disease. Infectious individuals have been previously infected and are now themselves contagious. Latent individuals are infected, however cannot yet transmit the disease. Recovered/removed include individuals that are incapable of transmitting the infection, and are currently recovering, have fully recovered or have deceased from the disease. In more complex models, the recovered/removed can return to the susceptible stage if they do not gain immunity after being infectious. The SIR model is based on the Kermack-McKendrick threshold theorem in a closed homogeneous population [13]. This threshold theorem is used as an epidemiological indicator to estimate the probability that an infectious disease will produce an epidemic in a completely susceptible population [3]. The model utilizes a closed population since it assumes the epidemic spreads sufficiently quickly that the changes introduced by births, deaths, and migration are negligible. The SLIR model is depicted by the following set of differential equations:

$$
\begin{aligned}
\tfrac{dS}{dt} &= -\beta IS \\
\tfrac{dL}{dt} &= \beta IS - \epsilon L \\
\tfrac{dI}{dt} &= \epsilon L - \gamma I \\
\tfrac{dR}{dt} &= \gamma I
\end{aligned}
\tag{1}
$$

In this model, the number of individuals in each compartment fluctuate through time depending on the latency rate (ϵ), the recovery rate (γ), and the force of infection (β). The latency rate, ϵ, is the proportion of individuals that depart the latent compartment to enter the infectious compartment. This rate is used to define the latent period. The latent period is the time an infected individual requires to become infectious after being infected. The recovery rate, γ, is the proportion of individuals that migrate from the infectious compartment to the recovered compartment. This rate is applied to define the infectious period (IP). The infectious period is the time extent that an individual from the infected compartment remains infectious. Finally, the force of infection is a transition rate that determines the number of susceptible individuals that are moved to the infectious compartment every

time step. This rate represents the number of interactions between infected and susceptible individuals that lead to transmission of disease.

The SIR model provides a simple framework to represent the spread of a disease. However, it does not provide sufficient detail to study outbreak dynamics and, consequently, lacks flexibility to be used as a policy or a planning tool. For example, the SIR model neither considers the regional geography or demographics of the population, nor does it consider differences between the individuals and their interactions.

Cellular automata (CA), agent-based models (ABM), and hybrid models, which combine two or more models, are some of the computational techniques that have been utilized to model the spread of infectious diseases in populations. The implementation of CA consists of a grid of cells, each cell representing an individual or a group of individuals from a population. At the beginning of this process, each cell is assigned a specific state representing its disease status. As the simulation progresses the system will examine the state of each cell and determine all possible state changes according to its own set of rules. Situngkir [87] utilized the cellular automata approach to simulate the spread of avian influenza in Indonesia. Mikler et al. [63] introduced an extended cellular automata paradigm, the global stochastic cellular automata (GSCA). The authors analyzed the interaction of cells based on the saturation of the neighborhood and the distance between elements to represent geographic, demographic, and migratory constraints. Gagliardi et al. [39] recreated the propagation of dengue in a population by overlapping two cellular automata grids, in which one grid represents the mosquito population and the other one the human population. The simulation computes the proximity between elements from different grids to symbolize the interactions between humans and mosquitoes.

Agent-based modeling has been widely used in epidemiological simulation [6][76][11]. This modeling technique represents each individual as an independent entity with properties or characteristics that can affect the spread of a disease. For instance, Barret et al. [15] developed a simulation environment, episimdemics, that is based on the agent-based paradigm. This model reproduces infectious outbreaks in large populations with elements that include

characteristics based on social-behavioral studies. Tian et al. [94] developed a computational model to simulate an infectious outbreak that incorporates a surveillance system. In this model, the severity of the pandemic is reduced by triggering the allocation of health care providers and drugs to the disaster areas once a threshold of infectious elements is reached. Bouden et al. [21] incorporated an agent-based approach to the simulation of the propagation of West Nile virus by replicating the interaction of mosquitoes and birds in different climates and different location distributions. Venkatachalam et al. [97] constructed a hybrid model, the global stochastic field model (GSFM), that included geographic and demographic information in the analysis of the spread of diseases. This model determines the probability of interactions between elements based on census information and their neighbors' status and characteristics.

Different models have been proposed to study the inclusion of differentiated populations in the simulation of an infectious outbreak. Wallinga et al. [99] concluded that contact patterns are heterogeneous and that incorporating non-homogeneous parameters into models could lead researchers to more realistic results. Some of the parameters considered in the model resulted in age stratification and a differentiation between local and global contacts. Mossong et al. [68] quantified various contact patterns according to different criteria, including age, gender, nationality, size of household, and day of the week. Other models focus on the geographic location of individuals, since the physical position alters the likelihood of contacting certain individuals. Avilov et al. [12] studied the spread of tuberculosis in different regions in Russia. The authors showed how location can affect the types of contacts, as well as the morbidity of the bacteria. Melnichenko et al. [59] incorporated socioeconomic parameters in similar geographical regions, indicating that the socioeconomic level of individuals will create affinities among them. Individuals status can affect the quality of health service they can obtain, hence changing the morbidity of the infection. Kress [49] applied a time-schedule driven simulation, in which the elements can be located in three types of places : households, meeting places, and schools, or workplaces. The simulation indicated that location determines the probability of contacts.

Studies have shown that particular age groups have different infection and mortality rates [7]. This has motivated the inclusion of differentiated age groups into the simulation of disease progression. Li et al. [54] developed a model to incorporate age-structured groups of elements to represent the spread of HIV/AIDS. This approach was further extended to incorporate two-age and multi-age group individuals in a malaria simulation. Pertsev et al. [78] created a mathematical and computational model that simulate the spread of diseases in populations, in which individuals are assigned specific locations, age, and a disease status.

2.3. Immunity and Infection Models

Immunological mathematical modeling is a relatively new research technique to attempt to understand the functionality of the immune system [30]. Multiple models have been developed to simulate immune processes before, during, and after infection. Forrest et al. [36] presented a review of the different modeling techniques that are used to model the immune system. These models can be divided in three groups:

1. Models of the immune system functionality.
2. Models of the immune response of a host during disease.
3. Models of the immune response of a host during a pathogen-specific infection.

The first group depict specific processes of the immune system. These models are implemented to obtain a greater understanding of specific functionalities of the immune system and its components. An example for this group of models is presented by Wilson et al. [106]. Wilson et al. developed a T-cell expansion/contraction model in which contraction is managed by adaptive regulatory T-cells instead of being a predetermined process. This model was used to augment a paradigm already utilized in other models [46]. Terry et al. [92] present another example of this type of models. The authors developed a mathematical model of a primary CD8 T-cell response during an infection in which the developing phases of the CD8 T-cells and their role during infection are replicated.

The second group of models replicate the processes during the interaction of an individual immune response with a disease. These models highlight different aspects of the

immune system based on the agent of study. The pathogen behavior in models of this group does not mimic a specific agent, but a type of organism or disease. Virus, bacteria, and parasites are some of the most commonly researched pathogens. For instance, Nowak et al. [73] presented a description of the basic viral dynamics model to further extended it to include T-cytotoxic lymphocytes (CTL) to eliminate infected cells. Further, Wodarz et al. [108] also expanded the basic viral dynamics model by introducing antibodies to block and destroy viruses. Additionally, Chao et al. [26] presented a stochastic model of cytotoxic T-cell responses to a viral infection. This model depicts the cellular interaction between immune and pathogen cells. The model considers the role of pathogen-independent and plasma cells, as well as possible effects of different drug therapies.

Wodarz [107] presented a review of different immune-viral models that incorporate lytic and non-lytic responses to defend against the pathogen. Lytic responses control viral growth by destroying infected cells, while non-lytic responses inhibit viral growth without killing the cell. In this review, combinations of both types of responses are studied to spotlight the optimal response based on the characteristics of the virus. An analysis of both cytopathic and non-cytopathic viruses is presented. Cytopathic viruses destroy cells during infection and they are represented in the model by equating the life-span of an infected cell with that of a susceptible cell. Analogously, non-cytophatic viruses in the model are implemented by assuming the natural death rate of infected cells to be equal to the death rate of uninfected target cells.

This type of model is not unique to infectious diseases, chronic diseases are also simulated. For instance, Bianca et al. [18] presented a mathematical model of the competition of the immune system and mammary cancerous cells. The immune system is represented by B-cells, T-helper cells, antibodies, Cytotoxic T-cells, antigen presenting cells, and Interleukin 12 and 2. The model incorporates the effects of three preventive vaccine (triplex) protocols.

The third type of models are used to study the processes during the interaction between the immune response and a specific pathogen during an infective process. These models incorporate specific processes of the immune system that are crucial to defend against

each respective pathogen. Some models replicate specific behaviors of the pathogen, while others study the illness afflicted to the host during infection. Influenza [53] and HIV are the most common pathogens modeled by immunologists. HIV models range from infection [89][77], treatment [102], co-infection [17], and testing strategies simulations [40], etc. Wodarz and Nowak [109] presented a review of different HIV viral dynamics, progression and therapy models. In this review, the authors describe a basic model of viral dynamics during infection without immune response to later expand it with anti-viral therapy effects. Lastly, a final model is introduced in which the viral replication in a host is simulated with two antagonists: CTL-mediated immune response and anti-viral therapy.

Multiple bacterial diseases have also been modeled, including chlamydia [105] and tuberculosis [101]. Wigginton et al. [104] presented a model to simulate the interaction between *Mycobacterium tuberculosis* and the host. This model is depicted at the cellular level and is divided in six cell populations, two bacterial populations and four types of cytokines. The mathematical model is divided in 12 differential equations with fifty six parameters. All the parameters were estimated from experimental data. The experiments consisted of solving the model to investigate different disease trajectories, damage induced by the immune system, and virtual deletion and depletion simulations.

2.4. Immune Response and Epidemiology Models

Epidemiologists have made use of different approaches to incorporate immunological factors into their models. Some models implement immunity as an effect of a vaccination strategy [60] or as a period in which the individual cannot be infected after a primary infection [51]. Moreover, other scientists have engaged in a comprehensive study of the relationship between immune response and the population. Hellriegel [45] presents a review of that relationship. In this review, the importance of integrating immunity and epidemiology (immunoepidemiology) is highlighted. The author defines three different approaches for this integration: within-host, between-host and individual-to-population dynamics. Furthermore, the author proposes different combinations of those approaches to assess the role of immunity in determining epidemiological patterns. Following, we present a description of

multiple epidemiological models that incorporate immunologic factors.

Dushoff [33] depicted a model in which the probability of the disease progressing in an infected individual is not only determined by its own characteristics, but by the level of disease in the population. That assumption is based on the suggestion that an exposure with low pathogen load may lead to immunity or short lasting infection and minimum disease transmission, while an exposure with high pathogen load may lead to a longer infection and to greater transmission of the disease. To implement that assumption, the model contains two individual classes: heavily infected and lightly infected. Individuals can depart from one class to another as a function of the force of infection.

Martcheva and Pilyugin [56] presented a susceptible-infectious-recovered model in which the immune status of the individuals increases during the infectious period. In this model, the initial absence of immunity of individuals sets all of them in the susceptible group. However, once an individual becomes infected its immune status increases over the course of infection. After an individual recovers from the disease, a possible reinfection its restricted by its immune status. However, the immunity may decrease as a function of time, increasing the probability of a secondary infection.

Vickers and Osgood [98] introduced a mathematical framework of population infection dynamics in which individuals mount an immune response in response to infection and the contacts between them are distributed in a simple contact network. The immune response is represented by a population of differentiated and non-differentiated cytotoxic immune cells. The social network is implemented by placing each individual in a Poisson distributed network such that the incoming viral load of an individual is proportional to the viral load of its neighbors. Each individual has a coefficient of connectedness to determine the weight of the connection with each of its neighbors.

2.5. Modeling Outbreaks in Geographic Context

Epidemiologists have introduced multiple methods to study the progression of epidemiological outbreaks in a geographic space. A major method to study that progression is contact network epidemiology. Contact network epidemiology is a framework that is used

to study epidemiological progression by the implementation of networks that reflect the associations between individuals involved in the outbreak [62]. The networks are constructed following the patterns of interactions within communities. Those networks are analyzed to determine their topological properties and, further, make epidemiological predictions or incorporate intervention strategies. Ultimately, this approach provides analytical tools to explicitly capture the diverse interactions that underline the spread of the disease in a population.

Newman [71] introduced an influential mathematical model in the contact network epidemiology area. This model is based on the SIR and its elements are individuals from a structured population that interact among themselves in a network. In this model, the interactions are designed to be non-homogeneous by allowing individuals to interact only along the links of their networks. Further, individuals in the network are assigned probabilities to determine the chances of their interactions. A simulation of this model results in analytic expressions that describe the size of the epidemic outbreak and the position of the epidemic threshold, as well as multiple network measures. Finally, the author presents a method to explore the significance of the analytic expressions to make specific suggestions for different strategies to control the epidemic.

Meyer et al. [62] presented a model in which the basic contact network is modified to incorporate geographic locations that mimic the areas in which individuals interact. Additionally, a thorough analysis of the role of the destination and direction of disease transmission (contact patterns) during an epidemic is presented. This analysis includes a comparison of the use of the basic reproductive number (R_0) to estimate the intensity of the epidemic as opposed to the contact patterns. The comparison consisted in using two different pattern networks of a patient zero while using the same R_0 during an outbreak. As a result, the outbreaks are significantly different resulting in inaccurate measurements of the prediction of the number of expected infections during the epidemic. Consequently, they propose the use of the estimated number of contacts during infection in conjunction with R_0, as a parallel method to the use of R_0 alone, to predict the expected number of cases at

each generation of transmissions.

Multiple computational models have also been proposed to study the progression of epidemiological outbreaks in a geographic space. Eubank et al. [35] introduced an agent-based model that includes population mobility such that each individual is assigned a position and activity every second. In this model, the characteristics of the population are calculated with the use of census, land-use, and transportation data. This information is used to construct a social network in which individuals and locations are represented with vertices and their interactions are the edges of the network. This network is used to simulate the spreading of a disease and to analyze the use of different mitigation methods to stop its progression. The simulation indicates that a more effective mechanism to stop the spread of an epidemic is the implementation of a program of targeted vaccination and early detection, as opposed to the traditional method of mass vaccination.

Borkowski et al. [20] outlined an agent-based epidemic simulator of the spread of a disease within an urban environment. The model is built using the discrete-space scheduled walker simulation engine. In this engine, the topology of a region is simulated based on geographic information. The agents are located in that region and their interactions are limited by their radius around their workplace or school. The radious varies depending on the day of the week. The model incorporates public and private transportation, agent schedules, and demographic information that determines the residence of the agent and its behavior. Ultimately, the model is used to simulate the spread of HIV in the Bronx in New York City.

Balcan et al. [14] introduced a data driven epidemiological simulation model called the global epidemic and mobility model. This model is a SIR implementation in a spatially structured geographic space that represents earth. The realism of this model is augmented by the use of multiple layers. The population layer amplifies the characteristics of the individuals of the population. These characteristics are determined based on demographics and scheduling data. Further, the mobility layer depicts a network in which the population interacts. This network includes populations (countries) that are divided in subpopulations

(cities). The nodes in the network are connected following airport and commuting network information. Finally, examples with seasonality, age-structure, and a specific disease are presented to show the functionality of the model.

2.6. Summary

In this chapter, we presented a summary of concepts and reviewed literature relevant to this research. We studied various epidemiological models in which the role of the geographic space and social networks is restricted to operate as a vehicle for the transmission of the disease. However, we expand that schema by implementing a method to use the geographic space to measure the progression of the disease and compute variations on the epidemic progression. The goal of this method is to provide an alternative measurement to traditional methods to analyze outbreak dynamics. Furthermore, we use those measurements to emphasize the use of heterogeneous characteristics in epidemiological models.

We also presented different models that imitate the dynamics of the progression of a pathogen in a population. The epidemiological models reviewed present an important variation on the characteristics assigned to the individuals of the population. In this research, we highlight the importance of consistently including heterogeneous characteristics to the population. we measure the variations caused on the outbreak dynamics by introducing changes in the social networks and interaction coefficients of a population.

Additionally, we reviewed multiple epidemiological models in which individuals are assigned an immune response. In these models, immunity is incorporated as a value that restricts the capacity to become infected or as a parameter that increases the capacity to infect other individuals. In reality, the immune system possesses many complex mechanisms that can be abstracted and incorporated into these models. We highlight the augmented role of the immune response as a function of the socio-demographic characteristics of the population and its prominent role during the epidemic progression.

Finally, we presented multiple epidemiological models that incorporate a realistic geographic space. In those models, the geographic space is used uniquely as the area in which individuals interact. However, in this research, we expand that role by analyzing

the effects demographics and immunological characteristics incur on the disease dynamics. Ultimately, we present a comparison between two distinct regions to highlight the crucial role of those characteristics.

CHAPTER 3

EPIDEMICS IN TIME AND SPACE

Capturing the underlying process of the trajectory of a epidemic in a population is a complex task. During an epidemic, the interaction of the four epidemiological factors lead to a process that progresses through time and space. In this process, the hosts determine the direction of the trajectory based on the interactions among them. Likewise the type of agent establishes the method of transmission and the environment provides the conditions for the epidemic to flourish. Therefore, the epidemic trajectory of an infectious disease progresses every time a host successfully transmits the disease to another. The trajectory will terminate when there are no more infectious individuals present in the population.

During an epidemic in which the pathogen is disseminated from interactions among hosts, the infectious individuals determines the trajectory of the epidemic. This leading role is assigned to the infectious hosts since they are the only individuals capable of transmitting the disease. Assuming that infectious individuals can infect at most a single individual during their infectious period, a trajectory would be a non-branching path. Figure 3.1.a depicts an example of this type of path. However, an infectious host is capable of transmitting the disease to multiple other hosts. This occurs since the infected host could interact with more than one person during its infectious period, hence causing the trajectory to diverge and separate into multiple paths. The generation of those numerous paths causes the trajectory to become a process that progresses into multiple branches simultaneously. Figure 3.1.b depicts an example of a multiple branching path.

Understanding the spatial progression of the epidemic requires the definition of the geographic space in which the population interacts. Such geographic space may represent a school, census block, country, etc. We assume the geographic location of an individual to be its permanent residence. Additionally, the residence of an individual is used as origin and destination of a transmission, independent of the location in which the interaction occurs. For example, let p_1 be an infectious individual that resides in region r_1 and let p_2 be a

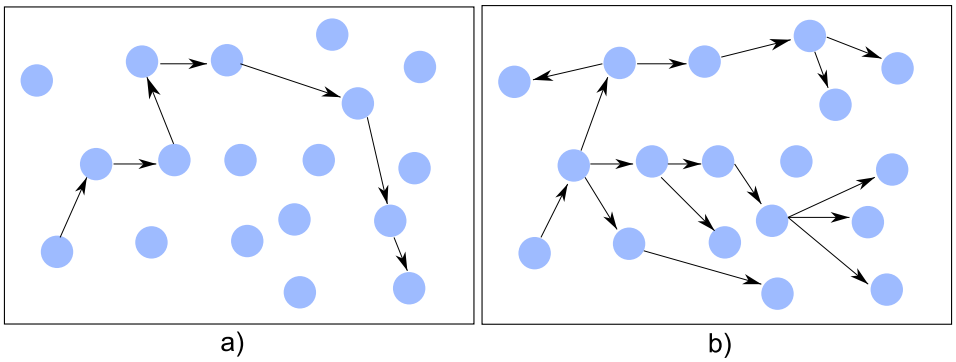

FIGURE 3.1. Single branch and multi-branch paths example

susceptible individual that resides in r_2. For an interaction in which p_1 infects p_2 in region r_3, we consider the origin of this transmission to be r_1 and the destination r_2, despite of the transmission having occurred in region r_3. This event represents the interaction of two individuals, that reside in different locations, at a social event outside their regions of residence.

Once the geographic space has been defined, the interaction space for the individuals of the population must be determined. The interaction space is the geographic area an individual can traverse from its geographic location to interact with other individuals. A naive assumption would allow individuals to only interact with individuals in their region and in regions which whom they share borders with. However, during the transmission of a disease this assumption is not precise since individuals are mobile entities and interactions may occur between individuals in regions that do not share a border. In a second interpretation, individuals are capable of transmitting the disease globally in the geographic space. Figure 3.2 depicts the trajectory of an epidemic following the two definitions of interaction areas previously described. Figure 3.2.a shows a single path trajectory with the naive interpretation and 3.2.b portrays a single path trajectory with the second interpretation.

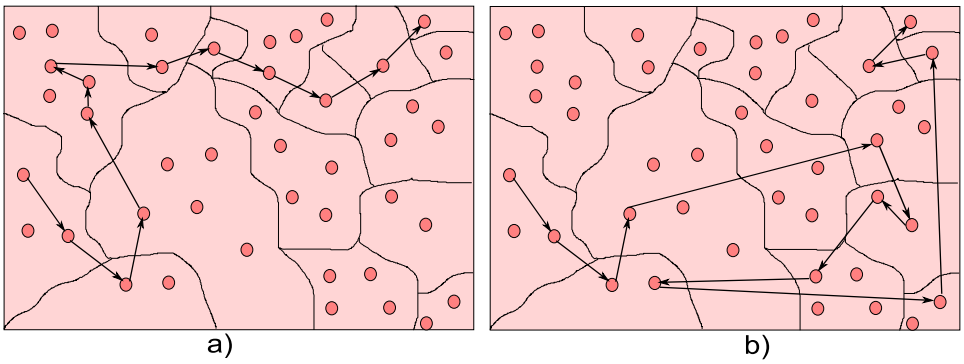

FIGURE 3.2. Different epidemic trajectories based on the interaction space interpretation

The second interpretation of the interaction space mimics the behavior of individuals with higher fidelity, but does not account for any restrictions in the movement of the individuals in the geographic space. Djist and Vidakovic [32] describe those restrictions as "potential action space" or the area in which individuals perform their daily activities. We incorporate that concept by integrating a fuzzy neighborhood. Mikler et al. [63] describe the fuzzy neighborhood as a network in which an individual from a region can reach each other regions based on a probability function. This function sets the likelihood for a region to reach any other region within a probability between 0 and 1. In general, we define the trajectory of an epidemic as the path, with multiple branches, of an contagion process that spreads in a fuzzy neighborhood. In the following section, we introduce a model and exemplify this definition.

3.1. Infectious Epidemic Progression

To understand the complex process of an infectious epidemic trajectory, we developed a model to simulate the spread of a pathogen in a population. This model mimics the mechanics of the susceptible-latent-infectious-recovered model (SLIR). As described in the previous section, the SLIR model divides the population in four compartments according to their disease status. In our model, individuals can transmit the disease by interacting with each other. Interactions in this model are represented following the interpretation of contact rate defined in Chapter 4. The contact rate follows the assumption that every individual

has a constant number of interactions or contacts per time interval. However, in this model the interactions are not assumed to be between individuals, but between regions; such that, individuals from every location contribute to the total number of interactions their region initiates. This assumption is based on the model proposed by Mikler et al. in [64].

In our model, the geographic space is represented as a rectangular grid. In this grid, every location represents a region in which individuals in groups of various sizes are accommodated. To represent unoccupied locations, we set the number of individuals in that region to be zero. Figure 3.3 depicts an example of a geographic space represented as a grid with 66 regions. In the figure, each location is assigned a random number to represent a non-homogeneous geographic distribution of the population. This random number is a value between 0 and 100 that is obtained from a uniform distribution. Therefore, each grid location hosts a region with a group of individuals of size between 0 and 100.

26	11	32	59	57	43
35	22	24	91	15	85
41	26	12	31	77	65
60	79	14	44	65	92
50	87	0	36	21	53
31	50	80	43	82	39
0	26	75	23	50	66
39	36	8	65	48	39
42	13	0	21	28	45
87	21	96	74	57	18
27	89	68	7	33	51

FIGURE 3.3. Example of a geographic space represented as a grid

As described previously, the contacts initiated by a region are calculated by adding all of the contact rate values of all of the individuals in that region. However, the computation of such calculation might result too expensive. For instance, if a region from a geographic space R has M individuals and they interact with a constant contact rate of CR_1, calculating all contacts of this region will generate a total of $(M * CR_1)/2$ contacts per time frame. Furthermore, if we assume each region R in the geographic space has N individuals and all interact with an equal contact rate CR_2, then calculating the number of interactions in all regions require $(|R| * N * CR_2)/2$ contacts per time frame. At the end of the epidemic (t_π),

the system will have computed a total of $(M * CR_1)/2 * t_\pi$ random contacts for one region and $(|R| * N * CR_2)/2 * t_\pi$ for the total geographic space. In a population of just 2,500 regions with 400 individual in each region and with a constant CR_2 of 40, the system would have to compute 20,000,000 contacts per time frame. The number of calls to the random number generator is a function of the number of contacts generated each time step. Hence in a simulation that takes 100 days we will have to generate 2×10^9 random numbers just for the purpose of contact generation. Consequently, in this model we only consider contacts that are initiated by an infectious individual, since those type of contacts are the only interactions that could lead to transmission of disease. Additionally, only considering the infectious contacts always results in a reduced number of computations than by calculating all contacts, except for the case in which the total population is infected at the same time [84].

In our model, the infectious contacts are distributed in the geographic space by implementing a fuzzy neighborhood. The fuzzy neighborhood in this model is represented with an interaction coefficient. The interaction coefficient is formally described in section 6.2.3. By implementing the interaction coefficient we obtain an array of matrices in which every element of the array is a matrix for an specific region. Each matrix contains the interaction probabilities for a region within all other regions based on their population and the distance among them. Therefore, whenever a region is set to seek a destination for one of its interactions, these probabilities are applied as bias in a random selection.

After an interaction between two regions has taken place, it is necessary to direct that interaction to an individual in contacted location. In order to determine the target of such interaction, we randomly select an individual from the population of the targeted location. If the selected individual is part of the latent, infectious, or recovered compartment we disregard that interaction since it cannot lead to transmission. However, if the interaction is directed to a susceptible individual we will determine if the interaction led to transmission of disease with a random experiment. This random experiment considers the infectivity of the disease to determine the probabilities of transmission. In the next section we introduce a methodology

that utilizes this model to quantitatively measure the properties of the progression of an epidemic.

3.2. Measuring the Properties of the Epidemic Trajectory

To measure the properties of the epidemic trajectory during an epidemic we analyze the geographic progression of a disease ϵ within the population P. Let r_a and r_b be two distinct regions in the geographic space R and let $\mathcal{M}_t$ be a matrix of order $|R|$ X $|R|$. Each element in the matrix $\mathcal{M}_t$ represents the number of interactions between two specific regions during an epidemic such that the number of interactions originating in region r_a to region r_b are stored in $\mathcal{M}_t(r_a, r_b)$. Following the methodology described in the previous section, once an individual from region r_a transmits the disease to an individual in region r_b we consider that the disease progressed from r_a to r_b. To capture the occurrence of such event, we increment the numerical value of $\mathcal{M}_t(r_a, r_b)$ by one. The resulting $\mathcal{M}_t$ is a sample trajectory $t \in T_{P,R,\epsilon}$ of the total possible trajectories $T_{P,R,\epsilon}$ for a specific disease within that population and geographic space.

To analyze the role of each region during the epidemic, the matrix $\mathcal{M}_t$ can be adapted into a complete directed weighted graph $\mathcal{G}(t)$. In this graph, each region of the geographic space $R = \{a, b, .., \zeta\}$ is a vertix such that the set of vertices $V = \{r_a, r_b, .., r_\zeta\}$. Since $\mathcal{G}(t)$ is a complete graph, each node has an edge within all other edges such that the set of ordered pairs $E = \{\{r_a, r_b\}, \{r_a, r_c\}, .., \{r_b, r_a\}, ..., \{r_\zeta, r_\zeta\}\}$. The weight of each edge is given by the number times the initial vertex transmitted the disease to the terminal vertex during the epidemic, e.g. for edge $\{r_a, r_b\}$ its weight is $\mathcal{M}_t(r_a, r_b)$. An example of a $\mathcal{M}_t$ represented as a graph $\mathcal{G}(t)$ is presented in Figure 3.4

Both $\mathcal{M}_t$ and $\mathcal{G}(t)$ are used to analyze the progression of the epidemic in the population by evaluating the interactions every region made with the rest of the population. We obtain two parameters from a resulting trajectory t: $\delta(t, i)$ and $\gamma(t, i)$. For any given matrix $\mathcal{M}_t$ and graph $\mathcal{G}(t)$ in which the regions $i, j = \{1, 2, 3, ..., \zeta\}$:

Region	A	B	C	D
A	2	1	33	0
B	12	1	1	2
C	3	21	14	12
D	0	8	14	0

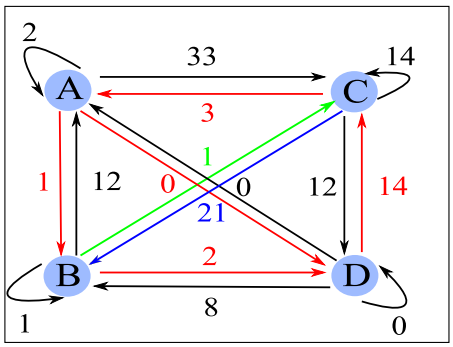

FIGURE 3.4. Example of a $\mathcal{M}_t$ represented as a $\mathcal{G}(t)$

$$\gamma(t,i) = \sum_{j=1}^{|R|} \mathcal{M}_t(r_i, r_j) = deg^+(i) \tag{2}$$

$$\delta(t,j) = \sum_{i=1}^{|R|} \mathcal{M}_t(r_i, r_j) = deg^-(j) \tag{3}$$

such that for a trajectory t, $\gamma(t,i)$ is the total number of interactions initiated by individuals from region r_i or the out-degree $deg^+(i)$ of vertix i. Similarly, $\delta(t,j)$ is the total number of interactions in which individuals from r_j were contacted or the in-degree $deg^-(j)$ of vertix j.

