Applications of Modeling to Hantavirus, Cholera, and COVID-19

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Morganne E Igoe

May 2022

To my family.

Acknowledgments

I would like to first thank my advisor, Suzanne Lenhart, for all of her mentorship and guidance throughout my career. Dr. Lenhart, thank you for giving me so much of your time, knowledge, and wisdom over the last few years and for helping me grow as a mathematician and a person. I would like to thank Dr. Agricola Odoi for sharing his advice on statistical modeling and for serving as a member of my committee. I would also like to thank Dr. Louis Gross for agreeing to serve on my committee and his support as I reach my career goals. Finally I would like to thank Dr. Olivia Prosper for serving on my committee and for her advice on my work.

To Mom, Dad, Nana, and Tt, thank you for always being there for me and being the most supportive family. To Tori, Christina, and Kalli, thank you for being the best sisters in the world, for always being there for me to vent to, and for having the same conversations over and over with me for the last 22 years. To Samantha, thank you for making coffee with me every morning in the office, for getting me through prelims, and always being there to keep my life in Tennessee together when everything else was falling apart. To Alex, thank you for always supporting and encouraging me and for including me in every event. To Kelly, Anna, and Wencel, thank you for all of your help and support during our classes and for always being willing to share new ideas and tips. You have all been so supportive and kind. I couldn't have done this without you.

Chapter 2, in full, is a reprint of the material as it appears in *Letters in Biomathematics* 2020. Authors: Morganne Igoe, E. Joe Moran, Theresa Sheets, Jeff DeSalu, Colleen B. Jonsson, Suzanne Lenhart, Robert D. Owen, Megan A. Rúa. The dissertation author was the primary investigator and author of this paper.

Chapter 3 was written in collaboration with Suzanne Lenhart and Giulio de Leo. The dissertation author was the primary investigator and author of this work.

Chapter 4, in full, is a reprint of the material as it appears in *BMC Public Health* 2022. Authors: Morganne Igoe, Praachi Das, Suzanne Lenhart, Alun L. Lloyd, Lan Luong, Dajun Tian, Cristina Lanzas, Agricola Odoi. The dissertation author was the primary investigator and author of this paper.

Abstract

Many rodent-borne hantaviruses are zoonotic pathogens that can cause disease in humans. To evaluate the prevalence of Jaborá virus (JABV) over time within its rodent reservoir, *Akodon montensis*, we formulated a discrete time model with multiple rodent age classes and unique infection class progression features. We parameterized the model with data collected from a survey of JABV harbored by *Akodon montensis* in the Mbaracayú Reserve in Paraguay. Our model incorporates three infection types and a recovered class. A new feature of the model allows transition from the latent to the persistently-infected class. With a more complete age and infection structure, we are better able to identify the driving forces of epidemiology of hantaviruses in rodent populations.

Optimal control (OC) applied to epidemic modeling can be a key tool in determining effective intervention strategies for infectious diseases. However, the common formulations of OC problems of infectious disease systems of ordinary differential equations do not always offer features needed by disease managers. These problems often minimize both the disease burden and the intervention cost over a time period to determine the optimal intervention strategy, not allowing for the incorporation of details such as budget constraints. We formulated and analyzed several OC formulations using a basic example of cholera, incorporating features such as a budget constraint or a desired epidemiological goal, to reframe OC problems applied to infectious disease models in a more accessible manner for public health managers.

There is evidence of geographic disparities in COVID-19 hospitalization risks that, if identified, could guide control efforts. Geographic disparities in distribution of socioeconomic, demographic and health-related predictors of ZCTA-level COVID-19 hospitalizations in St. Louis were investigated using choropleth maps. These predictors were then investigated

using global negative binomial and local geographically weighted negative binomial models. COVID-19 hospitalization risks were significantly higher in ZCTAs with high diabetes hospitalization and COVID-19 risks, black population, and populations with some college education. The impacts of the first three factors vary by geographical location, implying that a 'one-size-fits-all' approach may not be appropriate for management and control. These findings are useful for informing health planning and guiding vaccination efforts.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this dissertation, we show how mathematical and statistical models can be used to explore the drivers of different diseases and to determine optimal strategies for disease management. Two of our models, a discrete time age structured model and a negative binomial regression model, are connected to real world data, allowing us to more accurately investigate disease dynamics. We also employ a basic system of ordinary differential equations (ODE) modeling disease transmission to explore alternate formulations of optimal control problems incorporating the needs of disease managers, including budget constraints and epidemiological goals.

In Chapter 2, we aim to investigate prevalence patterns of hantaviruses in rodents. Hantaviruses have been observed to have large variation in prevalence in rodents over time, varying from less than 1% to as high as 25% in long term sampling in the United States [19]. When hantavirus prevalence in rodents is high, humans are at higher risk of coming into contact with the excreta of infected rodents, which can cause serious conditions such as hantavirus pulmonary syndrome, a condition with limited treatment options. To investigate these patterns, we evaluate the prevalence of a strain of hantavirus, Jaborá virus (JABV), over time within its rodent reservoir, *Akodon montensis*, and explore the drivers of hantavirus prevalence among rodents in Paraguay using an age structured model of hantavirus. This discrete time model consists of nine compartments incorporating three age classes, three types of infection, and a recovered class. The infection classes (acutely infected, persistently infected, and latent) are based on the infectiousness of the rodent. The latent class is a

new feature allowing rodents to lapse back into a persistently infected state, where they are capable of transmitting the virus, from a latent state, where they are unable to transmit the virus [88]. This class was proposed as a possible mechanism for the persistent low levels of hantavirus prevalence observed in rodent populations [19]. As this is a discrete time model, the order of events is carefully chosen. Using seroprevalence data on hantavirus from the Mbaracayú Reserve in Paraguay [28], we parameterize the model. We then perform a Latin Hypercube Sensitivity Analysis, a global sensitivity technique where all parameters are varied at one time, combined with Partial Rank Correlation Coefficients [72] to measure the impact of the parameters on disease prevalence over time. With these results to guide us, we perform further simulations to explore the effect of parameter changes and initial conditions on disease dynamics. By using a more complete infection structure and incorporating real world data, we are better able to identify the driving forces of hantavirus prevalence in rodent populations.

In Chapter 3, we develop and analyze new optimal control formulations applied to systems of ODEs modeling infectious disease transmission, with the goal of reframing optimal control problems in a manner more accessible and useful for public health managers. Oftentimes, optimal control problems applied to infectious disease systems are formulated to minimize both some measure of disease burden and the cost of intervention [3, 9, 79]. This formulation of the problem does not allow for the incorporation of issues such as budget constraints or specific epidemiological goals that public health departments may be trying to reach. It also requires researchers to place monetary value on human health and death, which leads to ethical considerations, particularly when considering diseases affecting marginalized groups or countries with developing economies and health care systems [11, 16, 48, 81, 96]. Therefore, we reframe optimal control problems to include these features, allowing the addition of budget constraints or epidemiological goals for a given system. We then apply these new formulations to a basic ODE model of cholera [108], which includes susceptible, infected, recovered classes as well as a compartment tracking the concentration of cholera vibrios in the water. In this model, the control is vaccination. We use an approach of obtaining the adjoints and optimal control characterization via Pontryagin's Maximum Principle [89] to construct the optimality system (adjoints, state equations, boundary conditions, and

optimal control characterization). We then solve this system using an iterative forward backward sweep method [68]. We deliberately use a simple model of cholera transmission and choose specific sets of model parameters to demonstrate key features of these optimal control formulations with numerical simulations.

In Chapter 4, we investigate ZIP Code Tabulation Area (ZCTA)-level drivers of COVID-19 hospitalization risks in the St. Louis area. There is evidence of geographic disparities in ZCTA-level COVID-19 hospitalization risks in this area. There are also geographic disparities in various ZCTA-level features, including COVID-19 cases, socioeconomics, demographics, health behaviors, and chronic disease conditions. We hypothesize that the disparities in these features are driving the observed disparities in ZCTA-level COVID-19 hospitalization risks in the St. Louis area. To investigate this, we employ global and local negative binomial regression models, where the outcome is ZCTA-level age-adjusted COVID-19 hospitalization risk and the predictors are ZCTA-level sociodemographic and economic factors, health behaviors, chronic disease conditions, and COVID-19 cases. Socioeconomic and demographic factors such as race, education level, and income were taken from the U.S. Census Bureau 2018 American Community Survey (ACS) 5-year estimates [109]. COVID-19 hospitalizations, health behaviors, and chronic disease conditions data were obtained from the Missouri Hospital Association, while data on confirmed COVID-19 cases were obtained from County Departments of Health. We first construct a global negative binomial regression model using backwards elimination. To determine if there is geographic variation in the predictors of our final global negative binomial regression model, we construct a local negative binomial regression model using Geographically Weighted Regression [92]. This regression method allows the coefficients in the model to vary by ZCTA, helping us to identify areas where the predictors in our model have differing impacts on our outcome. Utilizing both global and local regression models is crucial because it shows that a universal approach to health planning is insufficient in this area, as different factors are driving hospital risks in different regions, and that customized planning strategies must instead be determined for effective management.

We utilize several numerical and statistical approaches throughout this work. For the numerical simulations of our optimal control problems, we use the MATLAB ODE solver

ode45 (which is based on an explicit Runge-Kutta formula) coupled with an iterative method. The ZCTA-level global negative binomial regression model was built using a generalized linear model (GLM) framework in R. The GLM framework allows for linear models to be utilized when the outcome variable has a non-normal error distribution by relating the linear model to the outcome variable via a link function. The ZCTA-level geographically weighted negative binomial model was fit in SAS. This modeling approach fits a regression equation for each ZCTA in the dataset by incorporating only the data from the ZCTAs within a certain radius of the target ZCTA. This allows for the identification of local variation in the regression coefficients, showing the importance of specific predictors in different areas.

This work aims to use various modeling approaches and real world data to investigate the drivers of hantavirus and COVID-19 and employs a basic transmission model to explore new optimal control formulations for infectious diseases. In Chapter 2, we show that the addition of a carrier class to our model can explain some of the observed patterns in hantavirus prevalence in rodent populations. Understanding the drivers of hantavirus prevalence can contribute to the reduction of zoonotic transfer of the virus during outbreaks. In Chapter 3, we demonstrate how optimal control problems can be reframed to take into account the limitations and needs of disease managers, making optimal control more accessible as a management tool for public health officials. In Chapter 4, we identify local predictors of COVID-19 hospitalization in the St. Louis area, demonstrating which factors may be important to consider when developing hospital management strategies or control efforts. We also show the importance of utilizing both global and local modeling strategies to develop a more accurate understanding of disease drivers in a given area. All of these projects are joined by a common theme, to improve the management of infectious disease by leveraging the power of mathematical and statistical modeling while simultaneously considering the circumstances and limitations under which disease managers often function.

Chapter 2

A Discrete Age Structured Model of Hantavirus in a Rodent Reservoir in Paraguay

2.1 Introduction

Hantaviruses are endemic in most parts of the world, being harbored by rodents, bats, moles or shrews. Each mammalian reservoir carries a unique viral strain. Intraspecies transmission occurs through the inhalation of aerosolized virus excreted in urine or feces [88], or through an exchange of blood and saliva during aggressive encounters among conspecific individuals [37]. We note that hantavirus is not lethal in rodents and does not usually cause health issues. While not widely studied for most hantaviral reservoirs, viral infection of some rodent reservoirs may show lower survival rates, lower body mass, slower weight gains, and higher testosterone (for example, see [70, 71]). However, no pathology has been observed in bank voles infected with Puumala virus [119], in deer mice experimentally infected with Sin Nombre virus [15], or in cotton rats infected with Black Creek Canal virus [46]. But higher

A version of this chapter was previously published in Igoe M, Moran EJ, Sheets T, DeSalu J, Jonsson CB, Lenhart S, Owen RD, Rúa MA. *A Discrete Age Structured Model of Hantavirus in a Rodent Reservoir in Paraguay.* Letters in Biomathematics. vol. 7, 1: 127–142, 2020. Morganne Igoe was the leading author of this work.

doses of Black Creek Canal virus in the cotton rat can cause pneumonitis [45]. Researchers have reported that Sin Nombre virus (SNV) seroprevalence may vary from very low ($< 1\%$) to levels greater than 25% in long-term sampling in the western United States [19]. Field studies in Chile by Padula et. al [88] suggested that a latency period may be responsible for the low prevalence.

Rodent reservoirs of Jaborá hantavirus in *Akodon montensis* have been studied in the Interior Atlantic Forest of Paraguay, and these populations exhibited low prevalence levels [28, 87]. To mechanically explain the low prevalence of JABV over time within its rodent reservoir, *Akodon montensis*, we propose a mathematical model with rodent age classes and a unique 'latency' class (having the virus but without the ability to transmit it). While the model used in this paper is based on data from JABV found in a South American rodent reservoir, *Akodon montensis* in Paraguay, the model is generalizable to other rodent reservoirs of hantaviruses due to its structure with age and infection classes [49, 120].

Upon infection, the rodent reservoir can be characterized as having an acute infection as demonstrated by virus being shed in urine and feces. At some point, the rodent transitions to a persistent infection which is characterized by low or undetectable levels of virus being shed in urine or feces. However the rodent remains infected in many tissues such as the lung and heart (for example, see [15, 38]). Such persistently infected rodents do not clear the virus, as documented in [74], and this feature will be represented in our model. It is through these cycles that rodent-borne hantaviruses are maintained in rodent communities in nature.

Infection of humans by exposure to rodent-borne viruses is considered to be a dead end, since humans cannot transmit the virus to one another [62]. In the New World, some hantavirus species transmitted from rodents to humans cause hantavirus pulmonary syndrome, a serious condition with basic life support as the only medical treatment and mortality rates of 36% in the United States [20]. To better understand the drivers behind hantavirus spread, both intraspecies and interspecies transmission must be considered. Prevalence levels for hantavirus infection within rodent host populations vary seasonally and geographically, ranging from roughly 1 to 6% in the montane grass mouse, *Akodon montensis*, in Brazil and Paraguay, to 1 to 30% documented in the deer mouse, *Peromyscus maniculatus*, in the United States [23, 106]. A suite of factors drives viral prevalence in rodent

populations including host genetics, mating behaviors and environmental factors [28, 71]. For example, the species *A. montensis* does not enter torpor (a short period of reduced body temperature and metabolic activity), and can have up to three reproductive cycles each year, potentially leading to higher population density and/or higher dispersal rates [50, 87]. Both factors would likely increase contact and agonistic encounters with both conspecifics and other species. This contrasts with North American rodent species such as *P. maniculatus* which undergo torpor and have only a single reproductive cycle each year. Together these and other abiotic and biotic factors modulate viral prevalence observed in males and females. Environmental characteristics, such as seasonality and landscape structure, have also been linked to increased prevalence, as have anthropogenic factors such as ecological disturbance and its corresponding impact on biodiversity [61, 67, 71].

Previous theoretical work in this system has incorporated differences in transmission and environmental factors to understand changes in overall virus prevalence. Allen and collaborators have published several papers on epidemiological models of hantavirus in *A. montensis* in Paraguay [6, 8, 116]. These models include: a 2006 differential equations model followed by a stochastic model with SEIR compartments (susceptible, exposed, infected, recovered) to include random seasonal and environmental variations (both in [5]), a 2006 differential equations model featuring separate sex classes, followed by a stochastic differential equations model of similar structure (both in [6]), a 2009 model by Allen et al. [8] which used a habitat-based model for the spread of hantavirus between a reservoir and spillover species, and a 2010 model by Wesley et al. [117] that used an SI model which includes two infected classes and examines the role of viral-contaminated soil. Despite the extensive modeling work previously completed in this system, there has only been one discrete model developed to investigate the outbreak structure for viral transmission of hantavirus [116]. This model consists of a discrete-time rodent-hantavirus model structured by two infected classes (one for males and one for females) and three age classes. It includes juveniles, subadults, and adults, split into female and male groups, and susceptible and infected classes of adults. It further assumes that the juveniles and subadults could not become infected, and focuses on differences in contact rates based on the sex of the mice. Due to age structure in our data set, we chose to formulate a model discrete in time with three age classes and

to use three types of infected classes to investigate persistent low levels of prevalence. We include a 'latent' class, which cannot transmit the virus but carries it, which was suggested by the work of [88]. Our model also differs from other hantavirus models due to its strong connection to data from Paraguay [28].

Motivated by the extensive prevalence data for hantaviruses in rodents collected in August 2014 in the Mbaracayú Reserve in Paraguay [28], and by the fluctuations in prevalence but persistence of the virus at a low level in data collected from the region in 2005-2006 [87], we investigated possible mechanisms for the observed occurrences of the virus through modeling. Our model is an age structured, discrete model which incorporates latent and recovered classes, and our transition rates and prevalence levels are directly connected to this data [28] in Paraguay. In addition to the classical stages of infection, acute and persistent, we add a latent class to better examine the dynamics of the virus. In this model, rodents in the latent class do not shed virus, so they are not transmitting the virus, but are able to transition back into the persistently infected class. The latent class is a novel addition to our model (compared to previous models), which offers a possible explanation for the maintenance of the virus during normal periods and periods when outbreaks of virus are driven by biotic or abiotic factors. The age structure includes differing transmission coefficients between juvenile and adult populations. By incorporating the new latent class and an expanded age class structure into our model, we address the dynamics of hantavirus within rodent populations.

In the next section, we describe our data and our model. Then we follow with results and discussion, and in the last section, conclusions are given.

2.2 Methods

2.2.1 Our Data

Our model is based on data collected in August 2014 in the Reserva Natural de Bosque Mbaracayú (RNBM) in Paraguay [28]. The data collected at 22 sites from 2 -11 nights of trapping include 417 rodents. Recorded features included species, weight, total length, sex, reproductive status and categorical age. For all species, a combination of pelage (fur), size,

behavior and *gestalt* was used to designate age class. The presence of rodent-borne hantavirus antibodies and/or RNA and characteristic tail lesions (potentially due to Leishmania) were also collected. Because we were interested in rodents which serve as the main vector for hantavirus, we extracted species data for the most prevalent hantaviral reservoir species, *A. montensis* (sample size=274), which comprised about 70% of the collected rodents. Three of the 274 *A. montensis* were newborns so sex on these animals could not be determined and were therefore excluded from further analysis. Overall, we had 271 rodents in our dataset.

The demographic makeup of our data set [28] was 52% male, 48% female, 83% adults, and 17% juvenile. The overall hantavirus prevalence was 6%, which is close to that in other studies [28, 83, 87]. Of the entire population, 7% of the males had hantavirus, and 3% of the females had hantavirus. Within recorded hantavirus cases, 71% occurred in males, while 29% occurred in females. For the entire population, the prevalence among juveniles was 6% and the prevalence among adults was 5%. Among seropositive individuals, juveniles compose 21%, and adults compose 79%.

2.2.2 Our Model

In our model, the population is divided into three age classes (listed with lengths of time in each class): newborns B (23 days), juveniles J (37 days), and subadults and adults A (305 days). There was not enough data to adequately separate subadults and adults so they were combined into one class. These classes include males and females, homogeneously mixed. The susceptible population S is broken into two compartments: susceptible juveniles S_J and susceptible adults S_A. The literature suggests that newborns may be unable to become infected or may be delayed in their rate of infection due to maternal antibodies [12, 14, 25, 27, 35, 52–54, 113] and/or their decreased contact with infected individuals due to their limited movement from the den and mother, but the study by Hutchinson et al. did not observe the influence of maternal antibodies on infection in Black Creek Canal virus [46]. In our model, the newborns B are not susceptible and cannot be infected until they become juveniles. The acutely infected population is broken into two compartments: infected juveniles I_J and infected adults I_A which have been reported to have the highest probability of shedding virus in saliva and excreta. As with the majority of viral infections,

rodents within this class will not produce antibodies early during the course of infection, and begin in some reservoirs approximately ten days post-infection [98, 99]. Once antibodies are produced in response to the infection, fewer free virions will be available to infect other rodents.

The persistently-infected population is broken into two compartments: persistently-infected juveniles P_J and persistently infected adults P_A. Persistently-infected rodents shed fewer virions and are significantly less infectious than I_A and I_J populations and therefore have a lower transmission rate, and note that only the I_A class can move to recovered R. The latent class L and recovered class R consist only of adults. Rodents in the latent class L do not shed virus but are able to regress back into the persistently infected stage P_A. The recovered class R is not able to transmit the pathogen and is unable to regress back into an infected class. Figure 2.1 illustrates the possible infection status and age transitions of the model shown in equations (2.1)-(2.9). The periods of time in each infected class and age class are shown in Table 2.1 and Figure 2.2.

The model is stated in equations (2.1)-(2.9), where the adult population is $A(t) = S_A(t) + I_A(t) + P_A(t) + L(t) + R(t)$. The model is discrete in time, using time steps of days. The order of events in this model is: survival, age transition out of class, age transition into class, new infection, progression of infection out of class, and progression of infection into class. Transition from P to L_A happens last. The parameters of the model are defined in Table 2.2. All age transition rates are denoted by μ_B, μ_J and μ_A for the three classes (newborns, juveniles and adults).

We discuss each equation in the model in detail. Given a death rate, like d_B, then the corresponding survival rate is $(1 - d_B)$, and given a transition rate out of a compartment, like μ_B, the corresponding rate for remaining in that compartment is $(1 - \mu_B)$. Starting with the newborn class $B(t)$, the survival and the remaining in class have coefficients $(1 - d_B)$ and $(1 - \mu_B)$. Then lastly, the new births are added in with density dependence on the size of the adult population $A(t)$, which is shown with a baseline rate b_A and modified by

$$\frac{K}{K + A(t)}.$$

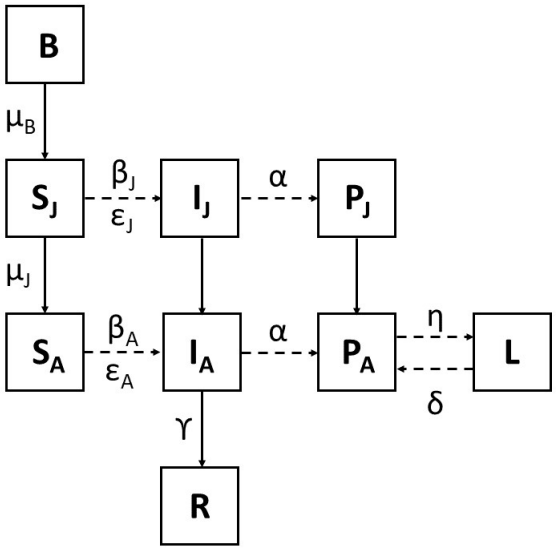

Figure 2.1: A flowchart of the model shown in equations (2.1)-(2.9). Vertical arrows denote age transitions and horizontal arrows denote infection transitions. Births from the adults, $S_A + I_A + P_A + L + R$, come into the newborn class B. Natural deaths occur from each compartment, but note that there are no pathogen induced deaths.

Table 2.1: Table of infection classes, transmission levels, and infection class length in days.

Class	Level of Infectiousness	Length (days)
Acutely Infected (I)	Shedding most virus	23
Persistently Infected (P)	Shedding little virus	37
Latent (L)	No Shedding	60

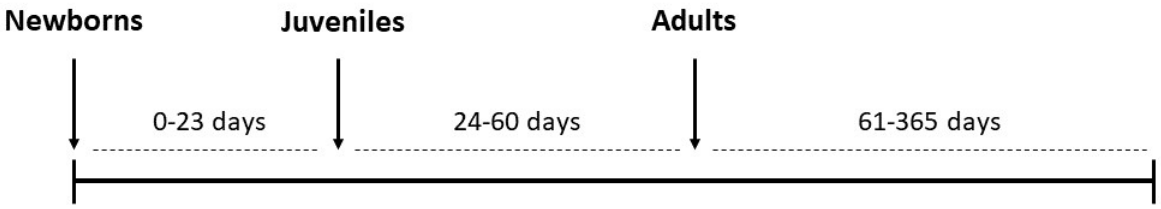

Figure 2.2: Timeline progression from newborn to adult. Time is measured in days post birth.

