Title

Modeling Contact and Spatial Heterogeneity in the Transmission of SARS-CoV-2 and Chagas Disease

Author

Roubenoff, Ethan Michael

Publication Date

2023

Modeling Contact and Spatial Heterogeneity in the Transmission of SARS-CoV-2 and Chagas Disease

By

Ethan Roubenoff

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Demography

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Ayesha Mahmud, Chair
Professor Dennis Feehan
Professor Joshua Goldstein

Summer 2023

Abstract

Modeling Contact and Spatial Heterogeneity in the Transmission of SARS-CoV-2 and Chagas Disease

by

Ethan Roubenoff

Doctor of Philosophy in Demography

University of California, Berkeley

Professor Ayesha Mahmud, Chair

The study of infectious disease is more than how pathogens transmit between hosts and cause illness and death. Infectious diseases are social, political, environmental, and infrastructural phenomena that humans have limited but equal power to eradicate or to exacerbate. Using modern disease modeling methods, researchers have the ability to test our assumptions about disease dynamics and to explore ways in which our behavior may help or hurt, and to imagine a better world with effective interventions. In this dissertation I focus on three applications of such models to two diseases: an evaluation of vaccine distribution and long-term seasonality of SARS-CoV-2, the pathogen causing the COVID-19 pandemic, and Chagas Disease, a parasitic infection in Brazil that despite its relative infrequency, has the potential to cause debilitating and fatal complications. In all studies, I focus on the heterogeneous risk of transmission and infection present in the population with an eye towards disease-averting interventions.

In Chapter 1, I find that distributing vaccines for COVID-19 in decreasing priority by age is a more effective strategy to limit deaths than prioritizing by high-contact employment status, consistent with the CDC's recommendations in early 2021. In the second chapter, I continue the analysis of SARS-CoV-2 using an Agent-Based Network Simulation, finding that recurrent outbreaks of the disease are likely to persist but, depending on the epidemiology of the pathogen, may be mitigated by annual booster doses distributed at certain times of year. Finally, I conclude in Chapter 3 by finding that cases of Acute Chagas Disease may be expected to increase over the ensuing decade, driven by an increasing population in affected areas and changing climate.

Dedication

This manuscript is dedicated to my first mother, Abby Shevitz, who taught me that our capacity to alleviate suffering is the greatest superpower, and to my second mother, Barri Falk, who taught me persistence, will, and execution.

Acknowledgements

It has been a pleasure to study in the UC Berkeley Department of Demography since 2018, when I initially entered intending to only complete a master's degree before falling in love with the discipline and staying for the PhD. There are many to whom I owe a infinite debt of gratitude. First, to Ayesha Mahmud, my advisor and committee chair, who has has been my steadfast support full of understanding, ideas, and (re)direction to guide me back from dozens of wrong turns. Ayesha: I could not have become the scholar such as I am without your care and encouragement. Thank you for introducing me to infectious disease modeling, for being patient with my missteps, and for allowing me the freedom to follow my passion. And, to my committee members Dennis Feehan, for whom I have had the pleasure of serving as Graduate Student Instructor and who taught me everything I know about networks and surveys; and Josh Goldstein, who has been a wonderful support intellectually and personally. More than anything, I am thankful for the Department's support for my interests in public health and spatial demography, and for giving me a playground to freely to follow my passion. I'd like to thank Ayesha and Dennis specifically for their assistance with Chapter 1 of this dissertation, which has appeared in publication at *Epidemics* as *Evaluating primary and booster vaccination prioritization strategies for COVID-19 by age and high-contact employment status using data from contact surveys* by Roubenoff, Feehan, and Mahmud (2023).

As a Graduate Student Researcher and Affiliate of the Berkeley Interdisciplinary Migration Initiative, I have also had the pleasure of working closely alongside Irene Bloemraad on the Mapping Spatial Inequality project. Together with Jasmijn Slootjes—who brought me on to the project mere days before the COVID-19 lockdown began—we published our work on developing the Service Accessibility Index (SAI) in *Socius* as Roubenoff, Slootjes, and Bloemraad (2023): *Spatial and Sociodemographic Vulnerability: Quantifying Accessibility to Health Care and Legal Services for Immigrants in California, Arizona, and Nevada.* Irene and I later published an Op-Ed in the LA Times: *No, California doesn't have a population crisis.* Thank you, Irene, for the pleasure of collaboration and for broadening my research horizons. As well, I would like to thank Perry de Valpine for serving on my Qualifying Exam committee; Magali Barbieri, who gave me my first research position with the Human Mortality Database; and Robert Chung, who helped me unlock the door to my intellectual and analytical creativity, and whose maxim I truly took to heart during my studies: "the point of modeling is not to fit the data; rather, it is to sharpen the questions."