For a given region $r_a \in R$, the value $\delta(t,a)$ represents the degree to which that region is attracting the rest of the population. Further, the set with the number of transmissions attracted by each region $\Delta(t) = [\delta(t,a), \delta(t,b), ..., \delta(t,\zeta)]$ represents the transmission trend of the epidemic for trajectory t. In this research, we use $\Delta(t)$ to study changes in the outbreak dynamics caused by the social/behavioral characteristics of the individuals since this parameters represents the transmission trend of the epidemic. Any changes in appeal or affinity of a region r_a could result in changes in the degree in which a population attracts other regions, thus, resulting in variations of $\delta(t,a)$ and consequentially in $\Delta(t)$.

The value $\gamma(t,a)$ represents the degree of contribution from region r_a towards the progression of the epidemic. The contribution of a region is measured with the number of

transmissions initiated by that region. Moreover, the set with the total number of transmissions initiated by each region is $\Gamma(t) = [\gamma(t,a), \gamma(t,b), ..., \gamma(t,\zeta)]^{\intercal}$. Since $\Gamma(t)$ reflects the force of infection of R, in this research we use this parameter to study changes in outbreak dynamics caused by physiological characteristics of the population and by the interaction coefficient and geographic distribution of individuals. As it will be detailed in Chapter 5, the physiological characteristics of an individual can affect the duration it is capable of transmitting the disease. As a result of those variations, the contribution of individuals to the epidemic is altered. Similarly, variations on the interaction space on which an individual is capable of transmitting the disease impact its contribution to the epidemic. Ultimately, both of these factors affect $\Gamma(t)$.

In this model, a disease ε spreads within a group of individuals P in a geographic space R. For a given combination of those three characteristics, the population of the properties of all possible trajectories is represented with $\tau(P,G,\varepsilon)$ and $\Upsilon(P,G,\varepsilon)$. The set $\tau(P,G,\varepsilon)$ consists of all possible sets $\Gamma(t)$ that result from an specific value for ε, P, and G. Similarly, the set $\Upsilon(P,G,\varepsilon)$ incorporates all feasible combinations of $\Delta(t)$ for the a given set of values for disease, population, and geographic space. In general, considering that $\tau(P,G,\varepsilon)$ and $\Upsilon(P,G,\varepsilon)$ are the population of all possible combinations for the properties of the disease trajectory for a given set of values for disease, population, and geographic space, $\Gamma(t)$ and $\Delta(t)$ are a sample trajectory from their respective sets for a given P, G, and ε. In a given experiment, both $\Delta(t)$ and $\Gamma(t)$ are the result from the same experiment; however, each value is analyzed independently. A naive approach to capture all possible $\tau(P,G,\varepsilon)$ and $\Upsilon(P,G,\varepsilon)$ for a P, ε, and G requires all possible trajectories. However, this is not a feasible approach since computing all those values is an stochastic process that would require a simulation that considers all random sequences. Thus, we require a different method to accurately capture those values without losing precision. This method captures the distribution of $\tau(P,G,\varepsilon)$ and $\Upsilon(P,G,\varepsilon)$ by the use of sampling statistics.

In general, for a given trajectory s, $\Gamma(s)$ and $\Delta(s)$ are a sample, respectively, of $\tau(P,G,\varepsilon)$ and $\Upsilon(P,G,\varepsilon)$ for a particular disease ε within population P and geographic

space G. To acquire the distribution of $\tau(P, G, \varepsilon)$ we capture $|E|$ number of out-degrees using the same set of parameters ε, P,G but with different random seeds. Each experiment $e_i \in E$ outputs a sample $\Gamma(i)$ such that its arithmetic mean $\bar{x}_i$ is an approximation to the population mean μ with a standard error[88]. The same process is used to obtain the population distribution of $\Upsilon(P, G, \varepsilon)$.

Figure 3.5 depicts three approaches to capture the distribution of 500 $\Gamma(i)$ for a particular disease ε within population P and geographic space G. In this figure, each line is the density curve of one $\Gamma(i)$. The density curve depicts the probability an out-degree takes in an interval [a,b] for any two number a and b with $a \leq b$. Additionally, the vertical dashed line superimposed in the curves displays the arithmetic mean of the distribution. For this set of experiments $|P| = 27000$, such that 300 individuals reside in each region. The geographic space G consists of 100 regions from which 10% of them have a population of zero individuals. Any unpopulated region has an out-degree and in-degree of zero such that those areas have no individuals to infect and, consequently, nobody could transmit the disease. Each experiment starts with the insertion of a random initial infectious individual in one of the regions. The rest of the simulation follows the methodology previously presented in 3.1. Due the random nature of the simulation some experiments may result in "no-outbreaks". A no-outbreak is an experiment that does not reflect the behavior of an epidemic. For this model, we consider a no-outbreak as an experiment in which less than 10% of the population is infected.

Figure 3.5.a depicts an approach that includes all $\Gamma(i)$ including those that represent no-outbreaks. We observe that a no-outbreak simulation results in out-degrees that converge close to zero. Since we are only interested in studying experiments that mimic an epidemic, we disregard those simulations. Figure 3.5.b depicts the distribution of $\Gamma(i)$ disregarding out-degrees that are extracted from experiments that resulted in no-outbreaks. We can observe in the figure, that regions that have zero individuals greatly influence the value of the mean. To remove that disturbance, we disregard all out-degree values that are smaller than 10. The result is presented in Figure 3.5.c. The data presented in Figure 3.5.c is

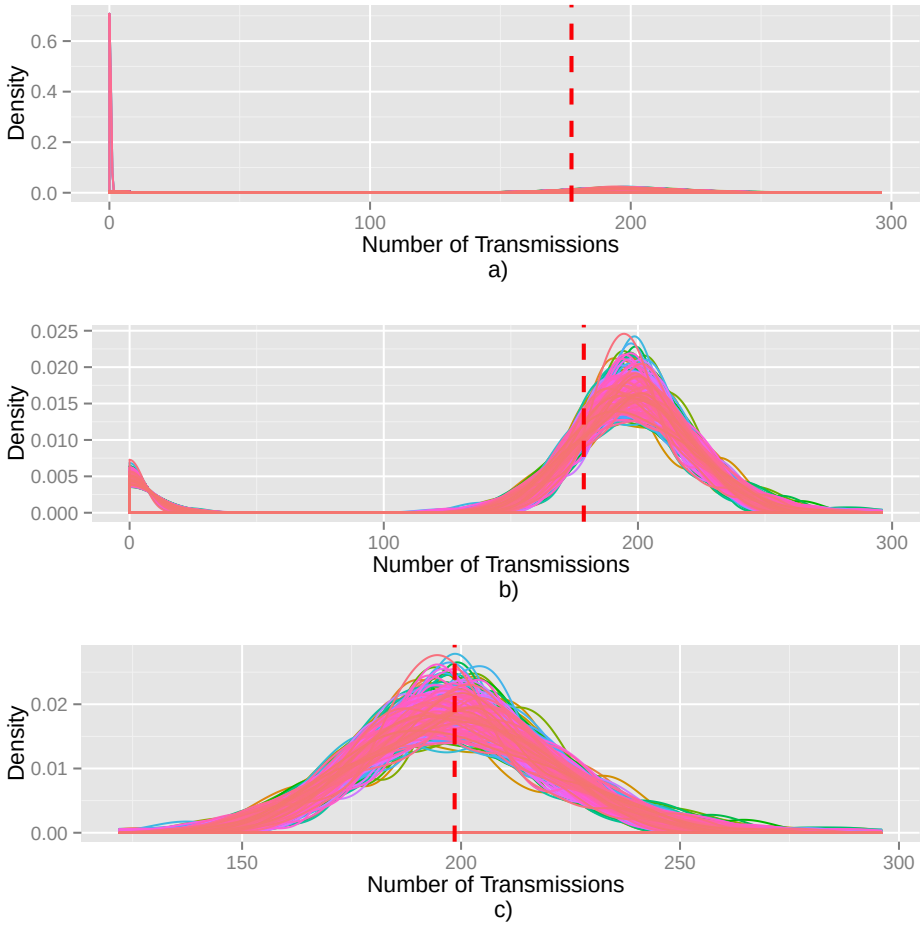

FIGURE 3.5. Distribution of 500 $\Delta(t)$ from $\tau(P, G, \varepsilon)$

subset $\pi(i) \subset \tau(P, G, \varepsilon)$. Since $\tau(P, G, \varepsilon)$ is the collection of values that represent the force of infection, we determine that $\pi(i)$ represents the properties of the force of infection of an epidemic trajectory. Similarly, the subset $\eta(i) \subset \Upsilon(P, G, \varepsilon)$ is obtained following the methodology used for $\tau(P, G, \varepsilon)$. Furthermore, given the characteristics represented by $\Upsilon(P, G, \varepsilon)$, the subset $\eta(i)$ represents the properties of the transmission trend of an epidemic trajectory.

Further, the distribution of the sample is obtained by capturing multiple sample means and, if $|E| \geq 30$, the distribution can be approximated to a normal distribution in accordance with the central limit theorem [41]. Figure 3.6 depicts the sample distribution of 100 sample means obtained from samples of size 100 from the same population of 500 $\Gamma(i)$

depicted in Figure 3.5.c. In the figure we observe that the mean of the sampling distribution of the mean is the mean of the population of the data sets. We are interested in comparing the population means of two data sets that originate from experiments that have differences in intrinsic characteristics between them. In the following section we describe a methodology to compare the properties of two trajectories.

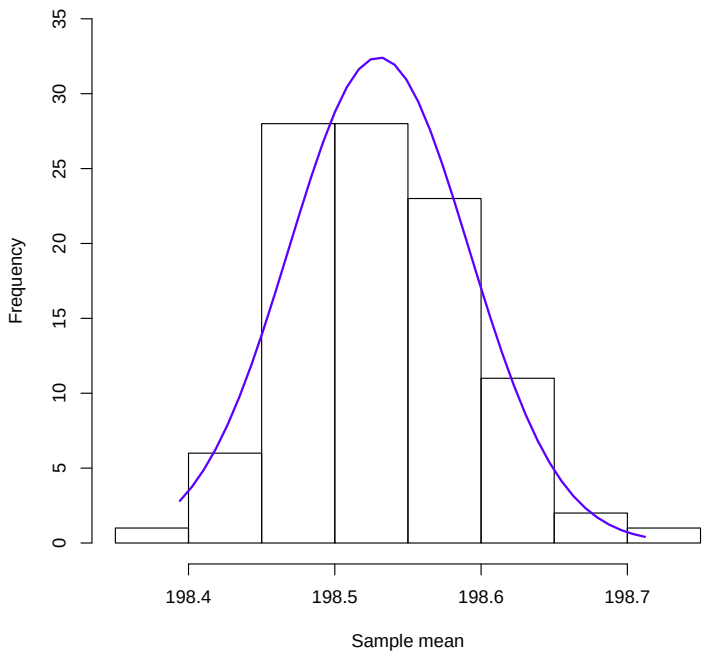

FIGURE 3.6. Distribution of the sample mean for 100 samples

3.3. Comparing the Properties of Two Epidemic Trajectories

We are interested in measuring the differences caused in the properties of the trajectories of two epidemics after modifying some of the characteristics that determine the epidemic trajectory in one of the epidemics. As described previously, the characteristics that determine the epidemic trajectory are: disease ε, characteristics of population P, and shape of the geographic space G. The methodology we present in this section refers to $\tau(P, G, \varepsilon)$, however, the same process is required to study differences in $\Upsilon(P, G, \varepsilon)$. From $\tau(P, G, \varepsilon)$ we are interested in its subset $\pi(t)$. From each $e_i \in E$ we calculate the arithmetic mean $\bar{x}_i$ of the

resulting subset $\pi(t)$ of that experiment. All arithmetic means from a group of experiments E for a trajectory t are contained in $X(E,t)$. Further, since, as previously described, $\pi(t)$ represents the properties of the force of infection of the epidemic trajectory, $\bar{x_i}$ portrays the mean force of infection for that experiment and $\overline{X(E,t)}$ represents the mean force of infection (MFI) of the trajectory. We are interested in the arithmetic mean since each sample mean is an approximation to the population mean. For a given trajectory, the population mean is the measurement of central tendency for the properties of the epidemic trajectory. Hence, the sample mean $\bar{x_i}$ is a sample of the properties for the epidemic trajectory in experiment e_i. Considering the relevance of the population distribution of sample means, significant variations in the distributions of two populations highlight the importance of analyzing the sources that caused the fluctuation. In general, if two populations are statistically similar after incorporating changes in one of them then, basically, it is unnecessary to incorporate the characteristics initially contemplated.

The methods used to measure the statistical variations are: Welch's t-test and Levene's test. Both methods are used to evaluate whether two unrelated groups are significantly different. The statistical significance represents the probability that an effect in a group is not caused by randomness [111]. To determine similarity, the null hypothesis, H_0, must be accepted with a level of significance. The null hypothesis is a statement that indicates no change between populations. The level of significance α is the probability of wrongly rejecting a null hypothesis H_0, if it is in fact true. The level of significance used in these experiments was set to 95% or $\alpha = .05$. This indicates that the result of a test was not due chance or randomness with a confidence of 95%.

The two tailed Welch t-test evaluates whether the population mean of two unrelated groups is significantly different. This test is an adaptation of the t-test for which homogeneity of variance of the samples is not required. The Welch t-test assumes independence and normal distribution of the data to compute accurate results. In our model, independence is established since each experiment represents a unique event that results from a simulation with a random seed. For two simulations, A and B, the computation of an event in A does

not change the probabilities of an event to occur in B. In general, a resulting $\tau(P_A, G_A, \varepsilon_A)$ does not affect the probability distribution of a $\tau(P_B, G_B, \varepsilon_B)$. Further, since $\pi(A)$ and $\pi(B)$ are the result of a simulation of specific $\tau(P_A, G_A, \varepsilon_A)$ and $\tau(P_B, G_B, \varepsilon_B)$, we conclude that the data is independent. In a Welch t-test, any violation to the assumption of normality results in little effects to the test given the robustness of the test[111]. Additionally, the test is relatively insensitive to any variations of normality since, in our tests, the size of each sample is greater than 30 and the sample sets are of the same size [74]. In general, the size of the samples is considered equal if the size of the larger group is not more than $1\frac{1}{2}$ times larger than the size of the smaller group[67]. The hypotheses for the Welch t-test are:

- H_0: The two samples have the same population mean ($\mu_1 = \mu_2$).
- H_a: The two samples do not have the same population mean ($\mu_1 \neq \mu_2$).

This test requires the calculation of a test statistic (t_w) to compare with the t distribution. The test statistic t_w is:

$$t_w = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{4}$$

Such that $\bar{x_i}$, s_i^2 ,and n_i are the sample mean, sample standard deviation and the sample size for sample i. To determine the t distribution used to compare with t_w, we calculate the degrees of freedom (v). In a Welch t-test the value of v is:

$$v = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{1}{n_1 - 1}(\frac{s_1^2}{n_1})^2 + \frac{1}{n_2 - 1}(\frac{s_2^2}{n_2})^2} \tag{5}$$

Typically v is not an integer and computer software is required to find the critical value $t(\frac{\alpha}{2}, v)$. The critical value is a boundary between the acceptance region and the rejection region for a given test. The value of α is divided in half since this is a two-tailed test. The critical value is calculated with the use of computer software and is used to determine the validity of a hypothesis. Additionally, the critical value is used calculate the confidence interval CI. The confidence interval shows the range in which the true mean is likely situated.

Furthermore, the width of the confidence interval determines the precision of the result. Obtaining a large width between the lower and upper bound indicates that the sample set is small and more samples are required [111]. The CI for a Welch t-test is:

$$CI = \bar{x_1} - \bar{x_2} \pm t(\frac{\alpha 2}{2}, v) * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{6}$$

Finally, a p−value is calculated with the software R using the statistic t_w and the degrees of freedom (v) as input. The H_0 is accepted if $p > \alpha$, or if the value 0 is within the limits of the confidence interval. If the H_0 is rejected, we assume, with a given confidence, that the differences on the population means were caused by the modifications incorporated to one of the populations and not by randomness. In the context of $\pi(t)$, if H_0 is rejected, we conclude that the variation in the force of infection is statistically significant. This indicates that the magnitude of the force of infection of the regions transmitting the disease or the number of transmitting regions has fluctuated between populations. This change occurs due to variations in the physiological characteristics of the population and in the interaction coefficient and geographic distribution of the individuals. Contrarily, accepting H_0 indicates that the force of infection of both populations is statistically similar. Further, since $\eta(i)$ represents the properties of the transmission trend, if H_0 is rejected, we conclude that the fluctuation in the transmission trend is statistically significant. This suggests that the affinity among regions during disease transmission of both populations was distinct. In general, this indicates that for one of the populations the transmissions attracted from its regions have fluctuated compared the other population. This variation is caused by changes in social/behavioral characteristics of the individuals. Otherwise, accepting H_0 indicates that the transmission trend of both populations is statistically similar.

Assessing fluctuations in the variances of two populations is another measure of statistical variation. Levene's test evaluates whether the variance of two samples come from populations with equal variances. We selected Levene's test given its robustness and low sensitivity to normality. Levene's test performs a two-sample t-test on the deviations of each

data point from its mean or median. The use of the median is preferred when the normality of the data is not guaranteed [111]. The hypotheses for this test are:

- H_0: The two samples have the same variance ($\sigma_1 = \sigma_2$).
- H_a: The two samples do not have the same variance ($\sigma_1 \neq \sigma_2$).

The Levene test requires the calculation of a test statistic, t_l, to compare with the t distribution. This statistic is calculated from modified samples estimated by transforming the original data. In the modified samples, Y_1 and Y_2, every data point from the original samples, X_1 and X_2, is subtracted the value of the median of that sample such that for every element i of sample s with median $\tilde{X}_s$, its value $y_{i,s}$ is:

$$y_{i,s} = |x_{i,s} - \tilde{X}_s| \tag{7}$$

The Levene test requires the sample sizes n_1 and n_2 to be even. If the sample sizes are odd, one $y_{i,s} = 0$ must be removed from the modified samples. Once Y_1 and Y_2 have been calculated, we use a two sided t-test on those new samples to test the hypothesis. The two sided t-test formula to obtain the test statistic t_l is:

$$t_l = \frac{\bar{Y_1} - \bar{Y_2}}{S_{Y_1 - Y_2}} \tag{8}$$

in which $\bar{Y_1}$ and $\bar{Y_2}$ are the sample means. Additionally, the standard error of the difference between the means, $S_{Y_1 - Y_2}$, is:

$$S_{Y_1 - Y_2} = \frac{s_p^2}{n_1} + \frac{s_p^2}{n_2} \tag{9}$$

such that n_1 and n_2 are the sample sizes of Y_1 and Y_2 and the value of s_p^2 is:

$$s_p^2 = \frac{SS_1 + SS_2}{v_1 + v_2} \tag{10}$$

in which $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ are the degrees of freedom of each sample, and SS_1 and SS_2 are the sum of squares of each sample. Finally, the confidence interval (CI)

is calculated using the same method as in the Welch test. First, we calculate the degrees of freedom (v) using Equation 5. The value v is used as input in the computer software to calculate the critical value $t(\frac{\alpha}{2}, v)$. Subsequently, the confidence interval is calculated using the critical value as described in Equation 6.

Similarly to the Welch test, a p−value is calculated in R using the statistic t_l and the degrees of freedom (v) as input. The H_0 is rejected if $p \leq \alpha$, or if the value 0 is not within the confidence interval. Whenever H_0 is rejected, the differences in the sample variances are unlikely to have occurred based on randomness. Further, when H_0 is rejected while comparing the variance of the forces of infection $\pi(t)$ of two populations, we establish that there is a significant variation in the degree of contribution from the regions towards the epidemic between populations. Dissimilarity in the variance indicates that the distribution of the force of infection among the participating regions fluctuated within the populations. Likewise, whenever two populations share the same variance, H_0 is accepted, we assume that their distribution of the force of infection was similar. Furthermore, when H_0 is rejected when comparing the variance of the transmission trend $\eta(i)$, we conclude a significant variation on the degree of spread from the regions that attract the transmissions during the epidemic. In general, this difference in the variances indicates a fluctuation in the distribution of the transmissions attracted by the regions within the populations. Contrarily, whenever H_0 is accepted indicates similarity in the distribution of the transmission trend for both populations.

To illustrate the concepts introduced in this chapter, the next section depicts a simulation of the model with an implementation of the methodology to compare the properties of two epidemic trajectories.

3.4. Experiments

Two experiments were conducted to examine the validity of the methodology to compare the properties of two epidemic trajectories. In the first experiment we analyze a base case for which P, G, and ε for both populations have the same characteristics. The base case highlights the correctness of the methodology whenever the populations are expected to be

practically equal. In the second experiment we modify the interaction coefficient of G from one of the populations while maintaining ε and P with the same values as the other population. This experiment is used to determine the capacity of the method to capture fluctuations within the populations. Both experiments are conducted using a perl implementation of a modified global stochastic contact framework (GSCF). In the GSCF, as defined in Chapter 6, each individual is assigned specific characteristics based on the demographic information from the U.S. census. Additionally, the demographic information is used to determine the geographic location in which the individuals are located. Contrarily, in the modified implementation, all of the individuals are assigned the same demographic characteristics and they are placed in the geographic space following a uniform distribution. These adjustments are included considering that the objective of these experiments is to analyze the nature of the methodology and, removing the demographic and geographic characteristics, eliminates biases.

A simulation of the GSCF is executed in timesteps, such that one timestep represents a period of 24 hours. During the simulation, at each timestep all infectious regions randomly select a number of regions to interact and possibly transmit the disease. The number of transmissions a region commences is related to the number of infectious individuals in that region. This value is calculated based on the contact rate, such that each infectious individual contributes to the number of infectious contacts originating from its region. The method to calculate the number of transmissions per region is formally defined in section 6.2. Once a contact has been established, the probability of a successful transmission of the pathogen is based on the transsmissibility (β) of the disease ε. The value of β for this simulation is set to 0.015.

A simulation of this implementation commences with the creation of a geographic space. The geographic space G consists of an area with a 100 regions from which 10% of them have a population of zero individuals. After the creation of the geographic space, the population and its characteristics are delineated. For this set of experiments $|P| = 27000$ and the characteristics of all individuals are set equal among the population. After defining

the population, the next step in the simulation is the placement of individuals in different locations of the geographic space. Individuals are placed in each region such that each location has an average of 3000 individuals. After the placement process, we randomly select one individual as the initial infectious case. The region that hosts that individual initiates the epidemic by directing its contacts to at least one destination. The selection of the destination region is determined by the interaction coefficient. The interaction coefficient is defined in section 6.2.3 as the likelihood of a region to interact with other areas. As defined in section 6.2.3, the likelihood for two regions to interact is calculated based on their population size and the distance between them. In this implementation, we establish the distance among regions to be equal. The incorporation of differentiated distances will impact the likelihood of disease spread, and thus, produce a bias in the epidemic.

Once a region has been selected as destination, the probability of a successful transmission of disease after an interaction between two regions is determined by conducting a random experiment. If the transmission is successful, we randomly select an individual from the destination region to become latent. This individual remains latent based on the duration of its latent period (LP). Once its latent period ends, the individual becomes infectious and starts contributing to the number of infectious contacts initiated by its region. An individual can contribute to the number of infectious contacts of its region as long as it is infectious. The length of that period is determined by the infectious period (IP). Once its IP ends, the individual becomes recovered and cannot be infectious anymore. The simulation ends whenever the number of infectious individuals reaches zero. In general, for each experiment we conducted $n = 500$ different simulations for each population. Additionally, all simulations that resulted in less than 10 percent of the population getting infected were discarded. Ultimately, the output of an experiment are the subsets $\pi(i)$ and $\eta(i)$ of each trajectory. Following, we present the results of two experiments that compare the properties $\pi(i)$ of the epidemic trajectory of two populations.

3.4.1. Experiment I

In Experiment I, we introduced two populations for which their values for P, G, and ε are the same. By setting the characteristics of the epidemic to be equal for both simulations, we expect that each resulting trajectory would be a sample of the same population. Further, if the trajectories are from the same population, the means and variations of both samples must be statistically similar. We define this experiment as the base case and its goal is to highlight the correctness of the methodology to compare the properties of two similar epidemic trajectories.

A summary of the data used in this experiment is presented in Table 3.1. As described in Section 3.2, the data presented in this table is calculated using the interactions of all the regions in the population. Due the random nature of the simulation, the interactions are distinct for every experiment. However, in the table we observe that the statistics of both populations are almost identical. To test the similarity of the distribution of the data of both samples we performed a Welch t-test and a Levene test.

TABLE 3.1. Summary of data used in Experiment I

Sample	n	MFI	Std. deviation	Std. error mean
1	492	198.5422	0.6633	0.0299
2	494	198.4932	0.6558	0.0295

To test the similarity of the mean force of infection (MFI) between the samples of each region, we conducted a Welch t-test with the use of the function "t.test" in R[82]. The input for the function were the subsets $\pi(i)$ from each simulation. The results are presented in table 3.2.

TABLE 3.2. Results from the Welch t-test in Experiment I

t	df	p	95 % confidence interval	
			Lower	Upper
1.1647	983.766	0.2444	-.0335	0.1313

By analyzing the results from table 3.2, we conclude that there was no significant difference between the MFIs of sample 1 (MFI=198.5422, SD=0.6633) and sample 2 (MFI=198.4932, SD=0.6558). The H_0 is accepted since $p = 0.2444 > \alpha = .05$. To test the consistency of this result we repeated this test for multiple samples of this population with similar results of acceptance of H_0 (data not shown).

We conducted a Levene test to determine the similarity in variances and, as described previously, the test required the conversion of the original data. The conversion was performed following equation 7. Table 3.3 depicts the data summary for the modified samples. Similarly to the data in table 3.1, the statistics for both samples are very similar.

TABLE 3.3. Summary of data used in Experiment I for the Levene test

Sample	n	MMFI	Std. deviation	Std. error mean
1	492	0.5258	0.4037	0.0182
2	494	0.5271	0.3900	0.0175

The Levene test was conducted with an implementation in R of the method described in section 3.3. In this implementation, the calculation of the test statistic is obtained with a two-sided t-test on the modified data. The summary of the data resulting from the conversion, including the modified mean force of infection (MMFI), are presented in table 3.4.

TABLE 3.4. Results from the Levene test in Experiment I

t_l	df	p	95 % confidence interval	
			Lower	Upper
-0.0528	984	0.9579	-0.0509	0.0482

From the results depicted in table 3.4, we conclude that there was no significant difference between the variances of sample 1 (MMFI=0.5258, SD=0.4037) and sample 2 (MMFI=0.5271, SD=0.3900). The H_0 is accepted given that $p = 0.9579 > \alpha = 0.05$. We also observe that the test statistic t_l is negative. This result occurred since the MMFI of the second sample is greater than the MMFI of the first sample. Despite the negative value, the

result of the test is not impacted. In general, the acceptance of the H_0 indicates that the degree of contribution towards the progression of the disease is similar for both samples.

In general, since the means and variances of both samples are statistically equivalent, we can conclude that there is strong evidence that both samples were extracted from the same population. Additionally, given the small width in the confidence intervals for both tests, we estimate that our results are precise. Furthermore, this experiment highlights the correctness of the method to compare the properties of two trajectories that result from an epidemic with similar characteristics.

3.4.2. Experiment II

In the second experiment, we simulate the spread of an epidemic in two distinct populations. The first population has the same epidemic characteristics as those from Experiment I. In the second population one of the characteristics of G is modified, while maintaining P, and ε the same as the first population. The difference incorporated in G is the modification of the interaction space of the regions. As defined previously, the interaction space between two regions is determined by the interaction coefficient. Further, the interaction coefficient of two regions is a probability calculated based on their population sizes and the distance between them. In the first experiment we assumed that the distances between all regions were equal. However, in this experiment the distance between regions is calculated using the euclidean metric. The euclidean distance between two regions is the length of the line segment connecting them. The metric in a two dimensional space (x, y) is calculated with the formula 11 [31].

$$D_{xy} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \tag{11}$$

By introducing the euclidean distance as a component of the interaction coefficient, we expect to observe a difference in the properties of the trajectory of the disease. The potential difference was measured by performing a Welch t-test and a Levene test on the data points. A summary of the data used for this experiment is presented in Table 3.5. In the

table we observe a variation on the central tendency parameters. However, this variations need to be statistically measured to reach any conclusions.