Table 2.2: Table of parameter definitions and values. All the rates have units of per day, the time step. The parameter K has units of number of rodents.

Parameter	Definition	Value
μ_i	Transition rate from age class i	$\mu_B = 1/23$ $\mu_J = 1/37$
d_i	Death rate for age class i	$d_B = 0.003521$ $d_J = 0.012123$ $d_A = 0.004$
α	Rate of transition from I to P class	$1/23$
γ	Rate of recovery	$1/10$
η	Rate of transition from P to L class	$1/37$
δ	Rate of transition from L to P class	$1/60$
β_{ij}	Transmission rate for I class	$\beta_{AA} = \beta_{JJ} = 0.00012$ $\beta_{AJ} = \beta_{JA} = 0.75 * \beta_{AA}$
ε_{ij}	Transmission rate for P class	$\varepsilon_{AA} = \varepsilon_{JJ} = 0.25 * \beta_{AA}$ $\varepsilon_{AJ} = \varepsilon_{JA} = 0.1875 * \beta_{AA}$
K	Density dependent effect on births	375
b_A	Birth rate	0.0143

Note the birth rate b_A will be one half of the birth rate for females in our population, assuming that there are equal numbers approximately of males and females. The order of events in the S_J equation begins with the individuals of $S_J(t)$ surviving with rate $(1-d_J)$ and remaining after age transition with rate $(1-\mu_J)$. This is followed by adding individuals due to age transition from newborn to juvenile, represented by $(1-d_B)\mu_B B(t)$. These actions in the S_J equation are followed by transmission, which moves a proportion of individuals from S_J to I_J, with corresponding β and ϵ terms in the exponential term, due to transmission from I_J, I_A, P_J, and P_A classes. The proportion of the S_J class remaining is

$$\exp\Big(-\big(\beta_{JJ}I_J\big(t\big)+\beta_{AJ}I_A\big(t\big)+\varepsilon_{JJ}P_J\big(t\big)+\varepsilon_{AJ}P_A\big(t\big)\big)\Big)$$

after the following proportion moves to I_J due to transmission

$$1-\exp\Big(-\big(\beta_{JJ}I_J\big(t\big)+\beta_{AJ}I_A\big(t\big)+\varepsilon_{JJ}P_J\big(t\big)+\varepsilon_{AJ}P_A\big(t\big)\big)\Big).$$

Then looking to the I_J class, we see the three factors $(1-d_J)(1-\mu_J)(1-\alpha)$ representing the proportion of $I_J(t)$ surviving with $(1-d_J)$, remaining after age transition with $(1-\mu_J)$ and remaining after transition to P with $(1-\alpha)$. Then, infecteds from S_J move into this class with the $(1-\exp(...))$ term. The last juvenile class P_J has the rates of surviving and remaining after age transition as $(1-d_J)(1-\mu_J)$. A proportion of the I_J individuals, who have survived and remain juvenile, make their transition into P_J with rate α, given by $I_J(t)(1-d_J)(1-\mu_J)\alpha$.

Similarly for the adult classes, we have S_A individuals surviving with rate $(1-d_A)$ and then individuals of the S_J class moving in due to age transition as $(1-d_J)\mu_J S_J$. The proportion remaining after transmission is represented by the $\exp(-...)$ with corresponding coefficients $\beta_{JA}, \beta_{AA}, \epsilon_{JA}$, and ϵ_{AA}. The I_A equation is similar to the I_J equation, except for transition in from the I_J compartment and followed by transition γ out to the R class.

The P_J equation has survival $(1-d_J)$ and the proportion $(1-\mu_J)$ remaining after the transition to P_A, followed by the transition from I_J with the term

$$I_J\big(t\big)(1-d_J)(1-\mu_J)\alpha.$$

The P_A equation has survival and the addition of the transitioning P_J individuals, followed by the transition to the L class, and then the transition from I_A (with rate α) and from L (with rate δ).

The L class has survival, transition in from P_A (with rate η) and then the remaining proportion $(1-\delta)$ after a proportion δ moves to P_A. Lastly the recovered class R survives, and then there is movement in from I_A with terms

$$\gamma(1-\alpha)\Big(I_A(t)(1-d_A)+I_J(t)(1-d_J)\mu_J\Big),$$

as can be seen in the last term in the system of equations.

$$
\begin{aligned}
B(t+1) &= B(t)(1-d_B)(1-\mu_B)+\frac{b_A K}{K+A(t)}A(t) && (2.1)\\
S_J(t+1) &= \Big(S_J(t)(1-d_J)(1-\mu_J)+B(t)(1-d_B)\mu_B\Big)\exp\Big(-(\beta_{JJ}I_J(t)+\beta_{AJ}I_A(t)+\varepsilon_{JJ}P_J(t)+\varepsilon_{AJ}P_A(t))\Big) && (2.2)\\
S_A(t+1) &= \Big(S_A(t)(1-d_A)+(1-d_J)\mu_J S_J(t)\Big)\exp\Big(-(\beta_{JA}I_J(t)+\beta_{AA}I_A(t)+\varepsilon_{JA}P_J(t)+\varepsilon_{AA}P_A(t))\Big) && (2.3)\\
I_J(t+1) &= I_J(t)(1-d_J)(1-\mu_J)(1-\alpha)+ && (2.4)\\
&\quad \Big(S_J(t)(1-d_J)(1-\mu_J)+B(t)(1-d_B)\mu_B\Big)\Big[1-\exp\Big(-(\beta_{JJ}I_J(t)+\beta_{AJ}I_A(t)+\varepsilon_{JJ}P_J(t)+\varepsilon_{AJ}P_A(t))\Big)\Big]\\
I_A(t+1) &= \Big(I_A(t)(1-d_A)+I_J(t)\Big(1-d_J\Big)\mu_J\Big)(1-\alpha)(1-\gamma) && (2.5)\\
&\quad +\Big(S_A(t)(1-d_A)+S_J(t)(1-d_J)\mu_J\Big)\Big[1-\exp\Big(-(\beta_{JA}I_J(t)+\beta_{AA}I_A(t)+\varepsilon_{JA}P_J(t)+\varepsilon_{AA}P_A(t))\Big)\Big]\\
P_J(t+1) &= P_J(t)(1-d_J)(1-\mu_J)+I_J(t)(1-d_J)(1-\mu_J)\alpha && (2.6)\\
P_A(t+1) &= \Big(P_A(t)(1-d_A)+(1-d_J)\mu_J P_J(t)\Big)(1-\eta)+\Big(I_A(t)(1-d_A)+I_J(t)(1-d_J)\mu_J\Big)\alpha && (2.7)\\
&\quad +\delta\Big[L(t)(1-d_A)+\eta\Big[P_A(t)(1-d_A)+P_J(t)(1-d_J)\mu_J\Big]\Big]\\
L(t+1) &= \Big[L(t)(1-d_A)+\eta\Big[P_A(t)(1-d_A)+P_J(t)(1-d_J)\mu_J\Big]\Big](1-\delta) && (2.8)\\
R(t+1) &= R(t)(1-d_A)+\gamma(1-\alpha)\Big(I_A(t)(1-d_A)+I_J(t)(1-d_J)\mu_J\Big) && (2.9)
\end{aligned}
$$

2.3 Results and Discussion

The parameters of the model are defined in Table 2.2. All age transition and infected class progression rates are in terms of days and are the reciprocal of the length of time an individual spends in each age or infected class (Figure 2.2, Table 2.1). Mortality rates were estimated

using those rates for a single season in [31, 34]. Mortality rates were then derived using the length of time of each age class and overall mortality rate for a season, with mortality in the newborn phase being low, juvenile high, and adult moderate [34]. Assuming half the adults are female, the baseline birth rate b_A would be one half of the birth rate from females. Note that $b_A = \frac{1}{70} = 0.0143$ births per day would result in about 10.5 newborns in year, usually coming from 2 or 3 litters per female per year [28, 87].

The parameters for the infected classes were derived from seroprevalence data showing days past infection [64]. The rates are the reciprocals of these infectious periods. Age class progression rates were derived from [34] by taking the reciprocal of the time within an age class. The parameter β_{AA} was manipulated to achieve a prevalence after one year of 6% to match our data set and what was observed in other *A. montensis* populations [106]. The other transmission rates are expressed as fractions of the baseline transmission rate β_{AA}.

To determine the baseline population from which simulations for the model were run, we first calculated the disease free equilibrium (DFE) (from equations (2.10)-(2.12)) by assuming that $I_J = I_A = P_J = P_A = L = R = 0$, and deriving expressions for the non-infected classes (B, S_J, and S_A) shown in equations (2.10)-(2.12). We then expressed the initial conditions for the infected classes as a proportion of the total population at the disease free equilibrium. The multipliers were chosen to achieve approximately a 6% overall prevalence in the initial population, as was observed in the data collected in the Mbaracayú Reserve. For example, we calculated from the data that approximately 60% of individuals in the infected classes belonged to the infected class I, while approximately 30% of the individuals were in the persistently infected class P. Additionally, we estimated that approximately 21% of the individuals in the infected classes were juveniles, while about 79% were adults. Lastly, we estimated that 10% of the infected population belonged to the latent class L based on the data set [28] and a prior study [87]. These approximations were used to create the initial conditions for the infected classes shown in equations (2.16)-(2.21). Any alterations to the model parameters in a simulation were recorded and the initial conditions were recalculated under these new conditions. This was done in order to standardize initial conditions for the parameter changes.

$$B^* = \frac{K[b_A(1-d_J)\mu_J(1-d_B)\mu_B - d_A(1-(1-d_B)(1-\mu_B))(1-(1-d_J)(1-\mu_J))]}{(1-(1-d_B)(1-\mu_B))(1-d_J)\mu_J(1-d_B)\mu_B} \quad (2.10)$$

$$S_J^* = \frac{B^*(1-d_B)\mu_B}{(1-(1-d_J)(1-\mu_J))} \quad (2.11)$$

$$S_A^* = \frac{(1-d_J)\mu_J S_J^*}{d_A} \quad (2.12)$$

$$B(0) = B^* \quad (2.13)$$

$$S_J(0) = S_J^* \quad (2.14)$$

$$S_A(0) = S_A^* \quad (2.15)$$

$$I_J(0) = (0.06 * 0.6 * 0.21)(B^* + S_J^* + S_A^*) \quad (2.16)$$

$$I_A(0) = (0.06 * 0.6 * 0.79)(B^* + S_J^* + S_A^*) \quad (2.17)$$

$$P_J(0) = (0.06 * 0.3 * 0.21)(B^* + S_J^* + S_A^*) \quad (2.18)$$

$$P_A(0) = (0.06 * 0.3 * 0.79)(B^* + S_J^* + S_A^*) \quad (2.19)$$

$$L(0) = (0.06 * 0.1)(B^* + S_J^* + S_A^*) \quad (2.20)$$

$$R(0) = 0 \quad (2.21)$$

The Next Generation Matrix Method for calculating the basic reproduction number R_0 for discrete-time models described by Allen and van den Driessche [7] was used to calculate R_0 for the model; we checked the needed assumptions to use this method on our model (meaning, the spectral radius of the Jacobian of the transitions between infected classes and of the Jacobian for the model without infection are both below 1.) Using this method, we take the infected classes to I_J, I_A, P_J, P_A, L and then R_0 takes the form $R_0 = \rho(F(I-T)^{-1})$, where F is the Jacobian matrix that results from differentiating the terms that describe new infections in the infected classes, T is the Jacobian matrix that results from differentiating all other transitions in the infected classes, and ρ is the spectral radius. This method allowed for the simplification of the system to 5 equations by limiting calculations to only equations (2.4)-(2.8); those pertaining to infected individuals. We used the block structure,

determinants, traces, and sum of minors, but we were not able to find a reasonable algebraic expression for R_0. Thus R_0 was only calculated for our particular parameter values, shown in Table 2.2, yielding $R_0 = 1.06$. This indicates that the DFE is unstable and the infection persists in our model.

We ran model simulations with the baseline parameters, which gave the initial conditions stated in equations (2.22) - (2.30). The total initial population is 652, of which 37 are infected and 616 non-infected. Figure 2.3 shows all the noninfecteds in the population. The susceptible adult S_A compartment decreases from 479 to 418, while the newborn B compartment increases. The susceptible juvenile S_J and recovered R compartments appear to be leveling off. In Figure 2.4, we see the five infected compartments with P_j, P_A and L increasing at first and with I_J and I_A both decreasing. The prevalence starts at about 5.6% and drops only slightly over time to 4.9% over a year.

$$B(0) = 64.3684 \tag{2.22}$$
$$S_J(0) = 71.8341 \tag{2.23}$$
$$S_A(0) = 479.4815 \tag{2.24}$$
$$I_J(0) = 4.6546 \tag{2.25}$$
$$I_A(0) = 17.5101 \tag{2.26}$$
$$P_J(0) = 2.2165 \tag{2.27}$$
$$P_A(0) = 8.7550 \tag{2.28}$$
$$L(0) = 3.6941 \tag{2.29}$$
$$R(0) = 0 \tag{2.30}$$

In order to assess the relationship between prevalence and the model parameters, we performed a Latin Hypercube Sensitivity Analysis [72]. This analysis randomly generates a matrix of 10,000 vectors of the model parameters within given bounds, which were chosen as 25% above and below the baseline parameters. Using parameters, $\beta_{AA}, b_A, d_B, d_J, d_A, u_J, \alpha, \gamma, \eta, \delta$, we verified that the monotonicity of the output measure

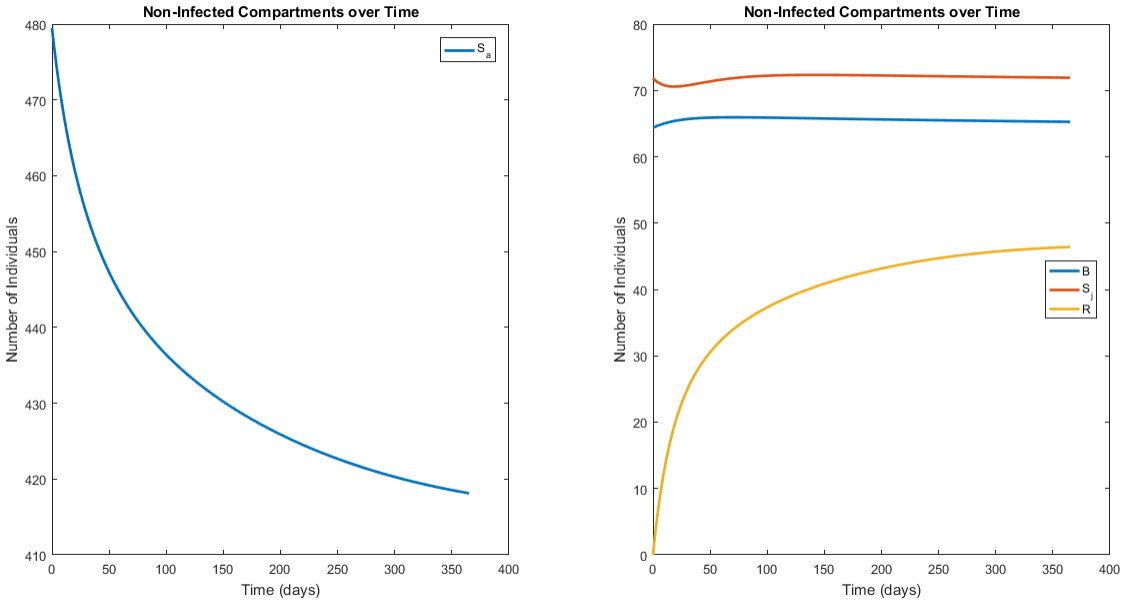

Figure 2.3: Plots showing the non-infected compartments with S_A on left and B, S_J and R on right over a 365 day period.

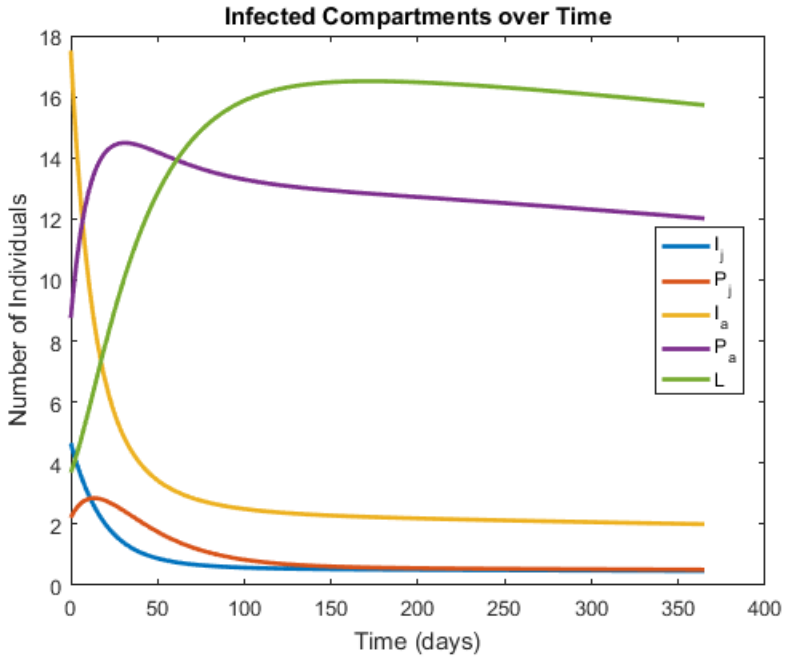

Figure 2.4: Plots showing the overall population in each infected compartment over a 365 day period.

of prevalence with respect to each parameter, which is needed to use the Partial Rank Correlation Coefficients (PRCC). Simulations were then run using each vector of the parameters, with the output measure of prevalence after 300 time steps. A PRCC and a p-value are then produced for each model parameter. The sign and magnitude of the correlation coefficient, as well as the p-value, were used to assess the impact and statistical significance of each parameter on the prevalence of the virus within the population. Using this analysis, we found that all parameters were statistically significant, having p-values below 0.05 (Figure 2.5). Note that the baseline birth and death rates, b_A and d_A, are known from this data and we do not need to further investigate its effect. If the rodents recover, then they do not get into the loop for P_A and L, and thus we will choose to vary the recovery rate γ to document its effect on prevalence. We will vary the transition rates, δ and η, between classes L and P_A.

To test how prevalence is affected by changes in the initial rodent population, the model was simulated at the baseline values. The model was simulated four times, each time varying the initial condition of one of the compartments B, L, I_A, S_j, at ten times its baseline initial condition value while keeping all other initial values at the baseline (Figure 2.6). For example, if the baseline value for infected adults was 1, a simulation would be run in which the initial number of infected adults was 10 and all other initial values were maintained at the baseline. This gives a comparison of how differences in the initial population of a particular class affect the prevalence of the virus. When varying the age structure of the population in our model, prevalence was found to be more sensitive to increases in the initial number of individuals in the younger age classes (Figure 2.6). When newborns and susceptible juveniles are the age class with the largest initial population, the prevalence of hantavirus spikes and eventually returns to the baseline. When the dominant initial age class is comprised of adults, the spike in prevalence is much lower and the prevalence returns to the baseline in a shorter amount of time than when the younger age classes are initially dominant. This is of particular importance during conditions that are conducive to newborn and juvenile success such as high resource availability or lack of predators.

In order to compare the effect of the infected (I) and latent (L) classes on prevalence, the model was run at the disease free equilibrium to get a baseline for comparison. Then, one

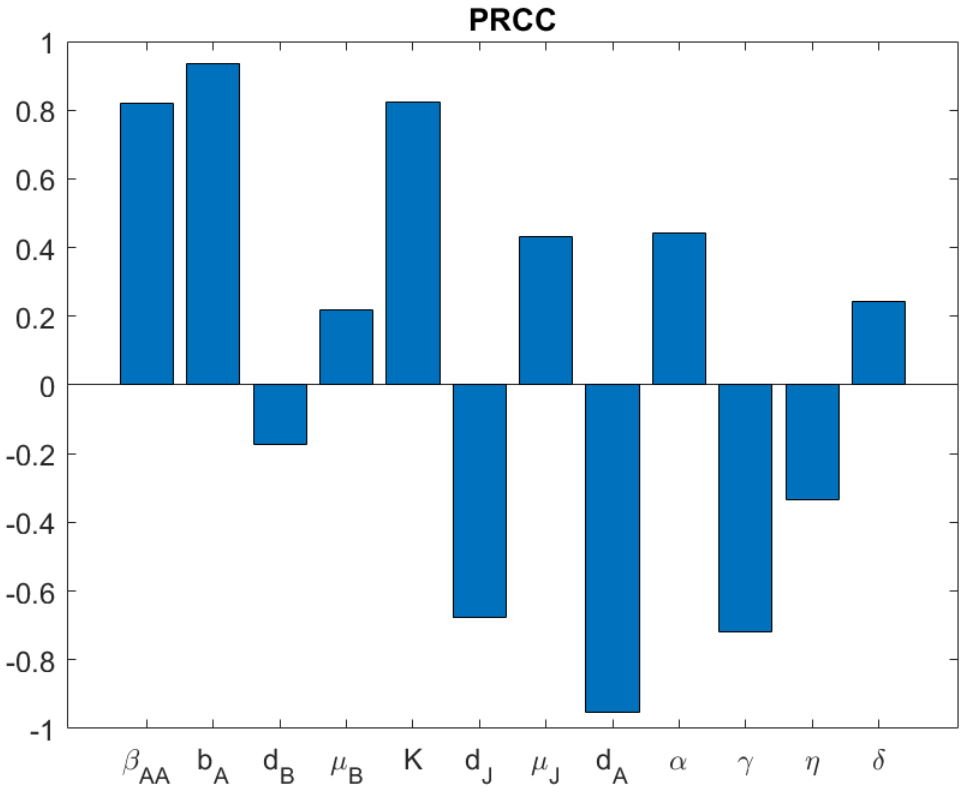

Figure 2.5: Chart showing the PRCC coefficients for model parameters determined using Latin Hypercube Sampling. All parameters had p-values less than 0.05 and were considered statistically significant.

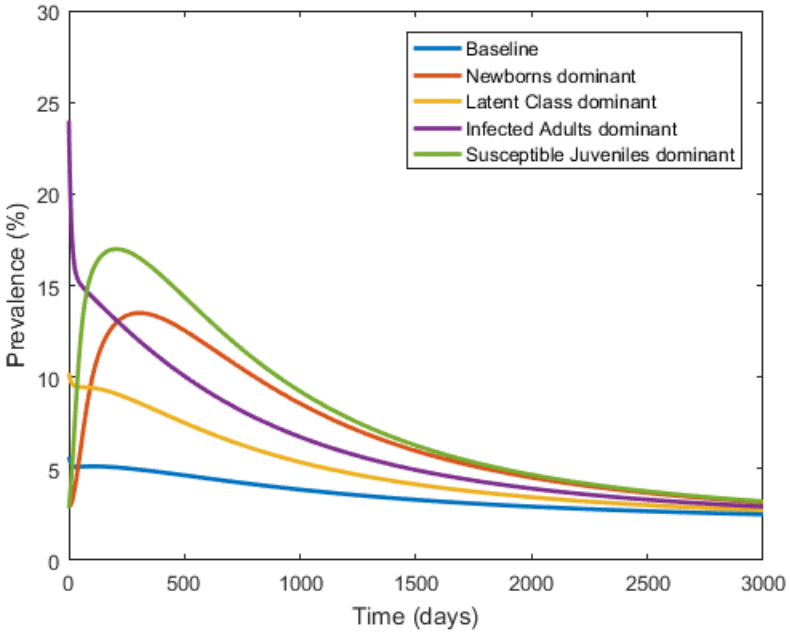

Figure 2.6: Graph showing the effect of different initial age structures on prevalence. The dominant initial class was the class that was set at 10 times its baseline value initially.

trial was run with the addition of 50 individuals in the I class and then with the addition of 50 individuals in the L class. Increases in the number of individuals in the latent class increased prevalence over the course of a year more than the acutely infected classes (Figure 2.7). This indicates that the latent class may be partially responsible for observed spikes in the number of infected individuals that have been observed in populations over time.