I am grateful for the Demography Department's and University's financial support of my studies: as a National Institute of Health Trainee under the Department's Training Grant from the National Institute on Aging; as a three-time Graduate Student Instructor with Profs. Jenna Johnson-Hanks, Dennis Feehan, and Ryan Edwards; and with a Doctoral Completion Fellowship from the UC Berkeley Graduate Division. Funding for the Berkeley Interpersonal Contact Survey (BICS) used in Chapters 1 and 2 included seed funding provided by a Berkeley Population Center pilot grant (NICHD P2CHD073964) and further funding was provided by the Hellman Fellows Program.

During my time at UC Berkeley, I had the honor of participating in two industry in-

ternships that shaped my career goals and trajectories. I'd like to thank Steve Fick, my manager at Pivot Bio and whose WorldClim dataset is cited in Chapter 3; and Jesse Shore, my manager on Meta's Demography and Survey Science team. Above all, these experiences taught me to take risks, to open my interests to things I would never have otherwise tried—especially corn fertilizer—and to have confidence in myself and my career future.

This dissertation is dedicated in part to my mother Abby Shevitz who passed when I was nine years old. I have spent the better part of my life trying to craft myself in her image, to carry her legacy forward, and live my life with the wisdom she could impart in our short time together. It has always been my goal to make you proud, and I hope that your influence comes through clear in my life and work—it was my greatest honor to include one of her publications on gender and HIV in the introduction to this dissertation. I am thankful for my uncle Danny, aunt Laura, and cousins Josh and Lily, who gave me a home away from home in New Mexico multiple times during the lockdown; and for my aunt Ellen, for our regular phone calls and her unwavering support of my studies.

The years that I was at Berkeley were some of the most personally difficult in my life. The COVID-19 pandemic began in the final months of my second year, and we were fully distributed through my fourth year. During this time, my adoptive mother Barri Falk, her brother Michael, and sister Lisa were diagnosed with ALS. Michael passed in 2021 and Barri a year later in 2022. One of the greatest pains has been coming to share the particular sorrow of losing a parent too young with my brother Austin, sister Talia, and cousin Collin—one I had hoped to never share, having lost my mother Abby at a tender age myself. The confluence of the isolation, loneliness, and uncertainty of the pandemic with our family's losses was so great that at times that I nearly succumbed to the pressure and gave up on my studies. Barri would not have approved; I am glad that I stuck it out, but my completion of this dissertation is not something I take for granted.

I could not have gotten through this without an incredible group of friends by my side: Amanda Glazer, my roommate, sounding board, and baseball buddy; my cohorts Felipe Menares, Andrea Miranda-Gonzalez, Casey Breen, Elizabeth Breen, Rae Willis-Conger, Payal Hathi; my dear friends, Jonathan DeBruin (without whom I might not have made it through Spring 2020), Zach and Courtney Person, Michael Henry, and my best friend since third grade, Alex Peltz.

Finally, I owe the most to my partner and fiancé Tristan, who has guided me through the worst lows, given me encouragement when I needed it most, who has stood by me every step of the way since our first date in 2019, through all of our struggles in California, then to our new life in Massachusetts. I am blessed to have had our union during the worst of times, and not a single day has passed that I am not grateful for your tolerance of my stress levels. And lastly, to my father and physician-turned-novelist Ronenn Roubenoff, who has been through the worst thrice over, an example of integrity, of perseverance, of wit and inspiration.