TABLE 3.5. Summary of data used in Experiment II

Sample	n	MFI	Std. deviation	Std. error mean
1	492	198.5422	0.6633	0.0299
2	455	223.1733	0.2015	0.0094

The Welch t-test was conducted using the same approach as in Experiment I. The results obtained from this calculation are presented in Table 3.6.

TABLE 3.6. Results from the Welch t-test in Experiment II

t	df	p	95 % confidence interval	
			Lower	Upper
-785.3409	587.563	2.2×10^{-16}	-24.6926	-24.5695

From the results depicted in table 3.6, we conclude that there was a significant difference between the MFIs of sample 1 (MFI=198.5422, SD=0.6633) and sample 2 (MFI=223.1733, SD=0.2015). The rejection of the H_0 is determined since $p = 2.2 \times 10^{-16} < \alpha = 0.05$. The difference in the MFI indicates that the distribution of the first population is different than the distribution of the second population. Considering that the MFI of each population is the central tendency for the force of infection, we can conclude that there was a variation of this parameter during the progression of the disease. More importantly, this result indicates that our methodology is capable of capturing that variation.

TABLE 3.7. Summary of data used in Experiment II for the Levene test

Sample	n	MMFI	Std. deviation	Std. error mean
1	492	0.5258	0.4037	0.0182
2	454	0.1612	0.1268	0.0059

Similarly to Experiment I, we converted the original data to conduct the Levene test. The conversion was performed following equation 7 and Table 3.7 depicts a summary for the modified samples. The results of this test are depicted in Table 3.8.

TABLE 3.8. Results from the Levene test in Experiment II

t_l	df	p	95 % confidence interval	
			Lower	Upper
18.42	944	2.2×10^{-16}	0.3257	0.4034

From Table 3.8, we conclude that there was a significant difference between the variances of sample 1 (MFI=0.5258, SD=0.4037) and sample 2 (MMFI=0.1612, SD=0.0182). The H_0 is rejected since $p = 2.2 \times 10^{-16} < \alpha = 0.05$. As previously described, variations in the variances between populations represent a variation in the degree of contribution from the regions towards the epidemic. This test does not indicate the magnitude or direction of the variation, however, this difference reveals fluctuations between the properties of the epidemic trajectories.

3.5. Summary

In this section, we introduced the concept of epidemic trajectory and studied some of its properties. An epidemic trajectory is defined as a path, with multiple branches, of a contagion process that spreads in a fuzzy neighborhood. Further, we considered that the properties that define the characteristics of that path are the force of infection and the transmission trend. The force of infection determines the degree on which individuals participated in the transmission of the disease. Similarly, the transmission trend depicts the appeal of the individuals from each area during the epidemic.

Both of these properties reflect important characteristics of the spread of the disease within a population. To analyze those properties, we developed a model that simulates the transmission of the disease in a population. In this model, the disease spreads at the regional level, such that the transmission of disease from a region A to region B is represented by a directed edge between A and B. This approach allows to represent all transmissions as

a complete directed weighted graph. From this graph, we calculate the force of infection of the epidemic by analyzing the out-degrees of each region. Similarly, we determine the transmission trend of the trajectory by evaluating the in-degrees.

Additionally, we developed a method to compare two epidemic trajectories in two populations. This method determines statistical significance of the variations in the properties of the trajectory of two epidemics. This methodology measures the variations of the means of the samples with the use of the Welch t-test and the variations in the variance with the Levene test. These tests are used to determine the statistical significance of the incorporating different variations in any of the characteristics of the epidemics. If the statistical tests determine statistic similarity in the characteristics of two epidemic trajectories, we conclude that there is strong evidence that both samples were extracted from the same trajectory. This result discourages the incorporation of the parameter that we expected would cause a significant variation in the epidemic. Contrarily, the rejection of the similarity by the test statistics, indicates that the variations were not caused by randomness and further investigation on the parameter that caused that fluctuation must be conducted.

Finally, we conducted two experiments to highlight the correctness of this method. In both experiments, we simulated two outbreaks and obtained the parameter that represents their force of infection as a result of each epidemic. In the first experiment, we compared two samples that were obtained by simulating an epidemic with the same epidemic characteristics. This experiment aimed to capture the efficiency of the methodology when the populations were practically equal. The results of this experiment indicated that both samples were obtained from the same distribution. This result implies that the method is capable of determining that two properties of epidemic trajectories are equal when they both are obtained from the same population. In the second experiment we introduced a variation in the epidemic characteristics of one simulation. This variation is used to measure the capacity of the method to capture fluctuations within the populations. The statistical tests used to measure the variation highlighted the statistically significant fluctuations in the force of infection between populations. These variations dictate that the changes introduced to the outbreak

simulation are responsible for the fluctuations and not the random nature of the experiment. In general, these experiments are a strong indication that the methodology is capable of revealing fluctuations between the properties of epidemic trajectories. Further, since the fluctuations are caused by different intrinsic components of the epidemic, this methodology could be used to highlight the importance of incorporating specific characteristics into an epidemiological study.

CHAPTER 4

MODELING SOCIAL/BEHAVIORAL CHARACTERISTICS *

Mathematical and computational models for different types of infectious diseases, including vector-borne and airborne transmitted illnesses, are based on the concept of interactions or contacts. For person-to-person transmitted diseases, a contact can be thought as an abstraction of human interaction, including attendance at the same social event, the same study groups or any other situation that could involve a close encounter between two individuals. Such relationships are developed in accordance to the structure of the social networks to which the individual belongs. Chances and frequencies of interactions between individuals will change according to the proximity of their social networks, their association and affinities [34],[44].

Early mathematical and computational models consider interactions between elements to be homogeneous, implying that all of the elements have the same likelihood to contact others and to be contacted. Real life behavior, however, indicates that this is not necessarily true; People create social clusters in which the chances of contacting others change according to affinities, geographic location, social position, etc.

A computational simulator was developed to study the effects of population heterogeneity on the outbreak dynamics during an infectious disease epidemic. The model is capable of representing both heterogeneous behavior of individuals as well as differentiated disease characteristics. The global stochastic contact model (GSCM) represents interactions between individuals in a given population. Every member of the population is assumed to have a constant number of interactions or contacts per time interval, i.e. days, hours, weeks. This number of interactions is referred to as its corresponding contact rate (CR). In the GSCM, the population is represented as an abstraction of individuals. This abstraction was necessary to be able to understand the underlaying processes involved in an infectious pro-

*Parts of this chapter have been previously published in part or in full, from Jorge Reyes-Silveyra, Armin R Mikler, Justin Zhao, and Angel Bravo-Salgado, Modeling infectious outbreaks in non-homogeneous populations, Journal of Biological Systems (2011). Reproduced with permission from the Journal of Biological Systems

cess. By removing the more realistic layers, we were able to quantify the effects the different social and behavioral characteristics in an outbreak.

The population P is represented as a set of individuals $p_i \in P$. The set P is mapped to 4 subsets: susceptible(S), latent(L), infectious(I), and recovered(R), thereby representing the components used in the SLIR. We assume a closed population P for which $S \cup L \cup I \cup R = P$, $S \cap L \cap I \cap R = \emptyset$, and $|P| = N$. The progression of the epidemic is represented by the movement of individuals between compartments during time instances $t_0, t_1, ..., t_\pi$, where π indicates the end of the epidemic. At every timestep t_k, all $p_i \in P$ satisfies their contact rate CR by interacting with individuals in P. Irrespectively, each interaction is a random experiment in which individual p_i chooses an arbitrary element $p_j \in P$. Individual p_i moves from S to L at some instance t_m when a contact with p_j is established and $p_j \in I$ is established and the disease has been transmitted successfully. At this point, p_i is assumed to be latent. Whether or not the disease is transmitted during such interaction is determined by yet another random experiment. This random experiment calculates the chance of infection based on the transmission probability of the disease. Transmission probability is the likelihood that an infectious agent enters p_i and multiply after a contact with p_j to produce infection or disease [37].

Once individual $p_i \in L$, it starts its latent period (LP), where LP is the amount of time necessary for p_i to fully develop capacity to infect other individuals. During this period, p_i is infected but not yet capable of transmitting the disease. Immediately after LP concludes, individual p_i will move from L to I and commence its infectious period IP. IP is the amount of time that p_i is capable of transmitting the disease to others. Upon reaching the end of IP, p_i is assumed to have recovered and hence is moved from I to R. Once $L \cup I = \emptyset$ at time $t = t_\pi$ the epidemic has concluded. As an example, the average infection timeline of the flu virus for an adult [27] is depicted in Figure 4.1.

Even though in the above model LP and IP are constant for all $p_i \in P$, in reality there may be significant variation in the amount of time an individual will spend in those periods [85]. These variations are due to physiological or behavioral characteristics, while

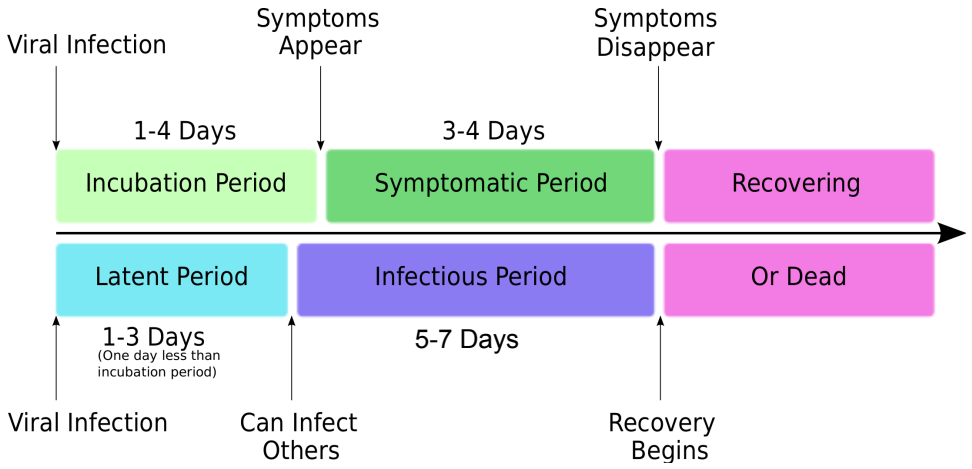

FIGURE 4.1. Infection timeline of influenza in humans

others are related to the amount of pathogen present at the moment of the infection [16]. Implementing this diversity and the causes that produce those distinctions will extend the current model towards representing outbreaks in differentiated populations.

If every individual in P interacts at a constant CR, the naive approach generates a total $(N * CR)/2$ contacts per time frame. At t_π the system has computed a total of $(N * CR)/2 * t_\pi$ random contacts. In a population of just 100,000 individuals with a CR of 40, the system would have to compute 2,000,000 contacts per time frame. The number of calls to the random number generator is a function of the number of contacts generated each time step. Hence in a simulation that takes 100 days we generated 2×10^8 random numbers just for the purpose of contact generation.

As the epidemic is driven by contacts between individuals in S and I, the naive model can be improved by only generating those interactions that involve at least one individual in I. Specifically, we limit the contacts to SI, IS, IL, LI, II, IR, and RI. Even though contacts between IL, LI, II, IR, and RI do not cause a transmission, disregarding them leads to inconsistencies in the contact accounting. This model is shown in Figure 4.2. The number of contacts generated by this new approach is always less than or equal to the number of contacts incurred by the naive model. The novel approach executes $\|I\| \times CR$ contacts per time period. Should all individuals become infected at the same t, the number of contacts

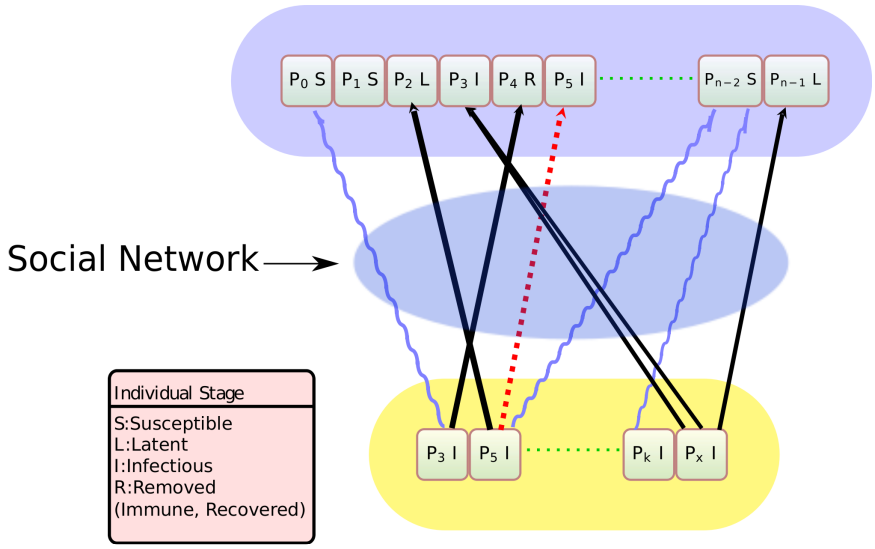

FIGURE 4.2. Contacts between I and P with superimposed social network

is the same as in the naive model. For all the other cases, the number of contacts is smaller.

As depicted in Figure 4.2, the interactions between individuals from I and individuals in P form a social network. This is as a result of $p_i \in I$ selecting its contacts according to its preferences and affinities. This is translated to the computational model by an affinity function $\mathcal{A}(c_i, c_j) = a_{i,j}$, in which c_i, c_j represent two clusters in the population, i.e. chances of an element from cluster i of contacting individuals from clusters i or j. The minimum value of the affinity function between two clusters is 0%. Similarly, the maximum value of the affinity function between two clusters is 100%. This is represented in the equation 12:

$$0 \leq \mathcal{A}(c_i, c_j) = a_{i,j} \leq 1 \tag{12}$$

In the model, the affinity function is represented as an affinity matrix. It consists of probabilities for interactions between different clusters. For M clusters in which $i, j = \{1, 2, .., M\}$,

$$\gamma_i = \sum_{j=1}^{M} \mathcal{A}(c_i, c_j) \; and \; \delta_j = \sum_{i=1}^{M} \mathcal{A}(c_i, c_j) \tag{13}$$

$\gamma_i = 1$ and $\delta_j \geq 0$. In any given affinity matrix, the affinity from a cluster to all clusters always equals 100% ($\gamma_i = 1$). Moreover, the affinity to a cluster from all clusters is equal or

greater than 0% ($\delta_j \geq 0$). An example of an affinity matrix can be observed in Table 4.1.

TABLE 4.1. Affinity matrix

		To				
		A	B	C	D	E
From	A	0.19	0.39	0	0.03	0.39
	B	0.18	0.02	0.31	0.18	0.31
	C	0.04	0.11	0.27	0.39	0.19
	D	0.13	0.18	0.04	0.05	0.6
	E	0.22	0.15	0	0.01	0.62
Affinity's sum		0.76	0.85	0.62	0.66	2.11

Additionally, individuals from each cluster may have different CRs to represent the diverse social skills, age, social status, etc. representing factors that may cause individuals to be involved in varied number of interactions per day. In the following section, we present multiple experiments to explore the effects produced in the outbreak dynamics by incorporating heterogeneous values for contact rates and affinities between regions in a population.

4.1. Experimental Results

Multiple experiments were conducted to analyze the sensitivity of disease dynamics to the differentiated population characteristics and to exemplify the importance of incorporating them into computational models. Experiment I explores the effects of incorporating a heterogeneous CR. A second group (Experiment II, III and IV) highlights the results of integrating differentiated affinity values ($a_{i,j}$). In all experiments, the population P is partitioned into 5 groups: $\Omega = \{A, B, C, D, E\}$. One group was selected as the pivot group ($\varrho \in \Omega$), implying that either CR or $\mathcal{A}$ is varied for a specific experiment. CR_{ϱ} is the contact rate assigned to the pivot group, and $\mathcal{A}(c_i, c_{\varrho})$ is the affinity towards it, where $i \in \Omega$. At all

times we maintain

$$CR_A + CR_B + CR_C + CR_D + CR_E = k \tag{14}$$

where k is constant; k for the series of experiments was set to 200. Additionally, for all experiments presented here, each cluster consists of a closed population of 300 individuals. All experiments start with a random index case inserted into one of the five clusters indistinctly. Note that affinities are not necessarily symmetric. For example, in the affinity matrix presented in Table 4.2 $\mathcal{A}(c_E, c_C) = 0$, however $\mathcal{A}(c_C, c_E) = 0.19$.

TABLE 4.2. Summary of CR used in Experiment I

Case	Contact rate	
	Pivot cluster	Other clusters
Standard control	40	40
Negative control	0	50
Positive control	200	0
T1	180	5
T2	160	10
T3	140	15
T4	120	20
T5	100	25
T6	80	30
T7	60	35
T8	20	45

Three base case experiments were conducted: positive control, negative control, and standard control. In Experiment I the positive control $CR_\varrho = k$, while in the negative control $CR_\varrho = 0$. During Experiment II $\mathcal{A}(c_i, c_\varrho) = 100\%$ for the positive control and inversely $\mathcal{A}(c_i, c_\varrho) = 0\%$ for the negative control. The standard control represents a homogeneous

population by uniformly distributing CR and $a_{i,j}$. Each experiment reports the average over 300 simulations. The experiments capture the following:

- Number of infected individuals of each group at every timestep
- Group membership of the individual that produced each infection.
- Duration of the outbreak.

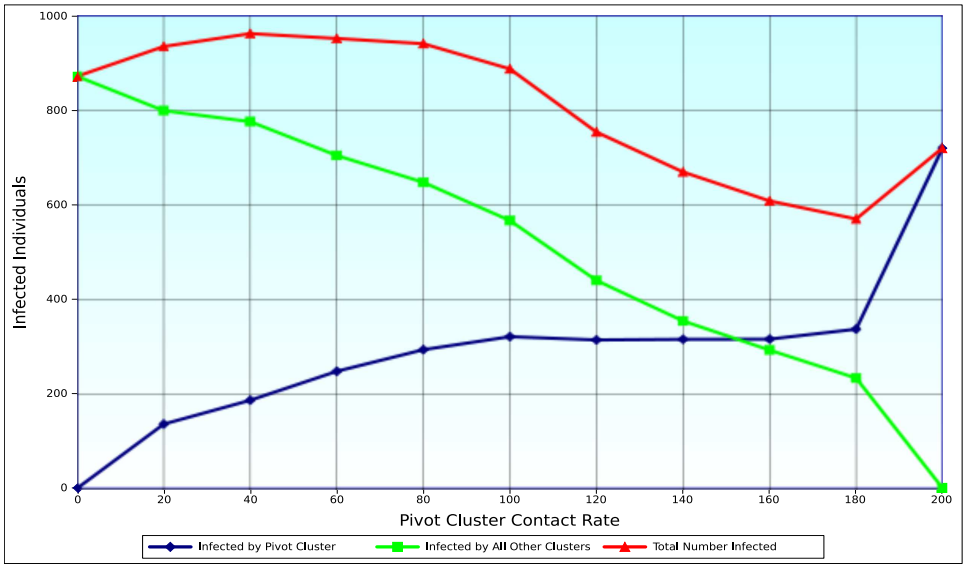

FIGURE 4.3. Infected individuals as a function of the CR of the pivot cluster

In reality, the number of contacts that every individual performs on a daily-basis is not uniform among all individuals of a population. Experiment I was constructed to explore the effects of varied CR on disease dynamics. In each experiment, CR_{ϱ} was incremented by 20 contacts and the rest of the clusters varied their value to maintain the sum of cluster specific CR equal to k. The parameters used in the experiments are shown in Table 4.2. For instance, in case T4: $CR_{\varrho} = 120$, and

$$CR_j = \frac{k - CR_{\varrho}}{\mid \Omega \mid -1} \, \forall j \neq \varrho \tag{15}$$

Figure 4.3 shows the results of this group of experiments. We observe that as CR_{ϱ} increases, so does the number of individuals infected by ϱ. At the same time, the individuals infected by the other clusters decreases. In the figure, the other clusters are represented as a sum.

Figure 4.3 also depicts CR_ϱ and the corresponding total number of infected individuals. The peak occurs whenever $CR_\varrho = CR_j \; \forall j \in \Omega$, causing the largest number of infections. Given the uniform distribution of the infectious individuals in Ω, the likelihood of contacts between individuals of S and I is maximized. As the value of CR_ϱ increases, the possibility for infectious individuals outside ϱ to transmit the disease diminishes.

TABLE 4.3. Affinities used for Experiment II

Cases	Affinity	
	Pivot cluster	Other clusters
Standard control	0.2	0.2
Negative control	0	0.25
Positive control	1	0
T1	0.8	0.05
T2	0.6	0.1
T3	0.4	0.15
T4	0.15	0.2125
T5	0.1	0.225
T6	0.05	0.2375

Experiment II demonstrates the effects of varying affinity from other clusters toward ϱ. This range of values are summarized in Table 4.3. For example in case T4: $\mathcal{A}(c_i, c_\varrho) = 0.15$ and $\mathcal{A}(c_i, c_j) = 0.2125$ for which $i \in \Omega, j \in \Omega - \varrho$. Figure 4.4 indicates that the number of infected individuals from ϱ gradually increases with the affinity towards that group. It is also noticeable that the total number of infected individuals reaches a maximum when $\mathcal{A}(c_i, c_\varrho) = \mathcal{A}(c_i, c_j) \forall j \in \Omega$. This is due to the fact that the distribution of the susceptible individuals is completely homogeneous at that point. Contrary to the experiments in which the affinity to the susceptible individuals is limited by $a_{i,j}$.

In Experiment III random affinities are assigned to each $\mathcal{A}(c_i, c_j)$. This experiment does not incorporate a pivot group neither does the values of $a_{i,j}$ change during the sim-

FIGURE 4.4. Infected individuals as a function of the affinity of the pivot cluster

ulations. It measures the variation in the distribution of infections during an outbreak, due to the alteration of the summed affinity(δ). The values used for this experiment are summarized in Table 4.3. For instance a δ_E of 2.11 represents the total affinity towards group E. The results depicted in Figure 4.5 indicate that the values of δ greatly impact the number of infections. The maximum number of infected individuals occurs in the cluster with the maximum δ (E). Inversely, the minimal number of infections occur in cluster C. Additionally, we observe that clusters C and D display the high sensitivity of the parameter. Notwithstanding $\delta_D - \delta_C = .04$, the infections in cluster D exceed the ones in cluster C.

Experiment IV highlights how incorporating heterogeneous affinity towards a single cluster vary the dynamics of an outbreak. When assigning a homogeneous affinity to all clusters, we observe that the duration and magnitude of outbreaks in each group are indistinguishable. Figure 4.6 depicts the number of infected individuals per day in each group in Ω and in P at large. Due to the similarity in outbreak dynamics in the individual clusters, the outbreak characteristics for the entire population is a scaled up version of each of the small outbreaks. When incorporating heterogeneous affinities towards a group $A \in \Omega$, the outbreak dynamics in cluster A change as a result. In this experiment, the distribu-

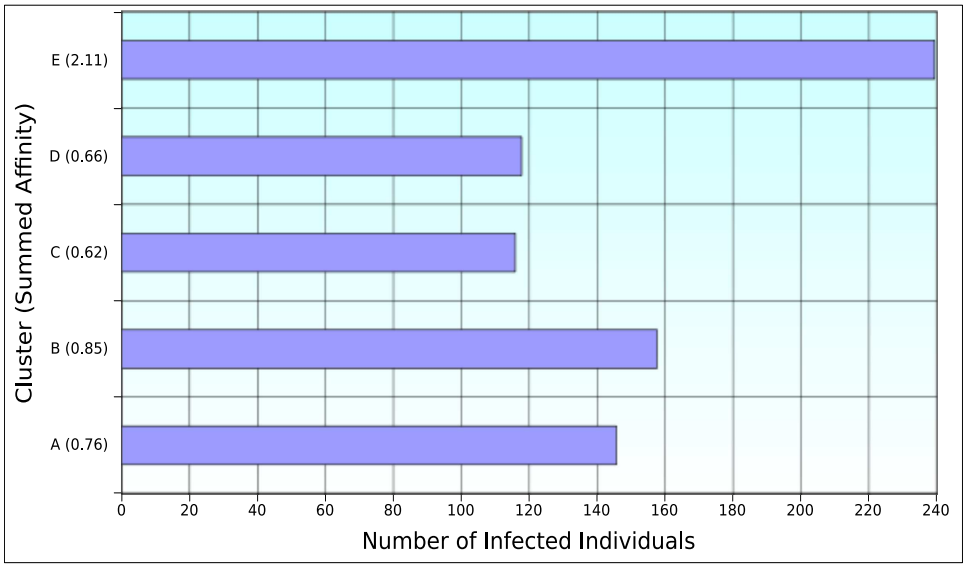

FIGURE 4.5. Infected individuals per cluster as a function of δ

tion of affinities has been modified to $\mathcal{A}(c_i, c_A) = .6$ in which $i \in \Omega$, and $\mathcal{A}(c_i, c_j) = .1$ for $i, j \in \Omega - A$. Figure 4.7 depicts the progression of infected individuals in each group in Ω and P over time. We observe that the $\Omega - A$ and P curves are similar to its counterparts in Figure 4.6, however, at reduced magnitudes. This reduction is due to the behavior limitation described in Experiment II. However, the outbreak dynamics in cluster A are different in both duration and magnitude. This variation is caused by the increase in the affinity towards cluster A, amplifying the possibility of infection of its individuals. The number of infected individuals from A exceeds those of other clusters at all times, causing A to be the most significant contributor for the outbreak observed for P. Consequently, a reduction in the capacity of infection of this cluster can drastically affect the spread of the disease in P.

4.2. Summary

In this section, we introduced an infectious-disease transmission model that is based on the global stochastic contact model. In this model, person-to-person contacts are limited to only those interactions that involve at least one infectious individual. Implicitly, the resulting contact activities create a social network between the infectious individuals and the population at large. Based on this social network, different person-to-person interaction

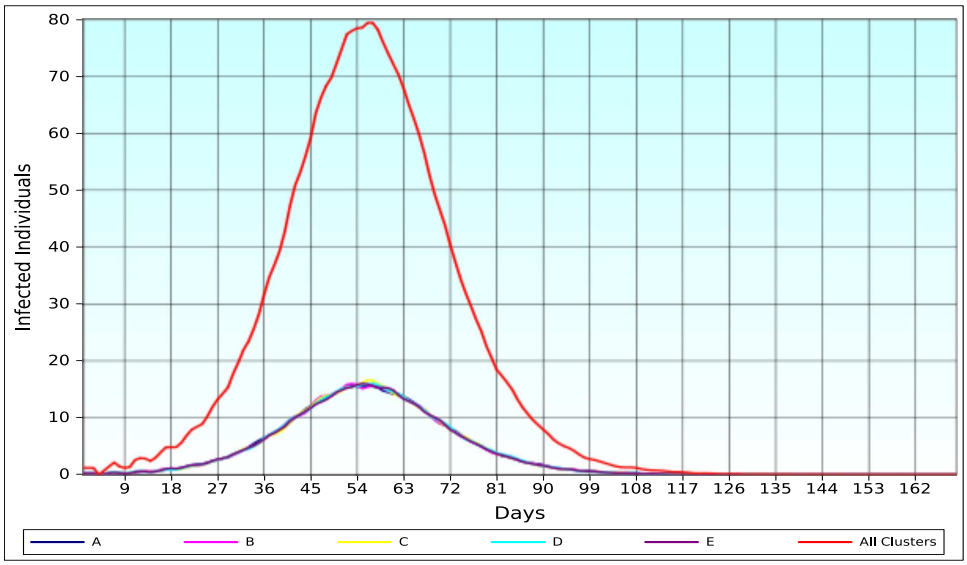

FIGURE 4.6. Infected individuals in Ω and P in a homogeneous population.

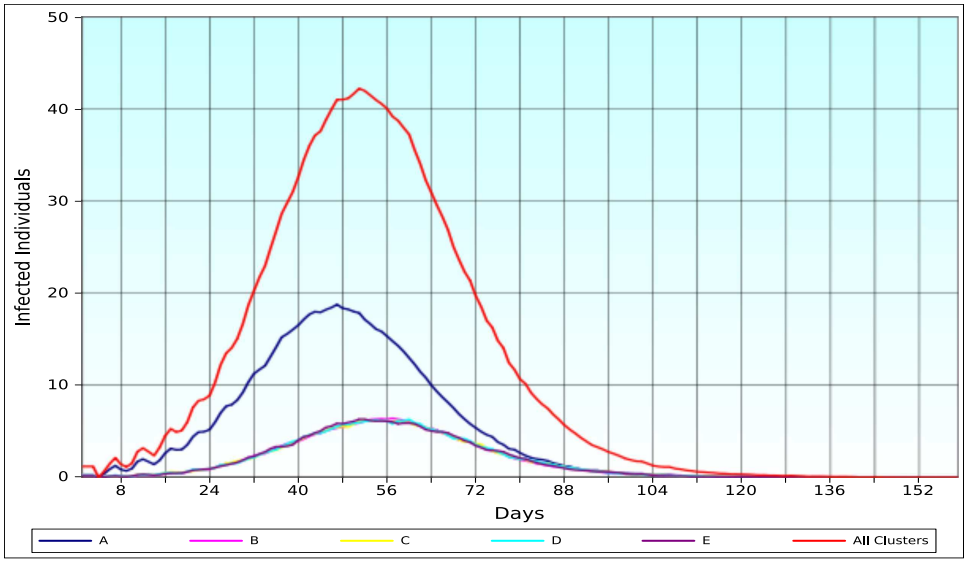

FIGURE 4.7. Infected individuals in Ω and P with heterogeneous affinity towards cluster A

probabilities can be expressed and hence integrated into the model. The model was further expanded to include clusters of individuals representing differentiated social groups. The social network determines the interaction behavior in each group. Group behaviors are expressed by their corresponding inter and intra cluster interaction frequency and affinity. The effects of incorporating differentiated behavior into the model and its relevance to the

outbreak has been assessed.