Next, we focus on transition rates in and out of our latent class. To study the effect of the transition rate δ from the latent class back to persistently infected class, simulations were run at different values of δ (Figure 2.8). Increases in δ caused significant increases in prevalence, as higher δ value results in a prevalence that is much higher than that corresponding to the baseline value $\delta = \frac{1}{60}$. This class has not been highly studied, however, our results demonstrate that it can have a large impact on the overall prevalence of hantavirus, which suggests it should be considered in future hantavirus models.

One can see the effect of the transition rate η from persistently infected class to the latent class in the results of simulations shown in Figure 2.9. As the rate η increases from $\eta = \frac{1}{56}$ to $\frac{1}{19}$, the prevalence doubles due to more individuals being able to transmit the virus.

In order to compare the effect of the recovered class, the recovery rate γ was altered 50% above and below the baseline conditions (Figure 2.10). The recovered class appears to have a large impact on the overall prevalence. When γ, the recovery rate, was varied by 50% above and below the baseline $\gamma = 1/10$ while keeping all other parameters constant, the overall prevalence ranged from 2% to 13%. When the rate of recovery was lowered to 1/15, the prevalence in the population was maintained at over twice the baseline value $\gamma = 1/10$. When it was increased to 1/5, the disease was eliminated from the population.

To understand the potential importance of the L latent compartment in this model, we ran a simulation without that class and compared the resulting prevalences. In Figure 2.11, the prevalence resulting from removing the L compartment is much higher than the prevalence found in our empirical data. Thus the L compartment may represent a reasonable mechanism in these dynamics.

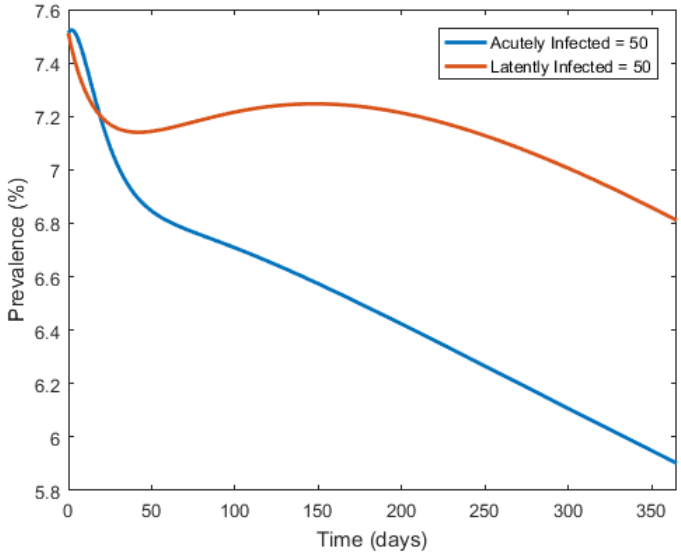

Figure 2.7: Graph comparing the effect of the latent and acutely infected classes on prevalence. These simulations were run at the disease free equilibrium, once with the addition of 50 individuals in the L class and once with the addition in the I class.

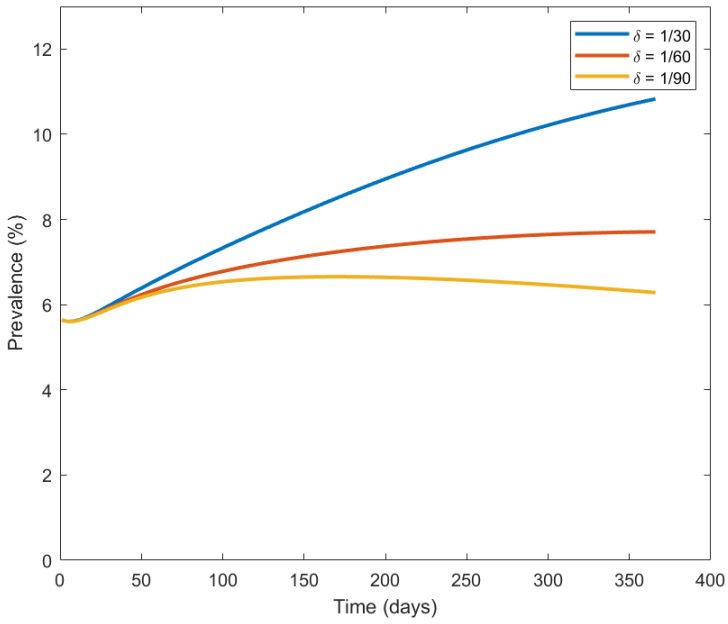

Figure 2.8: Graph showing the effect of the rate of transition δ on prevalence

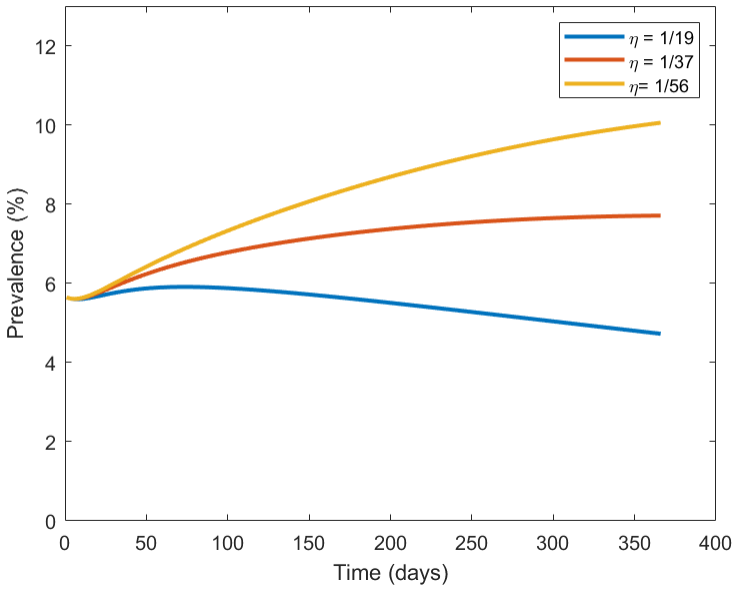

Figure 2.9: Graph showing the effect of the rate of transition η on prevalence

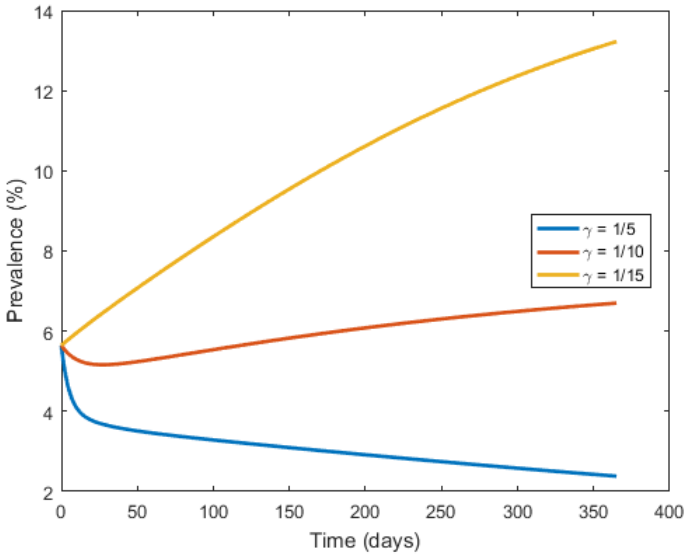

Figure 2.10: Graph showing the effect of varying the recovery rate γ on prevalence.

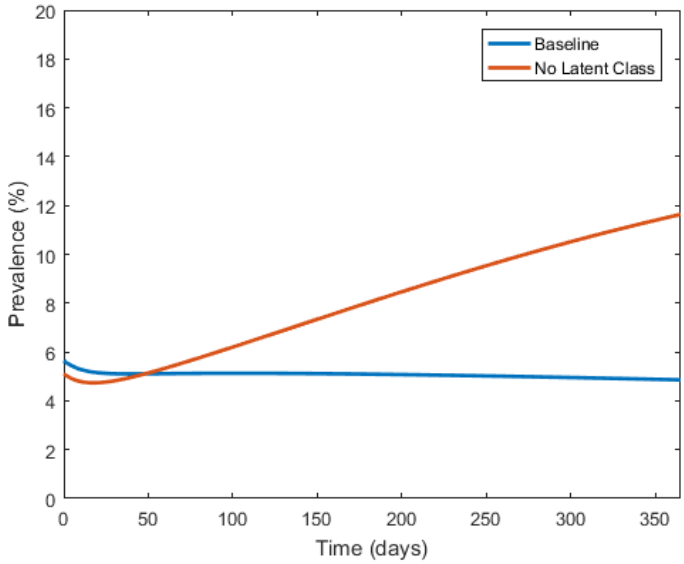

Figure 2.11: Graph comparing prevalence over a 365 day period for our model run at our baseline parameters from Table 2.2 and with the latent class removed.

2.4 Conclusions and Future Work

To understand the persistence of hantavirus in *A. montensis*, the rodent reservoir population represented by our data, we included a new non-infectious carrier class, L, into our model and investigated its effect on prevalence. In our model, individuals in this class (L) can transition back to the persistently infected, P_A, and begin to transmit the virus again. When the latent class is removed, the prevalence in the model is much higher (Figure 2.11) than that of our baseline model and the data set. The impact of the transition rates between latently infected and persistently infected adults is documented in Figures 2.8 and 2.9. Our simulation results indicate that this class may provide a potential explanation for the low prevalence endemicity observed in *A. montensis* populations in the Mybaracyú Reserve, and other locations throughout Paraguay and South America.

The addition of a carrier class can lead to different endemic equilibrium scenarios, depending upon the force of infection, as seen in certain communicable diseases, such as HBV [76]. When R_0 is high, infections are skewed to a younger age class, and in the opposing scenario with a lower R_0, infections are skewed to an older age class [55]. Provided an age structured model, such as ours, and different initial conditions; a variety of endemic prevalence values are possible given different initial conditions. Although we only consider the parameters derived from our data set (and our models robustness to those parameters (Figures 2.5, 2.7, 2.8, 2.9, 2.10), the literature documents a large range of observed prevalence values across various host species and environments [23, 106]. Motivated by low prevalence endemicity and modulations in prevalence, we created a model that incorporates both age structured population and carriers, which as previously noted, allows for endemicity with a wide array of parameter values. We also want to note that the transient dynamics observed when age structure is skewed suggest that factors affecting the age class distribution on a short-term timescale can heavily influence the prevalence. This result is very relevant to the observed endemicity at low percent prevalence (presumably when there's an adult age skew), and it also may explain prevalence's correlation to seasonal, annual, or semi-annual outbreaks, as temporal modulations affect the resources that facilitate high fecundity and survival of young *A. montensis* [73]. With an age structure that allows for sporadic

outbreaks and a carrier class that serves as a long term reservoir, we were able to evaluate the prevalence of JABV over time within its rodent reservoir, *Akodon montensis*, and offer insight into various other hantavirus systems.

Our work with discrete models, though primarily focused upon the epidemiological status of individuals as well as the age structure of our sample population, could be built upon by other modeling techniques and additional heterogeneities. Environmental characteristics such as seasonality and landscape structure have been linked to increased prevalence, as have anthropogenic factors such as ecological disturbance and its corresponding impact on biodiversity [61, 67, 71, 73]. Our model, which does not consider temporal variation, suggests that seasonal modulations are important for cyclical outbreaks (Figure 2.6), and it is well known that spatial arrangement of populations affects the endemicity and equilibrium prevalence of diseased populations [39, 42, 114]. Therefore, future work that considers spatial and temporal heterogeneity will complement our research, as well as past hantavirus modeling studies [6, 8, 116, 117]. Additional research that considers anthropogenic and climate change, as well as optimal control strategies to limit zoonotic transfer during outbreaks will also contribute to the understanding and management of one of the deadliest viruses in the world.

2.5 Data Access

Note that no new data was used in this paper, and the data used here were from [28].

2.6 Acknowledgements

This work was conducted during the 2016 Summer Research Experience (SRE) for undergraduates program at the National Institute for Mathematical and Biological Synthesis, an Institute sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville. Robert Owen was partially supported by the Programa Nacional de Incentivo a los Investigadores (CONACYT, Paraguay). Robert Owen and Colleen Jonsson were supported by NIH grant R01 AI103053. M. Igoe was supported by NIH/NIGMS-IMSD grant R25GM086761 and M.A. Rúa was supported by start-up funds from Wright State University. We would also

like to thank Dr. Linda Allen, Dr. Jason Bintz, and Eric Carr for their assistance on our project.

Chapter 3

Reframing Optimal Control Problems for Infectious Disease Management in Low Income Countries

3.1 Introduction

Due to the systematic non-linearities and pervasive heterogeneities characterizing contagious processes at the population level, epidemiological modeling is a key tool to explore a range of intervention strategies compared to approaches based on good sense, general principles, emotional pressures or, worse, by trial and error [18]. The question of what is the best timing and treatment effort for communicable diseases - whether through drug, antibiotics, vaccines, or non-pharmaceutical interventions - has been central in epidemiological modeling since the publication of the first malaria models by Ross and MacDonald between the end of the nineteenth and the beginning of the twentieth centuries[102].

Identification of optimal disease control strategies is based on the following rationale: in the absence of health intervention, an infectious disease follows its epidemic curve and causes substantial morbidity and mortality. By investing resources in control and treatment, it is possible to reduce the burden of disease. The higher the control and elimination effort, the lower the disease burden and the higher the costs of interventions to control the disease.

These relationships are often non-linear, with decreasing returns on the investment - i.e., health benefits per dollar invested - at high control effort. Accordingly, extreme control efforts, while desirable, might be infeasible or too expensive. There might be intermediate levels of control that minimize the overall costs accrued by the society, the sum of the intervention costs (which increase more than linearly with the control effort) and the cost associated with the burden of disease (which decrease with control effort).

Optimal Control (OC) theory has proved to be a particularly powerful mathematical tool to identify the best strategies for the control of infectious diseases. In contrast to most optimization approaches which assume control variables to be constant in time, the admissible control variables in OC (e.g., vaccination rate, drug treatment) can vary in time with the goal to minimize a suitable combination of the costs of intervention and the cost of the disease.

In the last couple of decades, an increasing number of papers have been published on OC applications for the control of infectious diseases of public health importance. With few exceptions, OC studies are notably missing from public health or medical journals where these studies should be expected to have the highest impact. The effect is that OC papers, typically published in journals of applied mathematics are considered too theoretical for public health scientists, managers, and practitioners.

Here we argue that an additional obstacle to the widespread uptake of OC theory for the control of communicable diseases is partially hidden in the typical framing of OC problems in epidemiology. In OC applications to disease control, the objective function to be optimized is generally formulated as the sum of intervention costs plus the cost associated with the burden of disease, assuming that both can be easily estimated monetarily with a reasonable degree of precision [3, 9, 51, 65, 79]. Yet, although routinely applied, the estimation of the monetary cost of disease burden is fraught with uncertainty and inherently plagued with ethical dilemmas that are even more relevant in low income countries and subsistence economies and in the lack of a comprehensive public health reporting system [11, 16, 48, 81, 96].

In this paper, we show that these limitations can be partially overcome by measuring epidemiological outcomes via well known health metrics and reformulating the goal and the

constraints on the OC problem to more closely reflect the framework in which governments and departments of health operate.

3.1.1 The challenges of estimating the monetary value for health outcomes

In industrialized countries with developed public health, welfare and demographic reporting systems, both the private and public sectors (i.e., health insurance and health departments/agencies) routinely assess the costs associated with diseases, accounting for the cost of over-the-counter and prescribed drugs, medical visits, hospital admissions, special treatments, loss of productivity, etc. This systematic monitoring and reporting effort provides a reasonable foundation to comprehensively estimate the monetary costs of disease outbreak and a range of public health interventions, which can be used to explore alternative control scenarios within an OC theory framework. However, this approach is not exempt from limitations. First, monetization of human health is ethically debatable, especially in the case of fatal diseases and when diseases disproportionately affect a specific segment of the population, such as underrepresented minorities, essential workers, or the homeless population [100].

Second, monetization of human health and/or loss of productivity becomes *particularly* problematic in low income countries, plagued by a wide range of Neglected Tropical Diseases (including several parasitic, viral, and bacterial diseases, such as schistosomiasis, Chagas disease, dengue fever, guinea worm disease, echinococcosis, human African sleeping sickness, and leishmaniasis) and other diseases of poverty, such as malaria, tuberculosis, cholera and HIV, affecting altogether more than 2.5 billion people worldwide [86]. For these countries, where a non-marginal fraction of the population is often engaged in a rural, cashless economy, using the same monetary currency to contrast the (usually observable) costs of public health interventions with the burden of disease - which may include also reduced capacity to perform physically engaging activities in subsistence agriculture, or the long term impact of chronic infections - is highly challenging and debatable. Some studies have explicitly recognized the challenges of estimating the cost of disease in monetary terms, and analyzed OC problems as a function of loosely defined "relative costs of disease burden" with respect to the cost

of intervention (see for instance, [13]). Yet the problem of integrating economic and public health metrics and indicators measured on vastly different units in a single monetary scale remains unsolved.

Cost-effectiveness analysis (CEA) partially circumvents these ethical shortcoming and practical challenges, as it abstains from measuring the monetary value of health outcomes and, instead, contrasts alternative control strategies in terms of public health metrics like Disability-Adjusted Life Years (DALYs) averted per dollar invested [69]. Although powerful, this approach is not exempt from limitations [115], and CEA does not typically leverage the power of OC theory to explore whether time-varying control strategies can be superior to time constant strategies.

3.1.2 Reframing optimal control problems

In this paper, we propose a simple reframing of optimization problems that allows applied mathematicians and epidemiologists to leverage the strengths of optimal control theory without incurring the challenges and limitations of assessing the monetary value of human health. Governments, health departments and governmental organizations do not operate in a vacuum but need to plan health interventions under budget constraints and/or to achieve specific health outcomes, such as to avoid overwhelming the health system by limiting the number of people requiring hospitalization [78]. Accordingly, OC problems should be reformulated to fully embrace this more empirical approach to the control of communicable diseases. Specifically, we argue that OC problems can be restated in one of the two following broad and more meaningful ways, namely:

- Alternative 1: minimize disease burden (the goal), measured in the most suitable, non-monetary units for the specific disease system under study for a given budget (the constraint) [82];
- Alternative 2: minimize the monetary cost of interventions (the goal) to achieve a specific epidemiological outcome (the constraint - set as a requirement of the control strategies to be admissible).

The first case recognizes that budget constraints are a significant factor in choosing control strategies in low income countries. In the second case, the emphasis is on achieving a

desired epidemiological target in the most efficient way, such as in the 2021 United Nations Political Declaration on HIV and AIDS where there is a goal of reducing annual new HIV infections to under 370,000 and annual AIDS-related deaths to under 250,000 by 2025 [10]. Control policies aimed at avoiding to exceed regional hospital capacity at the peak of a COVID-19 outbreak are also examples of the second case [78].

Here below, we use a classic cholera compartmental model that was kept deliberately simple for demonstration purposes to show that it is not only straightforward to reframe optimal control problems in these two alternative ways, but also more informative. In fact, we show that under the new frame (i) analyses can be easily performed to assess how the outcome changes with respect to, say, increasing budget (Alternative 1) or more ambitious epidemiological goals (Alternative 2) and (ii) the results of these analyses can be presented in terms of marginal utilities, that is, additional lives saved per extra dollar invested with respect to baseline.

The paper is structured as follows. After presenting a basic, mechanistic compartmental model for cholera with vaccination rate as the control variable, we first define the control problem as the minimization of the sum of the cumulative burden of disease*, estimated in monetary terms*, *and* of the cost of intervention during the simulation time, assuming the conversion factor from number of infected cases to dollars is known - referred here below as the "combined" OC problem.

Then, we reformulate the OC problem in the two alternative ways described above. We show that, if the conversion coefficient to assess the monetary value of disease burden can be estimated with reasonable precision, the outcome of the alternative OC formulations can be mapped back to the combined OC problem with no loss of information. Thirdly, we show that the two alternative formulations allow us to better explore the sensitivity of the OC solutions to changes in available budget or the desired epidemiological target. We conclude the paper by discussing future development and additional caveats and limitations of comprehensive cost assessment in epidemiology.

3.2 Cholera model

Cholera is a waterborne diarrhoeal illness caused by infection of the intestine with the bacterium *Vibrio cholerae* with both direct and indirect modes of infection and transmission, primarily through contaminated food and water supplies. The disease causes rapid dehydration and electrolyte imbalances. There are an estimated 3-5 million cholera cases and 100,000-120,000 deaths due to cholera every year [4, 66]. The most recent outbreak in Haiti caused over 10,000 deaths and over 820,000 cases [63, 121]. The disease has a short incubation period, between two hours and five days, affecting all ages. It is known that cholera persists in an aquatic reservoir and can exist in non-culturable, but viable, state for months to over a year [77].

There are two types of safe and effective oral cholera vaccines, Dukoral and Shanchol. Both provide short term protection (approx. 2 years) against *Vibrio cholerae* over all age groups, but Dukoral requires water for intake. Both are administered in two doses given between 7-14 days apart [101].

We illustrate our points using a basic model over a short time period. For comparisons of several models applied to outbreaks in Haiti and with vaccination scenarios, see the work of Lee et al. [63]. Tien and Earn (2010) [108] extended the classic Susceptible (S), Infected (I) and Recovered (R) SIR model by Codeco [24] to include a concentration of vibrios (W) in the water compartment. This allows for both direct (fecal-oral) and indirect (from water-borne pathogen) transmissions pathways for cholera at a per capita rate β_I and β_W respectively. The resulting $SIRW$ model with 4 state variables is described by the following differential equations:

$$
\begin{aligned}
S' &= \mu N - \beta_I SI - \beta_W SW - \mu S - v(t)S \\
I' &= \beta_W SW + \beta_I SI - \gamma I - \mu I - \delta I \\
R' &= \gamma I - \mu R + v(t)S \\
W' &= \xi(I - W)
\end{aligned}
$$

where $N = S + I + R$ represents the human population, μ is the natural mortality and birth rates of a human population, γ is the recovery rate from infected to recovered and resistant class, and δ is the death rate due to the disease. Note that the compartment W has been scaled to be in the same units as the I compartment (individuals). Therefore ξ is a parameter representing both per capita shedding rate of an infected individual I of viable infectious propagules in the water and the decay rate of the pathogen in the environment; see more detail in Tien and Earn [108] and Kelly et al. [57]. The term $\beta_I I$ is the force of infection, i.e., the rate at which susceptible individuals become infected via contaminated fecal-oral transmission; and lastly $\beta_W W$ is the force of infection via contaminated water. The parameters and their units are described in Table 3.1.

As in Kelly et al. (2016) [57], we assumed that susceptible individuals S are vaccinated and become resistant R at a rate $v(t)$, with $v(t)$ combining efficacy and rate of distribution. The basic reproduction number for this model in the absence of the control $v(t)$ is

$$R_0 = \frac{(\beta_I + \beta_W)S^*}{\gamma + \delta + \mu},$$

where S^* is the total susceptible population in the absence of disease.

The combined objective functional is the sum of the cumulative costs associated with the burden of the disease $J_b(v)$ and the cost of intervention $J_c(v)$, which are both functions of control effort $v(t)$, which, in turn, is a function of time, namely:

$$\min_{v \in V}(J_b(v) + J_c(v)) \tag{3.1}$$

with disease dynamics described by the above SIRW model. The control effort $v(t)$ is subject to constraints $0 \leq v(t) \leq v_{max}$, with v_{max} being the maximum vaccination rate that can be deployed during the simulation time $0 \leq t \leq T$.

The total cost of intervention is usually assumed to be proportional to the number of individuals that are vaccinated, $v(t)S(t)$, and to increase more than linearly with vaccination effort $v(t)$,

$$J_c(v) = \int_0^T [Bv^2(t) + Cv(t)S(t)]dt,$$

Table 3.1: Parameter descriptions with units.

Parameters	Description	Units
μ	natural birth/death rate	day^{-1}
β_I	direct transmission rate (P-P)	$\text{day}^{-1}\text{individuals}^{-1}$
β_W	indirect transmission rate (W-P)	$\text{day}^{-1}\text{individuals}^{-1}$
γ	transition rate from I to R	day^{-1}
ξ	decay rate of pathogen in water	day^{-1}
δ	death rate due to disease	day^{-1}

where B and C are positive constants to account for the cost of vaccination.

The cost associated with disease burden depends upon the specific disease system under study and the available information. It can be represented as

$$J_b(v) = \int_0^T A_b I(t) dt,$$

where A_b is a positive scaling coefficient accounting for the aggregated mean monetary value c_1 of each day of illness (including drug treatment and loss of productivity) and, in case of fatal disease, also for the value c_2 of human lives lost, namely:

$$A_b = c_1 + c_2 \delta,$$

where δI is the death rate due to the disease.

As an alternative, the monetary burden of disease might be also estimated as a flat rate associated to each infected case and, as such, derived from disease incidence, i.e., the total number of infected cases during the simulation time, namely:

$$J_n(v) = \int_0^T A_n[\beta_I S(t) I(t) + \beta_W S(t) W(t)] dt,$$

with the β terms giving the rate of new cases.

The precise form of optimal control is affected by the relative values of the coefficients A_b, A_n, B, C, c_1 and c_2. For simplicity, here we will compute disease burden as the total number of infections and set $A_n = 1$, since the relative sizes of the coefficents determine the optimal control. The optimal control can be found using the Pontryagin's Maximum Principle [89] as described in Appendix A.

3.3 Optimal Control Formulations

3.3.1 Combined Formulations

Combined Objective Functional with Disease Burden and Cost

The objective functional for minimizing both disease burden and cost is defined as

$$\min_{v \in V} \int_0^T [A_b I(t) + Bv^2(t) + Cv(t)S(t)]dt, \tag{3.2}$$

where the control set is

$$V = \{v \in L^2(0,T) \mid 0 \leq v(t) \leq v_{max} \ a.e.\}.$$

Combined Objective Functional with New Infections and Cost

The objective functional for minimizing both the total number of new infections and intervention cost is defined as

$$\min_{v \in V} \int_0^T [A_n(\beta_I S(t)I(t) + \beta_W S(t)W(t)) + Bv^2(t) + Cv(t)S(t)]dt. \tag{3.3}$$

3.3.2 Alternate Formulation 1: Minimize New Infections under Fixed Budget

Given a fixed budget G for the cost of the control interventions, we define the admissible control set (with constraint)

$$F_1 = \left\{ v \in V \mid \int_0^T (Bv^2(t) + Cv(t)S(t))dt \leq G \right\},$$

where S is the susceptible compartment of the population corresponding to the control v. The objective functional to minimize the new infections is defined as

$$\min_{v \in F_1} J_n(v) = \min_{v \in F_1} \int_0^T A_n[\beta_I S(t)I(t) + \beta_W S(t)W(t)]dt, \tag{3.4}$$

subject to the above SIRW model.

To illustrate the process of obtaining the adjoint functions and the optimal control characterization, we use Pontryagin's Maximum Principle [89]. To handle the budget constraint, we introduce a fifth state variable, x_5, that satisfies

$$x_5' = Bv^2 + CvS$$

with the boundary conditions $x_5(0) = 0$ and $x_5(T) = G$. Then, the Hamiltonian using this principle for our optimization is

$$\begin{aligned} H = {} & A_n\beta_I SI + A_n\beta_W SW + \lambda_1[\mu N - \beta_I SI - \beta_W SW - \mu S - vS] \\ & + \lambda_2[\beta_W WS + \beta_I SI - \gamma I - \mu I - \delta I] \\ & + \lambda_3[\gamma I - \mu R + v(t)S] \\ & + \lambda_4[\xi(I - W)] \\ & + \lambda_5[Bv^2 + CvS] \end{aligned}$$

with the following adjoint differential equations and boundary conditions

$$\begin{aligned} \lambda_1' &= -[A_n\beta_I I + A_n\beta_W W - \lambda_1\beta_I I - \lambda_1\beta_W W - \lambda_1 v + \lambda_2\beta_W W + \lambda_2\beta_I I + \lambda_3 v + \lambda_5 Cv] \\ \lambda_2' &= -[A_n\beta_I S + \lambda_1\mu - \lambda_1\beta_I S + \lambda_2\beta_I S - \lambda_2\gamma - \lambda_2\mu - \lambda_2\delta + \lambda_3\gamma + \lambda_4\xi] \\ \lambda_3' &= -[\lambda_1\mu - \lambda_3\mu] \\ \lambda_4' &= -[A_n\beta_W S - \lambda_1\beta_W S + \lambda_2\beta_W S - \lambda_4\xi] \\ \lambda_5' &= 0 \\ \lambda_1(T) = 0, \lambda_2(T) = 0, \lambda_3(T) = 0, \lambda_4(T) &= 0. \end{aligned}$$

The adjoint system is formed by $\lambda_i' = -\frac{\partial H}{\partial x_i}$, where x_i represents the i$^{\text{th}}$ state variable. Note λ_5 takes no boundary condition since x_5 takes two boundary conditions.

We can then characterize the optimal control on the interior of the control set using this optimality condition,

$$\frac{\partial H}{\partial v} = 2B\lambda_5 v + C\lambda_5 S - \lambda_1 S + \lambda_3 S = 0 \text{ at } v^*,$$

at the corresponding optimal states and adjoints and the control bounds to yield

$$v^*(t) = \min\left(v_{max}, \max\left(0, \frac{(\lambda_1(t) - \lambda_3(t) - C\lambda_5(t))S^*(t)}{2\lambda_5(t)B}\right)\right).$$

3.3.3 Alternate Formulation 2: Minimize Intervention Costs with Fixed New Infections

Assuming the desired outcome is the cost of number of new infection cases being bounded above by P, we define the admissible control set

$$F_2 = \left\{ v \in V \mid \int_0^T A_n[\beta_I S(t)I(t) + \beta_W S(t)W(t)]dt \leq P \right\}.$$

The objective functional, defined to minimize the costs of implementing the control, is

$$\min_{v \in F_2} J_c(v) = \min_{v \in F_2} \int_0^T [Bv^2(t) + Cv(t)S(t)]dt, \tag{3.5}$$

subject to the above SIRW model. To handle the constraint on the new cases, we introduce a new state variable, x_5 with

$$x_5' = A_n[\beta_I S(t)I(t) + \beta_W S(t)W(t)],\ x_5(0) = 0,\ x_5(T) = P$$

similar to the previous formulation.

Note that A_n is included in 3.4 and 3.5 to show the connection between the combined and alternate formulations and allow us to demonstrate how the results of the combined case can be recovered from the alternate cases. However, to achieve the goal of reformulating optimal control problems to avoid the monetization of human health, this constant can be redefined so that the objective function represents minimizing only the total number of new infections and not the monetary cost associated with these infections.

3.4 Results

3.4.1 OC solutions are superior to the best alternative with constant control

Modulating control variables in time may provide a more efficient and cost-effective intervention strategy than the best alternative with a constant control. This point is illustrated by using our cholera model and a combined objective functional (3.3) to minimize the cost of interventions plus the cost of the cumulative number of new cases. Our numerical results in this paper use the forward-backward sweep algorithm [68]. The results shown in Figure 3.1 and Table 3.2 show that the combined cost of intervention and new cases, $J_n(v) + J_c(v)$, is much larger in the case of no control than the cases of optimal or constant control. In the optimal control case, the control is time varying, staying at the maximum for approximately 40 days and then decreasing linearly to zero over the next 20 days. A constant vaccination rate $v \equiv 0.032$ leads to approximately the same intervention cost (black line in the plot) but with a higher number of cases. There is no constant value of vaccination rate for which the combined costs is lower than in the case of the optimal control solution.

3.4.2 Simulations

The optimal control policy is sensitive to how we estimate the monetary value of the burden of disease, which is notably fraught with uncertainty. For instance, following Castonguay et al. (2020), we can reasonably assume that each sick day entails a productivity loss equal to the daily per-capita Gross Domestic Product (GDP). However, in Africa, the per-capita GDP spans over a 15 fold difference between the richest countries (Gabon, Botsawa, Equatorial Guinea, Mauritius and Seychelles, all above $15,000 per year) and the poorest ones (Burundi, Somalia, Central African Republic, all below $1000/year) [107]. Keeping this in mind, to illustrate the implications of such a wide variation in the per-capita, daily productivity loss, we simulated cholera dynamics and derived the OC policy with the combined formulation (3.3) under two alternative assumptions, namely that the unit cost A_n of each new case is assumed to be equal to either 1 or 10 in suitable monetary units to measure the average productivity loss while sick. For the results shown in Figure 3.2 and Table 3.3, we use the

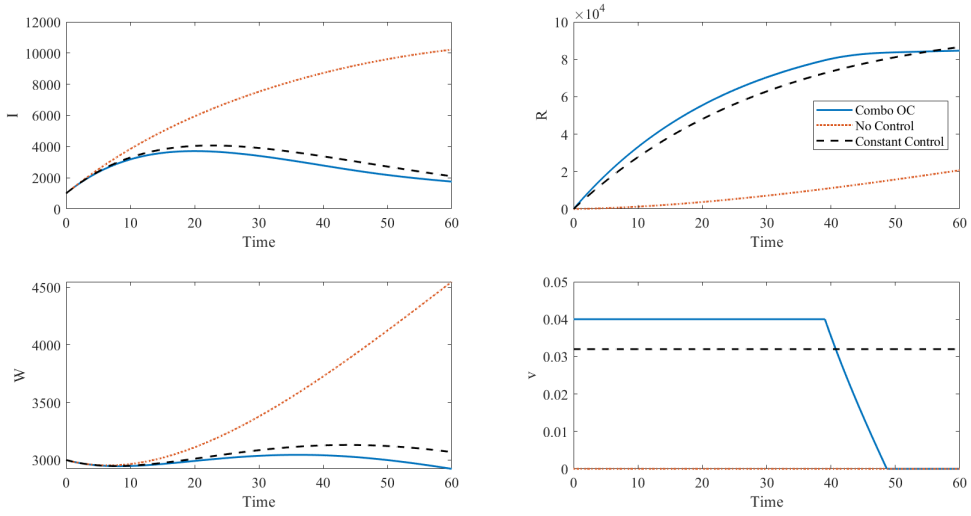

Figure 3.1: Optimal control and states for the combined objective functional (3.3) with new infections compared to a constant control with the same cost, a constant control at the maximum for all time, and the no control case, with parameters as in Table 3.2.

Table 3.2: Results of combined objective functional with time horizon of 60 days corresponding to Figure 3.1, using $A_n = 1$, $B = 8000$, $C = 0.0464$, $v_{max} = 0.04$, $\beta_I = 2.64e-7$, $\beta_W = 1.21e-6$, $\gamma = 0.05$, $\mu = 0.0001$, $\xi = 7.56e-3$, $\delta = 0.0005$, $S(0) = 99000$, $I(0) = 1000$, $R(0) = 0$, and $W(0) = 3000$. The R_0 for these simulations is 2.9.

	No Control	Comb. OC	$v \equiv 0.032$
Disease burden cost $J_n(v)$	30,219	9,337	10,829
Intervention costs $J_c(v)$	0	4,090	4,079
Total cost $J_n(v) + J_c(v)$	30,219	13,427	14,908
Max number of infected	10,219	3,712	4,072
Total number of deaths	208	85	96
Number of vaccinations	0	76,514	77,325

same β_I, β_W, μ, ξ, and δ parameters, initial conditions, and time horizon as in Table 3.2 and set $\gamma = 0.25$, $B = 100$, $C = 0.1625$, $v_{max} = 0.03$. The R_0 for these simulations is 0.59. Note that the optimal controls in these simulations are approximately bang-bang.

Simulations with our cholera model show that, when the unit cost of disease is undervalued by an order of magnitude (when $A_n = 1$), the optimal strategy entails 64% less vaccinations, a 71% reduction of number of days in which vaccines are deployed at the maximum rate, 40% more lives lost, and 51% more cases (Figure 3.2/Table 3.3). Reformulating the OC problem to minimize disease incidence under budget constraints or, in alternative, to minimize intervention costs with a cap on cumulative incidence, allows us to overcome the challenges in estimating the monetary value of disease burden, as well as the ethical perils and the uncertainties that plagues the combined formulation.

Do the alternative formulations systematically lead to different control policies than the OC policy derived by solving the combined formulation? Our analysis shows that, in the special (and unrealistic) case of perfect information - i.e., when the monetary value of disease burden can be derived with sufficient precision - budget constraints , or epidemiological constraints can be set so that the resulting OC policy of the alternative formulations is the same as the combined formulation (3.3). Specifically, we solved the OC problem for the first alternative formulation (i.e., minimize new infections under a budget constraint (3.4)) by setting the budget constraint to the intervention costs resulting from solving the combined OC problem (3.3). As shown in Table 3.4 and Figure 3.3, the OC policy and health outcomes (such as the maximum number of infected individuals, the number of deaths, and the number of vaccinations) are the same for the combined and the alternative formulations since the intervention cost budget is the same for both. We see the same for the second alternative formulation (i.e., minimize intervention costs to achieve a desired epidemiological outcome (3.5)) when the epidemiological outcome is constrained to the value derived by solving the combined OC problem (see Table 3.5 and Figure 3.4).

The simulations shown in Figure 3.3 (corresponding to Tables 3.4 and 3.6) and Figure 3.4 (corresponding to Tables 3.5 and 3.7) were performed using the same β_I, β_W, μ, ξ, and δ parameters, initial conditions, and time horizon as in Table 3.2 and setting $\gamma = 0.25$, $A_n = 1$, $B = 100$, $C = 0.0813$, $v_{max} = 0.03$.

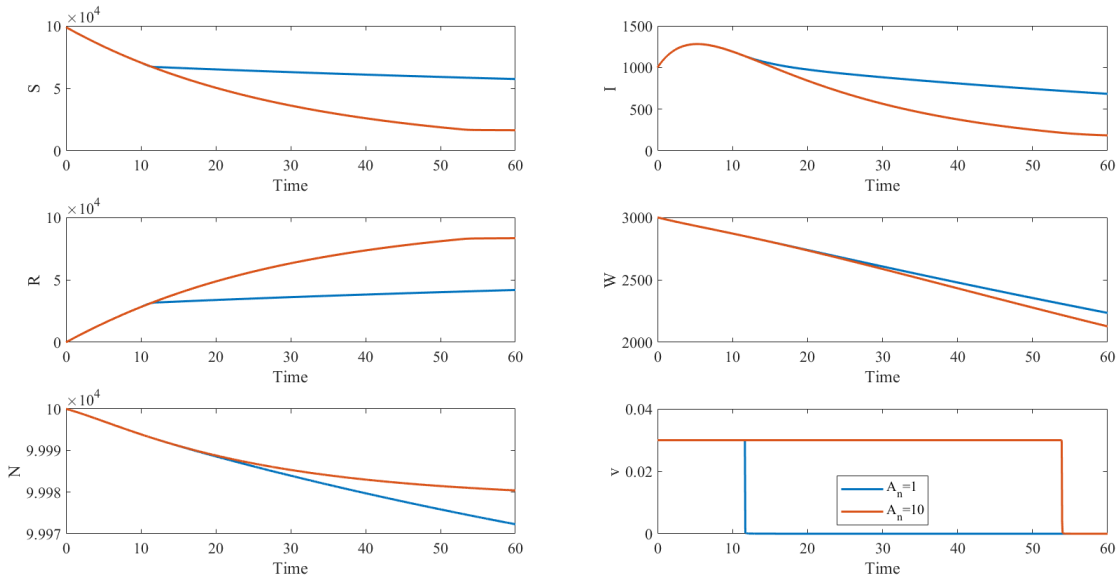

Figure 3.2: Optimal control and corresponding states for the combined objective functional for varying A_n (with 1 and 10) values corresponding to the results in Table 3.3.

Table 3.3: Epidemiological values corresponding to Figure 3.2 where the percent differences are measured from $A_n = 10$ case.

	$A_n = 10$	$A_n = 1$	% difference
Cumulative cases	9,011	13,580	+51%
Disease burden cost $J_n(v)$	\$90,105	\$13,580	-85%
Intervention costs $J_c(v)$	\$12,011	\$4,581	-62%
Total cost $J_n(v) + J_c(v)$	\$102,116	\$18,161	-82%
Total number of deaths	20	28	+40%
Number of vaccinations	73,654	26,634	-64%
Time spent at v_{max}	55	16	-71%

Table 3.4: Results of the objective functional with new infections under a budget constraint (3.4) compared to the case of no control and the combined objective functional with new infections (3.3), corresponding to the results in Figure 3.3. $ values are hypothetical and the $ symbol is used to clearly tease apart monetary values from health outcomes. The * symbol for the the alternative OC formulation indicates that the OC problem was solved by setting the intervention cost as a constraint. The † symbol indicates what value was minimized in the OC problem.

	No control	Comb. OC	Alt. OC
Intervention costs $J_c(v)$	\$0	\$4,746	$\$4,746^*$
Disease burden cost $J_n(v)$	\$19,080	\$9,814	-
Total costs $J_n(v) + J_c(v)$	\$19,080	$\$14,559^\dagger$	-
Cumulative cases	19,080	9,814	$9,814^\dagger$
Max number of infected	1,459	1,282	1,282
Number of deaths	38	21	21
Number of vaccinations	0	57,501	57,501

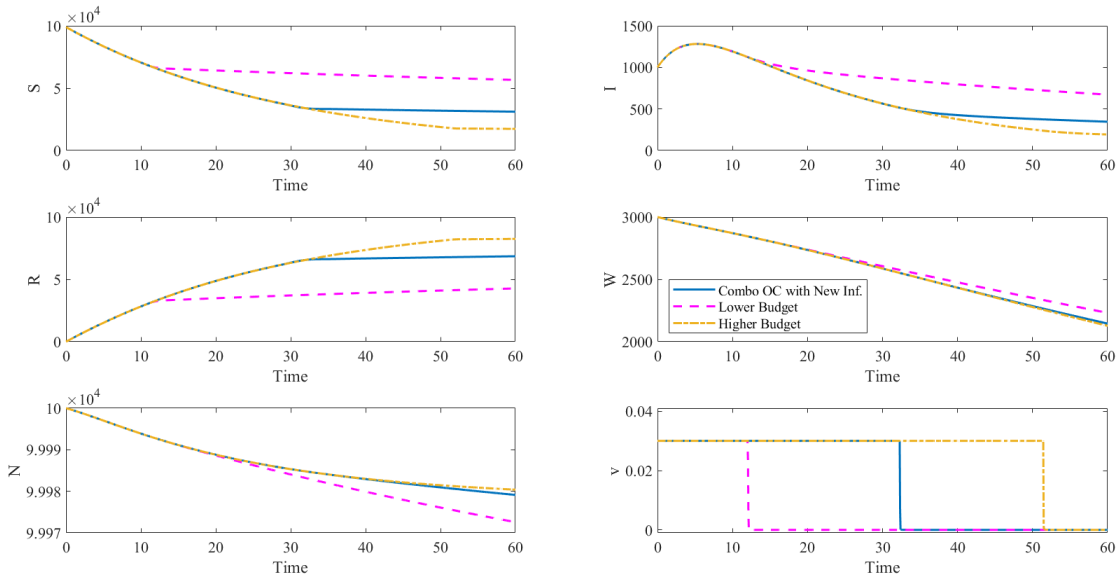

Figure 3.3: Optimal control and corresponding states for the combined objective functional with new infections (3.3) and the objective functional with new infections under a budget constraint (3.4), corresponding to the results in Tables 3.4 and 3.6.

Then what advantage do the alternative formulations provide? As commented above, in the majority of cases, the per-capita cost of disease is not known and cannot be estimated with any degree of confidence. Therefore, it is more meaningful and less debatable from the perspective of health equity to frame the optimization problem by explicitly acknowledging the fundamental reality any decision maker has to face. That is, to operate in conditions of limited budget, and/or, to have ethical, political, practical or humanitarian reasons to achieve a desired public health outcome. The alternative formulations allow us to incorporate these considerations much more straightforwardly than in the combined formulation. For instance, in the last two columns of Table 3.6 we report the breakdown for costs and epidemiological outcomes associated with a 50% budget decrease or a 25% increase, respectively, compared to the intervention cost found in the combined OC solution. The last two columns of Table 3.7 report the breakdown of the same information when the epidemiological target is 8% lower or 25% higher than the baseline number of new infections (as derived from the combined OC solution).

Note that, in the simulations shown in Tables 3.6 and 3.7, the maximum number of new infections does not change between the different budget and epidemiological constraints. This is because the peak of infections in those simulations occurs early in the simulation and the optimal controls in all of the cases shown are at the maximum control value from the beginning of the simulation until after the peak of infection. Therefore, in these cases, the maximum number of infected is not reduced by increasing budget or more ambitious epidemiological goal. To reduce the maximum number of infected individuals in these simulations, an increase in the maximum vaccination rate is required.

A type of sensitivity analysis can be conducted using alternative formulation 1 (minimizing new infections under a budget constraint (3.4)) by systematically exploring changes in health outcomes of interest over a wide range of budget constraints (see Figure 3.5). Our analysis shows that the relationship between budget and health outcomes are, for this model, approximately linear until one reaches the budget that uses the maximum vaccination rate over the entire simulation time. The analysis reveals that above this level, what is limiting is not vaccination budget but, rather, vaccination capacity (for instance, the number of vaccination centers). Therefore, above this limit, further health benefits could be achieved

Table 3.5: Results of the objective functional with intervention costs and an epidemiological constraint (3.5) compared to the no control case and the combined objective functional with new infections (3.3), corresponding to the results Figure 3.4. The * symbol for the the alternative OC formulation indicates that the OC problem was solved by setting the cumulative cases as a constraint. The † symbol indicates what value was minimized in the OC problem.

	No Control	Comb. OC	Alt. OC
Intervention costs $J_c(v)$	\$0	\$4,746	$\$4,746^{\dagger}$
Disease burden cost $J_n(v)$	\$19,080	\$9,814	-
Total costs $J_n(v) + J_c(v)$	\$19,080	$\$14,559^{\dagger}$	-
Cumulative cases	19,080	9,814	$9,814^{*}$
Max number of infected	1,459	1,282	1,282
Number of deaths	38	21	21
Number of vaccinations	0	57,501	57,501

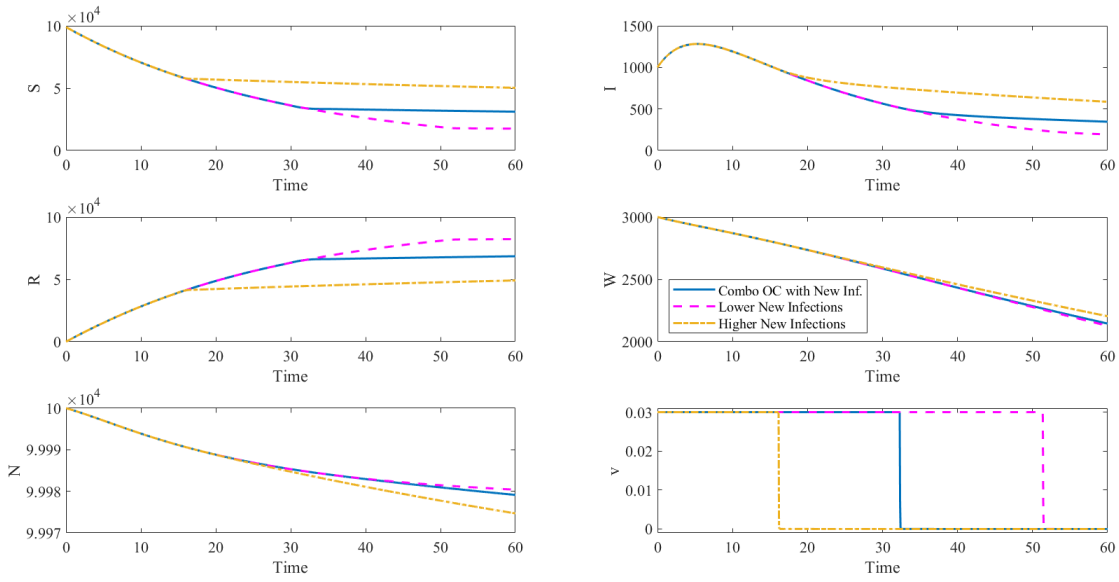

Figure 3.4: Optimal control and corresponding states for the combined objective functional with new infections (3.3) and the objective functional minimizing cost under an epidemiological constraint (3.5), corresponding to the results in Tables 3.5 and 3.7.