The results indicate that an increase in the number of contacts in a group of individuals increases their incidence of infection. It was further observed that the most critical outbreak occurs when the number of contacts is homogeneous across all groups. The results also reveal that the likelihood of exposing susceptible individuals to infectious individuals is greatly affected by group affinity. As affinity to a specific group increases, the susceptible individuals from the other groups reduce their chances to be contacted, thus reducing the size of the outbreak. Additionally, experiments exhibited a high sensitivity to affinity, indicating that even small variations result in significant changes in the duration and size of the outbreak. In general, we conclude that the inclusion of heterogeneity and diverse group behavior into an outbreak simulation is important to achieve a more realistic and complete prediction.

CHAPTER 5

MODELING PHYSIOLOGICAL DIFFERENCES

Scientists have proposed many models to represent the spread of infectious diseases in populations. The most recent models attempt to be more realistic by including differentiated populations and non-homogeneous contacts among individuals. Heterogeneity of contacts is obtained by incorporating social networks and social interactions. Non-homogeneity of individuals is modeled by assigning various characteristics such as immunity, age, and gender. In most of the epidemiological models, immunity is incorporated as a value that restricts the capacity to become infected or as a parameter that increases the capacity to infect other individual. In reality, the immune system possesses many complex mechanisms that can be abstracted and incorporated into these models. The immune system's responsibility is to stop the replication and proliferation of pathogens in the host. Age, gender, and fitness level are some of the key factors that determine the efficacy of the immune system. A more competent immune system will decrease the capacity of the invader to replicate. Ultimately, the quality of the immune response will determine the chances of infection and the amount of time an individual is capable of infecting others. Incorporating individual immune responses into a population during an outbreak simulation will affect the disease dynamics. In general, including these modifications will contribute to a better understanding of the progression of the disease in a population.

5.1. Immunity

The survival of an organism is highly correlated with the quality of its immune system [79][75]. The immune system provides two types of defense against invaders: innate and adaptive immunity. The innate response is non-specific to the pathogen and does not provide a long lasting immunity. The major components of the innate response are:

- Physical barriers, which include tears and skin, and whose main objective is to block the entrance of possible invaders.

- The complement system is composed by molecules that intensify the effect of other immune functions.
- Macrophages are responsible to phagocytose, digest and present pathogens.
- Natural killer cells induce cytotoxic apoptosis (cell death) to infected cells.

A pathogenic invasion occurs once viral or bacterial material passes these first lines of defense. Once they have crossed the innate defense, pathogens tend to migrate to suitable locations for occupation and multiplication. Foreign invasion activates an adaptive immune response that impedes the replication and migration of the pathogen to attempt to free the host from the external threat. The adaptive response is specific to each invader and is conducted by two main types of cells:

- T-cells : Lymphocytes maturated in the thymus.
- B-cells : Lymphocytes maturated in the bone marrow.

T-cells are highly specialized cells that not only coordinate (T-helper) and regulate (T-regulatory) the immune response, but also destroy infected cells (T-cytotoxic). B-cells secrete antibodies and perform pathogen presentation similar to the macrophages. Antibodies are proteins that can label an infected cell or a pathogen to facilitate its elimination. Additionally, antibodies can stop the replication of the pathogens by impeding their attachment to healthy cells. Both T-cells and B-cells provide immunity against a pathogen by producing memory cells (T-memory and B-memory) during an infection process. Similarly, immunity can be artificially induced by vaccination. Immunity against a pathogen heightens the immune response to prevent future infections.

The efficacy of the immune response is determined by multiple factors. Many of them are associated with the host, including age, physical fitness, gender, and nutrition [57]. Various studies report different causes for deterioration of the immune system throughout an individual's life-time. As the individual ages, limited capacity to defend against invaders is caused by multiple alterations in T-cell and B-cell functionality [23][55]. Likewise, nutrition is a critical factor for the quality of the immune response of an individual. Studies have shown a strong relationship between malnutrition and multiple immune response deficiencies. These

may include impairment of the complement system, cell mediated immunity, and phagocyte functionality [24]. Obesity [110], high-cholesterol levels [19], and low vitamin and mineral intake [25] are some of the nutrition related causes for those immune response deficiencies.

In addition to access to health care, types and frequency of social interactions, gender drastically affects the competence of the immune response [8][52]. This is primarily attributed to the blood levels of gonadal steroid hormones. Multiple studies have shown androgens as natural immunosuppressors, as well as estrogens as humoral immunity enhancers [28]. Moreover, physical fitness of individuals has a unique effect on the immune response. Although positive immuno-stimulatory activity is observed with moderate exercising, both lack of and excessive exercise produce an immuno-suppressive response [86][48].

5.2. Infection

Diseases in individuals develop in sequential phases [27]. It is known that once a pathogen invades a susceptible individual, he or she migrates through different infection stages: latent, infectious and recovered/removed. The progression from one stage to another occurs after the consummation of different time periods. These periods are known as latent period and infectious period. The latent period is the amount of time necessary for an individual to develop the capacity to infect others. Analogously, the infectious period is the amount of time during which an individual is capable of transmitting the disease to others.

An infectious agent replicates after it penetrates the host's basic defenses. The efficacy of the replication process can be quantified through the corresponding viral/bacterial load (vbl) [69]. The vbl is the concentration of virus or bacteria in plasma at a certain moment in time [47]. The vbl value is commonly used as an indicator for disease severity [29] and the host's capacity of transmitting it [80]. Since the immune system is responsible for controlling vbl in the host, we conceive a direct relationship between them. A stronger immune response restrains the growth of the external threat more efficaciously; Hence, we conclude that the quality of the immune response affects the quantity of vbl during infection.

We define 4 types of immune responses based on their quality: $\mathcal{R}^0$, $\mathcal{R}^+$, $\mathcal{R}^-$, $\mathcal{R}^*$. The quality of each response is determined by the severity of the infection and the length of the

infectious period of an individual with that type of response. A standard immune response $\mathcal{R}^0$ represents the average response in a healthy individual. An individual with this response remains infected for the same amount of time as the majority of the population. Individuals with this response usually recover after the infection is eradicated and rarely succumb to the disease. Individuals with the hyperimmune response $\mathcal{R}^+$ will stay infected for a shorter period of time than the majority of the population. $\mathcal{R}^+$ represent the immune response of those individuals with a superior count of pathogen specific immune cells as compared to individuals with an average immune response. This type of response guarantees the survival of the individual throughout infection. The hypoimmune response $\mathcal{R}^-$ represents the immune response of individuals with an immunocompromised immune system (eg. cancer, diabetes or HIV subjects). A person with this response is infectious for a longer period of time than the majority of the population. Individuals with this response have increased probabilities to succumb to the disease. $\mathcal{R}^*$ comprises individuals that are actively or passively immune against a specific pathogen. Elements of the population with this type of response will get infected for a short period of time, but never become symptomatic or infectious.

To illustrate these concepts and their relationship to vbl, Figure 5.1 portrays four contrasting scenarios of a primary viral infection. $\mathcal{R}^0$ results in consistent pathogen replication until the immune response is strong enough to overcome the infection and eliminate it. $\mathcal{R}^+$ produces a similar effect, but it is more effective than $\mathcal{R}^0$. In contrast, $\mathcal{R}^-$ is not capable of containing the infection. This will result in uncontrolled growth of the virus leading to chronic infection or death of the host. On the opposite side of the spectrum, $\mathcal{R}^*$ is more efficient at limiting pathogen replication than all the other responses. This results in smaller quantity of vbl in the host at all times.

The integration of vbl can be utilized as a threshold to determine the commencement and termination of the disease periods during an infection [10]. Figure 5.2 depicts the duration of the time periods for each scenario presented in Figure 5.1. The length of each period is determined by a transmission threshold. The transmission threshold (vbl^*) is established as the quantity of vbl that is necessary for an individual to become infectious.

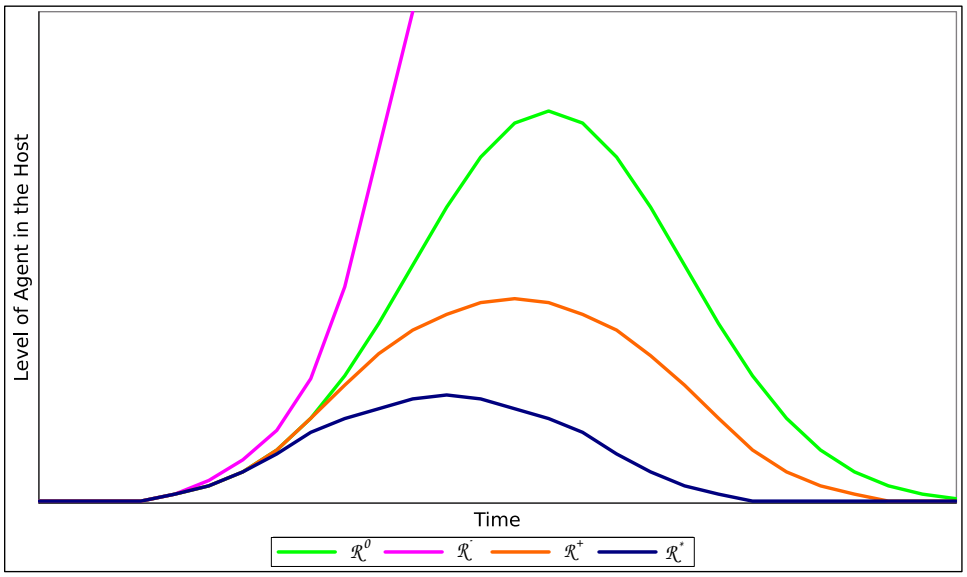

FIGURE 5.1. Viral load in an individual as a function of the quality of its immune response.

The duration of the latent stage comprises from the infection point to vbl surpassing vbl^*. Equivalently, the length of the infectious period starts at the end of the latent period and culminates once the vbl falls below vbl^*. For example, Figure 5.2.b portrays a hypoimmune response $\mathcal{R}^-$ during an infectious process. In this figure, the vbl growth rate at the beginning of infection is similar to the rest of the other responses. Once vbl exceeds vbl^*, the immune response is not capable of containing the pathogen proliferation. An individual with $\mathcal{R}^-$ may experience a long or a short infectious period depending on its capacity to fight the disease. If infection becomes chronic, the individual is infectious period will be longer than that of an individual with the standard immune response $\mathcal{R}^0$. Otherwise, if the individual succumbs to the infection, it will be moved to the recovered/removed population earlier than the average individuals, resulting in a short infectious period. An individual with a hyperimmune response $\mathcal{R}^+$ will have a shorter than average infectious period. Figure 5.2.c represents such a scenario. We observe that vbl in Figure 5.2.c exceeds the threshold for a brief amount of time, resulting in a short infectious period. Finally, since the $\mathcal{R}^*$ response does not exceeds vbl^*, it results in a long latent period and the individual never progresses to infection. This scenario is depicted in Figure 5.2.d.

In general, these scenarios illustrate how the length of the disease periods are a function of the viral load. It is possible to obtain those values from sources such as immunological mathematical or computational models. The model selected must simulate the interaction between the pathogen and the immune system. Ultimately, from the model, vbl is obtained to determine the length of the disease periods for each individual. In Section 5.4, we present an example to determine the infectious periods by simulating the disease trajectory in a host with the use of a mathematical model.

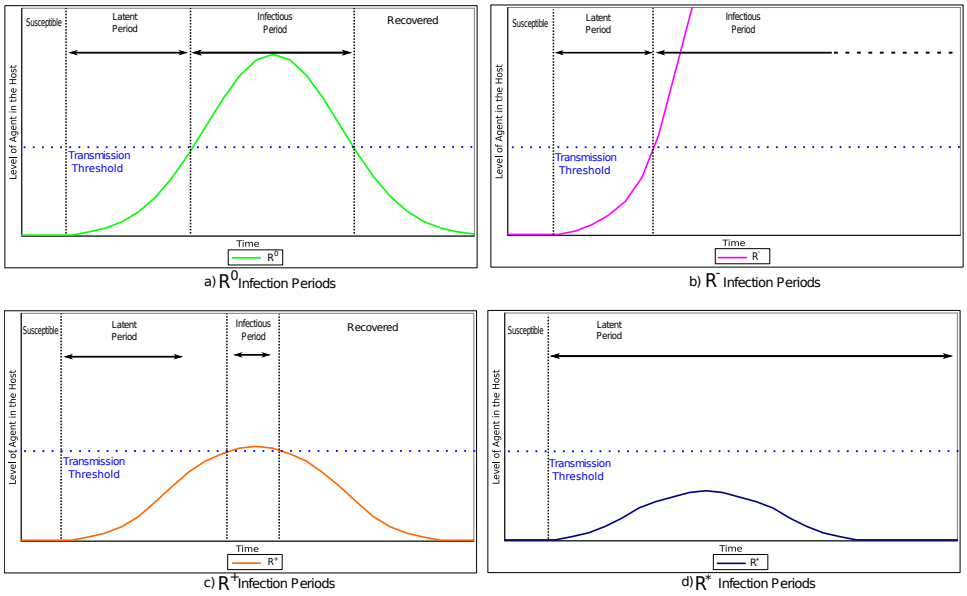

FIGURE 5.2. Four contrasting scenarios of infection length periods as a function of vbl.

5.3. Population Immune Response

Population immune response is a new concept that captures the collective immune response (IR) of individuals in a population represented by the superposition of individual immune responses. Figure 5.3 summarizes how the effects of PIR on outbreak dynamics can be exploited. The figure illustrates how the concept can be applied to create a computational model. The model is divided into three modules: population and disease database, immune competence and infectious disease outbreak simulation. A population and disease database is required to store the population and disease information. This database must

include demographic data for P that can be linked to the efficacy of an individual's immune response and, hence, determine the collective immune responses of P. The database information is exploited to categorize the individuals into multiple clusters. These clusters are based on the demographic characteristics of interest in the study(eg. age, gender). Additionally, the database contains information related to the disease itself, such as pathogenicity and pathogen growth rate. The disease information is used to estimate the value of the transmission threshold vbl^*.

Once the clusters have been created, the immune response upon infection of each group is determined by the immune competence module. In this research, we exemplify this calculation with a mathematical approach to model the immune response of an individual during a viral infection. The model presented was first introduced by Arnaout et al. [9] and expanded by Wodarz [107]. This model captures the development of the disease in the host by portraying the interaction between host cells, viruses and immune response. The interaction between vbl and the immune response is captured at a cellular level during the simulation of each cluster's immune response. The result of every simulation returns the values of vbl for each cluster. From every vbl value it is possible to obtain the length of the disease periods for each cluster by incorporating vbl^* as described in Section 5.2. Ultimately, the length of the disease period of each individual is calculated and incorporated into the simulation of the epidemic.

The infectious disease outbreak simulation simulates the spread of a disease in a population during time t. The simulation must incorporate every individual from P and simulate the interactions between them. Each individual needs to be assigned unique characteristics representative of his or her individuality and social activities. Some of these characteristics are the length of the disease periods and the number of social interactions per time interval. These values are usually calculated in a per day share. The length of the disease periods of every individual is assigned according to his or her cluster membership. The incorporation of different clusters in the simulations results in different disease dynamics. Since the variation of length of the disease periods between individuals is determined by their immune response,

we conjecture that PIR is a crucial driving force of the dynamics of an epidemic.

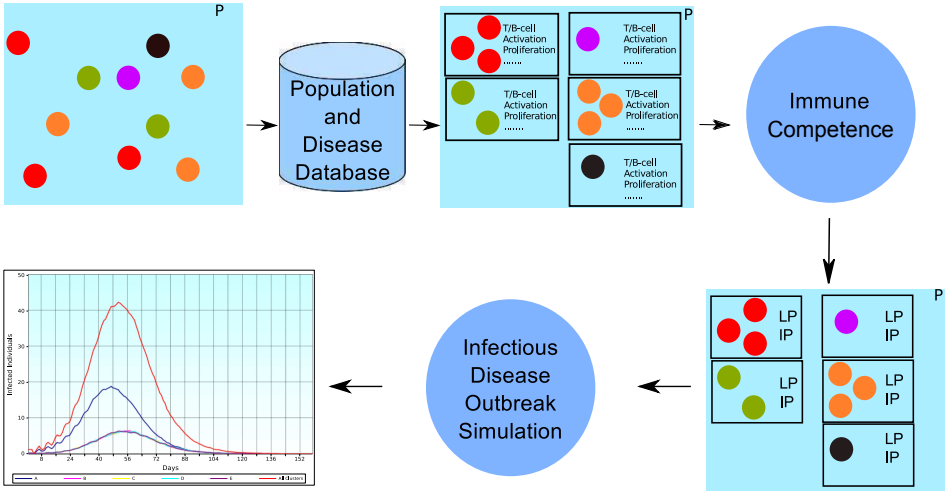

FIGURE 5.3. Exploring the effects of PIR on outbreak dynamics

5.4. Modeling the Immune Response

The module immune competence determines the infectious and latent period of individuals of every group from the population. To estimate those values it is possible to use models that capture the interaction between a pathogen and the immune response on a host. Multiple models have been proposed to simulate the immune system of an individual during infection [93]. In each model, different sets of parameters and interacting components are utilized to represent particular functions of the immune system. To illustrate the concept of PIR, we sought a mathematical model sufficiently complex to capture the essential components of the immune response and capably represent different diseases. Arnaout et al. [9] introduced a mathematical model that separates the immune response in two different types: lytic and non-lytic response. The flexibility of this model allows the study of the effects of each response type during different disease infections. This mathematical model depicts the interaction between infected cells (Y), viruses (V) and the immune response. The immune response is represented as the quantity of virus-specific CTL(Z) and the virus-specific antibodies (W). Wodarz [107] expanded the model of Arnaout et al. by incorporating uninfected

cells (X). The Wodarz model is represented by the set of differential equations 16, 17, 18, 19, and 20. A description of the parameters of the model is presented in Table 5.1.

$$\dot{X} = \lambda - dX - \beta XV \tag{16}$$

$$\dot{Y} = \beta XV - aY - pYZ \tag{17}$$

$$\dot{V} = kY - uV - qVW \tag{18}$$

$$\dot{W} = gVW - hW \tag{19}$$

$$\dot{Z} = cYZ - bZ \tag{20}$$

TABLE 5.1. Model parameters

Symbol	Definition
λ	Production rate of uninfected cells
d	Death rate of uninfected cells
β	Infection rate of uninfected cells by viruses
a	Death rate of infected cells
p	Lysis rate of infected cells by the CTL response
k	Production rate of virus by infected cells
u	Decay rate of viruses
q	Neutralization rate of viruses by antibodies
g	Development rate of antibodies in response of virus exposure
h	Decay rate of antibodies
c	Development rate of CTL in response to infected cells
b	Decay rate of CTL

To study the immune system efficacy, we focused on those parameters that determine the strength of the immune response. As described in Section 5.2, alterations of the immune system functionality are caused by multiple factors. For example, it is known that an individual will experience chronic involution of the thymus gland as he/she ages [43]. They involution of the thymus is considered one of the major reasons for the decline of immune response quality since the thymus is responsible for the production of naïve T-cells. Additionally, B-cell proliferation and efficacy are diminished due to the immunosenescence derived from aging [43]. Studies have shown a decrease of the B-cell population, reduction of antibody diversity, and decline of capacity to produce pathogen-specific antibodies as the individual ages [103, 2]. In the mathematical model, we represent this effect by incorporating individual age groups with different values for the immunological parameters. The immunological parameters of the model are: g, q, b, h, q, c, and p, as depicted in Table 5.1. Figure 5.4 depicts two simulation scenarios of the mathematical model for different values of the immunological variables. The simulation represents a standard immune response $\mathcal{R}^0$ and a hyperimmune response $\mathcal{R}^+$. The figure illustrates the performance of each response during infection and its impact on the viral load. Although there is evidence that the quality of those variables is affected by the age of the individual, determining its value only from age groups is not completely accurate. Multiple factors besides immunosenescence can be involved in establishing the strength of a specific immune response parameter; however, this is beyond the scope of this dissertation.

5.4.1. Modeling the Immune Response of the Population

As described in Section 5.4, the quality of the immune response of individuals is determined by multiple factors. However, given the scope of this research, we present an implementation that only considers the effects caused by immunosenescence with the goal to associate immunocompetence with demographic characteristics. This factor was selected given that its importance is highlighted in multiple studies [65][43][2][103]. Additionally, data for the length of the infectious periods of multiple pathogens is commonly reported for different age groups [70] [72]. In this implementation, the quality of the immune re-

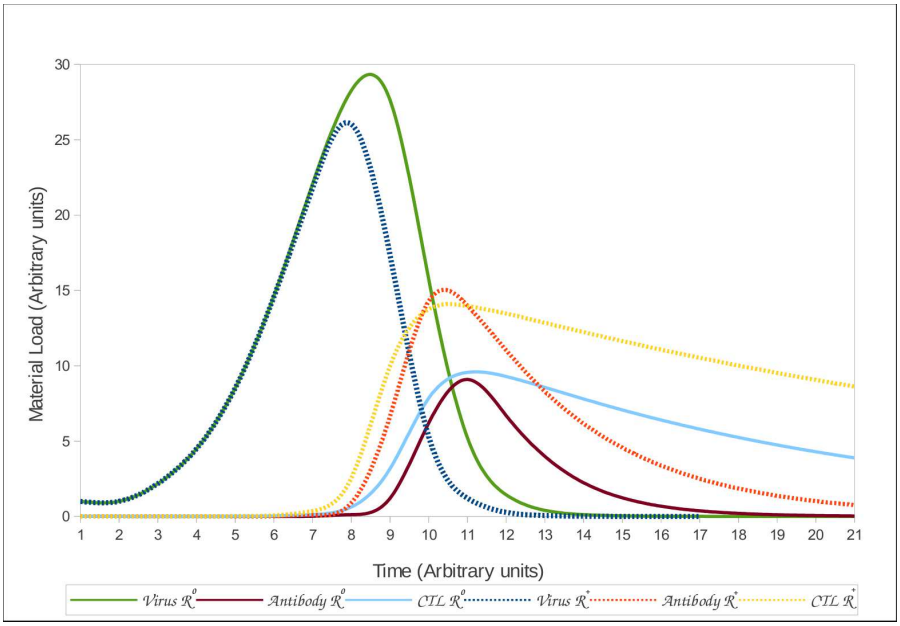

FIGURE 5.4. Viral load quantities and immune responses for 2 different values of immune response parameters.

sponse of individuals from different age groups was determined based on the reported data of disease periods for influenza. The World Health Organization (WHO) [42] reports that an infected adult (group A) is capable of transmitting the flu virus from 5 to 7 days. Similarly, infected children (group B) may infect others for up to 21 days with a median of 7-8 days and immunocompromised individuals (group C) could be infectious for weeks or months. The immunocompromised group includes adults 55 and older[50]. We included an additional group (group D) that represents individuals that have gained immunity to the disease either by natural or artificial immunization independent of their age. Considering that the infectious periods of each age group are determined based only on age, it is clear that the variability observed in the length of the infectious periods within the groups can be attributed to other factors such as gender, race, etc. Ultimately, each of the age groups is appointed to represent one type of immune response from the 4 types of immune response defined in Section 5.2, such that A = $\mathcal{R}^0$, B=$\mathcal{R}^{-1}$, C=$\mathcal{R}^{-2}$ and D=$\mathcal{R}^*$.

In Section 5.4, we exemplify the use of the Wodarz model as a method to determine

the length of the infectious and latent periods of an individual. In this model, its parameters can be modified to simulate different immune responses with distinct qualities. For the disease described below, each age group is associated with multiple levels of immune response efficacies and efficiencies. Further, as depicted in Section 5.4, the quality of the immune response can be represented by parameters g, q, b, h, q, c, and p from the mathematical model. Considering that the length of the infectious period varies for each age group, the parameters used to obtain the quality of the immune response of a group are not unique. Thus, the values for the parameters to simulate the immune response of an age group are calculated from a particular distribution of values. In table 5.2 we present some of the possible values used to simulate the immune response of individuals from each group. However, the values for the simulations presented do not include all possible infectious periods lengths nor all possible values for the parameter from their distribution. Figure 5.5 depicts the resulting viral load values for the simulations with the parameter values presented table 5.2. As described in Section 5.4, from Figure 5.5 we obtain the latent and infectious period for each group by evaluating its intersection points with vbl^*. Assuming the time intervals as days, the latent period (LP) is 5 days for all groups. Equivalently, the length of the infectious period (IP) for each group is: $IP_A = 5$, $IP_B = 7$, $IP_C = 11$, and $IP_D = 0$. Members of group D can be infected; however, considering the prime quality of their immune response, they will never develop the disease and thus never become contagious. A similar process is necessary to obtain all other infectious periods for each age group.

Determining the exact distributions for the parameters of the model in order to reflect the biological, sociological, and immunological characteristics of an individual is beyond the scope of this research. In this implementation, the lengths of the infectious period of each age group are calculated using probability distributions such that each distribution is an approximation to the data reported in the literature. Figure 5.6 depicts the probability distributions of the infectious period in each of the groups previously defined. Group A includes individuals between 14 and 55 years old and their infectious period is calculated with a normal distribution with mean $\mu = 6$ and standard deviation $\sigma = 1.5$. Group B

TABLE 5.2. Values used in the model to simulate each immune response group

Symbol	A	B	C	D
λ	30	30	30	30
d	0.1	0.1	0.1	0.1
β	0.01	0.01	0.01	0.01
a	0.5	0.5	0.5	0.5
p	0.2	0.1	0.05	0.1
k	0.4	0.4	0.4	0.4
u	2	2	2	2
q	0.006	0.01	0.0025	0.005
g	0.1	0.09	0.025	0.5
h	0.3	0.6	0.3	0.3
c	0.015	0.01	0.003	0.006
b	0.05	0.1	0.02	0.05

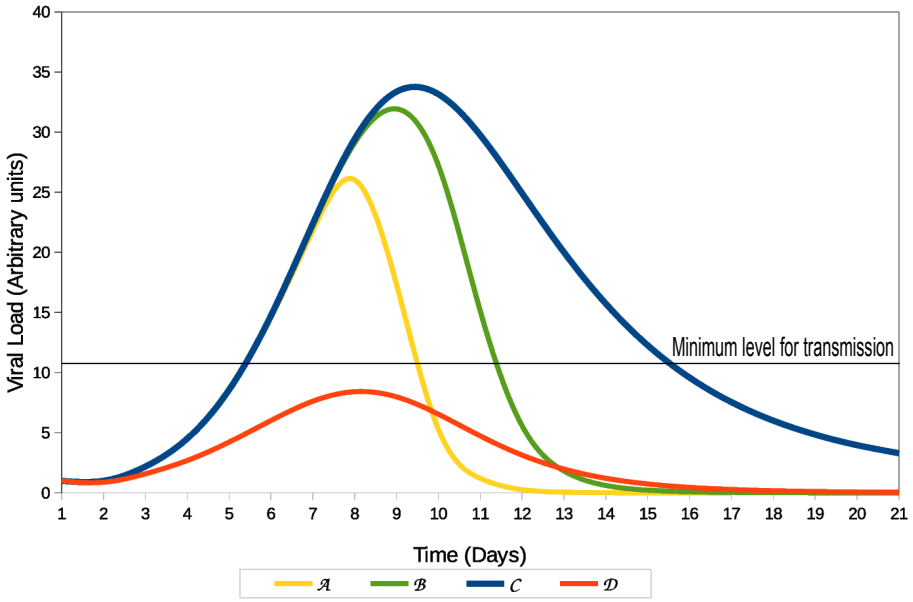

FIGURE 5.5. Viral load quantities for different values of the immune response parameters.

includes individuals with ages 0 to 14 and its represented with a Lévy distribution with location parameter $\mu = 7$ and scale parameter $c = 1$. The Lévy distribution is depicted in Equation 21. Finally, group C includes individuals 55 years old or more and their infectious period is calculated with a Lévy distribution with $\mu = 8$ and $c = 4$. Ultimately, in the simulation every individual from each age group is assigned an infectious period following the probability distribution of its group.

$$f(x;\mu;c) = \sqrt{\frac{c}{2\pi}} \frac{e^{-\frac{c}{2(x-\mu)}}}{(x-\mu)^{\frac{3}{2}}} \tag{21}$$

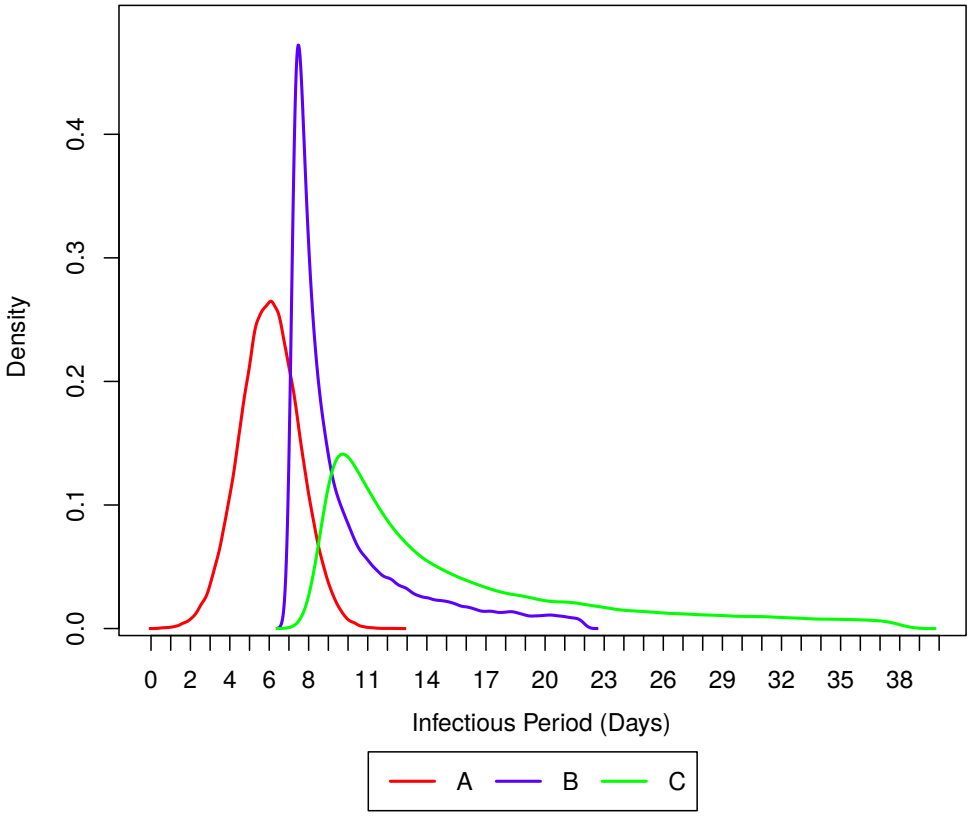

FIGURE 5.6. Probability distributions of the infectious period for the three age groups

In the next section, we present multiple experiments to analyze some of the effects of PIR on the disease dynamics.