Table 3.6: Results of the combined objective functional with new infections (3.3) and the objective functional with new infections and a fixed budget (3.4) corresponding to the results in Figure 3.3. The * symbol for the the alternative OC formulation indicates that the OC problem was solved by setting the intervention cost as a constraint. The † symbol indicates what value was minimized in the OC problem.

	Alt. OC	Lower budget	Higher budget
Intervention costs $J_c(v)$	$\$4,746^*$	$\$2,373^*$ (-50%)	$\$5,934^*$ (+25 %)
Cumulative cases	$9,814^\dagger$	$13,416^\dagger$ (+37%)	$9,028^\dagger$ (-8%)
Max number of infected	1,282	1,282 (0 %)	1,282 (0%)
Number of deaths	21	27 (+29%)	20 (-5 %)
Number of vaccinations	57,501	30,624 (-47 %)	72,653 (+26 %)

Table 3.7: Results of the combined objective functional (3.3) and the objective functional with intervention costs and an epidemiological goal (3.5), corresponding to the results in Figure 3.4. The * symbol for the the alternative OC formulation indicates that the OC problem was solved by setting the cumulative cases as a constraint. The † symbol indicates what value was minimized in the OC problem.

	Alt. OC	Lower New Inf.	Higher New Inf.
Intervention costs $J_c(v)$	$\$4,746^\dagger$	$\$5,921^\dagger$ (+25%)	$\$2,983^\dagger$ (-37%)
Cumulative cases	$9,814^*$	$9,030^*$ (-8%)	$12,275^*$ (+25%)
Max number of infected	1,282	1,282 (0%)	1,282 (0%)
Number of deaths	21	20 (-5%)	25 (+20%)
Number of vaccinations	57,501	72,507 (+26%)	37,846 (34%)

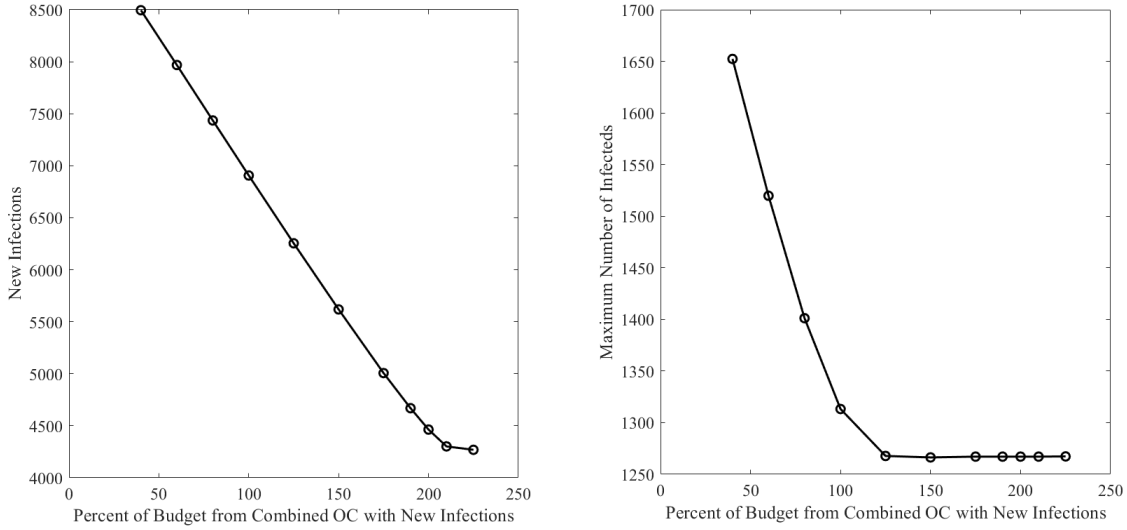

Figure 3.5: New infections and maximum number of infections as a function of available budget when minimizing new infections under a budget constraint (3.4), performed with $A_n = 1$, $B = 600$, and $C = 1$.

by a structural improvement of the vaccination system. The analysis reveals also that the peak incidence rate (new cases per day) level off at a much lower budget than the cumulative fatality cases (Figure 3.5).

For alternative formulation 2 (minimizing intervention cost to achieve a desired epidemiological outcome (3.5)), the sensitivity analysis shows that the total intervention cost and the cumulative deaths linearly increases/decrease, respectively, with more ambitious public health targets (percentage reduction in new cases with respect to the no control case - Figure 3.6). The maximum value of new daily infections drops linearly with more ambitious epidemiological outcome but levels off, in this case, when the required reduction in the cumulative number of cases exceed ca. 30% of that at baseline (no intervention).

The simulations shown in Figures 3.6 and 3.5 were performed using the same β_I, β_W, μ, ξ, and δ parameters, initial conditions, and time horizon as in Table 3.2 and setting $\gamma = 0.25$ and $v_{max} = 0.08$. The values of A_n, B, and C used are specified in the figure captions.

Simulations related to the optimal control formulations minimizing disease burden instead of new infections (3.2) can be found in Appendix B.

3.5 Discussion and Conclusions

Using a cholera model, we have illustrated the benefit of using time-varying controls to achieve the goals and incorporate the constraints of disease management. We have proposed valuable alternatives to the frequently-used objective function (the minimization of the sum of the total number of new cases (or total disease burden) and the cost of management actions) and showed the advantages of these alternatives in terms of outcomes (in costs, new cases and/or burden levels). Depending on the situation, one should consider minimizing the number of new cases given a fixed intervention budget or minimizing the intervention costs given a fixed epidemiological constraints (such as keeping the total number of new cases under a fixed level). These formulations can be valuable to apply to models of other diseases, such as schistosomiasis [103], West Nile virus [1], and Zika virus [80, 111].

We analyzed changes in health outcomes of interest over a range of budget constraints. This relationship between budget and health outcomes may be approximately linear until one reaches the budget level that allows the maximum the vaccination rate to be used over

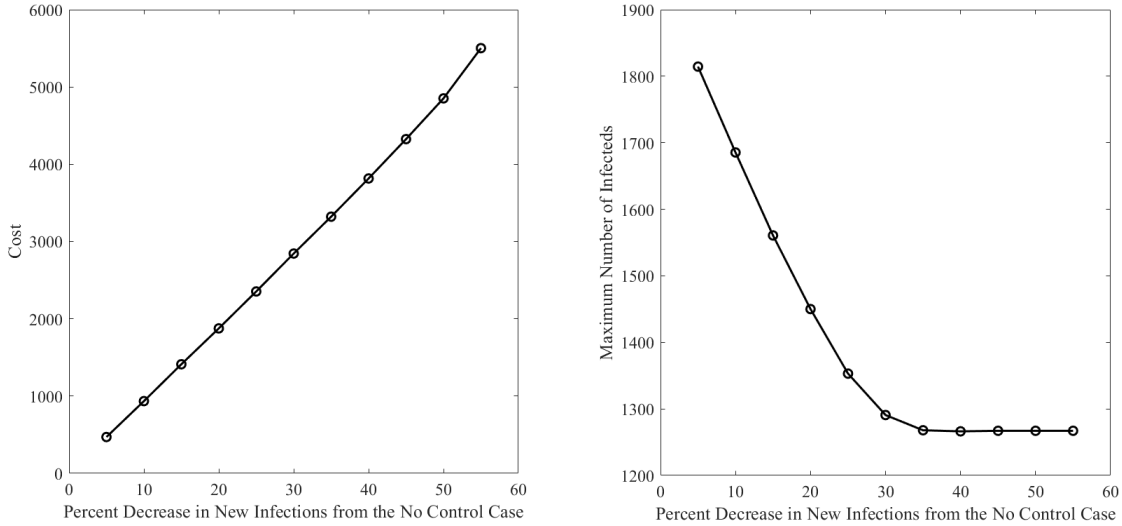

Figure 3.6: Intervention cost and maximum number of daily infections for the OC solution that minimizes intervention cost under an epidemiological constraint (3.5) as a function of the percent reduction in new infections from the no control case, performed with $A_n = 1$, $B = 600$, and $C = 1$.

the entire simulation time. The upper bound on the vaccination rate may be the limitation due to the capacity level (that is, the amount of vaccinations that can be completed per unit time given the current delivery system with a fixed number of workers). Similar results were obtained for changes in intervention costs over a range of fixed health outcomes. These results outline the power of reformulating combined optimal control problems and using health outcomes measured in metrics which public health decision makers are more familiar.

In some of our results, the differences in vaccination effort and epidemiological outcome merely reflect differences in the economic strengths of nations, and have little to do with the much less tangible value of the quality of life lost due to infection. This problem can be partially addressed by conducting a sensitivity analysis of how the OC solution changes as a function of the unit cost of disease, if such costs are known; see one such analysis in [13].

While the proposed reformulation of the optimization functional addresses some problems of the combined formulation and allows us to avoid the practical and ethical challenges of evaluating human health, it does not solve other problems inherent in the evaluation of health outcome and equity issues. For instance, Weintraub et al. [115] has clearly outlined several of these limitations for the cost-effectiveness analysis that apply also to optimal control problems in general and, specifically, to the newly proposed formulations. In addition, for practical and computational reasons, mechanistic models tend to simplify the complexity of real-world social systems, and optimal control theory cannot fix problems when models are oversimplified and lack relevant details (such as spatial structure), are not properly parameterized, or lack the real-world data to do so.

Without dismissing the power of OC applications, we are aware that simplified protocols with constant intervention rates might work equally well, if not better, than protocols that require continuous adjustment of the intervention effort. This may be especially true when it is difficult to monitor the health state of the system, testing capacity is low, or if there are delays in reporting infected cases in remote, rural areas in countries with limited health care and reporting systems. An optimal control can frequently be adjusted to give a close approximate control that is more feasible to implement [26].

Lastly, while assessing the monetary value of human health is fraught with uncertainties, even the estimation of intervention costs and its translation into model parameters present its

own challenges, especially when models are developed without a full commitment of health authorities to provide the necessary data.

OC applications to the control of communicable diseases are mostly confined in journals of applied mathematics with limited readership. While some barriers due to the inherent complexity of OC theory cannot be overcome, there is an untapped potential that should be leveraged to inform the decision making process on pressing public health problems. In spite of the limitations listed above, we believe that the scientific community proficient in optimal control theory could easily formulate optimization problems for the control of communicable diseases with our proposed approach. It would reflect more closely the framework in which public health decisions are usually taken in the full spirit of contemporary user-inspired research, as argued in [84, 105] .

Chapter 4

Geographic Disparities and Predictors of COVID-19 Hospitalization Risk in the St. Louis Area, Missouri (USA)

4.1 Background

There have been over 184 million confirmed Coronavirus Disease 2019 (COVID-19) cases and over 4 million deaths worldwide as of July 8, 2021, with over 33 million confirmed cases and over 600,000 deaths in the United States [118]. As of the same date, the state of Missouri has reported over 633,000 cases and 9,700+ deaths. The disease has overwhelmed many United States hospital systems, with large numbers of patients requiring critical care and interventions such as mechanical ventilation [41, 56, 58]. This surge of COVID-19 patients has put strain on hospital resources, potentially impacting the care available to both COVID-19 and non-COVID-19 patients [36].

There is evidence of geographic disparities in the severity of the disease with certain population groups experiencing more severe disease than others. These disparities might

A version of this chapter was previously published in Igoe M, Das P, Lenhart S, Lloyd AL, Luong L, Tian D, Lanzas C, Odoi A. *Geographic Disparities and Predictors of COVID-19 Hospitalization Risk in the St. Louis Area, Missouri (USA)*. BMC Public Health, 2022. Morganne Igoe was the leading author of this work.

be driven by population characteristics such as socioeconomic, demographic, and chronic disease factors [21, 29, 32, 33, 44]. For instance, there is evidence that Non-Hispanic American Indian, non-Hispanic Black, and Hispanic people have higher hospitalization risks than their non-Hispanic Asian and non-Hispanic White counterparts [22]. There are also reports that conditions such as diabetes mellitus, obesity, chronic lung conditions, renal disease, cancer, and cardiovascular disease may increase the severity of the condition and risk of hospitalizations among COVID-19 patients with these co-morbidities [33, 40, 94, 122]. Identifying geographic disparities in COVID-19 hospitalization risks and determinants of these disparities is important in providing information to guide hospital preparedness to handle the patient surge and for targeting resources for public health efforts to control the condition at the community level.

Improved understanding of the geographic disparities and predictors of COVID-19 hospitalization risk at the local level, such as the Zip Code Tabulation Areas (ZCTA), would help identify local geographic areas with higher needs for hospital beds and other healthcare resources. This is useful information for healthcare planning and service provision at the local level. It may also inform vaccination efforts by helping to identify areas where higher vaccination coverage may have the largest effect in reducing hospital burden. Therefore, the objective of this study is to identify geographic disparities and predictors of COVID-19 hospitalization risk in the St. Louis Area, Missouri, United States.

4.2 Methods

4.2.1 Study Area

This study was performed in an area of Missouri that includes 108 ZCTAs located in St. Charles county, St. Louis county, and St. Louis City as well as parts of Jefferson, Franklin and Warren counties (Figure 4.1). The area had a population of approximately 2 million people comprising 74% white, 20% black, 3% Hispanic/Latino, and 3% Asian. Forty-eight percent of the population was male, while 52% was female. Thirty-one percent of the population was younger than 25 years old, 53% was 25-64, and 16% was older than 65

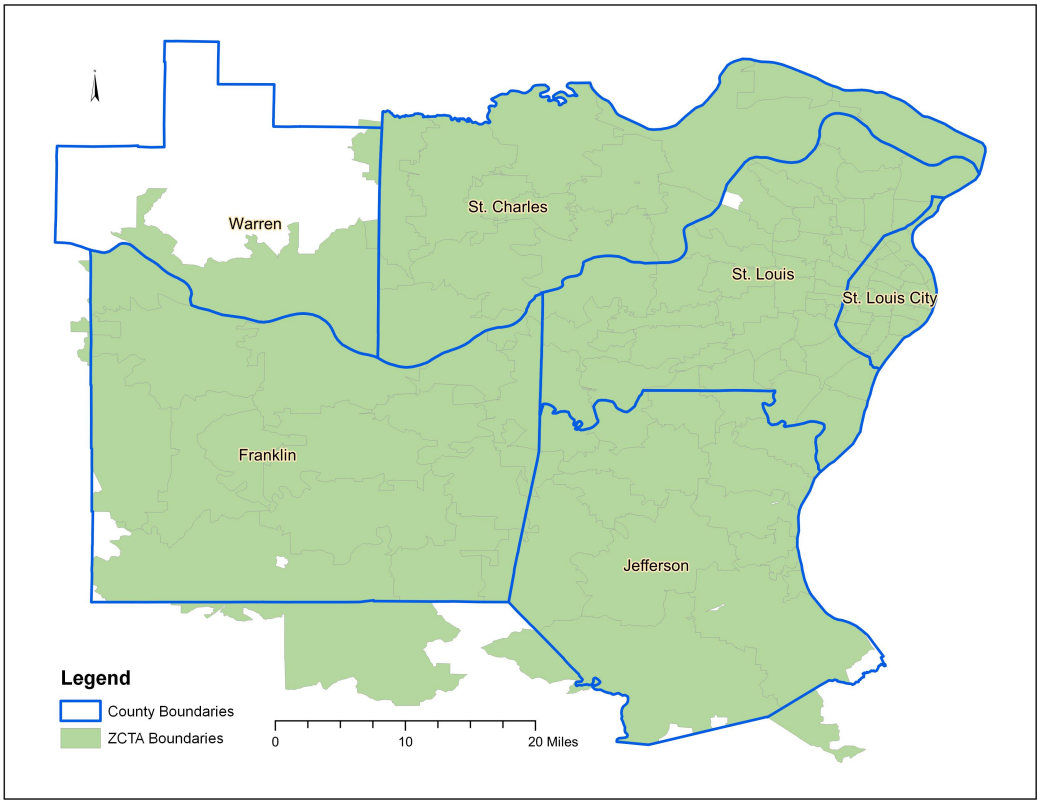

Figure 4.1: Map of study area showing geographic distribution of Zip Code Tabulation Areas and Counties

years of age. The ZCTA-level population density varied from 10 people per square mile in St. Charles County to 9,368 people per square mile in St. Louis City County.

4.2.2 Data Sources

COVID-19 and Chronic Disease Hospitalization Data

Data on COVID-19 and chronic condition hospitalizations were obtained from the Hospital Industry Data Institute, a not-for-profit organization founded by the Missouri Hospital Association that summarizes state level discharge data. The COVID-19 data provided information on the numbers of COVID-19 discharges by patient ZCTA and by age group from April 1, 2020 to September 30, 2020. In addition to the COVID-19 hospitalization data, data on the number of confirmed COVID-19 cases were obtained from the county Departments of Health. COVID-19 case data were included as one of the potential predictors of COVID-19 hospitalization risks. Data on chronic conditions were extracted for the time period 2019-2020 based on ICD-10 codes and included the following conditions/behaviors: obesity, tobacco use, cancer (breast, colorectal, prostate, lung, endometrial, leukemia and lymphoma), chronic obstructive pulmonary disease, chronic kidney disease, heart failure, and diabetes. These conditions were selected due to their potential association with COVID-19 severity. All data were aggregated to the ZCTA-level to facilitate subsequent ZCTA-level analyses. The ZCTA-level risks of these conditions were then computed and presented as disease-specific hospitalizations per 100 population.

Socioeconomic, demographic and base map data

ZCTA-level data on socioeconomic and demographic factors including age, sex, race, population density, education level, and median income, were obtained from the U.S. Census Bureau 2018 American Community Survey (ACS) 5-year estimates [109]. The cartographic boundary files, used for generating maps, were obtained from the US Census Bureau Website [110].

4.2.3 Descriptive Analyses

All descriptive analyses were performed in R version 4.1.0 [91] using RStudio version 1.4.1103 [95]. Normality of distribution of continuous variables was assessed using Shapiro-Wilk test.

Medians as well as 1st and 3rd quartiles were used as measures of central tendency and spread for all continuous variables since they were all non-normally distributed. The COVID-19 ZCTA-level hospitalization risks were directly age-standardized, in STATA version 16 [104], using the 2010 US population as the standard population.

4.2.4 Investigation of Predictors of COVID-19 Hospitalizations

Global Models

Univariable global Poisson models were used to assess simple associations between each of the potential predictors and COVID-19 hospitalization risk. As opposed to local models that estimate as many regression coefficients per predictor as the number of geographic units, global models assume constant relationships between each potential predictor and the outcome and therefore estimate one regression coefficient per potential predictor. These models were fit under the Generalized Linear Model (GLM) framework in R [91] specifying log link. The dependent variable was the expected count of COVID-19 hospitalizations in each ZCTA based on the age-adjusted/standardized number of COVID-19 hospitalizations and the offset was the natural log of the population in each ZCTA. A relaxed α of 0.15 was used to assess potentially significant predictors in univariable models. The linearity of the relationships between the log risk and potential predictor variables was assessed graphically.

Spearman rank correlation coefficients were computed to assess bivariate correlations among the potential predictor variables identified during the univariable analyses using R [91] . Variables with correlation coefficients greater than 0.7 were considered highly correlated. To avoid multicollinearity, only one of a pair of highly correlated variables was retained for further assessments using multivariable Poisson model. Biological and statistical considerations were used to determine which of a pair of highly correlated variables to retain. Non-correlated variables that had univariable $p \leq 0.15$ were assessed in a multivariable global Poisson model, built using manual backwards elimination and an α of 0.05. Variables were considered as confounders if their removal from the model resulted in changes of at least 20% of the coefficients of any of the variables in the model. Confounders were retained in the final main effects model. Biologically plausible two-way interactions between variables

in the final main effects multivariable model were assessed and the significant ones retained in the final model. The final Poisson regression model was then assessed for overdispersion.

Due to the presence of overdispersion in the final Poisson model (based on comparison of model deviance and degrees of freedom), a global negative binomial model was built. The process of building the negative binomial (NB) model was similar to that of the Poisson model described above. However, the glm.nb function of the MASS package [112] of R [91] was used instead of the glm function. The rest of the model specifications were similar to the Poisson model. Goodness-of-fit of the final global NB model was assessed using Pearson and Deviance chi-square goodness-of-fit tests.

Local Models

To assess if the association between each of the predictors and hospitalization risks varied by geographic location, a geographically weighted negative binomial (GWNB) model, proposed by Silva and Rodrigues [92], was fit to the data specifying the same outcome, link function, offset and significant predictors as in the final global multivariable NB model. However, unlike the global multivariable NB model that estimates one regression coefficient for each predictor and thus assumes a constant strength of association across all ZCTAs, the GWNB theoretically estimates as many regression coefficients as the number of ZCTAs. Essentially, GWNB evaluates a local model of the outcome by fitting a regression equation to every ZCTA in the dataset. These separate model equations are constructed by incorporating the outcome and predictor variables of the ZCTAs that fall within the neighborhood of each target ZCTA. Therefore, the GWNB allows identification of local variations in the strength of associations and therefore giving the importance of specific predictors in different local areas (ZCTAs). This implies that some factors may be more important predictors of hospitalization risk in some ZCTAs than others.

The local GWNB model was fit in SAS version 9.4 [97] using a SAS/IML macro [93]. The estimation of local regression coefficients was based on biquadratic kernel weighting function [93], while the bandwidth was estimated using the adaptive method which allows the size of the bandwidth to vary based on the density of observations. Bias-corrected Akaike Information Criteria (AICc) was used to determine the optimum kernel bandwidth and for

comparing the goodness-of-fit of the global NB and GWNB models. The better fitting model was identified as the one with the lower AICc value.

Stationarity of the GWNB coefficients was assessed using: (a) randomization non-stationarity test based on 999 replications [43]; (b) comparison of the interquartile range of the local GWNB model coefficients with the standard error estimates of the global NB model. Local coefficients whose interquartile ranges were larger than twice the standard error of the regression coefficient from the global NB model were considered non-stationary [17, 30].

4.2.5 Cartographic Displays

Choropleth maps showing the geographic distributions of ZCTA-level age-adjusted COVID-19 hospitalization risks, the socioeconomic, demographic and chronic disease factors as well as local regression coefficients from the GWNB models were generated using QGIS 3.16.6 [90]. Jenk's optimization classification scheme [2, 47, 75] was used to determine the critical intervals of the choropleth maps.

4.3 Results

4.3.1 Descriptive Statistics

The ZCTA-level median percentage of males was 48.6% while that of black and Hispanic populations were 3.7% and 2.2%, respectively (Table 4.1). For education variables, 38.2% of the population had high school education, 23% had some college education while 8.6% had associate's degree and 18% had bachelor's degree. The median household income was just over $59,400 with 9.5% of the ZCTA-level population living below the poverty line. Among chronic conditions investigated, Chronic Kidney Disease had the highest hospitalization risk (7.6%) followed by diabetes (7.5%) and heart failure had the lowest (3.2%) (Table 4.1). The ZCTA-level median number of confirmed cases of COVID-19 was 360 cases which was equivalent to 2.2% of the population at ZCTA-level (Table 4.1).