A series of experiments were conducted to exemplify the computational model proposed in the previous section. The simulation commences with the creation of a synthetic population. From this population, individuals are divided uniformly into four groups. Membership to three of the four groups is determined by the age of the individual. The three possible age groups are adult ($\mathcal{R}^{0_A}$, A), children ($\mathcal{R}^{-_B}$, B), and elderly ($\mathcal{R}^{-_C}$, C). Affiliation to the fourth group is independent from the age of the individual but related to its immune status ($\mathcal{R}^{*_D}$, D). All non-immune individuals are members of one of the other three groups (A, B, or C). Affiliation to a specific group was used to determine the quality of its immune response. We assume a disease in which the quality of the immune response of members from group A is more effective than those from group B and C. As described in 5.4.1, the immune response and, consequently, the infectious period for each group is determined by a probability distribution.

Once the lengths of the periods for each group have been determined, a computation of the spread of a disease within the groups is required. For this simulation, we utilized the global stochastic contact model (GSCM). The GSCM is a computational model that simulates the spread of an infectious disease in a population during an infectious outbreak [84]. The GSCM simulates the interactions between individuals in the population as the infection progresses. Based on multiple disease parameters, some of those interactions result in the transmission of the disease from individual to individual. In the model, multiple groups of individuals can be created. Each group is assigned values of specific disease parameters to represent heterogeneous populations. Some of the disease parameters include contact rate, transmissibility, affinity between clusters, and infectious and latent period lengths. For this simulation, we uniformly distributed the total population of 4000 individuals into the four groups previously described. Individuals from each group were assigned its respective latent and infectious period. Since this simulation is designed to explore the effects of the immune response, we assume all the other disease parameters to be identical for all groups.

5.5.1. Experiment I

The first experiment explores the importance of incorporating different immune responses into the population. This experiment is divided into four cases. Each case consists of the simulation of infectious outbreaks among homogeneous populations. The populations in each case consist of communities in which all individuals are members of only one of the immune response categories $\mathcal{R}$. In each case, individuals are assigned an infectious and latent period based on their immune response classification. All the other disease parameters are identical among all individuals for all cases. Each simulation started with the inclusion of a single infectious individual into the population. The result of every case is the average of 50 simulations. A run is considered an outbreak if more than 1% of the population is infected. Otherwise, herd immunity or a deficient pathogen transmission is assumed. The results are summarized in Table 5.3. The table depicts the average number of infected individuals, the average number of individuals infected at the peak of the outbreak, the day in which the average peak occurred, and the average day in which the outbreak ended. The table displays an increased value in the number of infected individuals as the quality of the immune response decrease. Additionally, the outbreaks with immunocompromised populations have an earlier peak and are shorter in duration due to the heightened count of infected individuals in those experiments. In general, these results display the existence of a variation in the outbreak dynamics by incorporating an immune response to individuals in a population.

TABLE 5.3. Experiment 1 results

$\mathcal{R}$	Average total infected	Average infected at peak	Day of peak	Average end of outbreak
$\mathcal{R}^0$	637.78	17.47	210	563.84
$\mathcal{R}^{-1}$	2979.29	384.24	128	301.16
$\mathcal{R}^{-2}$	3579.45	844.84	117	258.22
$\mathcal{R}^*$	1	1	1	1

5.5.2. Experiment II

The second experiment integrated groups A, B, C, and D to explore their effect on the disease dynamics. Since we are interested in measuring the effects of PIR in the disease dynamics, the rest of the characteristics of the population follow a homogeneous profile. To measure the effect of PIR, we computed the total number of infected individuals in each group and the proportion of the population that they infected. The results reported are the average over 50 simulations with different random seeds. As before, each simulation is initialized with a single infectious individual that is randomly assigned to one of the four groups. The results are summarized in Figure 5.7 and Table 5.4. In the table, we observe that the distribution of the infected individuals is almost uniform. In contrast, the distribution of the individuals *infected by* members of each cluster is biased towards members from group C.

TABLE 5.4. Experiment 2 results

Cluster	Infected members	*Infected by* members of this cluster
($\mathcal{R}^0$, A)	472.82	371.13
($\mathcal{R}^{-1}$, B)	472.86	599.68
($\mathcal{R}^{-2}$, C)	475.44	920.93
($\mathcal{R}^*$, D)	471.62	0

In Figure 5.7, we observe that members from group C are responsible for infecting the majority of individuals in the population. Individuals from this cluster also have the highest total infection rate at all times during the outbreak. These effects are a consequence of the incompetent immune response of the members of this group. This deteriorated response cannot control the viral growth, resulting in a larger than average vbl among its infected individuals. Due to the excessive quantity of pathogens, the level of vbl exceeds vbl^* for a longer period of time as compared to the other two groups of individuals. Consequently, members from this group have an increased opportunity to infect others due their longer infectious period. Although the number of infected individuals in group D is similar to that

of the other groups, members from this group present a special behavior. Since members from group D cannot transmit the infection to others, the number of individuals *infected by* members of this group is zero. Equation 19 depicts the relationship between infection rates among all clusters:

$$\int_0^{\pi} inf\_by\_D \, \mathrm{d}t < \int_0^{\pi} inf\_by\_A \, \mathrm{d}t < \int_0^{\pi} inf\_by\_B \, \mathrm{d}t < \int_0^{\pi} inf\_by\_C \, \mathrm{d}t \tag{22}$$

where π denotes the end of the outbreak.

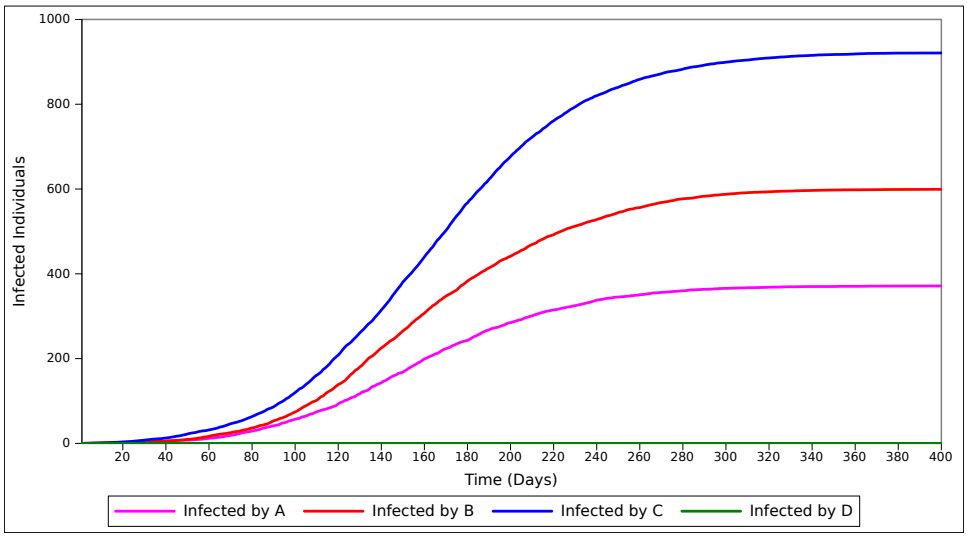

FIGURE 5.7. Cumulative number of individuals *infected by* members of every cluster of the population per day.

5.5.3. Experiment III

The third experiment measures the role of PIR in a non-homogeneous population during an infectious outbreak. In this experiment, the population is divided to represent different demographic distributions. The population is divided in 3 groups: A ($\mathcal{R}^0$), B ($\mathcal{R}^{-1}$), and C ($\mathcal{R}^{-2}$). Individuals from group D are disregarded from this experiment since the role of immunized individuals is analyzed in Experiment IV. In this experiment, we varied the number of individuals assigned to each group in each simulation to represent multiple distributions of the population. In both cases, the experiment commences by assigning all

4000 individuals from the population into group A ($\mathcal{R}^0$). Consequentially, individuals are added to the other immune response groups ($\mathcal{R}^{-1}$ and $\mathcal{R}^{-2}$) in increments of 5% to each group until $|A|$=0. Figure 5.8 depicts the number of infected individuals in the simulations. Additionally, Table 5.5 displays the percentage of individuals from the population that was infected in each simulation. The results presented in the figure and the table are an average of 50 runs with different seeds for each simulation.

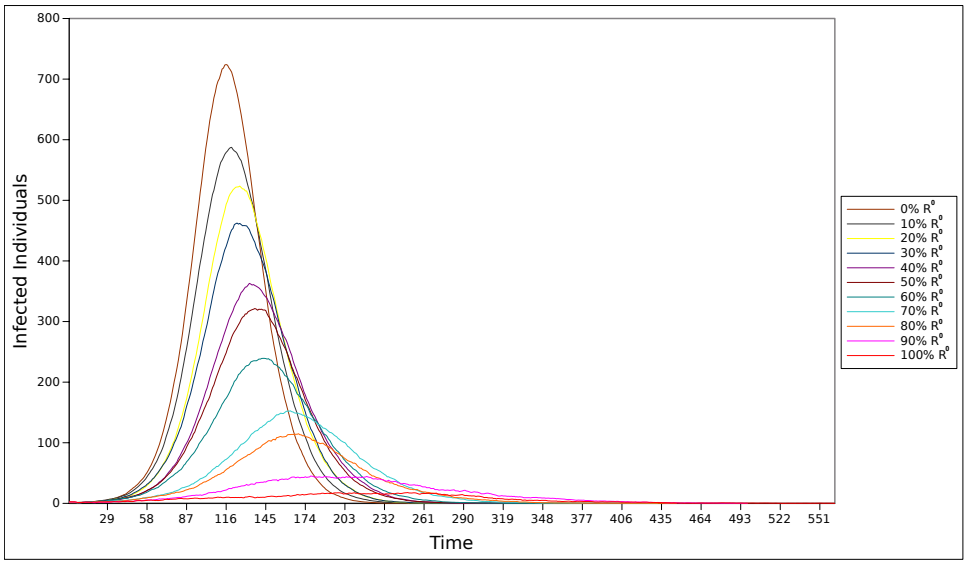

FIGURE 5.8. Number of infected individuals per day in a population with different distributions of $\mathcal{R}^0$, $\mathcal{R}^{-1}$ and $\mathcal{R}^{-2}$.

Figure 5.8 depicts the variations in the outbreak dynamics for different combinations of $\mathcal{R}^0$, $\mathcal{R}^{-1}$ and $\mathcal{R}^{-2}$. We observe that the magnitude of the outbreak increases as more individuals from $\mathcal{R}^0$ are moved to the other two groups. Individuals from those groups present a higher infectious period resulting in increased probabilities of infecting other individuals from the population. The figure also depicts a variation in the length of the outbreak based on the number of individuals from $\mathcal{R}^{-1}$ and $\mathcal{R}^{-2}$. The duration of the outbreak is reduced as the number of individuals in those groups increases. This effect is produced since individuals from those groups infect the susceptible individuals at a faster rate. This results in an exhaustion of susceptible individuals earlier in the outbreak. Further, Table 5.5 shows that the proportion of infected individuals increases as a larger number of individuals from $\mathcal{R}^{-1}$

TABLE 5.5. Experiment 3 results

Percentage of R^0	Percentage of population infected
0	88.68
10	84.90
20	82.26
30	81.09
40	74.08
50	71.80
60	64.47
70	51.16
80	44.18
90	30.42
100	15.94

and $\mathcal{R}^{-2}$ are introduced to the population. These variations indicate that the presence of fewer individuals from $\mathcal{R}^0$ not only affect the duration and magnitude of the outbreak, but the percentage of infections as well.

5.5.4. Experiment IV

The fourth experiment attempts to measure the effect individuals from group C have on the outbreak at large under the assumption of different vaccination strategies. Individuals from group C were selected since the previous experiments showed that members from this group cause the majority of the infections in the population. Three vaccination strategies were implemented to represent different intervention scenarios. Each strategy consists of removing a percentage of individuals from group C and incorporating them into the immunized group (D). Each vaccination strategy removed 0%, 10% and 20% of the members from C, respectively. The 0% strategy represents a zero-intervention scenario. This scenario is simulated utilizing the same population and values of disease parameters as in Experiment II. The other two intervention strategies also use the same values of disease parameters for each

group, but with a modified population distribution. Similar to the previous experiments, a single infectious individual was randomly assigned to a group at the beginning of each simulation. The final result was obtained by averaging the outbreak dynamics over 300 runs for every strategy. The results from this experiment are divided in two parts: total number of infected individuals in P and total number of individuals *infected by* C. The results are depicted in Figure 5.9 and Figure 5.10, respectively.

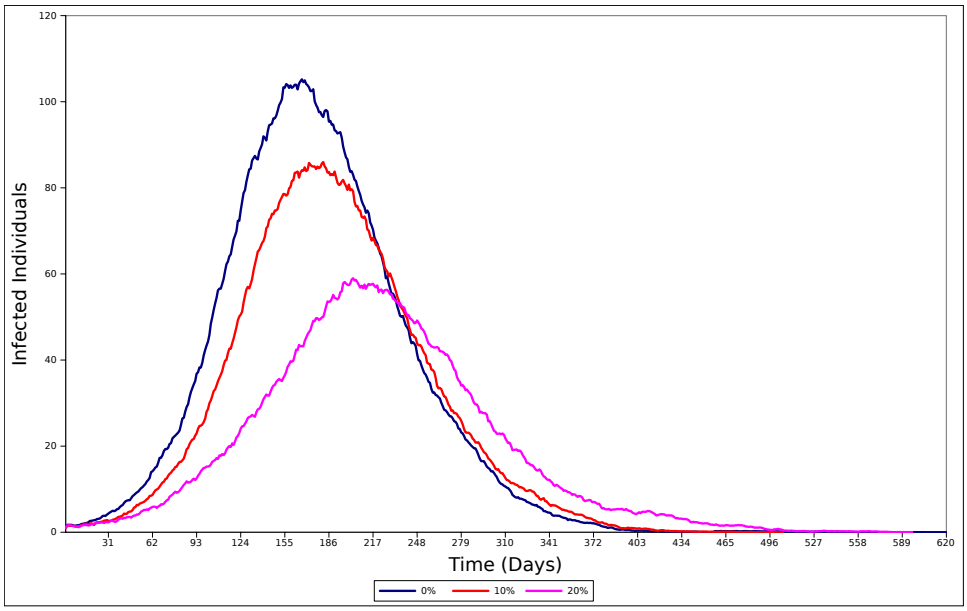

FIGURE 5.9. Total number of infected individuals from the population per day after different vaccination strategies.

Figure 5.9 illustrates the different outbreaks in P after each vaccination strategy is implemented. We observe that the 0% strategy produces the greatest number of infected individuals and results in the highest infection rate per day compared to the other strategies. On the other hand, the 20% strategy is more effective at limiting the progression of the disease due to its high number of members in group D. This results in the lowest number of infected individuals among all strategies. Further, we observe a variation in the length of the outbreak for each strategy. The 0% strategy results in the shortest outbreak compared to the other two strategies. This effect is caused by the increased number of individuals with weak immune response in that strategy. The further reduction of individuals from C, in the other two strategies, results in a reduced opportunity to infect susceptible individuals that

could, potentially, spread the disease even further. Consequently, the number of infections is reduced and the pathogen spread less aggressively, providing a larger population of susceptible individuals for future transmissions as opposed to populations with a more aggressive progression.

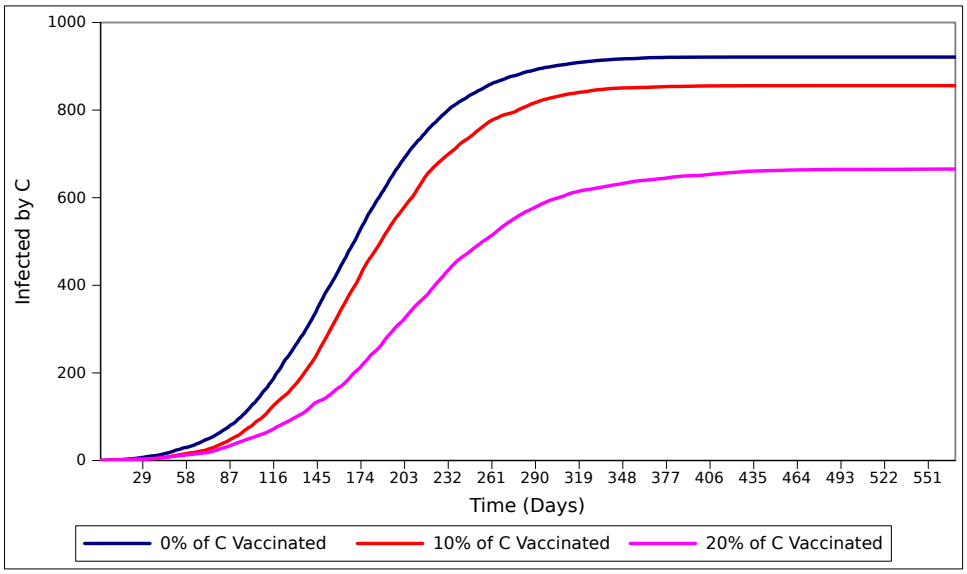

FIGURE 5.10. Cumulative number of individuals *infected by* members from group C after different vaccination strategies.

Figure 5.10 illustrates the cumulative number of individuals *infected by* members from group C for each vaccination strategy. Here, the number of individuals *infected by* group C behaves similarly to the total number of infected individuals in Figure 5.9. We observe that the no-intervention strategy causes the greatest number of individuals *infected by* members from group C. Furthermore, the 10% and 20% strategies decrease that count according to the intervention type, resulting in a reduced number of individuals *infected by* group C. More importantly, this result confirms the strong relationship between this group and the general outbreak. We observe that Figure 5.9 and Figure 5.10 display similar patterns for each respective outbreak. More realistic strategies can be applied to study the effect of high-risk groups to the outbreak at large, but this is beyond the scope of this research.

5.6. Summary

In this chapter, we highlighted the importance of integrating the viral/bacterial load during an infectious outbreak simulation. This value is utilized to measure the severity of the disease and the host's capacity to transmit it. More importantly, we emphasized the direct relationship between the quality of the immune response and the quantity of the viral/bacterial load. A stronger immune response controls the growth of the pathogen more efficaciously. Further, the viral/bacterial load is utilized to determine the length of the infectious periods. A minimum transmission threshold is established as the minimum quantity of viral/bacterial load required for an individual to become contagious. Once the viral/bacterial load surpasses this threshold, the individual becomes contagious and capable to infect others until his or her load falls below that threshold.

This research introduces the concept of population immune response (PIR). PIR captures the collective immune response of individuals in a population represented by the superposition of individual immune responses. A computational model that captures the effects of PIR on the outbreak dynamics has been presented. This model is divided into three compartments:

(1) Population and disease database

(2) Immune competence

(3) Infectious disease outbreak simulation

The population and disease database includes demographic information that is used to determine the efficacy of the immune response of the individuals. The database also contains disease-specific data utilized to establish the transmission rate and the value of the minimum infection threshold. The immune competence module is used to determine the quality of the the immune response of each individual of the population upon infection. In this chapter, we exemplified this concept by describing a model that simulates the interaction, at a cellular level, between the immune cells and the pathogen during an infectious process. This mathematical model determines the value of the viral/bacterial load in each iteration. This value is used to establish the length of the infectious periods by implement-

ing the minimum infection threshold. The infectious disease outbreak model simulates the spread of an infectious disease among a population. The population is created based on the data obtained from the other two compartments. This model must incorporate unique characteristics for every individual and simulate the interactions between them. The execution of this model captures the disease dynamics during an infectious outbreak.

Finally, multiple experiments were conducted to analyze some of the effects PIR has on the disease dynamics. The first experiment depicted the existence of a variation in the outbreak dynamics by incorporating the same immune response to all individuals in a homogeneous population. The next set of experiments divided the population into three age groups and an immunized group. Each group was characterized by a unique immune response quality and thus a different length for its immune periods. A simulation was conducted to study the spread of a disease within the groups. The results show that individuals with weak immune responses and those who are immune to the pathogen make a relevant contribution to the general outbreak. In general, our results suggest that it is essential for the public health establishment to increase their understanding of the characteristics of regional demographics in order to monitor and mitigate ongoing epidemics effectively.

CHAPTER 6

MODELING OUTBREAKS IN GEOGRAPHIC CONTEXT

In chapter 3 and 5, we introduced the concepts of epidemic trajectory and population immune response, respectively. In this chapter, we will implement those concepts to study the effects of PIR in the epidemic trajectory. To capture the progression of a disease in a geographic location, we expanded Mikler et al. global stochastic contact framework (GSCF) [64]. The GSCF was implemented as a computational model to simulate the spread of a disease within a population P with elements $p_i \in P$ in which $i = 1, 2, 3, ..., |P|$. All elements p_i are spatially distributed in geographical regions $r_l \in R$ such that $l = 1, 2, 3, .., |R|$. The individuals are assigned a value r_l based on their location of residence. $\mathcal{S}(i)$ is a function that returns the residence r_l of individual i. Additionally, elements of the population are divided in groups $g_k \in G$ in which $k = 1, 2, 3, .., |G|$. Individuals are assigned a group g_k based on their socio-physiological characteristics, such as age or gender. The function $\mathcal{G}(i)$ returns the group membership g_k of individual p_i. Finally, the function $\mathcal{M}(g_k, r_l)$ returns the set of individuals that are members of group g_k and are located in region r_l.

Figure 6.1 depicts a population from the global stochastic contact framework. The population from the figure is divided into 4 groups distributed over 4 regions. Further, the quantity of individuals $|\mathcal{M}(g_k, r_l)|$ for a given g_k varies for all r_l. Furthermore, not all groups of individuals of the population from Figure 6.1 may have members in all regions.

In this framework, the groups from each region are assigned unique characteristics to represent a heterogeneous population. In the following section, we describe four socio-physiological characteristics that can be assigned to individuals, and consequently their groups, in the global stochastic contact framework.

6.1. Role of Socio-Physiological Characteristics of Individuals

Studies have shown that the socio-physiological characteristics of individuals, such as age [100] or gender [8], will directly impact their social-behavioral interactions and immune response quality. Our model represents these relationships through social-behavioral param-

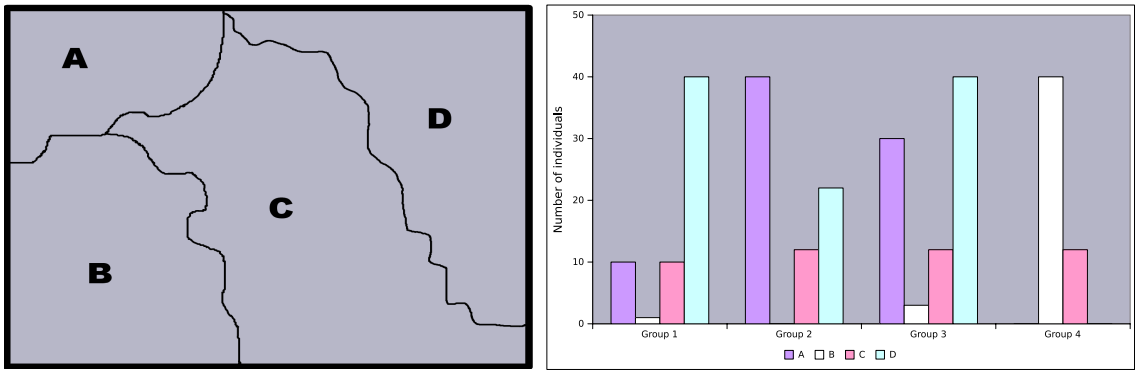

FIGURE 6.1. Example of population from the global stochastic contact framework

eters and a parameter that describes the quality of the immune response. The different social-behavioral parameters considered in this framework are:

- Contact rate
- Mobility
- Group affinity

The contact rate (CR) is the number of contacts an individual initiates every timestep $t \in T$. In this model, CR_i is the unique CR of element p_i. For two individuals p_i and p_j if $\mathcal{G}(i) = \mathcal{G}(j)$ then $CR_i = CR_j$. In this framework, this equality represents the property that individuals with similar socio-demographic characteristics have the same CR regardless of their location. In general, since all individuals from a group g_k have the same contact rate, we define CR_{g_k} as the contact rate of individuals that, independent of their region, are members of group g_k.

The mobility m_i, $0 \leq m_i \leq 1$, is the proportion of contacts p_i directed to individuals in region $\mathcal{S}(i)$. If $m_i = 0$, all contacts occur in the region of p_i. Otherwise, if $m_i = 1$, all contacts are distributed in $R - \{\mathcal{S}(i)\}$. For two individuals p_i and p_j, if $\mathcal{G}(i) = \mathcal{G}(j)$ then $m_i = m_j$. In this framework, this property represents that, since mobility is determined by socio-demographic characteristics, all individuals from the same group have the same mobility, regardless of their location. In general, since all individuals from group g_k have the same mobility, we define m_{g_k} as the mobility of individuals that, independent of their

location, are members of group g_k.

Group affinity is the likelihood of social interaction between members of two individual groups. In this model, affinity among groups is determined by the socio-demographic characteristics of their members and not by their location. Affinity is represented with the function $\mathcal{A}(g_k, g_e) = a_{g_k,g_e}$ i.e. the affinity between any individual from group g_k and any individual in group g_e. For any $\mathcal{A}(g_k, g_e)$:

$$0 \leq \mathcal{A}(g_k, g_e) = a_{g_k,g_e} \leq 1$$

In this framework, the affinity function is represented as group affinity matrix $\mathcal{GA}(g_k, g_e) = a_{g,e}$, in which g_k, g_e represent two groups in the population. It consists of probabilities for interactions between different groups. The sum of probabilities for a group g_k to contact all other groups, ν_{g_k}, is defined in equation 23. Equivalently, the affinity to a group g_k from all groups, σ_{g_k}, is defined in equation 24. In any given affinity matrix, $\nu_{g_k} = 1$ and $\sigma_{g_k} \geq 0$.

$$\nu_{g_k} = \sum_{j=1}^{|G|} \mathcal{GA}(g_k, g_j) \tag{23}$$

$$\sigma_{g_k} = \sum_{j=1}^{|G|} \mathcal{GA}(g_j, g_k) \tag{24}$$

The parameter that describes the quality of the immune response, IR_i, is assigned to every element p_i following the framework defined in Chapter 5. For two individuals p_i and p_j if $\mathcal{G}(i) = \mathcal{G}(j)$ then $IR_i = IR_j$ i.e. The quality of the immune response of two individuals is the same, if they are members of the same group. Since all individuals from a group g_k have the same immune response quality value, we define I_{g_k} as the immune response quality value of individuals that, independent of their region, are members of group g_k. The value I_{g_k} represents the PIR of a specific group of the population, i.e. if g_1 consists of individuals whose socio-demographic characteristics shape their immune response as immunocompromised, then I_{g_1} is defined by $\mathcal{R}^-$.

In the next section, we present a detailed description of the GSCF contact model and all of its components.

6.2. Contact Model

In the contact model, the set P is mapped to 4 subsets: susceptible(S), latent(L), infectious(I), and recovered(R), thereby representing the components used in the SLIR. We assume a closed population P for which $S \cup L \cup I \cup R = P$, $S \cap L \cap I \cap R = \emptyset$, and $|P| = N$. The progression of the epidemic is represented by the movement of individuals between compartments during time instances $t \in T$, where T is the duration of the simulation. In the global stochastic contact framework, contacts are modeled at the regional level at every timestep t. Each individual from a group g_k and region r_l contributes to the total number of contacts originating from the region r_l. However, only contacts that originate from an infectious individual may result in a possible transmission of disease. Let $\mathcal{IN}(g_k, r_l)$ be a function that returns the set of infectious individuals in group g_k and region r_l. Since all the infectious individuals of a region r_l contribute to the total number of possible transmissions originating from their region, the infectious contacts in a given time t, $ICo(r_l, t)$ are:

$$ICo(r_l, t) = \sum_{b=1}^{|G|} |\mathcal{IN}(g_b, r_l)| * CR_{g_b} \tag{25}$$

To emulate the dynamics of the SIR, we consider all the regions with infectious individuals as a single force of infection instead of separate efforts. Let $\kappa(t)$ be a function that returns the set of regions that have at least one infectious individual at timestep t. Then the total number of infectious contacts on timestep t, $\xi(t)$, is:

$$\xi(t) = \sum_{s \in \kappa(t)} ICo(r_s, t) \tag{26}$$

Another feature that we implemented to emulate the SIR mechanics in this model is the attenuation rate of the force of infection as the number of susceptible individuals decreases. The attenuation rate represents the reduction in the probability to transmit the disease for infectious individuals as the epidemic progresses. This variation is caused by the the reduction in the number of susceptible individuals in the population as the number of

infectious and recovered individuals increases. Since the demographics of each region are unique, we calculate the regional attenuation rate $\lambda(r_m, t)$ for region r_m as:

$$\lambda(r_m, t) = \frac{|S(r_m, t)|}{|P_{r_m}|} \tag{27}$$

In which $S(r_m, t)$ and P_{r_m} represent the set of susceptible individuals and the set of total individuals in region r_m at time t, respectively.

All regional attenuation rates are averaged to obtain the general attenuation rate $\eta(t)$ with value:

$$\eta(t) = \frac{\sum_{s=1}^{|R|} \lambda(r_s, t)}{|R|} \tag{28}$$

The value $\eta(t)$ represents the proportion of susceptible individuals in the population at time t. Further, not all infectious contacts on timestep t, $\xi(t)$, results in transmission of the disease. The value β represents the chances of disease to be transmitted after a contact between an infectious and a susceptible individual has occurred. In general, the total number of transmissions in a given time t, $\Omega(t) \in \mathbb{N}$ are:

$$\Omega(t) = \xi(t) * \beta * \eta(t) \tag{29}$$

The value $\Omega(t)$ represents the total force of infection of the population at time t. However, since in this model contacts are modeled at the regional level, $\Omega(t)$ needs to be distributed between all regions. To determine the number of infectious contacts initiated from each region, the force of infection $\Omega(t)$ is proportionately distributed between all regions. Let $\mathcal{IR}(r_l)$ be a function that returns the set of infectious individuals in region r_l. The force of infection $\Omega(t)$ is distributed among all regions that have infectious members ($\kappa(t)$) proportionally to the size of their set $\mathcal{IR}(r_l)$. This distribution is calculated based in the regional distribution rate $D(r_l)$. Let $D(r_l)$ be:

$$D(r_l) = \frac{|\mathcal{IR}(r_l)|}{\sum_{s \in \kappa(t)} |\mathcal{IR}(r_s)|} \tag{30}$$

Since the total force of infection $\Omega(t)$ is distributed among all regions based on $D(r_l)$, the total number of transmissions a region r_l initiates at timestep t, $\Gamma(r_l, t)$, is

$$\lfloor \Gamma(r_l, t) \rfloor = \Omega(t) * D(r_l) \tag{31}$$

6.2.1. Origin of the Transmission

For every region that is initiating a transmission, it is necessary to determine the group membership g_k of the individual p_i that its commencing the infectious contact. To select the g_k of p_i, we distribute the force of infection of region r_l, $\Gamma(r_l, t)$, among all of its population. To distribute the transmissions from region r_l initiated by individuals from group g_k at timestep t, we calculate the group rate $E(r_l, g_k, t)$:

$$E(r_l, g_k, t) = \frac{|\mathcal{IN}(g_k, r_l)|}{\sum_{b=1}^{|G|} \mathcal{IN}(g_b, r_l)|} \tag{32}$$

Since $\Gamma(r_l, t)$ is distributed among all groups from a region based on $E(r_l, g_k, t)$, the number of transmissions from region r_l initiated by an individual from group g_k at timestep t, $\Psi(r_l, g_k, t)$, is:

$$\lfloor \Psi(r_l, g_k, t) \rfloor = \Gamma(r_l, t) * E(r_l, g_k, t) \tag{33}$$

The value of $\Psi(r_l, g_k, t)$ is calculated for all groups present in r_l and is used to populate the array $\rho_{r_l,t}$. The array $\rho_{r_l,t}$ is populated with the group membership of the initiators of all transmissions commenced by r_l. For group g_k there are $\Psi(r_l, g_k, t)$ infectious transmissions originating from it; then the array $\rho_{r_l,t}$ contains $\Psi(r_l, g_k, t)$ elements with value g_k.

6.2.2. Types of Transmissions

In this model, transmissions are divided in two types:

(1) Local transmissions $L\Gamma(r_l, t)$, which are the number of transmissions that originate in region r_l and are directed to region r_l in time t.

(2) Global transmissions $G\Gamma(r_l, t)$, which are the number of transmissions that originate in region r_l and are directed to all regions besides r_l in time t.

To determine the types of transmission each region commences, we define γ_{r_l}, $0 \leq \gamma_{r_l} \leq 1$, as the proportion of transmissions initiated by individuals from region r_l that are directed to individuals within region r_l. Additionally, let $\varrho(r_l)$ be a function that returns the set of individuals that are located in region r_l. Let γ_{r_l}:

$$\gamma_{r_l} = \frac{\sum_{i \in \varrho(r_l)} m_i}{|\varrho(r_l)|} \tag{34}$$

The value γ_{r_l} is used to calculate the number of global and local transmissions that each region initiates:

- The number of local transmissions $L\Omega(r_l, t)$ is:

$$L\Gamma(r_l, t) = \gamma_{r_l} * \Gamma(r_l, t) \tag{35}$$

- The number of global transmissions $G\Gamma(r_l, t)$ is:

$$G\Gamma(r_l, t) = (1 - \gamma_{r_l}) * \Gamma(r_l, t) \tag{36}$$

6.2.3. Interaction Coefficient

For all global transmissions, multiple factors determine the likelihood of a region interacting with all other regions. In this model, that interaction is defined by the interaction coefficient. The interaction coefficient as proposed by Mikler et. al. in [64] is defined as the product of the populations of the regions divided by the distance between them. The interaction coefficient represents the notion that populous regions have a higher interaction rate between themselves than with more sparsely populated regions. Also, the likelihood of

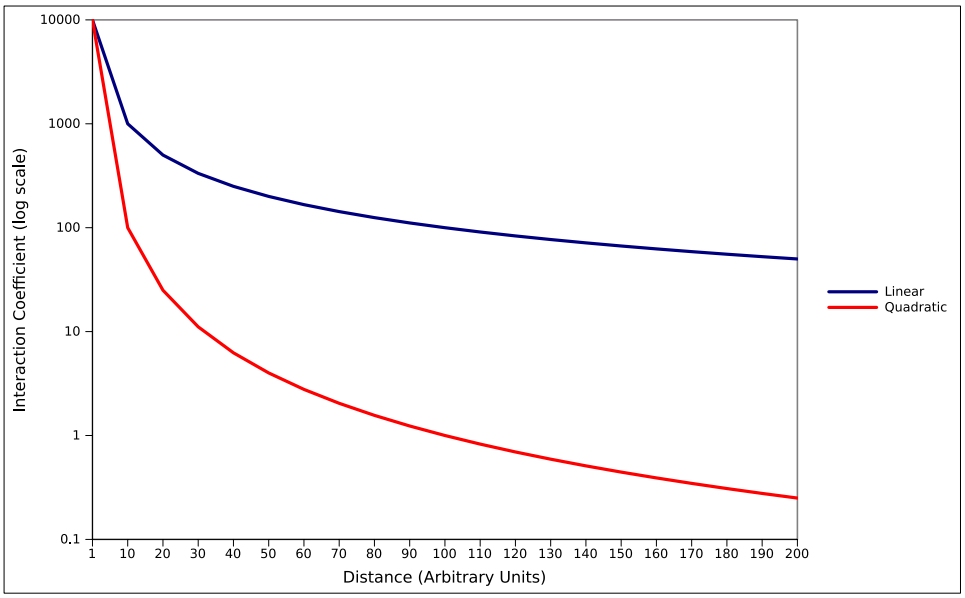

FIGURE 6.2. Values for the linear and quadratic interaction coefficients in a uniform population

interaction is directly impacted by the distance between regions. For any two regions r_l and r_m, the most frequently used interaction coefficients $IC(r_l, r_m)$ are:

$$IC(r_l, r_m) = \frac{|\varrho(r_l)| * |\varrho(r_m)|}{dist(r_l, r_m)} \tag{37}$$

$$IC(r_l, r_m) = \frac{|\varrho(r_l)| * |\varrho(r_m)|}{dist(r_l, r_m)^2} \tag{38}$$

A higher value for $IC(r_l, r_m)$ represents an increased likelihood of interaction between region r_l and r_m. Both interaction coefficients equations consider the impact of the distance between regions, however, equation 38 attributes a greater weight to the distance. The difference between both equations is depicted in Figure 6.2. In the figure, the population is assumed to be homogeneous in all regions. Independently of the interaction coefficient method, the values for population density and distances between regions can differ significantly between pairs of regions. Normalizing this value will reduce the variations among them. By normalizing the original values, the interaction coefficient is assigned a value between 0 and 1. The normalized interaction coefficient $ICN(r_l, r_m)$ is:

$$ICN(r_l, r_m) = \frac{IC(r_l, r_m)}{\sum_{k=0, k \neq l}^{|R|} IC(r_l, r_k)} \tag{39}$$

The function ICN is represented using a ICN matrix E. In this matrix, element e_{r_l, r_m} represents the interaction coefficient between region r_l and r_m. The sum of interaction coefficients from a region r_l to all other groups η_{r_l} is defined in equation 40. Equivalently, the interaction coefficient from all regions to a region r_l, $\sigma(r_l)$, is defined in equation 41. In any given ICN matrix, $\eta_{r_l} = 1$ and $\sigma_{r_l} \geq 0$.

$$\eta_{r_l} = \sum_{j=1}^{|R|} E_{r_l, r_j} \tag{40}$$

$$\sigma_{r_l} = \sum_{j=1}^{|G|} E_{r_j, r_l} \tag{41}$$

6.2.4. Distribution of Infectious Transmissions and Disease Progression in Individuals

Once the transmissions from a region r_l have been divided in local and global, it is necessary to determine the location r_m to which one or more of those transmissions will be directed to. If the transmission is local then the destination region $r_m = r_l$. Otherwise, the normalized interaction coefficient of r_l with all other regions is used to randomly select a r_m. The selection of r_m is determined by a random experiment in which a higher normalized interaction coefficient value between regions r_l and r_s results in increased probabilities to select r_s as the destination region.

After, selecting r_m as the destination region, an individual p_j from that region must be selected as the receiver of the transmission. However, since affinities between individuals are determined at the group level, it is necessary to select the group g_e of p_j to later select p_j. To select the group destination, it is necessary to use the array $\rho_{r_l, t}$. As described in subsection 6.2.1, the array $\rho_{r_l, t}$ is populated with the group membership of the initiators of all transmissions commenced by r_l. We extract one element from this array to obtain the group membership g_k of the initiator of an specific transmission. The group membership of

g_e is used to establish the group affiliation g_l of p_j with the use of the group affinity matrix $\mathcal{GA}(g_e, g_k)$. A higher value in the group affinity matrix between groups g_l and g_s results in increased probabilities to select g_s as the destination group. Ultimately, the selection of g_l is conducted by a weighted random experiment in which the group affinity matrix produces a bias in the experiment.

Once a group is selected, a random experiment is used to select a susceptible individual p_j from within S in group g_l and change its status to latent (L). Following, we determine the destination of the next transmission that originates from r_l. This process is repeated until the number of transmissions for r_l in time t are exhausted. After all transmissions from r_l have been distributed, we continue with the next region. Finally, once all transmissions have been distributed, it is necessary to update the infection status for all individuals. Individuals in the latent compartment remain in that compartment until their latent period terminates. Once individuals terminate that period, they immediately move from L to I and commence infecting others. Individuals in I remain infectious for the duration of their infectious period and then migrate to the recovered compartment. Once the infection status of all individuals has been updated, we repeat the general process and start the computation of day $t + 1$. This process is performed as long as there are infectious individuals.

In the next section, we present a series of experiments that use a computational implementation of the global stochastic contact framework.

6.3. Experiments

Four experiments were conducted to explore the effects of multiple geographic, immunological, and socio-demographic characteristics of the population have on the disease dynamics. To investigate those effects we conducted simulations in a model that mimics the concepts of the global stochastic contact framework in the counties of Sumter, Florida and Denton, Texas. Those counties were selected considering the divergence of the age demographic characteristics between each other. For instance, Table 6.1 depicts the contrasting distributions of the populations in Sumter and Denton. For instance in Denton TX., 16 percent of individuals are over 55 years and 23.2 percent under 15[95] whereas 63.12 percent

of the population in Sumter, FL. are over 55 and 7.4 under 15 [96]. In the model, the populations of each county were generated and placed in the geographic space using census data from 2010. Further, the census data was used to separate the populations into the three age groups.

TABLE 6.1. Population distribution in Sumter, FL and Denton, TX.

Group (age range)	Sumter, FL.	Denton, TX.
Adult (15-55)	27,494	402,402
Children (0-14)	6,951	153,741
Elderly (56+	58,975	106,471
Total	93,420	662,604

In this research, we feature the function of the age demographic characteristics of the population in an epidemic since its role in determining the quality of the immune response of individuals has been highlighted in multiple studies [65] [43] [2] [103]. Additionally, data for the length of the infectious periods of multiple pathogens is commonly reported for different age groups [70] [72]. In this experiments, the quality of the immune response of individuals from different age groups was determined based on the reported data of disease periods for influenza. The World Health Organization (WHO) [42] reports that an infected adult (group A, 14 and 55 years old) is capable of transmitting the flu virus from 5 to 7 days. Similarly, infected children (group B, 0 to 14 years old) may infect others for up to 21 days with a median of 7-8 days and immunocompromised individuals (group C) could be infectious for weeks or months. The immunocompromised group includes adults 55 and older[50].

In addition to the variances in the age demographic distribution, Sumter and Denton also present different population densities and geographic distributions. Figure 6.3 and Figure 6.4 depict the population distribution and population density of Sumter, FL and Denton, TX respectively. In Figure 6.3, we observe that the county of Sumter accommodates half of its population in 69.1 sq km of its territory and half in 1347.4 sq km. Furthermore, the densely populated area consists almost entirely of elderly population. In contrast, the remainder of the county hosts a more diverse population distribution.

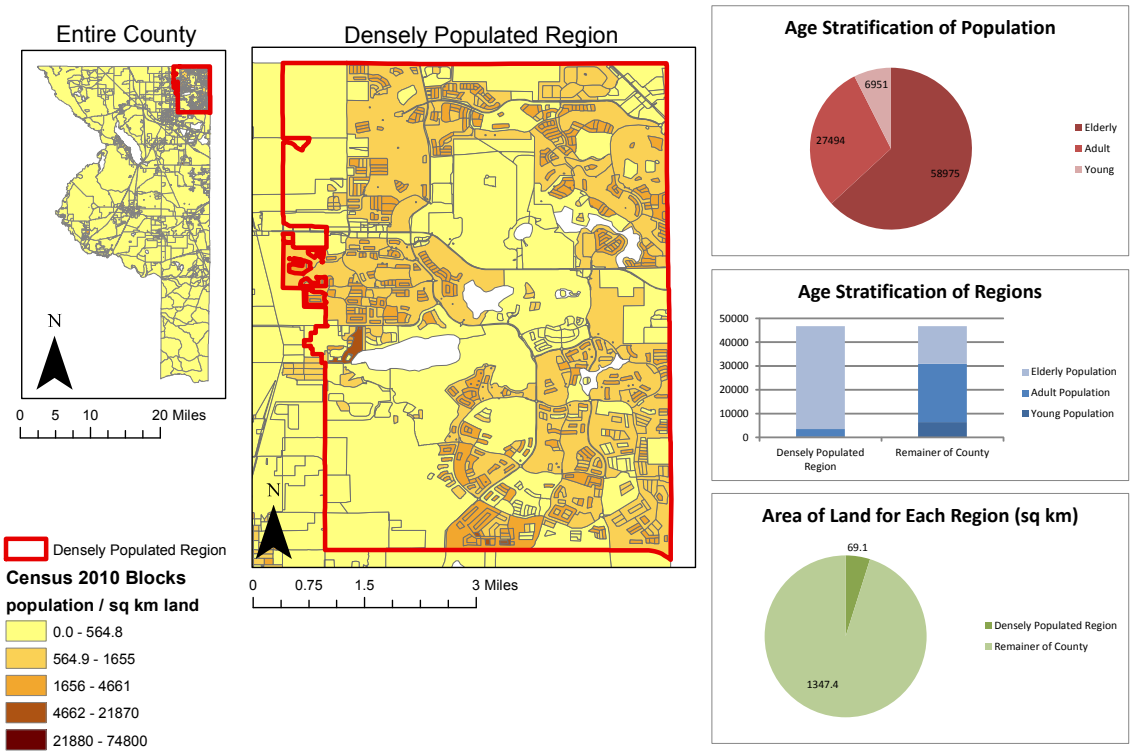

FIGURE 6.3. Distribution of the entire population and for each age group for Sumter, FL.

In Figure 6.4, the census blocks are divided in categories based on the number of individuals that reside in each block. The categories are designed to accommodate similar proportions of the population, i.e. category 1-20 includes a similar number of census blocks as category 92-2004. Given the large quantity of census blocks with population of zero individuals, those census blocks are assigned a unique category to avoid them from distorting the proportions of the other categories. Furthermore, we compare the distribution between age groups in the entire region by using the same categories to determine the population distribution for each age group and for the entire population. In the figure, we observe that the entire population and the individuals from each group in Denton county are sparsely distributed in its territory. Additionally, we observe that all age groups have a similar presence in all regions. In general, these figures highlight the differences in the population

distribution and population density between the two counties.

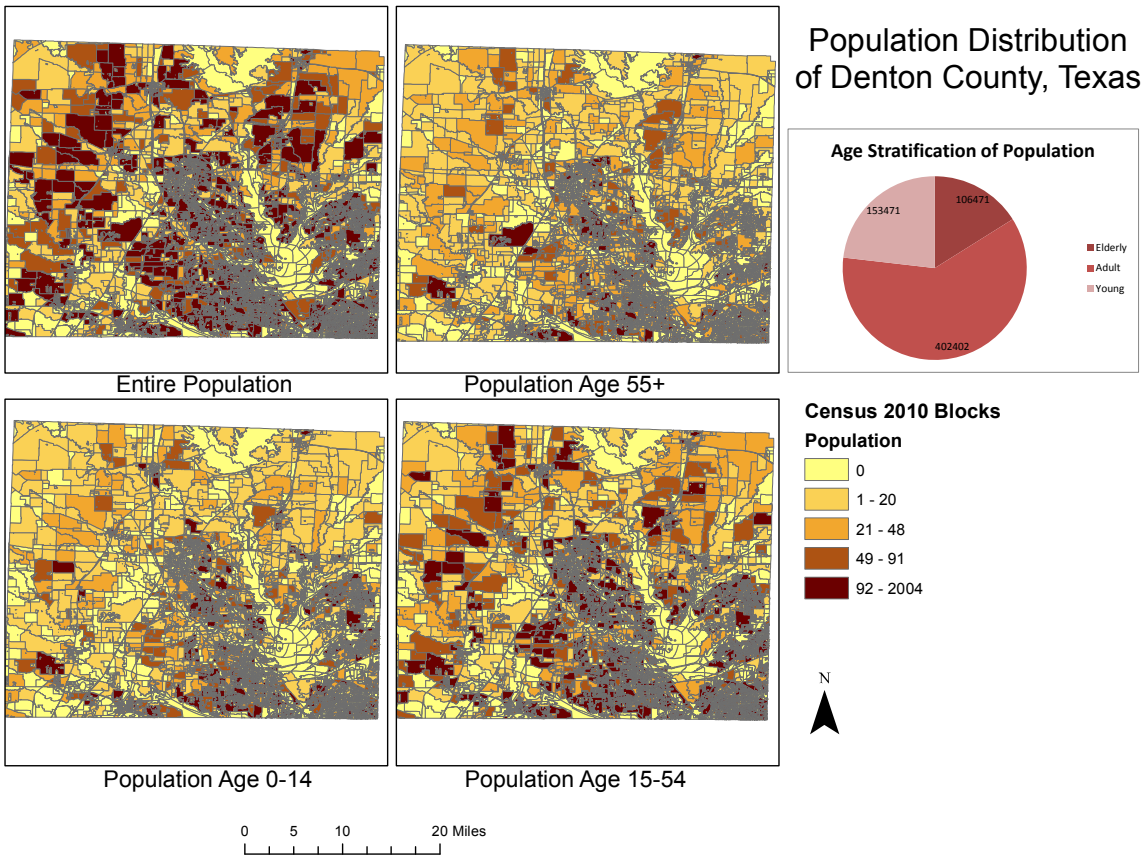

FIGURE 6.4. Distribution of the entire population and for each age group for Denton, TX.

For this experiments, the lengths of the infectious period of each age group are calculated using probability distributions such that each distribution is an approximation to the data reported in the literature. Figure 6.5 depicts the probability distributions of the length of the infectious period in each of the groups previously defined. Group A includes individuals between 14 and 55 years old and their infectious period is calculated with a normal distribution with mean $\mu = 6$ and standard deviation $\sigma = 1.5$. Group B includes individuals with ages 0 to 14 and its represented with a Lévy distribution with location parameter $\mu = 7$ and scale parameter $c = 1$. The Lévy distribution is depicted in equation 21. Finally, group C includes individuals 55 years old or more and their infectious period is calculated with a Lévy distribution with $\mu = 8$ and $c = 4$. Ultimately, during the simulation, every individual

from each age group is assigned an infectious period following the probability distribution of its group.

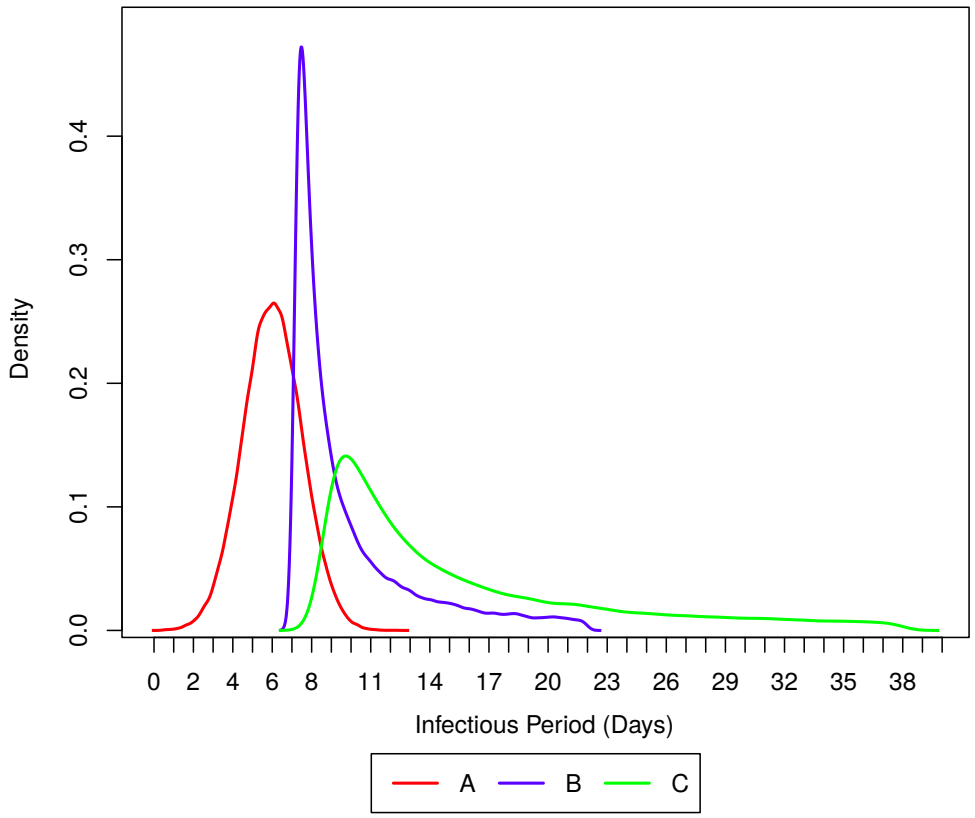

FIGURE 6.5. Probability distributions of the infectious period for the three age groups

The output from every simulation is the properties of the epidemic trajectory for that population, disease and geographic region. We compare the properties of two trajectories following the methodology defined in 3.3. Since these experiments are designed to explore the effects of multiple geographic, immunological, and socio-demographic characteristics of the population, we assume all other disease parameters to be identical for all groups.

6.3.1. Experiment I

The first experiment aims to display the consistency of the methodology to compare the properties of similar epidemic trajectories in Sumter, FL and Denton, TX, respectively. We conducted 2 sets of 50 outbreak simulations in each population in which all individuals are considered to be adults (group A). The length of their infectious period is set to 6, such

that the duration of their IP is the mean of the probability distribution of this age group. In this experiment, the first set of 50 samples from a region were compared with the second set of the same region. This comparison aims to highlight the correctness of the methodology whenever the trajectories are expected to be practically equal.

In Table 6.2 we present a summary of the data from each set of simulations. In the table we observe that the statistics of both samples are almost identical among each region. To test the similarity of the distribution of the data of both samples we performed a Welch t-test and a Levene test.

TABLE 6.2. Summary of data of Experiment I for Sumter and Denton.

	Sumter				Denton			
Sample	n	MFI	Std. Dev	Std. Err. Mean	n	MFI	Std. Dev	Std. Err. Mean
1	50	41.1962	1.0973	0.1551	50	50.1951	0.2272	0.0321
2	50	40.9184	1.0903	0.1541	50	50.1955	0.2500	0.0353

To test the similarity of the mean force of infection (MFI) between the samples of each region, we conducted a Welch t-test with the use of the function "t.test" in R[82]. The input for the function were the subsets $\pi(i)$ from each simulation from each region. The results are presented in table 6.3.

TABLE 6.3. Results from the Welch t-test in Experiment I for Sumter and Denton

Sumter					Denton				
t	df	p	95 % CI		t	df	p	95 % CI	
			Lower	Upper				Lower	Upper
1.2698	97.996	0.2072	-0.1563	0.7119	-0.0084	97.122	0.9933	-0.0952	0.0944

By analyzing the results from table 6.3, we conclude that there was no significant difference between the MFIs of Sumter and Denton, respectively. In Sumter, the H_0 is accepted since $p = 0.2072 > \alpha = .05$ for sample 1 (MFI=41.1962, SD=1.0973) and sample 2 (MFI=40.9184, SD=1.0903). Similarly in Denton, the H_0 is accepted since $p = 0.9933 > \alpha = .05$ for sample 1 (MFI=50.1951, SD=0.2272) and sample 2 (MFI=50.1955, SD=0.2500).

Additionally, we conducted a Levene test to determine the similarity in variances. As described in section 3.3, the test required the conversion of the original data. The conversion was performed following equation 7. Table 6.4 depicts the data summary for the modified samples, including the modified mean force of infection (MMFI), for Sumter and Denton. Similarly to the data in table 6.3, the statistics of the samples of both regions are similar among counties.