Table 4.1: Descriptive Statistics of ZCTA-level Potential Predictors of COVID-19 Hospitalization Risks in the St. Louis Area, Missouri

Variable	Median	First Quartile	Third Quartile
Demographic Factors			
% male population	48.6	47.4	50.3
% black population	3.7	0.9	34.2
% Hispanic/Latino population	2.2	1.0	3.2
Educational Variables			
% with $\leq$ high school education	38.2	23.9	47.6
% with some college	23	19.8	25.6
% with associate's degree	8.6	6.6	10.6
% with bachelor's degree	18	9.4	25.9
Economic Variables			
% below poverty level	9.5	5.8	16.8
median household income	59768.5	46005.3	77504
Health Behavior (ZCTA-level number of hospitalized patients that use tobacco per 100 Population)			
% tobacco[a]	10.4	6.8	14.5
Co-morbidities (ZCTA-level number of hospitalized patients with specific condition per 100 Population)			
% obesity	7	5.6	8
% diabetes	7.5	6.5	9.2
% cancer	3.8	3.3	4.3
% COPD[b]	4	3.1	4.9
% CKD[c]	7.6	6.4	8.9
% heart failure	3.2	2.6	3.8
COVID-19 Cases			
total cases	360	118	607
cases per 100 population	2.2	1.9	2.5

[a]ZCTA-level % of COVID-19 hospitalized patients that were tobacco users
[b]Chronic Obstructive Pulmonary Disease
[c]Chronic Kidney Disease

4.3.2 Predictors of COVID-19 Hospitalization Risks

Global Model

A number of variables had univariable associations with COVID-19 hospitalization risks at a relaxed $p < 0.2$ (Table 4.2). Of the assessed demographic variables, only percentage of black population had a univariable association ($p < 0.001$) with hospitalization risk. By contrast, all the assessed educational, economic, health behavior and chronic disease variables had significant univariable associations with COVID-19 hospitalization risks (Table 4.2). The ZCTA-level risk of confirmed COVID-19 cases had a significant ($p < 0.001$) univariable association with COVID-19 hospitalization risk but the raw count of COVID-19 cases per ZCTA did not ($p = 0.9$) (Table 4.2).

Based on the final global multivariable NB model, the ZCTA-level hospitalization risks were higher in ZCTAs that were high in the following predictors: percentage of black population ($p = 0.0416$), percentage of population with some college education ($p = 0.0005$), percentage of individuals hospitalized with diabetes ($p < 0.0001$), and number of ZCTA-level COVID-19 cases per 100 population ($p = 0.0001$) (Table 4.3). A map of the distribution of age-adjusted COVID-19 hospitalization risks and each of the significant predictors is shown in Figure 4.2. High age-adjusted COVID-19 hospitalization risks tended to occur in the Northeast of the study area and included ZCTAs in parts of St. Charles, St. Louis and Louis City counties. These areas also had high percentages of black population, individuals with some college education, those with high diabetes hospitalization risks as well as high risks of COVID-19 cases (Figure 4.2). High hospitalization risks were also evident in the Southwest areas of the study area that included parts of Warren and Franklin counties. It is worth noting that these areas also tended to have high diabetes hospitalization risks and percentages of the population with some college education (Figure 4.2). Since the global model did not show evidence of good fit based on both Deviance ($p = 0.03$) and Pearson ($p = 0.01$) goodness-of-fit tests, stationarity of the regression coefficients was assessed using a local GWNB model.

Table 4.2: Univariable Associations of Sociodemographic, Economic, and Chronic Disease Potential Predictors of COVID-19 Hospitalization Risk in the St. Louis Area, Missouri

Variable	**Coefficient**	**95 % Confidence Interval**	**p-values**
Demographic Factors			
% male population	0.004	-0.015, 0.024	0.727
% black population	0.007	0.005, 0.009	< 0.0001
% Hispanic/Latino population	-0.021	-0.056, 0.016	0.292
Educational Variables			
% with $\leq$ high school education	0.015	0.010, 0.020	< 0.0001
% with some college	0.035	0.022, 0.049	< 0.0001
% with associate's degree	0.028	0.0002, 0.055	0.036
% with bachelor's degree	-0.021	-0.028, -0.014	< 0.0001
Economic Variables			
% below poverty level	0.017	0.009, 0.024	< 0.0001
median household income	-5.00E-06	-0.000008, -0.000003	< 0.0001
Health Behavior (ZCTA-level number of hospitalized patients that use tobacco per 100 Population)			
% tobacco[a]	0.046	0.034, 0.058	< 0.0001
Co-morbidities (ZCTA-level number of hospitalized patients with specific condition per 100 Population)			
% obesity	0.138	0.104, 0.171	< 0.0001
% cancer	-0.029	-0.096,0.032	0.377
% COPD[b]	0.11	0.055, 0.164	< 0.0001
% CKD[c]	0.1	0.068, 0.131	< 0.0001
% heart failure	0.194	0.123, 0.265	< 0.0001
% diabetes	0.11	0.083, 0.136	< 0.0001
COVID-19 Cases			
total cases	1.40E-05	-0.0002, 0.00023	0.9
cases per 100 population	0.277	0.192, 0.3621	< 0.0001

[a]ZCTA-level % of COVID-19 hospitalized patients that were tobacco users
[b]Chronic Obstructive Pulmonary Disease
[c]Chronic Kidney Disease

Table 4.3: Final Global Negative Binomial Model Showing Significant Determinants of COVID-19 Hospitalization Risk in the St. Louis area

Name	**Coefficient**	**95 % Confidence Interval**	**p-value**
% black population	0.0014	0.0001, 0.0027	0.0416
% with some college education	0.0180	0.0078, 0.0281	0.0005
% diabetes[a]	0.0628	0.0397, 0.0860	< 0.0001
Confirmed COVID-19 cases per 100 population	0.2623	0.2027, 0.3218	< 0.0001

[a]Number of hospitalized patients with diabetes per 100 Population

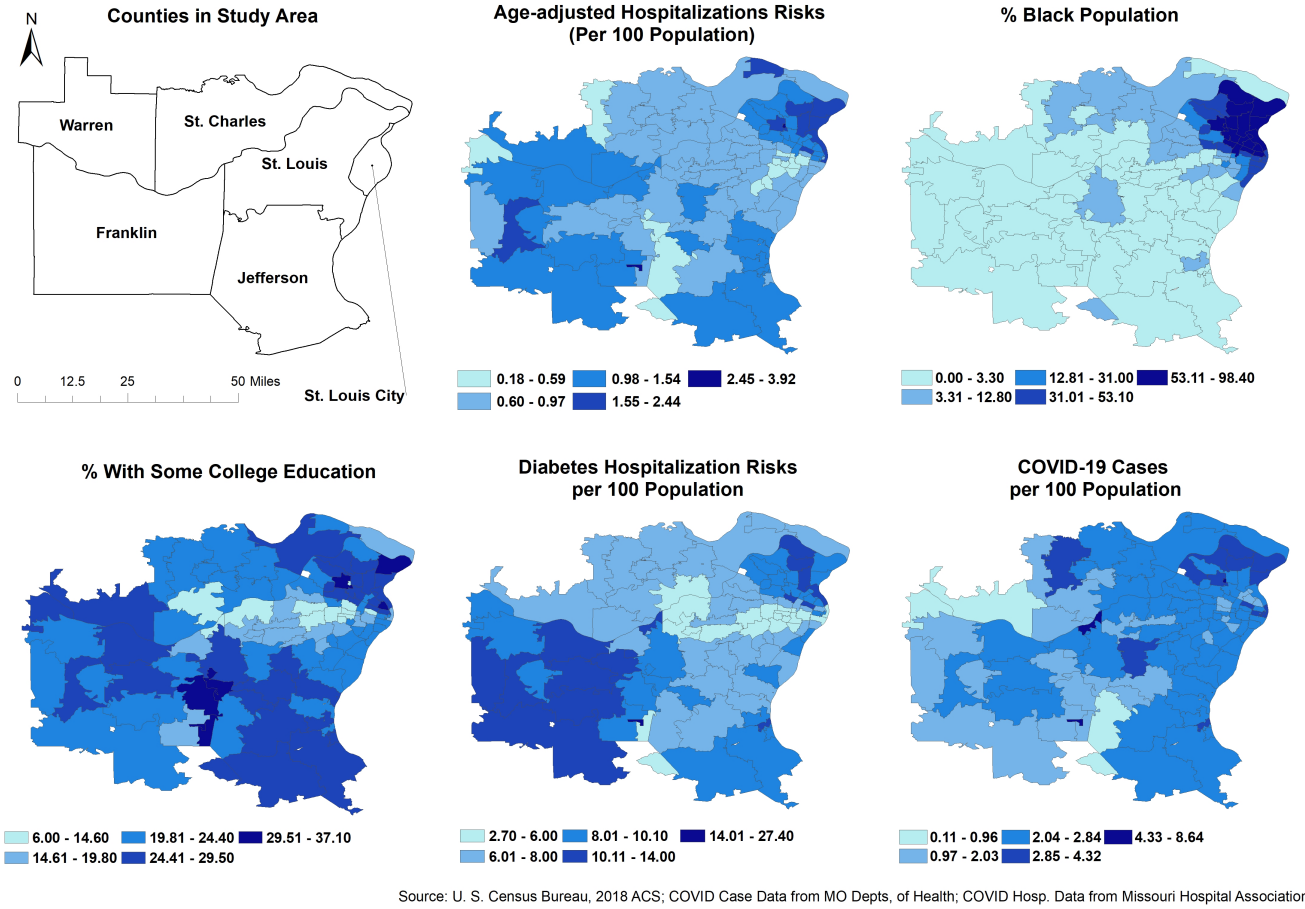

Figure 4.2: Geographic distribution of ZCTA-level COVID-19 age-adjusted hospitalization risks and its significant predictors in the St. Louis area, Missouri

Local Model

The p-values of the stationarity tests from the GWNB model indicate that the coefficients for the association between COVID-19 hospitalization risks and percentage of black population ($p = 0.001$) and number of hospitalized patients with diabetes per 100 population ($p = 0.032$) were non-stationary (Table 4.4). Additionally, comparison of the 2×SEs of the global coefficients and IQR of the local coefficients showed evidence of non-stationarity of the coefficients of the above two predictors as well as the population adjusted cases of COVID-19 (Table 4.4).

The spatial distribution of the local coefficients of the three predictors whose relationship with COVID-19 hospitalizations were identified as non-stationary provides visual evidence for variability of the local relationships (Figure 4.3). Thus, associations of COVID-19 hospitalization risks with percentage of black population, diabetes hospitalization risks and COVID-19 adjusted cases varied considerably across the study area, with a strong East-West gradient. The association between percentage of black population and COVID-19 hospitalization risks, for instance, was positive in the Northeast and negative in the West and Southwest. Moreover, the strength of the association was higher in ZCTAs in the West compared to those near the center of the study area. The strength of association between COVID-19 hospitalization risks with diabetes hospitalization risks was also higher in the Northeast and lower in the West. All ZCTAs, except one (63025), showed evidence of positive association between COVID-19 hospitalization risks and diabetes hospitalization risks. Finally, the association between COVID-19 hospitalization and population adjusted cases of COVID-19 increased from West to East, with the association staying positive in all ZCTAs of the study area. It is worth noting that the local GWNB model had a much better goodness-of-fit ($AICc = 986.4$) than the global model ($AICc = 1002.7$).

4.4 Discussion

The goal of this study was to investigate geographic disparities and identify predictors of ZCTA-level COVID-19 hospitalization risks in the St. Louis area. The findings of this study can be used to identify areas where the population is at higher risk of hospitalization

Table 4.4: Results of assessment of stationarity of the coefficients of the predictors of the COVID-19 hospitalization risks in the St. Louis Area, Missouri

	Global SE[a]	**Global SE**[a]**x2**	**IQR**[b] **of Local Coefficient**	**IQR-2(SE)**	**Stationarity test p-value**	**Is Coefficient Stationary?**
% black population	0.0007	0.0014	0.0148676	0.0134676	0.001	No[c]
% with some college education	0.0052	0.0104	0.0092401	-0.0011599	0.406	Yes
% diabetes[d]	0.0118	0.0236	0.0416903	0.0180903	0.032	No[c]
Confirmed COVID-19 cases per 100 population	0.0304	0.0608	0.1164299	0.0556299	0.109	No[e]

[a]Standard Error
[b]Interquartile Range
[c]Coefficients are non-stationary based on both the p-value of the stationarity test and IQR-2(SE) assessment
[d]Number of hospitalized patients with diabetes per 100 Population
[e]Coefficient is non-stationary based on IQR-2(SE) assessment

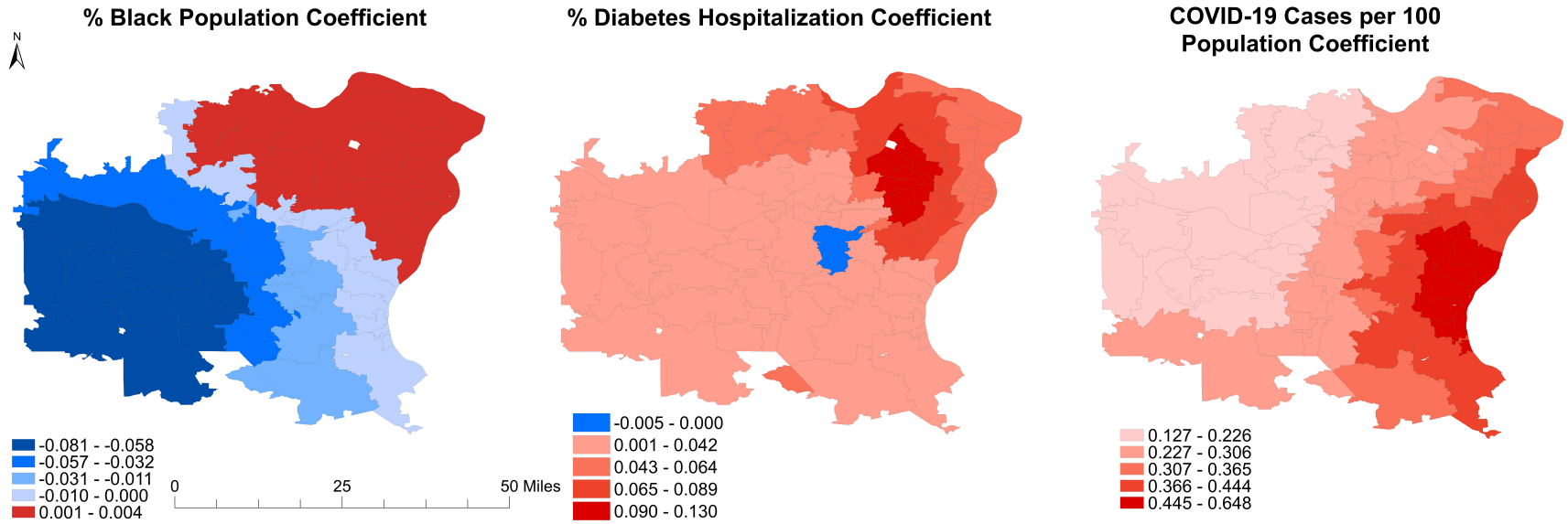

Figure 4.3: Geographically Varying Coefficients of the Local Geographically Weighted Negative Binomial Model Predicting COVID-19 Age-adjusted Hospitalization Risks in the St. Louis area

due to COVID-19 in order to guide planning and control efforts and to reduce potential overburdening of hospitals during COVID-19 surges.

There was evidence of geographic disparities in COVID-19 age-adjusted hospitalization risks in the study area. Urban ZCTAs in St. Louis City and St. Louis County exhibited high hospitalization risks. These ZCTAs also have high percentages of the population that are black, some as high as 98.4%. Some of these ZCTA also had high diabetes hospitalization risks. It is worth noting that some rural ZCTAs in Franklin and Warren counties had high COVID-19 hospitalization risk but very low percentages of the population that were black. However, these ZCTAs had high diabetes hospitalization risks implying that the COVID hospitalization risks in these rural ZCTAs was more driven by diabetes burden than demographic factors. Thus, although COVID-19 hospitalization risks in the more urban areas seems to be driven by the demographic composition of the population, the risks in the more rural areas seem to be driven more by diabetes burden. Geographic areas with intermediate to high COVID-19 hospitalization risks tended to have high percentages of the population with some college education and included ZCTAs in Franklin, Jefferson, St. Louis City and St. Louis County. It is worth noting that St. Louis City and St. Louis County tend to have similar restrictions, but often time, the restrictions from the other counties are more laxed. Previous ecological and individual level studies have not considered education level as a predictor of hospitalization risk. The level of education may be a proxy of occupation and other sociodemographic factors that may impact both the risk of infection and the resulting severity of the disease and hence hospitalization risks.

The above findings are consistent with those from previous ecological studies that investigated risk factors and predictors of COVID-19 hospitalization. For instance, an ecological study by Nguyen et al. reported that diabetes was a significant predictor of increased COVID-19 hospitalization risk at the county level in Georgia, USA even after controlling for sociodemographic and economic factors [85]. On the contrary, an Iranian study conducted at the provincial level did not find a significant association when only controlling for chronic disease factors [59]. This may indicate differences of relationships in different geographic areas and populations as well as the importance of controlling for

sociodemographic factors when evaluating the impact of chronic disease variables on COVID-19 hospitalization risk. Considering the fact that ZCTAs with high diabetes hospitalization risks also tended to have high COVID-19 hospitalization risks, COVID-19 mitigation efforts may need to be targeted to these ZCTAs to reduce the potential burden of the disease.

Interestingly, the study by Nguyen et al., which also considered sociodemographic and economic factors as well as comorbidities, did not find a significant association between percentage of black population with COVID-19 hospitalization risk [85]. However, it did identify other socioeconomic factors that this study did not consider, including: percentage of children in poverty and percentage of the population with severe housing problems. Although not directly comparable with the current ecological study, previous individual level studies have also identified associations between black race and COVID-19 hospitalization risk [32, 40, 60].

The local GWNB model allowed modeling of geographically varying associations between COVID-19 hospitalization risks and its predictors instead of assuming constant associations across the study area. Although the coefficients of percentage of black population, diabetes hospitalization risk and risk of COVID-19 infections varied spatially, the coefficients of percentage of the population with some college education did not and hence was modeled as stationary. This suggests that the coefficients for the percentage of the population with some college education are generalizable to all ZCTAs in the study area. In contrast, geographic variations in the associations between COVID-19 hospitalization risks and the percentage of black population, diabetes hospitalization risks and risks of COVID-19 infections suggest that global coefficients do not adequately describe the associations between COVID-19 hospitalization risks and these predictors across the study area. These findings have health planning and service provision implications. For instance, a "one size fits all" approach would not be suitable for addressing geographic disparities in COVID-19 hospitalization risks across the study area since some predictors are more important in some locales than others. Thus, different locales may require slightly different strategies depending on the most important predictors driving COVID-19 hospitalization risk in the location. Therefore, planning for hospital capacity and other disease management and control efforts will need

to use evidence-based approaches informed by empirical evidence from both global and local models.

4.4.1 Strengths and limitations

This study used both global and local models to investigate geographic disparities and identify predictors of COVID-19 hospitalization risks in St. Louis region of Missouri. Unlike studies that use only global statistical models and no GIS and/or local models, this study has been able to identify how the associations between the outcome (hospitalization risks) and each of the identified significant predictors change based on geographic location. The use of local models to investigate stationarity of regression coefficients of significant predictors and model non-stationary coefficients is a key strength of the study. This approach is particularly important in guiding local health planning since the importance of different predictors are not constant across the study area implying that different management and control strategies may need to be used in different areas. For example, while race (% black) may be a more important driver of hospitalization risks in some locations, other factors (such as diabetes) play more important roles in other areas. This is very useful information for guiding local health planning and disease control/prevention strategies and suggests that a one-size-fits-all strategy might not be the best approach for disease control. Therefore, modeling approaches that use both global and local models help to better understand the relationships between the outcome and predictors and may be more useful in guiding control efforts at the local level. However, the study is not without limitations. The hospital data has limitations associated with diagnostic classifications of COVID-19 in situations when the patient has co-morbidities that may have contributed to hospitalization. The age-standardization of the hospitalization risks was done using the 2010 US census data which was the most recent US census data available at the time of the study. Additionally, there may be geographic differences in COVID-19 case ascertainment and reporting.

4.5 Conclusions

There is evidence of geographic disparities in COVID-19 hospitalization risks in the St. Louis area of Missouri. These disparities are driven by socioeconomic, demographic and

health-related factors. The impacts of these factors vary by geographical location with some factors being more important predictors of COVID-19 hospitalization risk in some locales than others. This demonstrates the importance of using not only global but also local models to investigate determinants of geographic disparities in health outcomes and utilization of health services. This study's findings are useful for informing healthcare system planning to identify geographic areas likely to have high numbers of individuals needing hospitalization as well as in guiding vaccination efforts.

4.6 Acknowledgments

This work was supported by CDC U01CK000587-01M001. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank the St. Louis Comparative Modeling Network for facilitating access to data and useful discussions.