TABLE 6.4. Summary of modified data used in Experiment I for the Levene test for Sumter and Denton

	Sumter				Denton			
Sample	n	MMFI	Std. Dev	Std. Err. Mean	n	MMFI	Std. Dev	Std. Err. Mean
1	50	0.9084	0.6016	0.0850	50	0.1955	0.1199	0.0169
2	50	0.8093	0.7233	0.1022	50	0.1994	0.1511	0.0213

The Levene test was conducted with an implementation in R of the method described in Section 3.3. In this implementation, the calculation of the test statistic is obtained with a two-sided t-test on the modified data of each county. The results for Sumter and Denton are presented in table 6.5.

TABLE 6.5. Results from the Levene test in Experiment I for Sumter and Denton

Sumter					Denton				
t_l	df	p	95 % CI		t	df	p	95 % CI	
			Lower	Upper				Lower	Upper
0.745	98	0.4581	-0.1649	0.3631	-0.1437	98	0.886	-0.0580	0.0502

From the results depicted in table 6.5, we conclude that there was no significant difference in the variances of Sumter and Denton. In Sumter, the H_0 is accepted since $p = 0.4581 > \alpha = .05$ for sample 1 (MMFI=0.9084, SD=0.6016) and sample 2 (MMFI=0.8093, SD=0.7233). Similarly in Denton, the H_0 is accepted since $p = 0.886 > \alpha = .05$ for sample 1 (MMFI=0.1955, SD=0.1199) and sample 2 (MMFI=0.1994, SD=0.1511). Essentially, the

acceptance of the H_0 indicates that the degree of cooperation towards the progression of the disease is similar for both samples.

In general, since the means and variances of both samples for both regions are significantly equivalent, we conclude that there is strong evidence that both samples were extracted from the same population. Further, this experiment highlights the consistency of the method to compare the properties of two trajectories that result from an epidemic with similar characteristics.

6.3.2. Experiment II

To expand the analysis presented in Experiment I, we conducted a similar experiment in which the length of the infectious period of each individual is based on its age group and the quality of the immune response of its group. Similarly to Experiment I, we conducted 2 sets of 50 outbreak simulations for Sumter and 2 sets of 50 outbreak simulations for Denton. In this experiment, the IP of each individual is determined by its group membership and the probability distribution described in section 6.3. Finally, the first set of each region was compared with the second set of the same region to highlight the correctness of the methodology to capture the equality of two similar trajectories in spite of the effects of PIR. The summary of the data used in this experiment, for both Sumter and Denton, is presented in Table 6.6.

TABLE 6.6. Summary of data used in Experiment II for Sumter and Denton

	Sumter				Denton			
Sample	n	MFI	Std. Dev	Std. Err. Mean	n	MFI	Std. Dev	Std. Err. Mean
1	50	67.1596	0.9270	0.1311	50	85.5206	0.2965	0.0419
2	50	66.9981	0.8792	0.1243	50	85.5711	0.2396	0.0338

The Welch t-test was conducted using the same approach as in Experiment I. The results obtained from this calculation are presented in Table 6.7.

TABLE 6.7. Results from the Welch t-test in Experiment II for Sumter and Denton

Sumter					Denton				
t	df	p	95 % CI		t	df	p	95 % CI	
			Lower	Upper				Lower	Upper
0.894	97.727	0.3735	-0.1970	0.5201	-0.936	93.873	0.3517	-0.1575	0.0565

By analyzing the results from table 6.7, we conclude that there was no significant difference in the MFIs of Sumter and Denton. In Sumter, the H_0 is accepted since $p = 0.3735 > \alpha = .05$ for sample 1 (MFI=67.1596, SD=0.9270) and sample 2 (MFI=66.9981, SD=0.8792). Similarly in Denton, the H_0 is accepted since $p = 0.3517 > \alpha = .05$ for sample 1 (MFI=85.5206, SD=0.2965) and sample 2 (MFI=85.5711, SD=0.2396).

TABLE 6.8. Summary of data used in Experiment II for the Levene test for Sumter and Denton

	Sumter				Denton			
Sample	n	MMFI	Std. Dev	Std. Err. Mean	n	MMFI	Std. Dev	Std. Err. Mean
1	50	0.7449	0.5672	0.0802	50	0.2357	0.1864	0.0263
2	50	0.6906	0.5351	0.0756	50	0.1780	0.1603	0.0226

Similarly to Experiment I, we converted the original data to conduct the Levene test. The conversion was performed following equation 7 and Table 6.8 depicts a summary for the modified samples. The results of this test are depicted in Table 6.9.

TABLE 6.9. Results from the Levene test in Experiment II for Sumter with PIRBoth

Sumter					Denton				
t_l	df	p	95 % CI		t	df	p	95 % CI	
			Lower	Upper				Lower	Upper
0.4924	98	0.6236	-0.1645	0.2731	1.6566	98	0.1008	-0.0114	0.1266

From the results depicted in table 6.9, we conclude that there was no significant

difference in the variances of Sumter and Denton. In Sumter, the H_0 is accepted since $p = 0.6236 > \alpha = .05$ for sample 1 (MMFI=0.7449, SD=0.5672) and sample 2 (MMFI=0.6906, SD=0.5351). Similarly in Denton, the H_0 is accepted since $p = 0.1008 > \alpha = .05$ for sample 1 (MMFI=0.2357, SD= 0.1864) and sample 2 (MMFI=0.1780, SD=0.1603). Essentially, the acceptance of the H_0 indicates that the degree of cooperation towards the progression of the disease is similar for both samples.

Overall, considering that the means and variances of both samples for both regions are significantly equivalent, we conclude that there is strong evidence that both samples were obtained from the same population. Further, this experiment emphasizes the consistency of the method to compare the properties of two trajectories that represent epidemics with similar characteristics in spite of the incorporation of PIR.

6.3.3. Experiment III

The third experiment aims to identify variations in the properties of the trajectories of an epidemic in a region with a homogeneous population and a region with a population with differentiated PIR. By introducing the effects of PIR in the population, we expect to observe a difference in the properties of the trajectory of the epidemic. The potential difference was measured by performing a Welch t-test and a Levene test on the data points. A summary of the data used for this experiment is presented in Table 6.10. In the table we observe a variation on the central tendency parameters. However, this variations need to be statistically measured to reach any conclusions.

TABLE 6.10. Summary of data used in Experiment III for Sumter and Denton in a homogeneous population and a population with differentiated PIR

	Sumter				Denton			
Sample	n	MFI	Std. Dev	Std. Err. Mean	n	MFI	Std. Dev	Std. Err. Mean
1	50	41.1962	1.0973	0.1551	50	50.1951	0.2272	0.0321
2	50	66.9981	0.8792	0.1243	50	85.5711	0.2396	0.0338

Using the same approach as in Experiment I and II, we conducted the Welch t-test on the data summarized in Table 6.10. The results from this test are presented in Table 6.11.

TABLE 6.11. Results from the Welch t-test in Experiment III for Sumter and Denton in a homogeneous population and a population with differentiated PIR

Sumter					Denton				
t	df	p	95 % CI		t	df	p	95 % CI	
			Lower	Upper				Lower	Upper
-129.7	93.554	2.2×10^{-16}	-26.1	-25.4070	-757.2978	97.727	2.2×10^{-16}	-35.4687	-35.2833

From the results depicted in table 6.11, we conclude that there was a significant difference between the MFIs of Sumter and Denton. In Sumter, the H_0 is rejected since $p = 2.2\times10^{-16} < \alpha = 0.05$ for sample 1 (MFI=41.1962, SD=1.0973) and sample 2 (MFI=66.9981, SD=0.2015). Similarly in Denton, the H_0 is rejected since $p = 2.2 \times 10^{-16} < \alpha = 0.05$ for sample 1 (MFI=50.1951, SD=0.2272) and sample 2 (MFI=85.5711, SD=0.2396). The difference in the means indicates that the distribution of the first population is different than the distribution of the second population. Considering that the MFI of each sample is the central tendency value for the force of infection, we can conclude that there was a variation of this parameter during the progression of the disease. Additionally, the small width of the confidence intervals for Sumter (0.7897) and Denton (0.1854), highlight the high precision of the result.

To compare the variances of the two samples of each population we conducted the Levene test. We converted the data following equation 7 and Table 6.12 depicts a summary for the modified samples. The modified samples were used to perform the test and the results obtained are presented in Table 6.13.

From the results depicted in table 6.13, we conclude that there was no significant difference in the variances of Sumter and Denton. In Sumter, the H_0 is accepted since $p = 0.4581 > \alpha = .05$ for sample 1 (MMFI=0.9084, SD=0.6016) and sample 2 (MMFI=0.8093,

TABLE 6.12. Summary of data used in Experiment III for Sumter and Denton in a homogeneous population and a population with differentiated PIR

	Sumter				Denton			
Sample	n	MMFI	Std. Dev	Std. Err. Mean	n	MMFI	Std. Dev	Std. Err. Mean
1	50	0.9084	0.6016	0.0850	50	0.1955	0.1199	0.0169
2	50	0.6906	0.5351	0.0756	50	0.1780	0.1603	0.0226

TABLE 6.13. Results from the Levene test in Experiment III for Sumter and Denton in a homogeneous population and a population with differentiated PIR

Sumter					Denton				
t_l	df	p	95 % CI		t	df	p	95 % CI	
			Lower	Upper				Lower	Upper
1.9128	98	0.0586	-0.0081	0.4438	0.6165	98	0.539	-0.0387	0.0736

SD=0.7233). Similarly in Denton, the H_0 is accepted since $p = 0.886 > \alpha = .05$ for sample 1 (MMFI=0.1955, SD=0.1199) and sample 2 (MMFI=0.1994, SD=0.1511). The acceptance of the H_0 indicates that the shape of the distribution did not change between samples. Further, the similarity reflects a comparable degree of contribution towards the epidemic from the census blocks in both populations. Furthermore, the small width in the confidence intervals for Sumter (0.4357) and Denton (0.0349), indicate a high precision in the results.

The results from the tests revealed that the shape of the distribution of the properties of the trajectories is similar between samples for homogeneous populations and populations with differentiated PIR in both regions. However, the results also exhibited a shift in mean force of infection between the samples of those aforementioned populations. Hence, this result exposed fluctuations in the properties of the epidemic trajectories produced by the incorporation of a differentiated PIR into the population.

6.3.4. Experiment IV

The final experiment depicts and measures the effects PIR produced in the outbreak dynamics in the county of Sumter compared with the effects it generated in Denton county. To examine those effects, we conducted simulations in homogeneous populations and populations with differentiated PIR in both Sumter and Denton. Since, these experiments explore the effects PIR produced in the outbreak dynamics, all individuals follow a homogeneous contact profile. Figure 6.6 depicts the average number of infected individuals per day in both homogeneous populations and populations with differentiated PIR in Sumter and Denton. For ease of visualization, this figure does not portray the end of those epidemics and only presents results until day 1200. In the figure, we observe that the shape of the epidemic in the simulations in which the populations are homogeneous follows a traditional bell shape curve. However, the epidemics in which we incorporate PIR, the peak is followed by a long tail. The long tail is caused by the longer infectious periods of the elderly and children populations. The prolonged epidemics are caused by individuals with a prolonged infectious period since they are presented with an increased opportunity to infect individuals in regions that, otherwise, will not be reached.

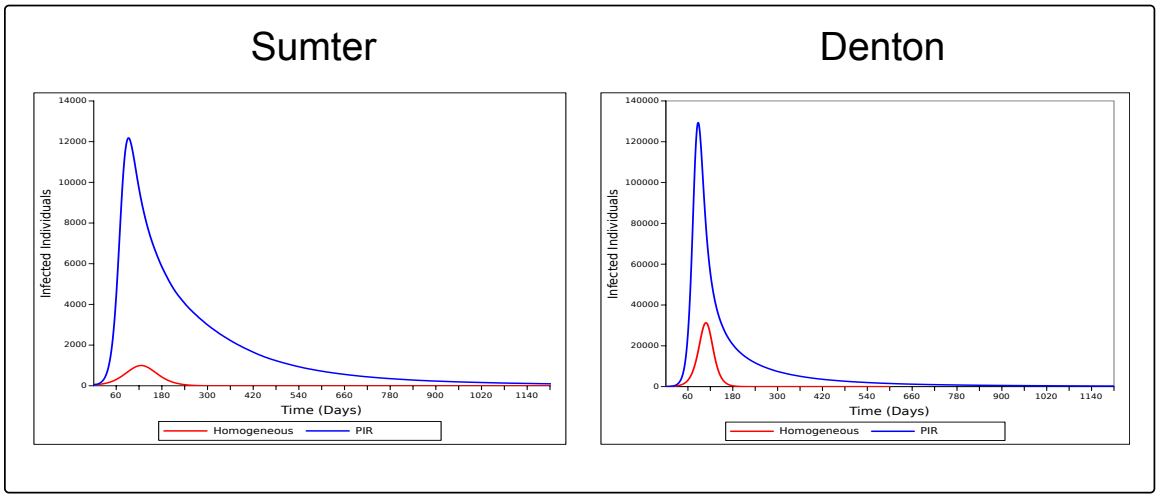

FIGURE 6.6. Number of infected individuals per day in homogeneous populations and populations with PIR in Sumter and Denton

To quantitatively estimate the effects PIR caused in a particular region, we measured

the variation of the mean force of infection of the epidemic in the experiments with homogeneous populations and those with differentiated PIR. In order to contrast the effects of PIR in both regions, we compared the variation in Denton with the variation in Sumter. In table Table 6.14 we present a summary of the information depicted in Table 6.10 of the results from the experiments with homogeneous populations and populations with differentiated PIR for Sumter and Denton.

TABLE 6.14. Mean force of infection for Sumter and Denton in a homogeneous population and a population with differentiated PIR

	Sumter	Denton
Experiment	MFI	MFI
Homogeneous	41.1962	50.1951
PIR	66.9981	85.5711

However, since both populations are different, aiming to compare the effects of PIR in the two regions using the data from Table 6.14 is not accurate. In our model, each interaction, and, consequently, the mean force of infection (MFI), is driven by three main factors: contact rate, population, and infectious period. Those factors are unique in each census block since they are calculated based on the demographic characteristics of the population of every block. Furthermore, the characteristics of a census block might change in experiments with a homogeneous population and in experiments that include PIR. Subsequently, in order to compare two regions, it is necessary to normalize the MFI of each experiment to a common scale. The force of infection is converted to the normalized force of infection, $\widehat{x}$, with the use of equation 42

$$\widehat{x_1} = \frac{\bar{x}}{CR_A * P_A * IP_A + CR_C * P_C * IP_C + CR_E * P_E * IP_E} \tag{42}$$

in which P_A, P_C, P_E, CR_A, CR_C, CR_E, IP_A, IP_C, IP_E are the population, contact rate, and infectious period of the adults, children, and elderly, respectively. However, since

these experiments were constructed with the goal of exploring the effects PIR produced in the outbreak dynamics, all individuals follow a homogeneous contact profile. Consequently, the value of the contact rate is equal for all groups and equation 42 can be simplified to equation 43.

$$\widehat{x_2} = \frac{\bar{x}}{P_A * IP_A + P_C * IP_C + P_E * IP_E} \tag{43}$$

The values for the population of each group are obtained from Table 6.1. Further, the infectious period for each age group is the rounded mean of 20000000 values extracted from each age probability distribution presented in Figure 6.5. The value of the infectious period of each age group are: IP_A= 6, IP_C= 10, and IP_E= 15. Furthermore, for a homogeneous population, in which all individuals are considered to be adults, the value for IP is six for all individuals. Table 6.15 depicts the normalized values for the mean force of infection (NMFI) for each experiment. The normalized values for both experiments were calculated using equation 43.

TABLE 6.15. Normalized mean force of infection for Sumter and Denton in a homogeneous population and a population with differentiated PIR

	Sumter	Denton
Experiment	NMFI	NMFI
Homogeneous	0.0000734964	0.00001262553
PIR	0.0000568708	0.00001513098
Variation	0.0000166252	0.000002505453

We observe in Table 6.15 that the variation caused by incorporating PIR in Sumter is six times greater than the variation produced in Denton. The increased variation in Sumter indicates that the effects of incorporating PIR in regions with an increased elderly population are greater than those in which the population is more diverse. In general, these results indicate that the age demographics, and subsequently the immune response quality, of

a region greatly influences the properties of the trajectory of the epidemic and, consequently, the outbreak dynamics.

6.4. Summary

In this chapter, we introduced a formal definition of the global stochastic contact framework (GSCF) and explored the effects of the population immune response (PIR) in the outbreak dynamics. We conducted that study in the counties of Sumter, FL and Denton, TX. with the use of a computational implementation of this framework. In the GSCF, the population is spatially distributed in geographic regions and individuals are assigned a group memberships based on their socio-physiological characteristics. To reflect those characteristics, we use the group membership to establish the value of the contact rate, mobility, group affinity, and the quality of the immune response of the individuals.

In this framework, interactions are modeled at the regional level such that each contagious individual from a region will contribute to the number of possible transmissions its region initiates. However, since this framework aims to mimic the dynamics of the SIR, the transmissions from all regions are combined and, subsequently, distributed among the entire set of regions proportionally to their number of infectious individuals. Once a region is selected as the origin of the transmission, the infection can be directed to an individual of the same region or be aimed to an individual in an external region. If a transmission is aimed to an external region, we determine its destination by the use of an interaction coefficient.

In this research, we define the interaction coefficient as the product of the populations of the regions divided by the distance between them. The interaction coefficient represents the notion that populous regions have a higher interaction rate between themselves than with more sparsely populated regions. Also, the likelihood of interaction is directly impacted by the distance between regions. Once the destination region has been selected, it is necessary to determine the individual to whom the transmission will be directed to. This resolution is determined by the group membership of the individual originating the infection and the affinity of its group to the other groups. This process is repeated until there are no more infectious individuals.

We conducted four experiments that highlighted the crucial role of the demographic and immunological characteristics of the population during the progression of an epidemic in a geographic space. All experiments simulate the spread of an influenza epidemic in the counties of Sumter, Florida and Denton, Texas. Those counties were specifically selected for their differences in age demographics. In these experiments, we simulated the progression of the disease in homogeneous populations and populations with differentiated PIR. In the homogeneous populations all individuals are assumed to be adults. However, in the populations with PIR the age of the individuals is determined by the use of the census data of 2010. Consequently, the age group of the individual is used to determine the length of its infectious period.

In the first and second experiment, the statistical method clearly indicated consistency while comparing the properties of similar trajectories. The trajectories compared originated from either two regions with homogeneous populations or two regions with populations with PIR. The results indicated the correctness of the method to compare the properties of two trajectories that result from an epidemic with similar characteristics. Further, the third experiment investigated the variations in the properties of the trajectories of an epidemic in a region with a homogeneous population and a region with a population with differentiated PIR. The results from this experiment highlighted the change in the force of infection after incorporating PIR into the population.

Finally, the last experiment depicted and measured the effects PIR produced in the outbreak dynamics in Sumter county compared with the effects it generated in Denton. We observed that the variation caused by incorporating PIR in Sumter is six times greater than the variation produced in Denton. The increased variation in Sumter indicates that the effects of incorporating PIR in regions with an increased elderly population are greater than those in which the population is more diverse. In general, these results indicated that the age demographics and the immune response quality of individuals in a geographic region greatly influence the outbreak dynamics.

CHAPTER 7

SUMMARY AND FUTURE WORK

7.1. Summary

Computational epidemiology aims to provide new and robust tools to public health officials for policy and decision making. In this research, we highlighted the importance of developing tools that incorporate social/behavioral and physiological characteristics to the populations of study. Furthermore, we developed methods and models that integrate those characteristics and measure the effects they inflict on the outbreak dynamics. In the following section we present a summary of the methods and results from each chapter.

7.1.1. Epidemics in Time and Space

In Chapter 3, we introduced the concept of epidemic trajectory and studied some of its properties. An epidemic trajectory is defined as a path, with multiple branches, of a contagion process that spreads in a fuzzy neighborhood. Further, we considered that the properties that define the characteristics of that path are the force of infection and the transmission trend. To analyze those properties, we developed a model that simulates the transmission of the disease in a population. In this model, the disease spreads at the regional level, such that the transmission of disease from a region A to region B is represented by a directed edge between A and B. This approach allows to represent all transmissions as a complete directed weighted graph. Furthermore, we developed a method to compare two epidemic trajectories in two populations.

This method determines statistical significance of the variations in the properties of the trajectory of two epidemics and asserts the statistical significance of incorporating variations to the population characteristics. By means of statistical tests, we contrasted the characteristics of two epidemic trajectories. If statistical similarity is observed, we conclude that there is strong evidence that both samples were extracted from the same trajectory. This result discourages the incorporation of the parameter that we expected would cause a significant variation in the epidemic. Contrarily, the rejection of the similarity by the

test statistics, indicates that the variations were not caused by randomness and further investigation on the parameter that caused that fluctuation must be conducted.

7.1.2. Modeling Social/Behavioral Characteristics

In Chapter 4, we introduced an infectious-disease transmission model in which the contact patterns between individuals create a social network in which the infectious individuals and the population at large interact. Based on this social network, different person-to-person interaction probabilities can be expressed and hence integrated into the model. The model was further expanded to include clusters of individuals representing differentiated social groups. The social network determines the interaction behavior in each group. Group behaviors are expressed by their corresponding inter and intra cluster interaction frequency and affinity. Finally, we assessed the effects of incorporating differentiated behavior into the model and its relevance to the outbreak by implementing and constructing multiple experiments. The results indicate that the inclusion of heterogeneity and diverse group behavior into an outbreak simulation is important to achieve a more realistic and complete prediction.

7.1.3. Modeling Physiological Differences

In Chapter 5, we highlighted the importance of integrating the viral/bacterial load during an infectious outbreak simulation. This value is utilized to measure the severity of the disease and the host's capacity to transmit it. More importantly, we emphasized the direct relationship between the quality of the immune response and the quantity of the viral/bacterial load. A stronger immune response controls the growth of the pathogen more efficaciously. Further, the viral/bacterial load is utilized to determine the length of the infectious periods. A minimum transmission threshold is established as the minimum quantity of viral/bacterial load required for an individual to become contagious. Once the viral/bacterial load surpasses this threshold, the individual becomes contagious and capable to infect others until his or her load falls below that threshold.

We defined the concept of population immune response (PIR). PIR captures the collective immune response of individuals in a population represented by the superposition of

individual immune responses. A computational model that captures the effects of PIR on the outbreak dynamics has been presented. This model is divided into three compartments: population and disease database, immune competence, and infectious disease outbreak simulation. The population and disease database includes demographic information that is used to determine the efficacy of the immune response of the individuals. The database also contains disease-specific data utilized to establish the transmission rate and the value of the minimum infection threshold. The immune response upon infection of each individual is determined by the immune competence module. In this research, we exemplified this calculation with a mathematical approach that models the immune response of an individual during viral infection. The mathematical model determines the value of the viral/bacterial load in each iteration. This value is used to establish the length of the infectious periods by implementing the minimum infection threshold. The infectious disease outbreak model simulates the spread of an infectious disease among a population. The population is created based on the data obtained from the other two compartments. This model must incorporate unique characteristics for every individual and simulate the interactions between them. The execution of this model captures the disease dynamics during an infectious outbreak.

Last but not least, we analyzed some of the effects PIR has on the disease dynamics. Our results suggested that it is essential for the public health establishment to increase their understanding of the augmented role of the immune response as a function of the socio-demographic characteristics of the population and its prominent role during the epidemic progression.

7.1.4. Modeling Outbreaks in Geographic Context

In Chapter 6, we introduced a formal definition of the global stochastic contact framework (GSCF) and explored the effects of the population immune response (PIR) in the outbreak dynamics. We conducted that study in the counties of Sumter, FL and Denton, TX with the use of a computational implementation of this framework. In the GSCF, the population is spatially distributed in geographic regions and individuals are assigned a group memberships based on their socio-physiological characteristics. To reflect those characteris-

tics, we used the group membership type to establish the value of the contact rate, mobility, group affinity, and the quality of the immune response of the individuals.

In this framework, interactions are modeled at the regional level such that each infectious individual from a region contributes to the number of possible transmissions its region initiates. However, since this framework emulates the dynamics of the SIR, the transmissions from all regions are combined and, subsequently, distributed among the entire set of regions proportionally to their number of infectious individuals.

We conducted four experiments that highlighted the crucial role of the demographic and immunological characteristics of the population during the progression of an epidemic in a geographic space. All experiments simulate the spread of an influenza epidemic in the counties of Sumter, Florida and Denton, Texas. These counties were specifically selected for their great differences in age demographics. In these experiments, we simulated the progression of the disease in homogeneous populations and populations with differentiated PIR.

The results highlight the change in the force of infection after incorporating PIR into the population. Furthermore, we conducted an experiment that compared the effects PIR produced in the outbreak dynamics in Sumter county compared with the effects it generated in Denton county. We observed that the variation caused by incorporating PIR in Sumter is six times greater than the variation produced in Denton. The increased variation in Sumter indicates that the effects of incorporating PIR in regions with an increased elderly population are greater than those in which the population is more diverse.

In general, this research highlights the importance of incorporating behavioral changes and immunological characteristics when investigating the spread of diseases in populations. We presented multiple methods and simulations to support the inclusion of those factors into mathematical and computational models. Based on the evidence presented, we encourage public health officials to carefully examine the effects of demographics and corresponding physiological characteristics when preparing for an infectious disease epidemic.

7.2. Future Work

Although the results presented in this document have highlighted the importance of incorporating behavioral changes and immunological characteristics during the study of the spread of diseases in populations, we recognize that this work can be further extended in the following ways:

7.2.1. Study of Clustering of Regions during the Epidemic Progression

When we introduced the concept of trajectory of an epidemic, we defined a methodology that analyzes the progression of the disease from one region to another. However, we consider that our methodology could be expanded to capture the infectious clustering of regions during an epidemic. An infectious cluster of regions represents a group of regions that mostly transmit the disease among themselves. However, some of the transmissions from this group are directed to regions outside the cluster. Identifying and studying this clusters is an important effort since it can lead to the implementation of strategies that could diminish the spreading of a pathogen in the population.

7.2.2. Effects in the Transmission Trend

The methodology to compare the properties of two epidemics presented in this research provides an analysis of the changes caused in the force of infection between two trajectories. However, the effects caused in the transmission trend are not analyzed. We consider that studying the transmission trend is important, since any changes in appeal or affinity of a region could result in changes in the degree to which regions interact. Future studies will focus in different characteristics of the population such as affinity and interaction coefficient.

7.2.3. Estimation of the quality of the immune response

In this research, we introduced a methodology to calculate the immune response quality of the individuals from the population with the use of demographics and an immune response model. However, given the scope of this research and the limited data available, we calculated the population immune response quality with the use of probability distributions

and the reported data from the World Health Organization. We consider that our methodology could be further augmented with the use of realistic data. However, such data has yet to be collected. Furthermore, considering that our study highlighted the augmented role of the demographics of a population in shaping the outbreak dynamics, we encourage public health officials to capture the required information.

7.2.4. Study of combination of heterogeneous parameters

In this research, we presented a profound analysis of the importance of incorporating behavioral changes and immunological characteristics during the study of the spread of diseases. We highlighted the effects produced by those characteristics in the epidemic trajectory and, eventually, the outbreak dynamics. However, we consider that this study could be expanded by introducing variations in the rest of the disease and population parameters. Some of the parameters that could be expanded are the contact rate, mobility, and transmissibility. In general, future studies would analyze the effects of multiple combinations of factors in the epidemic progression and, ultimately, the disease dynamics.

APPENDIX

LIST OF SYMBOLS

TABLE A.1. List of symbols for Chapter 2

Symbol ⟨Name⟩	Definition
S	Susceptible
L	Latent
I	Infectious
R	Recovered
ϵ ⟨epsilon⟩	Latency rate
γ ⟨gamma⟩	Recovery rate
β ⟨beta⟩	Force of infection

TABLE A.2. List of symbols for the methodology of Chapter 3

Symbol ⟨Name⟩	Definition
P	Population
p_i	Individual from the population
ϵ ⟨epsilon⟩	Disease type
R	Geographic space
r_a	Region from the geographic space
$\mathcal{M}_t(r_a, r_b)$	Numerical value of transmissions commenced by r_a and directed to r_b
$\mathcal{M}_t$	Matrix that stores a sample trajectory
$T_{P,R,\epsilon}$	Total possible trajectories for a P, R, and ϵ
$\mathcal{G}(t)$	Graph representation of $\mathcal{M}_t$
$\gamma(t,i)$ ⟨gamma⟩	Total number of interactions initiated by individuals from region r_i
$\delta(t,j)$ ⟨delta⟩	Total number of interactions in which individuals from r_j were contacted
$deg^+(i)$	Out-degree of vertix i in $\mathcal{G}(t)$
$deg^-(j)$	In-degree of vertix i in $\mathcal{G}(t)$
$\Delta(t)$ ⟨Delta⟩	Transmission trend of the epidemic for trajectory t
$\Gamma(t)$ ⟨Gamma⟩	Force of infection of R
$\tau(P,G,\varepsilon)$ ⟨tau⟩	Set of all sets of $\Gamma(t)$ for a P, R, and ϵ
$\Upsilon(P,G,\varepsilon)$ ⟨Upsilon⟩	Set of all sets of $\Delta(t)$ for a P, R, and ϵ
$\|E\|$	Number of experiments
$\pi(i)$	Properties of the force of infection
$\eta(i)$ ⟨eta⟩	Properties of the transmission trend
$\bar{x_i}$	Mean force of infection of a region
$X(E,t)$	Group of all $\bar{x_i}$ of a trajectory
$\overline{X(E,t)}$	Mean force of infection (MFI)

TABLE A.3. List of symbols for the comparison methods of Chapter 3

Symbol ⟨Name⟩	Definition
H_0	Null hypothesis
α ⟨alpha⟩	Level of significance
H_a	Alternative hypothesis
t_w	Welch t-test t statistic
$\bar{x_i}$	Sample mean for sample i
s_i^2	Sample standard deviation for sample i
n_i	Sample size for sample i
v	Degrees of freedom
$t(\frac{\alpha}{2}, v)$	Critical value of Welch t-test
CI	Confidence interval
t_l	Levene test t statistic
X_s	Original sample
$x_{i,s}$	Element from X_s
$\tilde{X_s}$	Median of sample X_s
Y_s	Modified sample
$y_{i,s}$	Element from Y_s
$S_{Y_1-Y_2}$	Standard error of the difference between the means
LP	Latent period
IP	Infectious period

TABLE A.4. List of symbols for Chapter 4

Symbol ⟨Name⟩	Definition
P	Population
p_i	Individual from the population
N	Population size
S	Susceptible
L	Latent
I	Infectious
R	Recovered
CR	Contact rate
t_i	Timestep in the epidemic
t_π	End of the epidemic
LP	Latent period i
IP	Infectious period
c_i	Cluster of the population
Ω ⟨Omega⟩	All groups from the population
$\mathcal{A}(c_i, c_j)$	Affinity function between two clusters
δ_j ⟨delta⟩	Affinity to a cluster from all clusters
γ_i ⟨gamma⟩	Affinity from a cluster to all clusters
ϱ ⟨rho⟩	Pivot group
CR_ϱ	Contact rate of the pivot group
$\mathcal{A}(c_i, c_\varrho)$	Affinity to the pivot group
CR_j	Contact rate of group j

TABLE A.5. List of symbols for Chapter 5

Symbol ⟨Name⟩	Definition
$\mathcal{R}^0$	Average immune response
$\mathcal{R}^+$	Hyperimmune response
$\mathcal{R}^-$	Hypoimmune response
$\mathcal{R}^*$	Immune
vbl	Viral/bacterial load
vbl^*	Transmission threshold
PIR	Population immune response
LP	Latent period i
IP	Infectious period
IP_A	Infectious period of group A

TABLE A.6. Model parameters for the Wodarz model in Chapter 5

Symbol ⟨Name⟩	Definition
λ ⟨lambda⟩	Production rate of uninfected cells
d	Death rate of uninfected cells
β ⟨beta⟩	Infection rate of uninfected cells by viruses
a	Death rate of infected cells
p	Lysis rate of infected cells by the CTL response
k	Production rate of virus by infected cells
u	Decay rate of viruses
q	Neutralization rate of viruses by antibodies
g	Development rate of antibodies in response of virus exposure
h	Decay rate of antibodies
c	Development rate of CTL in response to infected cells
b	Decay rate of CTL

TABLE A.7. List of symbols for the socio-physiological characteristics of individuals in Chapter 6

Symbol ⟨Name⟩	Definition
P	Population
p_i	Individual from the population
R	Geographic space
r_a	Region from the geographic space R
$\mathcal{S}(i)$	Residence r_l of individual i
G	Groups in the population
g_k	A group from the set of groups G
$\mathcal{G}(i)$	Group membership g_k of individual p_i
$\mathcal{M}(g_k, r_l)$	Set of individuals from group g_k and reside in region r_l
CR_i	CR of element p_i
CR_{g_k}	Contact rate of members of group g_k
m_i	Proportion of contacts p_i will direct to individuals in its own region (mobility)
m_{g_k}	Mobility of individuals that are members of group g_k
$\mathcal{A}(g_k, g_e)$	Affinity between any individual from group g_k and any individual in group g_e
$\mathcal{GA}(g_k, g_e)$	Group affinity matrix
ν_{g_k} ⟨nu⟩	Sum of probabilities for a group g_k to contact all other groups
IR_i	Quality of the immune response of individual p_i
I_{g_k}	Immune response quality value of individuals that are members of group g_k.

TABLE A.8. List of symbols for the contact model in Chapter 6

Symbol ⟨Name⟩	Definition
S	Susceptible
L	Latent
I	Infectious
R	Recovered
CR	Contact rate
$\mathcal{IN}(g_k, r_l)$	Set of infectious individuals in group g_k and region r_l
$ICo(r_l, t)$	Number of infectious contacts originating from region r_l in a given time t
$\kappa(t)$ ⟨kappa⟩	Set of regions that have at least one infectious individual at timestep t
$\xi(t)$ ⟨xi⟩	Total number of infectious contacts on timestep t
$\lambda(r_m, t)$ ⟨lambda⟩	Regional attenuation rate of region r_m
$S(r_m, t)$	Set of susceptible individuals in region r_m at time t
P_{r_m}	Set of total individuals in region r_m at time t
$\eta(t)$ ⟨eta⟩	General attenuation rate
β ⟨beta⟩	Transmissibility
$\Omega(t)$ ⟨Omega⟩	Total number of transmissions in a given time t
$\mathcal{IR}(r_l)$	Set of infectious individuals in region r_l
$D(r_l)$	Regional distribution rate of $\Omega(t)$
$\Gamma(r_l, t)$ ⟨Gamma⟩	Total number of transmissions a region r_l initiates at timestep t

TABLE A.9. List of symbols for the contact model in Chapter 6 cont.

Symbol ⟨Name⟩	Definition
g_k	Group of the individual p_i that its commencing the infectious contact
$E(r_l, g_k, t)$	Group rate of $\Gamma(r_l, t)$
$\Psi(r_l, g_k, t)$ ⟨Psi⟩	Number of transmissions from region r_l initiated by an individual from group g_k at timestep t
$\rho_{r_l,t}$ ⟨rho⟩	Array with the group membership of the initiators of all transmissions commenced by r_l
$L\Gamma(r_l, t)$ ⟨Gamma⟩	Number of transmissions that originate in region r_l and are directed to region r_l in time t
$G\Gamma(r_l, t)$ ⟨Gamma⟩	Number of transmissions that originate in region r_l and are directed to all regions besides r_l in time t
γ_{r_l} ⟨gamma⟩	Proportion of transmissions initiated by individuals from region r_l that are directed to individuals within region r_l
$\varrho(r_l)$ ⟨rho⟩	Set of individuals that are located in region r_l.
$IC(r_l, r_m)$	Interaction coefficient between region r_l and r_m
$ICN(r_l, r_m)$	Normalized interaction coefficient between region r_l and r_m
E	Normalized interaction coefficient matrix
e_{r_l,r_m}	Normalized interaction coefficient between region r_l and r_m in matrix E
η_{r_l} ⟨eta⟩	Sum of interaction coefficients from a region r_l to all other groups
$\sigma(r_l)$ ⟨sigma⟩	Interaction coefficient from all regions to a region r_l
r_l	Region that originates the transmission
r_m	Destination region
g_k	Group membership of the originator of the transmission
g_e	Group membership of the individual that the disease is transmitted
p_j	Individual to whom the disease is transmitted
MFI	Mean force of infection
MMFI	Modified mean force of infection

BIBLIOGRAPHY

[1] FikriM. Abu-Zidan, AlaaK. Abbas, AshrafF. Hefny, HaniO. Eid, and Michal Grivna, *Effects of seat belt usage on injury pattern and outcome of vehicle occupants after road traffic collisions: Prospective study*, World Journal of Surgery 36 (2012), no. 2, 255–259 (English).

[2] Shradha Agarwal and Paula J. Busse, *Innate and adaptive immunosenescence*, Annals of Allergy, Asthma and Immunology 104 (2010), no. 3, 183 – 190.

[3] E.S. Allman and J.A. Rhodes, *Mathematical models in biology: an introduction*, Cambridge University Press, 2004.

[4] S. Altizer, A. Dobson, P. Hosseini, P. Hudson, M. Pascual, and P. Rohani, *Seasonality and the dynamics of infectious diseases*, Ecology Letters 9 (2006), no. 4, 467–484.

[5] Sonia Altizer, Richard S. Ostfeld, Pieter T. J. Johnson, Susan Kutz, and C. Drew Harvell, *Climate change and infectious diseases: From evidence to a predictive framework*, Science 341 (2013), no. 6145, 514–519.

[6] E. Amouroux, S. Desvaux, and A. Drogoul, *Towards virtual epidemiology: an agent-based approach to the modeling of H5N1 propagation and persistence in North-Vietnam*, Intelligent Agents and Multi-Agent Systems (2008), 26–33.

[7] R.M. Anderson and R.M. May, *Infectious diseases of humans: dynamics and control*, Oxford University Press, USA, 1992.

[8] M. Anker, *Addressing sex and gender in epidemic-prone infectious diseases*, Geneva: World Health Organization Report (2007).

[9] Ramy A. Arnaout and Martin A. Nowak, *Competitive coexistence in antiviral immunity*, Journal of Theoretical Biology 204 (2000), no. 3, 431 – 441.

[10] J.L. Aron, *Mathematical modeling: the dynamics of infection*, Infectious disease epidemiology: Theory and practice (2000).

[11] A.H. Auchincloss and A.V. Diez Roux, *A new tool for epidemiology: the usefulness*

of dynamic-agent models in understanding place effects on health, American journal of epidemiology 168 (2008), no. 1, 1.

[12] K.K. Avilov and A.A. Romanyukha, *Mathematical modeling of tuberculosis propagation and patient detection*, Automation and Remote Control 68 (2007), no. 9, 1604–1617.

[13] N.T.J. Bailey, *The mathematical theory of epidemics*, Hafner Publishing Company, New York USA 194 (1957).

[14] Duygu Balcan, Bruno Gonçalves, Hao Hu, José J Ramasco, Vittoria Colizza, and Alessandro Vespignani, *Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model*, Journal of computational science 1 (2010), no. 3, 132–145.

[15] C.L. Barrett, K.R. Bisset, S.G. Eubank, X. Feng, and M.V. Marathe, *EpiSimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks*, Proceedings of the 2008 ACM/IEEE conference on Supercomputing, IEEE Press, 2008, pp. 1–12.

[16] A.S. Benenson, *Control of communicable diseases manual: an official report of the American Public Health Association*, American Public Health Association, 1995.

[17] CP Bhunu, W. Garira, and Z. Mukandavire, *Modeling hiv/aids and tuberculosis coinfection*, Bulletin of mathematical biology 71 (2009), no. 7, 1745–1780.

[18] Carlo Bianca, Ferdinando Chiacchio, Francesco Pappalardo, and Marzio Pennisi, *Mathematical modeling of the immune system recognition to mammary carcinoma antigen*, BMC Bioinformatics 13 (2012), 1–15 (English).

[19] Mathieu Blanc, Wei Yuan Hsieh, Kevin A Robertson, Steven Watterson, Guanghou Shui, Paul Lacaze, Mizanur Khondoker, Paul Dickinson, Garwin Sing, Sara Rodríguez-Martín, et al., *Host defense against viral infection involves interferon mediated down-regulation of sterol biosynthesis*, PLoS biology 9 (2011), no. 3, e1000598.

[20] Maciej Borkowski, Blake W Podaima, and Robert D McLeod, *Epidemic modeling with discrete-space scheduled walkers: extensions and research opportunities*, BMC Public Health 9 (2009), no. Suppl 1, S14.

[21] M. Bouden, B. Moulin, and P. Gosselin, *Epidemic propagation of west nile virus using a multi-agent geo-simulation under various short-term climate scenarios*, Proceedings of the 2008 Spring simulation multiconference, Society for Computer Simulation International, 2008, pp. 98–105.

[22] P. Braitstein, M. W. Brinkhof, F. Dabis, M. Schechter, A. Boulle, P. Miotti, R. Wood, C. Laurent, E. Sprinz, C. Seyler, D. R. Bangsberg, E. Balestre, J. A. Sterne, M. May, and M. Egger, *Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries*, Lancet 367 (2006), no. 9513, 817–824.

[23] S.C. Castle, *Clinical relevance of age-related immune dysfunction*, Clinical Infectious Diseases 31 (2000), no. 2, 578–585.

[24] R.K. Chandra, *Nutrition and the immune system: an introduction*, American Journal of Clinical Nutrition 66 (1997), no. 2, 460S.

[25] RK Chandra, *Nutrition and the immune system from birth to old age.*, European journal of clinical nutrition 56 (2002), S73.

[26] Dennis L. Chao, Miles P. Davenport, Stephanie Forrest, and Alan S. Perelson, *A stochastic model of cytotoxic t cell responses*, Journal of Theoretical Biology 228 (2004), no. 2, 227 – 240.

[27] J. Chin, *Control of Communicable Diseases Manual, 17*, Auflage. American Public Health Association, Washington, DC (2000).

[28] M. Cutolo, A. Sulli, S. Capellino, B. Villaggio, P. Montagna, B. Seriolo, and RH Straub, *Sex hormones influence on the immune system: basic and clinical aspects in autoimmunity*, Lupus 13 (2004), no. 9, 635.

[29] M.D. de Jong, C.P. Simmons, T.T. Thanh, V.M. Hien, G.J.D. Smith, T.N.B. Chau, D.M. Hoang, N.V.V. Chau, T.H. Khanh, V.C. Dong, et al., *Fatal outcome of human influenza a (h5n1) is associated with high viral load and hypercytokinemia*, Nature medicine 12 (2006), no. 10, 1203–1207.

[30] C DeLisi, *Mathematical modeling in immunology*, Annual Review of Biophysics and Bioengineering 12 (1983), no. 1, 117–138, PMID: 6870214.

[31] M.M. Deza and E. Deza, *Dictionary of distances*, Elsevier Science, 2006.

[32] Martin Dijst and Velibor Vidakovic, *Travel time ratio: the key factor of spatial reach*, Transportation 27 (2000), no. 2, 179199 (English).

[33] J. Dushoff, *Incorporating immunological ideas in epidemiological models*, Journal of theoretical biology 180 (1996), no. 3, 181–187.

[34] K.T.D. Eames, *Modelling disease spread through random and regular contacts in clustered populations*, Theoretical population biology 73 (2008), no. 1, 104–111.

[35] Stephen Eubank, Hasan Guclu, V.S. Anil Kumar, Madhav Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Want, *Modelling disease outbreaks in realistic urban social networks*, Nature 429 (2004), no. 6988, 180–183.

[36] Stephanie Forrest and Catherine Beauchemin, *Computer immunology*, Immunological Reviews 216 (2007), no. 1, 176–197.

[37] P.J. Fos and D.J. Fine, *Designing health care for populations: applied epidemiology in health care administration*, Jossey-Bass, 2000.

[38] Sebastian Funk, Marcel Salath, and Vincent A. A. Jansen, *Modelling the influence of human behaviour on the spread of infectious diseases: a review*, Journal of The Royal Society Interface (2010).

[39] H. Gagliardi, F. da Silva, and D. Alves, *Automata network simulator applied to the epidemiology of urban dengue fever*, Computational Science–ICCS 2006 (2006), 297–304.

[40] Reuben M Granich, Charles F Gilks, Christopher Dye, Kevin M De Cock, and Brian G Williams, *Universal voluntary hiv testing with immediate antiretroviral therapy as a strategy for elimination of hiv transmission: a mathematical model*, The Lancet 373 (2009), no. 9657, 48 – 57.

[41] Charles Miller Grinstead and James Laurie Snell, *Introduction to probability*, American Mathematical Soc., 1998.

[42] World Health Organization Writing Group, *Nonpharmaceutical interventions for pandemic influenza, international measures*, January 2005.

[43] AL Gruver, LL Hudson, and GD Sempowski, *Immunosenescence of ageing*, The Journal of Pathology 211 (2007), no. 2, 144–156.

[44] M.S. Handcock, A.E. Raftery, and J.M. Tantrum, *Model-based clustering for social networks*, Journal of the Royal Statistical Society: Series A (Statistics in Society) 170 (2007), no. 2, 301–354.

[45] B. Hellriegel, *Immunoepidemiology–bridging the gap between immunology and epidemiology*, Trends in parasitology 17 (2001), no. 2, 102–106.

[46] PeterS. Kim, PeterP. Lee, and Doron Levy, *Emergent group dynamics governed by regulatory cells produce a robust primary t cell response*, Bulletin of Mathematical Biology 72 (2010), 611–644 (English).

[47] T.J. Kindt, R.A. Goldsby, B.A. Osborne, and J. Kuby, *Kuby immunology*, WH Freeman, 2007.

[48] E.J. Kovacs and K.A.N. Messingham, *Influence of Alcohol and Gender on Immune Response.*, Alcohol Research & Health 26 (2002), no. 4, 257–265.

[49] M. Kress, *The Effect of Social Mixing Controls on the Spread of Smallpox A Two-Level Model*, Health Care Management Science 8 (2005), no. 4, 277–289.

[50] Rajesh Kumar and Edith A Burns, *Age-related decline in immunity: implications for vaccine responsiveness*, (2008).

[51] Yuliya N. Kyrychko and Konstantin B. Blyuss, *Global properties of a delayed sir model with temporary immunity and nonlinear incidence rate*, Nonlinear Analysis: Real World Applications 6 (2005), no. 3, 495 – 507.

[52] R.L. Langley, *Sex and gender differences in health and disease*, Carolina Academic Press, 2003.

[53] H.Y. Lee, D.J. Topham, S.Y. Park, J. Hollenbaugh, J. Treanor, T.R. Mosmann, X. Jin, B.M. Ward, H. Miao, J. Holden-Wiltse, et al., *Simulation and prediction of the adaptive*

immune response to influenza a virus infection, Journal of virology 83 (2009), no. 14, 7151.

[54] J. Li and F. Brauer, *Continuous-time age-structured models in population dynamics and epidemiology*, Mathematical Epidemiology (2008), 205–227.

[55] T. Makinodan and M.M.B. Kay, *Age influence on the immune system*, Adv Immunol 29 (1980), no. 287, 287.

[56] M. Martcheva and S.P. Sergei, *An epidemic model structured by host immunity*, Journal of Biological Systems 14 (2006), no. 02, 185–203.

[57] T.W. McDade, *The ecologies of human immune function*, Annu. Rev. Anthropol. 34 (2005), 495–521.

[58] William Hardy McNeill, *Plagues and peoples*, Anchor, New York, 1989.

[59] A. Melnichenko and A.A. Romanyukha, *A model of tuberculosis epidemiology: Data analysis and estimation of parameters*, Mathematical Models and Computer Simulations 1 (2009), no. 4, 428–444.

[60] Xinzhu Meng and Lansun Chen, *The dynamics of a new sir epidemic model concerning pulse vaccination strategy*, Applied Mathematics and Computation 197 (2008), no. 2, 582 – 597.

[61] R. M. Merril and T. C. Timmreck, *Introduction to epidemiology*, Jones and Bartlett Publishers, 2009.

[62] Lauren Ancel Meyers, Babak Pourbohloul, M. E J Newman, Danuta M Skowronski, and Robert C Brunham, *Network theory and SARS: predicting outbreak diversity*, Journal of Theoretical Biology 232 (2005), no. 1, 71–81.

[63] A.R. Mikler, S. Venkatachalam, and K. Abbas, *Modeling infectious diseases using global stochastic cellular automata*, Journal of Biological Systems 13 (2005), no. 4, 421–440.

[64] Armin R Mikler, Sangeeta Venkatachalam, and Suhasini Ramisetty-Mikler, *Decisions under uncertainty: a computational framework for quantification of policies addressing infectious disease epidemics*, Stochastic Environmental Research and Risk Assessment 21 (2007), no. 5, 533–543.

[65] Richard A. Miller, *The aging immune system: Primer and prospectus*, Science 273 (1996), no. 5271, 70–74.

[66] A. Morabia, *A history of epidemiologic methods and concepts*, Birkhäuser Basel, 2004.

[67] G.A. Morgan, N.L. Leech, G.W. Gloeckner, and K.C. Barrett, *Ibm spss for introductory statistics: Use and interpretation, fourth edition*, Taylor & Francis, 2011.

[68] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds, *Social contacts and mixing patterns relevant to the spread of infectious diseases*, PLoS Med 5 (2008), no. 3, e74.

[69] N. Nathanson and R. Ahmed, *Viral pathogenesis and immunity*, Academic Pr, 2007.

[70] Children's Health Network, *Infections: Incubation and contagious periods*, 2013.

[71] M. E. J. Newman, *Spread of epidemic disease on networks*, Physical Review E 66 (2002), no. 1.

[72] NHS, *How long is someone infectious after viral infection*, 2013.

[73] M.A. Nowak and C.R.M. Bangham, *Population dynamics of immune responses to persistent viruses*, Science 272 (1996), no. 5258, 74.

[74] R. Pagano, *Understanding statistics in the behavioral sciences*, Available Titles Aplia Series, Cengage Learning, 2008.

[75] A. Pape Møller and N. Saino, *Immune response and survival*, Oikos 104 (2004), no. 2, 299–304.

[76] P. Patlolla, V. Gunupudi, A. R. Mikler, and R. Jacob, *Agent-Based Simulation Tools in Computational Epidemiology*, Innovative Internet Community Systems (2006), 212–223.

[77] A.S. Perelson, *Modelling viral and immune system dynamics*, Nature Reviews Immunology 2 (2002), no. 1, 28–36.

[78] N.V. Pertsev and B.Y. Pichugin, *An individual-based stochastic model of the spread of tuberculosis*, Journal of Applied and Industrial Mathematics 4 (2010), no. 3, 359–370.

[79] S. Phuphanich, J. Rudnick, R. Chu, M. Mazer, H. Wang, N. Serrano, J. Richardson,

C. Wheeler, K. Black, M. Singh, et al., *Immune response correlation with progression-free survival in glioblastoma following dendritic cell immunotherapy (ICT-107)*, The 46th Annual Meeting of the American Society of Clinical Oncology, Chicago, Illinois, June, vol. 6, 2010.

[80] T.C. Quinn, M.J. Wawer, N. Sewankambo, D. Serwadda, C. Li, F. Wabwire-Mangen, M.O. Meehan, T. Lutalo, and R.H. Gray, *Viral load and heterosexual transmission of human immunodeficiency virus type 1*, New England Journal of Medicine 342 (2000), no. 13, 921–929.

[81] Diel R., Vandeputte J., de Vries G., Stillo J., Wanlin M., and Nienhaus A., *Costs of tuberculosis disease in the eu - a systematic analysis and cost calculation*, The European respiratory journal (2013).

[82] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0.

[83] J. Repace, B. Zhang, S. J. Bondy, N. Benowitz, and R. Ferrence, *Air quality, mortality, and economic benefits of a smoke free workplace law for non-smoking ontario bar workers*, Indoor Air 23 (2013), no. 2, 93–104.

[84] Jorge Reyes-Silveyra, Armin R Mikler, Justin Zhao, and Angel Bravo-Salgado, *Modeling infectious outbreaks in non-homogeneous populations*, Journal of Biological Systems (2011).

[85] P.E. Sartwell, *The distribution of incubation periods of infectious disease*, American Journal of Epidemiology 51 (1950), no. 3, 310.

[86] E.M.S. Silveira, M.F. Rodrigues, M.S. Krause, D.R. Vianna, B.S. Almeida, J.S. Rossato, L.P. Oliveira Jr, R. Curi, and P.I.H. de Bittencourt Jr, *Acute exercise stimulates macrophage function: possible role of NF-κB pathways*, Cell biochemistry and function 25 (2007), no. 1, 63–73.

[87] H. Situngkir, *Epidemiology through cellular automata: case of study avian influenza in Indonesia*, Bandung Fe Institute, Tech. Rep (2004).

[88] M.R. Spiegel and L.J. Stephens, *Schaum's outline of statistics*, Schaum's Outline Series, McGraw-Hill Education, 2007.

[89] Max A Stafford, Lawrence Corey, Yunzhen Cao, Eric S Daar, David D Ho, and Alan S Perelson, *Modeling plasma virus concentration during primary hiv infection*, Journal of Theoretical Biology 203 (2000), no. 3, 285–301.

[90] Thomas Lathrop Stedman, *Stedman's medical dictionary for the health professions and nursing*, illustrated 5th ed. ed., Lippincott Williams and Wilkins Baltimore, MD, 2005 (English).

[91] PS Teng and R.C. Close, *Effect of temperature and uredinium density on urediniospore production, latent period, and infectious period of puccinia hordei otth*, New Zealand Journal of Agricultural Research 21 (1978), no. 2, 287–296.

[92] Emmanuelle Terry, Jacqueline Marvel, Christophe Arpin, Olivier Gandrillon, and Fabien Crauste, *Mathematical model of the primary cd8 t cell immune response: stability analysis of a nonlinear age-structured system*, Journal of Mathematical Biology 65 (2012), 263–291 (English).

[93] J. Thakar, M. Poss, R. Albert, G.H. Long, R. Zhang, et al., *Dynamic models of immune responses: what is the ideal level of detail?*, Theoretical Biology and Medical Modelling 7 (2010), 35–35.

[94] C. Tian, X. Zhang, W. Ding, and R. Cao, *Epidemic alert & response framework and technology based on spreading dynamics simulation*, ICCS (3) (Y. Shi, G. D. van Albada, J. Dongarra, and P. M. A. Sloot, eds.), Lecture Notes in Computer Science, vol. 4489, Springer, 2007, pp. 1032–1039.

[95] United States Census Bureau, *Denton county Texas quickfacts*, http://quickfacts.census.gov/qfd/states/48/48121.html, November 2002.

[96] ———, *Sumter county Florida quickfacts*, http://quickfacts.census.gov/qfd/states/12/12119.html, November 2002.

[97] S. Venkatachalam and A.R. Mikler, *Modeling infectious diseases using global stochastic*

field simulation, Granular Computing, 2006 IEEE International Conference on, May 2006, pp. 750–753.

[98] David Vickers and Nathaniel Osgood, *A unified framework of immunological and epidemiological dynamics for the spread of viral infections in a simple network-based population*, Theoretical Biology and Medical Modelling 4 (2007), no. 49.

[99] J. Wallinga, W.J. Edmunds, and M. Kretzschmar, *Perspective: human contact patterns and the spread of airborne infectious diseases.*, Trends Microbiol 7 (1999), no. 9, 372–7.

[100] Jacco Wallinga, Peter Teunis, and Mirjam Kretzschmar, *Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents*, American Journal of Epidemiology 164 (2006), no. 10, 936–944.

[101] Christina Warrender, Stephanie Forrest, and Frederick Koster, *Modeling intercellular interactions in early Mycobacterium infection*, Bulletin of Mathematical Biology 68 (2006), 2233–2261 (English).

[102] Lawrence M. Wein, Rebecca M. D'Amato, and Alan S. Perelson, *Mathematical analysis of antiretroviral therapy aimed at hiv-1 eradication or maintenance of low viral loads*, Journal of Theoretical Biology 192 (1998), no. 1, 81 – 98.

[103] Marc E. Weksler and Paul Szabo, *The effect of age on the b-cell repertoire*, Journal of Clinical Immunology 20 (2000), 240–249, 10.1023/A:1006659401385.

[104] J.E. Wigginton and D. Kirschner, *A model to predict cell-mediated immune regulatory mechanisms during human infection with Mycobacterium tuberculosis*, The Journal of Immunology 166 (2001), no. 3, 1951–1967.

[105] DP Wilson, P. Timms, and DLS McElwain, *A mathematical model for the investigation of the th1 immune response to Chlamydia trachomatis*, Mathematical biosciences 182 (2003), no. 1, 27–44.

[106] S.N. Wilson, P. Lee, and D. Levy, *A mathematical model of the primary t cell response with contraction governed by adaptive regulatory t cells*, 26th Southern Biomedical Engineering Conference SBEC 2010, April 30 - May 2, 2010, College Park, Maryland, USA

(KeithE. Herold, Jafar Vossoughi, and WilliamE. Bentley, eds.), IFMBE Proceedings, vol. 32, Springer Berlin Heidelberg, 2010, pp. 209–212.

[107] D. Wodarz, *Mathematical models of immune effector responses to viral infections: Virus control versus the development of pathology*, Journal of computational and applied mathematics 184 (2005), no. 1, 301–319.

[108] D. Wodarz, R.M. May, and M.A. Nowak, *The role of antigen-independent persistence of memory cytotoxic T lymphocytes*, International immunology 12 (2000), no. 4, 467.

[109] Dominik Wodarz and Martin A. Nowak, *Mathematical models of hiv pathogenesis and treatment*, BioEssays 24 (2002), no. 12, 1178–1187.

[110] H. Yang, Y.H. Youm, B. Vandanmagsar, J. Rood, K.G. Kumar, A.A. Butler, and V.D. Dixit, *Obesity accelerates thymic aging*, Blood 114 (2009), no. 18, 3803.

[111] J.H. Zar, *Biostatistical analysis*, Prentice-Hall/Pearson, 2010.