Bibliography

[1] Abdelrazec, A., Lenhart, S., and Zhu, H. (2016). Optimal Control of West Nile Virus in Mosquito, Birds and Humans with Seasonality. *Canadian Applied Math Quarterly*, 23(3):355–375. 60

[2] AD Cliff and P Haggett (1988). *Atlas of Disease Distributions: Analytic Approaches to Epidemiological Data*. Basil Blackwell Ltd, Oxford. 70

[3] Agusto, F., Pantha, B., and Elmojtaba, I. (2020). Optimal control applied to a visceral leishmaniasis model. *Electronic Journal of Differential Equations*, 80:1–24. 2, 35

[4] Ali, M., Nelson, A. R., Lopez, A. L., and Sack, D. A. (2015). Updated global burden of cholera in endemic countries. *PLOS Neglected Tropical Diseases*, 9(6):1–13. 39

[5] Allen, L. J. S., Allen, E. J., and Jonsson, C. B. (2006a). The Impact of Environmental Variation on Hantavirus Infection in Rodents. In C. Castillo-Chavez and D. P. Clemence, and A. B. Gumel, editor, *Mathematical Studies on Human Disease Dynamics: Emerging Paradigms and Challenges*, number 410 in AMS Contemporary Mathematics (CONM), pages 1–15. AMS, Providence, RI. 7

[6] Allen, L. J. S., McCormack, R. K., and Jonsson, C. B. (2006b). Mathematical models for hantavirus infection in rodents. *Bulletin of Mathematical Biology*, 68(3):511–524. 7, 32

[7] Allen, L. J. S. and P. van den Driessche (2008). The basic reproduction number in some discrete-time epidemic models. *Journal of Difference Equations and Applications*, 14(10–11):1127–1147. 18

[8] Allen, L. J. S., Wesley, C. L., Owen, R. D., Goodin, D. G., Koch, D., Jonsson, C. B., Chu, Y. W., Hutchinson, J. M. S., and Paige, R. L. (2009). A habitat-based model for the spread of hantavirus between reservoir and spillover species. *Journal of Theoretical Biology*, 260:510–522. 7, 32

[9] Asano, E., Gross, L. J., Lenhart, S., and Real, L. A. (2008). Optimal control of vaccine distribution in a rabies metapopulation model. *Mathematical Biosciences and Engineering*, 5(2):219–238. 2, 35

[10] Assembly, U. N. G. (2021). Political declaration on HIV and AIDS: Ending inequalities and getting on track to end AIDS by 2030. 38

[11] Baltussen, R. (2006). Priority setting of public spending in developing countries: Do not try to do everything for everybody. *Health Policy*, 78(2-3):149–156. 2, 35

[12] Bernshtein, A. D., Apekina, N. S., Mikhailova, T. V., Myasnikov, Y. A., Khlyap, L. A., Korotkov, Y. S., and Gavrilovskaya, I. N. (1999). Dynamics of Puumala hantavirus infection in naturally infected bank voles (*Clethrinomys glareolus*). *Archives of Virology*, 144(12):2415–2428. 9

[13] Bolzoni, L., Tessoni, V., Groppi, M. V., and De Leo, G. A. (2014). React or wait: which optimal culling strategy to control infectious diseases in wildlife. *Journal of Mathematical Biology*, 69. 37, 62

[14] Borucki, M. K., Boone, J. D., Rowe, J. E., Bohlman, M. C., Kuhn, E. A., DeBaca, R., and Jeor, S. C. S. (2000). Role of maternal antibody in natural infection of *Peromyscus maniculatus* with Sin Nombre virus. *Journal of Virology*, 74(5):2426–2429. 9

[15] Botten, J. and Fix, E. A. (2000). Experimental infection model for Sin Nombre hantavirus in the deer mouse (*Peromyscus maniculatus*). *Proceedings National Academy of Science USA*, 97(19):10578–10593. 5, 6

[16] Brock, D. and Wikler, D. (2006). *Disease Control Priorities in Developing Countries*, chapter 14: Ethical Issues in Resource Allocation, Research, and New Product Development, pages 259–270. Oxford University Press, 2 edition. 2, 35

[17] Brunsdon, C., Fotheringham, A., and Charlton, M. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4):281–298. 70

[18] Bubar, K. M., Reinholt, K., Kissler, S. M., Lipsitch, M., Cobey, S., Grad, Y. H., and Larremore, D. B. (2021). Model-informed covid-19 vaccine prioritization strategies by age and serostatus. *Science*, 371(6532):916–921. 34

[19] Calisher, C. H., Wagoner, K. D., Amman, B. R., Root, J. J., Douglass, R. J., Kuenzi, A. J., Abbott, K. D., Parmenter, C., Yates, T. L., Ksiazek, T. G., Beaty, B. J., and Mills, J. N. (2007). Demographic factors associated with prevalence of antibody to Sin Nombre virus in deer mice in the western United States. *Journal of Wildlife Diseases*, 43(1):1– 11. 1, 2, 6

[20] CDC (2016). Reported Cases of Hantavirus Disease. http://www.cdc.gov/hantavirus/surveillance. 6

[21] CDC COVID-19 Response Team (2020). Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. Technical report, CDC. 65

[22] Centers for Disease Control and Prevention (2021). COVID-19 Racial and Ethnic Health Disparities - CDC. 65

[23] Childs, J. E., Ksiazek, T. G., Spiropoulou, C. F., Krebs, J. W., Morzunov, S., Maupin, G. O., Gage, K. L., Rollin, P. E., Sarisky, J., Enscore, R. E., Frey, J. K., Peters, C. J., and Nichol, S. T. (1994). Serologic and Genetic Identification of *Peromyscus maniculatus* as the Primary Rodent Reservoir for a New Hantavirus in the Southwestern United States. *Journal of Infectious Diseases*, 169(6):1271–1280. 6, 31

[24] Codeco, C. T. (2001). Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC Infectious Diseases*, 1:1–14. 39

[25] Dearing, M. D., Previtali, M. A., Jones, J. D., Ely, P. W., and Wood, B. A. (2009). Seasonal variation in sin nombre virus infections in deer mice: preliminary results. *Journal of Wildlife Diseases*, 45(2):430–436. 9

[26] Demir, M. and Lenhart, S. (2020). Optimal sustainable fishery management of the black sea anchovy with food chain modeling framework. *Natural Resource Modeling*, 33(2):1–29. 62

[27] Dohmae, K. and Nishimune, Y. (1998). Maternal transfer of Hantavirus antibodies in rats. *Laboratory Animal Science*, 48(4):395–397. 9

[28] Eastwood, G., Camp, J. V., Chu, Y. K., Sawyer, A. M., Owen, R. D., Cao, X., Taylor, M. K., Valdivieso-Torres, L., Sage, R. D., Yu, A., Goodin, D. G., Martinez-Bruyn, V. J., McAllister, R. C., Rodriguez, L., William, E. P., and Jonsson, C. B. (2018). Habitat, species richness and hantaviruses of sigmodontine rodents within the Interior Atlantic Forest, Paraguay . *PLoS ONE*, 13(8):e0201307. 2, 6, 7, 8, 9, 17, 32

[29] Fadl, N., Ali, E., and Salem, T. (2021). COVID-19: Risk Factors Associated with Infectivity and Severity. *Scandinavian journal of immunology*, 93(3):e13039. 65

[30] Fotheringham, A., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons Ltd. 70

[31] Galiano, D., Kubiak, B. B., Marinho, J. R., and Ochotorena de Freitas, T. R. (2011). Population dynamics of *Akodon montensis* and *Oligoryzomys nigripes* in an Araucaria forest of Southern Brazil . *Mammalia*, 77(2):173–179. 17

[32] Gao, Y., Ding, M., Dong, X., Zhang, J., Azkur, A. K., Azkur, D., Gan, H., Sun, Y., Fu, W., Li, W., Liang, H., Cao, Y., Yan, Q., Cao, C., Gao, H., Brüggen, M., van de Veen, W., Sokolowska, M., Akdis, M., and Akdis, C. (2021). Risk factors for severe and critically ill COVID-19 patients: A review. *Allergy: European Journal of Allergy and Clinical Immunology*, 76(2):428–455. 65, 80

[33] Garg, S., Kim, L., Whitaker, M., O'Halloran, A., Cummings, C., Holstein, R., Prill, M., Chai, S. J., Kirley, P. D., Alden, N. B., Kawasaki, B., Yousey-Hindes, K., Niccolai, L., Anderson, E. J., Openo, K. P., Weigel, A., Monroe, M. L., Ryan, P., Henderson, J., Kim, S., Como-Sabetti, K., Lynfield, R., Sosin, D., Torres, S., Muse, A., Bennett, N. M., Billing, L., Sutton, M., West, N., Schaffner, W., Talbot, H. K., Aquino, C., George, A., Budd, A., Brammer, L., Langley, G., Hall, A. J., and Fry, A. (2020). MMWR - Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020. Technical report, CDC. 65

[34] Gentile, R., D'Andrea, P. S., Cerqueira, R., and Maroja, L. S. (2000). Population dynamics and reproduction of marsupials and rodents in a Brazilian rural area: a five-year study. *Studies on Neotropical Fauna and Environment*, 35(1):1–9. 17

[35] Georges, C. G., Artunc, F., Weyrich, P., Friedrich, B., and Wolf, S. C. (2008). Nephropathia epidemica as the result of a Puumala virus infection in a pregnant patient. *Deutsche Medizinische Wochenschrift*, 133(37):1830–1832. 9

[36] Ghanchi, H., Takayanagi, A., Savla, P., Hariri, O., Tayag, E., Schiraldi, M., Jorgensen, L., and Miullig, D. (2020). Effects of the COVID-19 Pandemic on Stroke Patients. *Cureus*, 12(9):e9995. 64

[37] Glass, G. E., Childs, J. E., Korch, G. W., and Leduc, J. W. (1988). Association of Intraspecific Wounding with Hantaviral Infection in Wild Rats (*Rattus norvegicus*).

Epidemiology and Infection, 101:459–472. 5

[38] Green, W. and Fix, E. (1998). Tissue distribution of hantavirus antigen in naturally infected humans and deer mice. *Journal of Infectious Disease*, 177(6):1696——1700. 6

[39] Gurarie, D. and Seto, E. Y. W. (2008). Connectivity sustains disease transmission in environments with low potential for endemicity: modelling schistosomiasis with hydrologic and social connectivities . *Journal of the Royal Society Interface*, 6:495–508. 32

[40] Harlem, G. (2020). Descriptive analysis of social determinant factors in urban communities affected by COVID-19. *Journal of Public Health (United Kingdom)*, 42(3):466–469. 65, 80

[41] Hassan, E. and Mahmoud, H. (2021). Impact of multiple waves of COVID-19 on healthcare networks in the United States. *PLoS ONE*, 16(3). 64

[42] Hess, G. (1996). Disease in Metapopulation Models: Implications for Conservation . *Ecology*, 77:1617–1632. 32

[43] Hope, A. (1968). A Simplified Monte Carlo Significance Test Procedure. *J R Stat Soc Series B Stat Methodol*, 30(3):582–598. 70

[44] Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., and Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395(10223):497–506. 65

[45] Hutchinson, K. L. and Fix, E. A. (1998). Pathogenesis of a North American hantavirus, Black Creek Canal virus, in experimentally infected *Sigmodon hispidus*. *American Journal of Tropical Medicine and Hygiene*, 59(1):58——65. 6

[46] Hutchinson, K. L., Rollin, P. E., and Peters, C. J. (2000). Transmission of Black Creek Canal virus between cotton rats. *Journal of Medical Virology*, 60(1):70——76. 5, 9

[47] Jenks, G. (1967). The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography*, 7:186–190. 70

[48] Jiang, S., Chen, Z., Wu, J., Zang, X., and Jiang, Y. (2020). Addressing methodological and ethical issues in practicing health economic evaluation in china. *J Glob Health*, 10(2). 2, 35

[49] Jonsson, C. B., Figueiredo, L. T., and Vapalahti, O. (2010). A global perspective on hantavirus ecology, epidemiology, and disease. *Clinical Microbiology Review*, 23(2):412—421. 6

[50] Jordão, J. C., Ramos, F. N., and da Silva, V. X. (2010). Demographic parameters of *Akodon montensis* (Mammalia: Rodentia) in an Atlantic Forest remnant of Southeastern Brazil. *Mammalia*, 74:395–400. 7

[51] Jung, E., Iwami, S., Takeuchi, Y., and Jo, T.-C. (2009). Optimal control strategy for prevention of avian influenza pandemic. *Journal of Theoretical Biology*, 260(2):220–229. 35

[52] Kallio, E. R., Begon, M., Henttonen, H., Koskela, E., Mappes, T., Vaheri, A., and Vapalahti, O. (2010). Hantavirus infections in fluctuating host populations: the role of maternal antibodies. *Proceedings of the Royal Society B: Biological Sciences*, 277(1701):3783–3791. 9

[53] Kallio, E. R., Henttonen, H., Koskela, E., Lundkvist, A., Mappes, T., and Vapalahti, O. (2013). Maternal antibodies contribute to sex-based difference in hantavirus transmission dynamics. *Biology Letters*, 9(6):20130887.

[54] Kallio, E. R., Poikonen, A., Vaheri, A., Vapalahti, O., Henttonen, H., Koskela, E., and Mappes, T. (2006). Maternal antibodies postpone hantavirus infection and enhance individual breeding success. *Proceedings of the Royal Society B: Biological Sciences*, 273(1602):2771–2776. 9

[55] Keeling, M. J. and Rohani, P. (2008). *Modeling Infectious Disease in Humans and Animals* . Princeton University Press. 31

[56] Keene, A., Shiloh, A., Eisen, L., Berger, J., Karwa, M., Fein, D., Orsi, D., and Gong, M. (2021). Critical Care Surge During the COVID-19 Pandemic: Implementation and Feedback From Frontline Providers. *J Intensive Care Med*, 36(2):233 – 240. 64

[57] Kelly, M. R., Tien, J. H., Eisenberg, M. C., and Lenhart, S. (2016). The impact of spatial arrangements on epidemic disease dynamics and intervention strategies. *Journal of Biological Dynamics*, 10(1):222–249. 40

[58] Khera, R., Jain, S., Lin, Z., Ross, J. S., and Krumholz, H. (2020). Evaluation of the anticipated burden of COVID-19 on hospital-based healthcare services across the United

States. 64

[59] Khorrami, Z., Nili, S., Sharifi, H., Eybpoosh, S., and Shokoohi, M. (2020). Association of cigarette smoking, obesity, and underlying medical conditions with COVID-19 hospitalization and mortality in Iran: A nationwide retrospective ecological study. *Medical journal of the Islamic Republic of Iran*, 34:133. 79

[60] Ko, J., Danielson, M., Town, M., Derado, G., Greenlund, K., Kirley, P., Alden, N., Yousey-Hindes, K., Anderson, E., Ryan, P., Kim, S., Lynfield, R., Torres, S., Barney, G., Bennett, N., Sutton, M., Talbot, H., Hill, M., Hall, A., Fry, A., Garg, S., Kim, L., and COVID-NET Surveillance Team (2021). Risk Factors for COVID-19-associated hospitalization: COVID-19-Associated Hospitalization Surveillance Network and Behavioral Risk Factor Surveillance System. *Clin Infect Dis*, 72(11):e695–e703. 80

[61] Langlois, J. P., Fahrig, L., Merriam, G., and Artsob, H. (2001). Landscape structure influences continental distribution of hantavirus in deer mice. *Landscape Ecology*, 16(2):255–266. 7, 32

[62] LeDuc, J. W., Childs, J. E., and Glass, G. E. (2002). The Hantaviruses, etiologic agents of hemorrhagic fever with renal syndrome: a possible cause of hypertension and chronic renal disease in the United States. *Annual Review of Public Health*, 13:79–98. 6

[63] Lee, E., Dennis L Chao, D., Lemaitre, J., Laura Matrajt, L., Pasetto, D., Perez-Saez, J., Finger, F., Rinaldo, A., Sugimoto, J., Halloran, M., Longini Jr, I., Ternier, R., Vissieres, K., Andrew S Azman, A. S., Lessler, J., and Ivers, L. (2020a). Achieving coordinated national immunity and cholera elimination in Haiti through vaccination: a modelling study. *Lancet*, 8:1–9. 39

[64] Lee, H. W., Baek, L. J., and Johnson, K. M. (1982). Isolation of Hantaan virus, the etiologic agent of Korean hemorrhagic fever, from wild urban rats . *The Journal of Infectious Diseases*, 146(5):638–644. 17

[65] Lee, S., Baek, O., and Melara, L. (2020b). Resource allocation in two-patch epidemic model with state-dependent dispersal behaviors using optimal control. *Processes*, 8(9). 35

[66] Legros, D. (2018). Global cholera epidemiology: Opportunities to reduce the burden of cholera by 2030. *The Journal of Infectious Diseases*, 218, suppl 3:S137–S140. 39

[67] Lehmer, E. M., Clay, C. A., Pearce-Duvet, J., St.Jeor, S., and Dearing, M. D. (2008). Differential regulation of pathogens: the role of habitat disturbance in predicting prevalence of Sin Nombre virus. *Oecologia*, 155:429–439. 7, 32

[68] Lenhart, S. and Workman, J. (2007). *Optimal control applied to biological models*. Chapman and Hall/CRC, New York, NY. 3, 46

[69] Lo, N. C., Gurarie, D., Yoon, N., Coulibaly, J. T., Bendavid, E., Andrews, J. R., and King, C. H. (2018). Impact and cost-effectiveness of snail control to achieve disease control targets for schistosomiasis. *Proceedings of the National Academy of Sciences*, 115(4):E584–E591. 37

[70] Luis, A. D., Douglas, R. J., Hudson, P. J., Mills, J. N., and Bjornstad, O. N. (2012). Sin Nombre hantavirus decreases survival of male deer mice. *Oceolgia*, 169(1):431–439. 5

[71] Luis, A. D., Douglass, R. J., Mills, J. N., and Bjornstad, O. N. (2010). The effects of seasonality, density and climate on the population dynamics of the Montana deer mice, important reservoir hosts for the Sin Nombre hantavirus. *Journal of Animal Ecology*, 79:462–470. 5, 7, 32

[72] Marino, S., Hogue, I. B., Ray, C. J., and Kirschner, D. E. (2009). A Methodology For Performing Global Uncertainty And Sensitivity Analysis In Systems Biology. *Journal of Theoretical Biology*, 254(1):178–196. 2, 19

[73] Martínez, J. P. S. (2017). Relación entre disturbios humanos ocasionados en el paisaje y la abundancia y otras medidas poblaciones de *Akodon montensis*, un hospedero de *Hantavirus* en Paraguay. Master's thesis, Universidad Nacional de Asunción, San Lorenzo, Paraguay. 31, 32

[74] McAllister, R. C. and Jonsson, C. B. (2014). Hantaviruses: past, present and future. *Future Virology*, 9(1):87 – 99. 6

[75] McMaster, R. (1997). In Memoriam: George F. Jenks (1916–1996). *Cartography and Geographic Information Science*, 24(1):56–59. 70

[76] Medley, G. F., Lindop, N. A., Edmunds, W. J., and Nokes, D. J. (2001). Hepatitis-B virus endemicity: heterogeneity, catastrophic dynamics and control . *Nature Medicine*, 7(5):619–624. 31

[77] Melbourne, editor, E. L. (2011). *Cholera : Symptoms, Diagnosis and Treatment*. Nova Science Publishers. 39

[78] Miller, I., Becker, A., Grenfell, B., and Metcalf, C. (2020). Disease and healthcare burden of COVID-19 in the United States. *Nature Medicine*, 26:1212–1217. 37, 38

[79] Miller Neilan, R., Schaefer, E., Gaff, H., Fister, K. R., and Lenhart, S. (2010). Modeling optimal intervention strategies for cholera. *Bulletin of Mathematical Biology*, 1 (4):379–393. 2, 35

[80] Miyaoka, T., Lenhart, S., and Meyer, J. (2019). Optimal control of vaccination in a vector-borne reaction–diffusion model applied to Zika virus. *Journal of Mathematical Biology*, 79(3):1077–1104. 60

[81] Morrow, R. and Byrant, J. (1995). Health policy approaches to measuring and valuing human life: Conceptual and ethical issues. *American Journal of Public Health*, 85(10):1356–1360. 2, 35

[82] Murray, C., Vos, T., and Lozano et al., R. (2012). Disability-adjusted life years (dalys) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the global burden of disease study 2010. *Lancet*, 380:2197–223. 37

[83] Muylaert, R. L., Bovendorp, R. S., Jr., G. S.-S., Prist, P. R., Melo, G. L., de Fatíma Priante, C., Wilkinson, D. A., Ribeiro, M. C., and Hayman, D. T. S. (2019). Hantavirus host assemblages and human disease in the Atlantic Forest . *PLOS Neglected Tropical Diseases*, 13(8). 9

[84] National Research Council (2008). *Research and Networks for Decision Support in the NOAA Sectoral Applications Research Program*. The National Academies Press. 63

[85] Nguyen, T., Shah, G., Schwind, J., and Richmond, H. (2021). Community Characteristics and COVID-19 Outcomes: A Study of 159 Counties in Georgia, United States. *Journal of Public Health Management and Practice*, 27(3):251–257. 79, 80

[86] Ntuli, editor, M. M. (2020). *Ending the neglect to attain the Sustainable Development Goals: A road map for neglected tropical diseases 2021–2030 for neglected tropical diseases*. World Health Organization. 36

[87] Owen, R. D., Goodin, D. G., Koch, D. E., Chu, Y. K., and Jonsson, C. B. (2010). Spatiotemporal variation in *Akodon montensis* (*Cricetidae: Sigmodontinae*) and

hantaviral seroprevalence in a subtropical forest ecosystem. *Journal of Mammology*, 91(2):467–481. 6, 7, 8, 9, 17

[88] Padula, P., Figueroa, R., Navarrete, M., Pizarro, E., Cadiz, R., Bellomo, C., Jofre, C., Zaror, L., Rodriguez, E., and Murúa, R. (2004). Transmission Study of Andes Hantavirus Infection in Wild *Sigmodontine* Rodents. *Journal of Virology*, 78(21):11972–11979. 2, 5, 6, 8

[89] Pontryagin, L. S., Boltyanskii, V. G., Gamkrelize, R. V., and F.Mishchenko, E. (1962). *The Mathematical Theory of Optimal Processes*. Wiley. 2, 42, 44

[90] QGIS.org (2021). QGIS Geographic Information System. 70

[91] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. 67, 68, 69

[92] Ricardo Da Silva, A. and Rodrigues, T. (2014). Geographically Weighted Negative Binomial Regression-incorporating overdispersion. *Statistics and Computing*, 24(5):769–783. 3, 69

[93] Ricardo Da Silva, A. and Rodrigues, T. (2016). A SAS Macro for Geographically Weighted Negative Binomial Regression. In *SAS Global Forum 2016*. SAS Institute. 69

[94] Robilotti, E., Babady, N., Mead, P., Rolling, T., Perez-Johnston, R., Bernardes, M., Bogler, Y., Caldararo, M., Figueroa, C., Glickman, M., Joanow, A., Kaltsas, A., Joo Lee, Y., Lucca, A., Mariano, A., Morjaria, S., Nawar, T., Papanicolaou, G., Predmore, J., Redelman-Sidi, G., Schmidt, E., Seo, S., Sepkowitz, K., Shah, M., Wolchok, J., Hohl, T., Taur, Y., and Kamboj, M. (2020). Determinants of COVID-19 disease severity in patients with cancer. *Nature Medicine*, 26(8):1218–1223. 65

[95] RStudio Team (2021). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com. 67

[96] Rutstein, S., Price, J., Rosenberg, N., Rennie, S., Biddle, A., and Miller, W. (2017). Hidden costs: the ethics of cost-effectiveness analyses for health interventions in resource-limited settings. *Glob Public Health*, 12(10):1269–1281. 2, 35

[97] SAS Institute (2016). Statistical Analysis System (SAS) Version 9.4. Cary, NC, USA. SAS Institute Inc. 69

[98] Schountz, T., Acuña-Retamar, M., Feinstein, S., Prescott, J., Torres-Perez, F., Podell, B., Peters, S., Ye, C., Black IV, W. C., and Hjelle, B. (2012). Kinetics of immune responses in deer mice experimentally infected with Sin Nombre virus. *Journal of Virology*, 86(18):10015–10027. 10

[99] Schountz, T., Quackenbush, S., Rovnak, J., Haddock, E., Black IV, W. C., Feldmann, H., and Prescott, J. (2014). Differential lymphocyte and antibody responses in deer mice infected with Sin Nombre hantavirus or Andes hantavirus. *Journal of Virology*, 88(15):8319–8331. 10

[100] Scovronick, N., Ferranna, M., Dennig, F., and Budolfson, M. (2020). Valuing health impacts in climate policy: Ethical issues and economic challenges. *Health Affairs*, 39(12):2105–2112. 36

[101] Sèvére, K., Rouzier, V., Anglade, S. B., Bertil, C., Joseph, P., Deroncelay, A., Mabou, M. M., Wright, P. F., Guillaume, F. D., and Pape, J. W. (2016). Effectiveness of Oral Cholera Vaccine in Haiti 37 Month Follow Up. *The American Society of Tropical Medicine and Hygiene*, 94(5):1136 – 1142. 39

[102] Smith, D. L., Battle, K. E., Hay, S. I., Barker, C. M., Scott, T. W., and McKenzie, F. E. (2012). Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLoS Pathogens*, 8(4):1–13. 34

[103] Sokolow, S., Huttinger, E., Jouanard, N., Hsieh, M., Lafferty, K., Kuris, A., Riveau, G., Senghor, S., Thiam, C., N'Diaye, A., Faye, D., and De Leo, G. (2015). Reduced transmission of human schistosomiasis after restoration of a native river prawn that preys on the snail intermediate host. *PNAS*, 112(31):9650–9655. 60

[104] StataCorp. (2019). Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC. 68

[105] Stokes, D. E. (1997). *Pasteur's Quadrant: Basic Science and Technological Innovation*. Brookings Institution. 63

[106] Teixeira, B. R., Loureiro, N., Strecht, L., Gentile, R., Oliveira, R. C., Guterres, A., Fernandes, J., Mattos, L. H. B. V., Raboni, S. M., Rubio, G., Bonvicino, C. R., C. N. Duarte dos Santos, Lemos, E. R. S., and D'Andrea, P. S. (2014). Population Ecology of Hantavirus Rodent Hosts in Southern Brazil. *The American Journal of Tropical Medicine*

and Hygiene, 91(2):249–257. 6, 17, 31

[107] The World Bank (2021). GFP per capita for different countries 2021. https://data.worldbank.org/indicator/NY.GDP.PCAP.CD. 46

[108] Tien, J. H. and Earn, D. J. D. (2010). Multiple transmission pathways and disease dynamics in a waterborne pathogen model. *Bulletin of Mathematical Biology*, 72(6):1506–1533. 2, 39, 40

[109] United States Census Bureau (2018). ACS Demographic and Housing Estimates, 2018: 5-Year Estimates. 3, 67

[110] United States Census Bureau (2019). TIGER/Line Shapefiles. 67

[111] Valega-Mackenzie, W. and Ríos-Soto, K. (2018). Can vaccination save a Zika virus epidemic? *Bulletin of Mathematical Biology*, 80(3):598–625. 60

[112] Venables, W. and Ripley, B. (2002). *Modern Applied Statistics with S*. Springer, 4 edition. 69

[113] Voutilainen, L., Kallio, E. R., Niemimaa, J., Vapalahti, O., and Henttonen, H. (2016). Temporal dynamics of Puumala hantavirus infection in cyclic populations of bank voles. *Scientific Reports*, 6:21323. 9

[114] Wang, W. and Mulone, G. (2003). Threshold of disease transmission in a patch environment. *Journal of Mathematical Analysis and Applications*, 285:321–335. 32

[115] Weintraub, W. S. and Cohen, D. J. (2009). The limits of cost-effectiveness analysis. *Circulation: Cardiovascular Quality and Outcomes*, 2(1):55–58. 37, 62

[116] Wesley, C. L., Allen, L. J. S., Jonsson, C. B., Chu, Y. K., and Owen, R. D. (2009). A discrete-time rodent-hantavirus model structured by infection and developmental stages. *Advanced Studies in Pure Mathematics*, 53:1–12. 7, 32

[117] Wesley, C. L., Allen, L. J. S., and Langlais, M. (2010). Models for the Spread and Persistence of Hantavirus Infection in Rodents with Direct and Indirect Transmission. *Mathematical Biosciences and Engineering*, 7(1):195–211. 7, 32

[118] World Health Organization (2021). WHO Coronavirus (COVID-19) Dashboard. http://covid19.who.int/. 64

[119] Yanagihara, R., Amyx, H. L., and Gajdusek, D. (1985). Experimental infection with Puumala virus, the etiologic agent of *nephropathia epidemica*, in bank voles (*Clethrionomys*

glareolus). *Journal of Virology*, 55(1):34–38. 5

[120] Yates, T. L., Mills, J. N., Parmenter, C. A., Ksiazek, T. G., Parmenter, R. R., Vande-Castle, J. R., Calisher, C. H., Nichol, S. T., Abbott, K. D., Young, J. C., Morrison, M. L., Beaty, B. J., Dunnum, J. L., Baker, R. J., Salazar-Bravo, J., and Peters, C. J. (2002). The Ecology and Evolutionary History of an Emergent Disease: Hantavirus Pulmonary Syndrome. *BioScience*, 52(11):989–998. 6

[121] Zarocostas, J. (2017). Cholera outbreak in Haiti—from 2010 to today. *The Lancet*, 389(10086):2274–2275. 39

[122] Zhou, Y., Chi, J., Lv, W., and Wang, Y. (2021). Obesity and diabetes as high-risk factors for severe coronavirus disease 2019 (Covid-19). *Diabetes Metabolism Research and Reviews*, 37(2). 65

Appendices

A Hamiltonians, Adjoints, and Optimal Control Characterizations

A.1 Combined Objective Functional with Disease Burden and Cost

The objective functional combining disease burden and intervention cost is formulated as

$$\min_{v \in V} \int_0^T [A_b I(t) + Bv^2(t) + Cv(t)S(t)]dt,$$

where the control set is

$$V = \{v \in L^2(0,T) \mid 0 \leq v(t) \leq v_{max} \ a.e.\}.$$

The Hamiltonian is then

$$\begin{aligned} H = {} & A_b I + Bv^2 + CvS + \lambda_1[\mu N - \beta_I SI - \beta_W SW - \mu S - vS] \\ & + \lambda_2[\beta_W WS + \beta_I SI - \gamma I - \mu I - \delta I] \\ & + \lambda_3[\gamma I - \mu R + v(t)S] \\ & + \lambda_4[\xi(I - W)] \end{aligned}$$

with the following adjoints and boundary conditions:

$$\begin{aligned} \lambda_1' &= -[Cv - \lambda_1\beta_I I - \lambda_1\beta_W W - \lambda_1 v + \lambda_2\beta_W W + \lambda_2\beta_I I + \lambda_3 v] \\ \lambda_2' &= -[A_b + \lambda_1\mu - \lambda_1\beta_I S + \lambda_2\beta_I S - \lambda_2\gamma - \lambda_2\mu - \lambda_2\delta + \lambda_3\gamma + \lambda_4\xi] \\ \lambda_3' &= -[\lambda_1\mu - \lambda_3\mu] \\ \lambda_4' &= -[-\lambda_1\beta_W S + \lambda_2\beta_W S - \lambda_4\xi] \\ \lambda_1(T) = 0, \lambda_2(T) &= 0, \lambda_3(T) = 0, \lambda_4(T) = 0. \end{aligned}$$

We can then characterize the optimal control, using this optimality condition, on the interior of the control set.

$$\frac{\partial H}{\partial v} = 2Bv + CS - \lambda_1 S + \lambda_3 S = 0 \text{ at } v^*.$$

Note $\frac{\partial^2 H}{\partial v^2} = 2B > 0$, since $B > 0$. Therefore, using the control bounds gives

$$v^* = \min\left(v_{max}, \max\left(0, \frac{(\lambda_1 - \lambda_3 - C)S}{2B}\right)\right).$$

A.2 Combined Objective Functional with New Infections and Cost

The objective functional combining new infections and intervention costs is formulated as

$$\min_{v \in V} \int_0^T [A_n(\beta_I S(t)I(t) + \beta_W S(t)W(t)) + Bv^2(t) + Cv(t)S(t)], dt$$

where the control set V is the same as above.

The Hamiltonian is then

$$\begin{aligned} H = {} & A_n(\beta_I SI + \beta_W SW) + Bv^2 + CvS + \lambda_1[\mu N - \beta_I SI - \beta_W SW - \mu S - vS] \\ & + \lambda_2[\beta_W WS + \beta_I SI - \gamma I - \mu I - \delta I] \\ & + \lambda_3[\gamma I - \mu R + v(t)S] \\ & + \lambda_4[\xi(I - W)] \end{aligned}$$

with the following adjoints and boundary conditions:

$$\lambda_1' = -[A_n\beta_I I + A_n\beta_W W + Cv - \lambda_1\beta_I I - \lambda_1\beta_W W - \lambda_1 v + \lambda_2\beta_W W + \lambda_2\beta_I I + \lambda_3 v]$$

$$\lambda_2' = -[A_n\beta_I S + \lambda_1\mu - \lambda_1\beta_I S + \lambda_2\beta_I S - \lambda_2\gamma - \lambda_2\mu - \lambda_2\delta + \lambda_3\gamma + \lambda_4\xi]$$

$$\lambda_3' = -[\lambda_1\mu - \lambda_3\mu]$$

$$\lambda_4' = -[A_n\beta_W S - \lambda_1\beta_W S + \lambda_2\beta_W S - \lambda_4\xi]$$

$$\lambda_1(T) = 0, \lambda_2(T) = 0, \lambda_3(T) = 0, \lambda_4(T) = 0.$$

We can then characterize the optimal control, using this optimality condition, on the interior of the control set.

$$\frac{\partial H}{\partial v} = 2Bv + CS - \lambda_1 S + \lambda_3 S = 0 \text{ at } v^*.$$

Note $\frac{\partial^2 H}{\partial v^2} = 2B > 0$, since $B > 0$. Therefore, we find that

$$v^* = \min\left(v_{max}, \max\left(0, \frac{(\lambda_1 - \lambda_3 - C)S}{2B}\right)\right).$$

A.3 Minimize Disease Burden under Fixed Budget

Given a fixed budget G, the cost of the control interventions, we define the control sets:

$$V = \{v \in L^2(0,T) \mid 0 \leq v(t) \leq v_{max} \ a.e.\}$$

and

$$F_1 = \left\{v \in V \mid \int_0^T (Bv^2(t) + Cv(t)S(t))dt \leq G\right\},$$

where S is the susceptible compartment of the population corresponding to the control v. The alternate objective functional to minimize the disease burden is defined as

$$\min_{v \in F_1} \int_0^T A_b I(t)dt.$$

To show the process of obtaining the adjoint functions and the optimal control characterization, we use Pontryagin's Maximum Principle. To handle the budget constraint, we introduce a fifth state variable, x_5, that satisfies

$$x_5' = Bv^2 + CvS$$

with the boundary conditions $x_5(0) = 0$ and $x_5(T) = G$.

The Hamiltonian is

$$\begin{aligned} H = {} & A_b I + \lambda_1[\mu N - \beta_I SI - \beta_W SW - \mu S - vS] \\ & + \lambda_2[\beta_W WS + \beta_I SI - \gamma I - \mu I - \delta I] \\ & + \lambda_3[\gamma I - \mu R + vS] \\ & + \lambda_4[\xi(I - W)] \\ & + \lambda_5[Bv^2 + CvS] \end{aligned}$$

with the following adjoint differential equations and boundary conditions

$$\begin{aligned} \lambda_1' &= -[-\lambda_1\beta_I I - \lambda_1\beta_W W - \lambda_1 v + \lambda_2\beta_W W + \lambda_2\beta_I I + \lambda_3 v + \lambda_5 Cv] \\ \lambda_2' &= -[A_b + \lambda_1\mu - \lambda_1\beta_I S + \lambda_2\beta_I S - \lambda_2\gamma - \lambda_2\mu - \lambda_2\delta + \lambda_3\gamma + \lambda_4\xi] \\ \lambda_3' &= -[\lambda_1\mu - \lambda_3\mu] \\ \lambda_4' &= -[-\lambda_1\beta_W S + \lambda_2\beta_W S - \lambda_4\xi] \\ \lambda_5' &= 0 \\ \lambda_1(T) = 0, \lambda_2(T) &= 0, \lambda_3(T) = 0, \lambda_4(T) = 0. \end{aligned}$$

Note λ_5 takes no boundary condition since x_5 takes two boundary conditions.

We can then characterize the optimal control, using this optimality condition, on the interior of the control set,

$$\frac{\partial H}{\partial v} = 2B\lambda_5 v + C\lambda_5 S - \lambda_1 S + \lambda_3 S = 0 \text{ at } v^*,$$

and then obtain

$$v^*(t) = \min\left(v_{max}, \max\left(0, \frac{(\lambda_1(t) - \lambda_3(t) - C\lambda_5(t))S(t)}{2\lambda_5(t)B}\right)\right).$$

A.4 Minimize Intervention Costs with Fixed New Infections

Given an upper bound P on the number of new infections, we define the control set F_2 as

$$F_2 = \left\{ v \in V \mid \int_0^T A_n[\beta_I S(t)I(t) + \beta_W S(t)W(t)]dt \leq P \right\},$$

where S is the susceptible compartment, I the infected compartment, and W the environmental contamination compartment of the population corresponding to the control v.

The alternative objective functional minimizing cost is defined as

$$\min_{v \in F_2} \int_0^T [Bv^2(t) + Cv(t)S(t)]dt.$$

To handle the constraint on the new cases, we introduce a new state variable, x_5 with

$$x_5' = A_n[\beta_I S(t)I(t) + \beta_W S(t)W(t)],\ x_5(0) = 0,\ x_5(T) = P.$$

The Hamiltonian is then

$$\begin{aligned} H = {} & Bv^2 + CvS + \lambda_1[\mu N - \beta_I SI - \beta_W SW - \mu S - vS] \\ & + \lambda_2[\beta_W WS + \beta_I SI - \gamma I - \mu I - \delta I] \\ & + \lambda_3[\gamma I - \mu R + v(t)S] \\ & + \lambda_4[\xi(I - W)] + \lambda_5[A_n\beta_I SI + A_n\beta_W SW] \end{aligned}$$

with the following adjoints and tranversality conditions:

$$\lambda_1' = -[Cv - \lambda_1\beta_I I - \lambda_1\beta_W W - \lambda_1 v + \lambda_2\beta_W W + \lambda_2\beta_I I + \lambda_3 v + A_n\lambda_5\beta_I I + A_n\lambda_5\beta_W W]$$

$$\lambda_2' = -[\lambda_1\mu - \lambda_1\beta_I S + \lambda_2\beta_I S - \lambda_2\gamma - \lambda_2\mu - \lambda_2\delta + \lambda_3\gamma + \lambda_4\xi + A_n\lambda_5\beta_I S]$$

$$\lambda_3' = -[\lambda_1\mu - \lambda_3\mu]$$

$$\lambda_4' = -[-\lambda_1\beta_W S + \lambda_2\beta_W S - \lambda_4\xi + A_n\lambda_5\beta_W S]$$

$$\lambda_5' = 0$$

$$\lambda_1(T) = 0, \lambda_2(T) = 0, \lambda_3(T) = 0, \lambda_4(T) = 0.$$

Note that λ_5 takes no boundary condition because x_5 takes two boundary conditions, $x_5(0) = 0$ and $x_5(T) = P$.

We can then characterize the optimal control, using this optimality condition, on the interior of the control set.

$$\frac{\partial H}{\partial v} = 2Bv + CS - \lambda_1 S + \lambda_3 S = 0 \text{ at } v^*$$

Note that $\frac{\partial^2 H}{\partial v^2} = 2B > 0$, since $B > 0$. Therefore, we find that

$$v^*(t) = \min\left(v_{max}, \max\left(0, \frac{(\lambda_1(t) - \lambda_3(t) - C)S(t)}{2B}\right)\right).$$

A.5 Minimize New Infections under Fixed Budget

Given a budget constraint G, we use the same control set as Section A.3,

$$F_1 = \left\{v \in V \mid \int_0^T (Bv^2(t) + Cv(t)S(t))dt \leq G\right\}.$$

The alternative objective functional minimizing the new infections is then defined as

$$\min_{v \in F_1} \int_0^T A_n[\beta_I S(t)I(t) + \beta_W S(t)W(t)]dt.$$

To handle the budget constraint, we introduce a fifth state variable, x_5, defined by

$$x_5' = Bv^2 + CvS$$

with the boundary conditions $x_5(0) = 0$ and $x_5(T) = G$.

The Hamiltonian is then

$$\begin{aligned} H = {} & A_n\beta_I SI + A_n\beta_W SW + \lambda_1[\mu N - \beta_I SI - \beta_W SW - \mu S - vS] \\ & + \lambda_2[\beta_W WS + \beta_I SI - \gamma I - \mu I - \delta I] \\ & + \lambda_3[\gamma I - \mu R + v(t)S] \\ & + \lambda_4[\xi(I - W)] \\ & + \lambda_5[Bv^2 + CvS] \end{aligned}$$

with the following adjoints and boundary conditions:

$$\begin{aligned} \lambda_1' &= -[A_n\beta_I I + A_n\beta_W W - \lambda_1\beta_I I - \lambda_1\beta_W W - \lambda_1 v + \lambda_2\beta_W W + \lambda_2\beta_I I + \lambda_3 v + \lambda_5 Cv] \\ \lambda_2' &= -[A_n\beta_I S + \lambda_1\mu - \lambda_1\beta_I S + \lambda_2\beta_I S - \lambda_2\gamma - \lambda_2\mu - \lambda_2\delta + \lambda_3\gamma + \lambda_4\xi] \\ \lambda_3' &= -[\lambda_1\mu - \lambda_3\mu] \\ \lambda_4' &= -[A_n\beta_W S - \lambda_1\beta_W S + \lambda_2\beta_W S - \lambda_4\xi] \\ \lambda_5' &= 0 \\ \lambda_1(T) = 0, \lambda_2(T) &= 0, \lambda_3(T) = 0, \lambda_4(T) = 0. \end{aligned}$$

Note λ_5 takes no boundary condition since x_5 takes two boundary conditions.

We can then characterize the optimal control, using this optimality condition, on the interior of the control set.

$$\frac{\partial H}{\partial v} = 2B\lambda_5 v + C\lambda_5 S - \lambda_1 S + \lambda_3 S = 0 \text{ at } v^*$$

Note that $\frac{\partial^2 H}{\partial v^2} = 2B\lambda_5 > 0$ if $\lambda_5 > 0$.

Therefore, we find that

$$v^* = \min\left(v_{max}, \max\left(0, \frac{(\lambda_1 - \lambda_3 - C\lambda_5)S}{2\lambda_5 B}\right)\right).$$

B Combined OC: minimize the sum of disease burden and the costs of interventions

Minimizing the cost of interventions and the integral of the infected compartment, scaled appropriately to represent the cost of disease burden, is commonly seen as the goal in management of disease models. The case of a combined objective functional with burden (Equation 3.2) is illustrated here in Figure 4 with model parameters listed in the caption of Table 5. In that table, one can see the positive effects of implementing the optimal control on the disease are striking. Since the number of susceptible individuals has become low about half way through the time interval, the constant control (set at the maximum control value, 0.04) has similar effects to the optimal control resulting from the combined objective functional (1% reduction in disease burden), but has a slightly higher intervention cost compared to the optimal control (11% increase). These simulations show that considering a time varying control allows for a better result (i.e. a smaller objective functional) than a constant control at the maximum will allow, by 1.4% in this case.

Using the parameters in Table 6 and a combined objective functional with burden (Equation 3.2), we compare the optimal control to a constant control with same cost and to the no control case. The plots shown in Figure 5 do not show much difference in the dynamics of the compartments when comparing the optimal control case and a constant control of 0.13. The two controls result in similar intervention costs but have 10% difference in the overall combined cost.

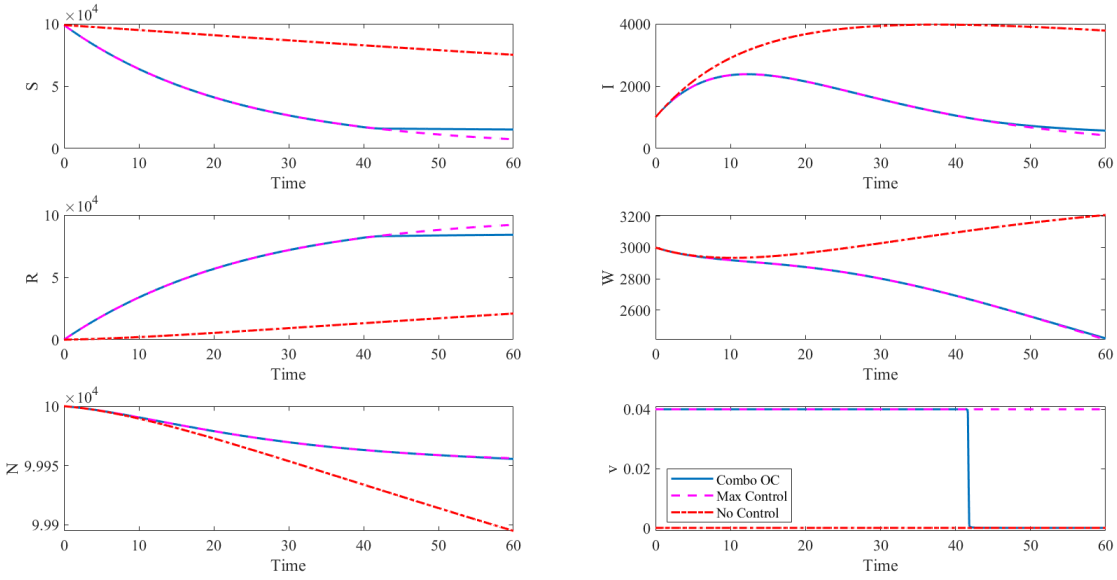

Figure 4: Optimal control and corresponding states for the combined objective functional with disease burden compared to the no control case and a constant control with the above parameters, corresponding to the results in Table 5.

Table 5: Results of the combined objective functional with disease burden compared to the no control case and a constant control of 0.16 for a time horizon of 60 days and with $\beta_I = 2.64e-7$, $\beta_W = 1.21e-7$, $\gamma = 0.1$ and $A_b = 1$, $B = 1000$, and $C = 0.325$. The initial population was set to $S = 99000$, $I = 1000$, $R = 0$, and $W = 3000$. The R_0 for this simulation is 1.47.

	No Control	Comb. OC with Burden	Constant control
Intervention costs $J_c(v)$	0	24,648	27,314
Disease burden cost $J_b(v)$	210,159	88,907	87,802
Total cost $J_b(v) + J_c(v)$	210,159	113,555	115,116
% difference in total objective functional from constant control	83	-1.4	-
Max number of infected	3,975	2,383	2,383
Number of deaths	105	44	43

Table 6: Results of combined objective functional with disease burden with time horizon of 60 days and $A_b = 1$, $B = 40$, $C = 0.0464$, $v_{max} = 0.16$, $\beta_I = 2.64e-6$, $\beta_W = 1.21e-5$, $\gamma = 0.05$, $S(0) = 9700$, $I(0) = 300$, $R(0) = 0$, and $W(0) = 300$. The R_0 for this simulation is 2.9.

	No Control	Comb. OC	$v \equiv 0.13$
Disease burden cost $J_b(v)$	47,933	10,576	11,658
Intervention costs $J_c(v)$	0	479	477
Total cost $J_b(v) + J_c(v)$	47,933	11,055	12,135
Max number of infected	1,069	363	376
Number of deaths	24	5	6
Number of vaccinations	0	9,460	9,422

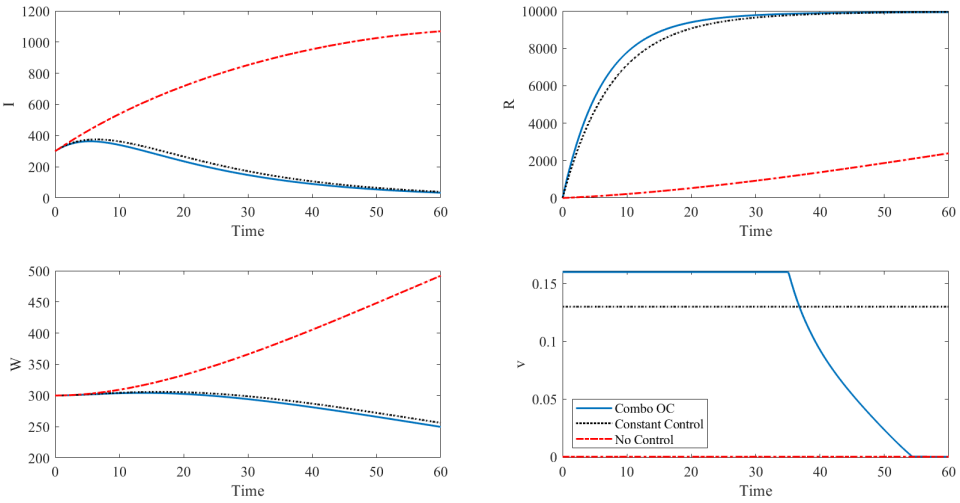

Figure 5: Optimal control and states for the combined objective functional with burden compared to a constant control with the same intervention cost and the no control case, corresponding to results in Table 6.

Vita

Morganne Igoe was born in Cleveland, Ohio to Tammy and Bob Igoe and grew up in Lorain, Ohio with one older sister, Tori. She attended Marion L. Steele High school in Amherst, Ohio, graduating in 2013. She then moved to Minneapolis, MN to attend the University of Minnesota-Twin Cities, where she earned a BS in Mathematics with a minor in ecology in 2017. Following this, she moved to Knoxville, TN to pursue a PhD in Mathematics at the University of Tennessee-Knoxville. While there, she worked with her advisor Suzanne Lenhart to seek out a variety of research opportunities to learn how to be a stronger mathematical biologist and better collaborator. She earned her MS in Mathematics in 2019. Morganne finished her PhD in Mathematics in 2022.