Contents

List of Figures

List of Tables

Introduction

Social inequalities including gender and racial discrimination, information and opportunity inequality, and financial and economic disparities are widely known to have detrimental effects on individuals' health outcomes, livelihood, and mortality (Berkman and Syme 1979; Braveman and Gottlieb 2014; Galea et al. 2011; House, Kessler, and Herzog 1990; Link and Phelan 1995; Cassel 1976; Krieger 2001; McMichael 1999; Rose 1985, among others). Social epidemiology, the branch of public health that focuses on the social causes of mortality and morbiditiy, has grown substantially since since the early 1990s. Early theories such as Cassel's (1976) *Host Resistance* were the first to posit that over decades, the social environment would stress and weaken a body's ability to fight disease (see also Segerstrom and Miller 2004 for a review of the immunological effects of stress). Early social epidemiology focused largely on the effect of poverty on heart disease and obesity—among so-called 'lifestyle' factors—such that the opportunities presented to poor and minority people were inherently more unhealthy (Krieger 2001; McMichael 1999). Link and Phelan (1995) were among the first to fully crystallize the concept of social conditions as fundamental causes of disease, summarizing from a litany of research that socioeconomic status itself was a cause of health disparities, rather the reverse pathway. Writing just over a decade after the emergence of HIV—the most significant public health event of the 20th century—Link and Phelan opened the door for inquiry in explicit and structural discriminatory mechanisms behind infectious disease. In the decades since, scholars have shown that health inequalities among mothers affect children even before birth and persist over generations (Currie 2011).

Research in socio-demographic inequality in public health and infectious disease ultimately aims to identify the necessary starting point for instituting restorative change. My dissertation research focuses on infectious diseases in heterogeneous populations, emphasizing inequality in risk of infection along social and spatial divides. Identifying the causes and effects of health inequality is an important step towards effectively implementing a disease-averting intervention—which is realistically constrained by supply, distribution, or other barrier to access—to result in the most benefit, like priority allocation of early COVID-19 vaccines. To that end, my research uses statistical and mathematical modeling as a means to study population inequality and produce actionable recommendations by focusing and expanding on risk factor heterogeneity. Simple compartmental mathematical models for infectious disease—despite maintaining accurate disease etiology—may fail to capture important dynamics within a heterogeneous population or increase estimate bias and uncertainty. However, overly complex models run the risk of being too multidimensional for easy interpretation. The projects outlined in this dissertation aim to balance complexity and simplicity.

I focus primarily on disease inequality driven by contact and spatial heterogeneity. Simple disease-transmitting models assume that risk of contracting an infectious disease is equally distributed throughout a population (Blackwood and Childs 2018). This homogeneity assumption may hold with certain simple diseases in small populations; however, population heterogeneity can have substantial effects on disease risk, with important nuances for disease modeling, prevention and intervention, and social equity[1]. Within the context of the ongoing COVID-19 pandemic, I focus on how inequality in contact frequency and network size itself can drive outbreak dynamics (Berkman and Krishna, Aditi 2014; Chang et al. 2021; Feehan and Mahmud 2021). Some high contact individuals' behavior is voluntary, such as social gatherings or vacations; however, involuntary, potentially COVID-transmitting contact such as that required by service-sector employment has been shown to mirror racial employment inequality (Wrigley-Field et al. 2020; Selden and Berdahl 2020; Baker, Peckham, and Seixas 2020). As well, I focus on spatial inequality in Chagas disease, a rare parasitic infection that primarily affects one region in northeastern Brazil. Chagas disease is most frequently caused by contact with *Triatomine* insects, and people who live within the geographic distribution of *Triatomines* are at risk. Chagas is known to be a disease of poverty (Choi et al. 2020; Dias 1987; Medone et al. 2015; Sosa-Estani and Segura 2015); residential preventative measures include removal of rotting wood and use of insecticides, which can be cost-prohibitive for households to implement. Therefore, the true risk of contracting Chagas is an intersection between spatial location and social conditions.

In Chapter 1, which has appeared in publication in *Epidemics* as Roubenoff, Feehan, and Mahmud 2023, I conduct simulations to determine optimal vaccination prioritization for COVID-19 by age and employment status. Throughout the COVID-19 pandemic, there has been substantial discussion of the inherent increase in risk of infection that in-person workers face as a result of their employment obligations. As many essential workers are people of color and low-income, economically-mandated potential exposure is one driver of the increasingly observed disparity in COVID-19 infection and mortality among vulnerable groups. We utilize data from the Berkeley Interpersonal Contact Survey (BICS; Feehan and Mahmud 2021) to parameterize a SIR-type compartmental model that highlights heterogeneity in face-to-face contact frequency among adults working in-person jobs—such as essential workers, healthcare workers, and retail employees—and children, adults not working in person, and seniors. We unambiguously find that in the first wave of vaccination—before the circulation of highly infectious variants that easily evade vaccine-derived immunity—prioritizing seniors limits the most deaths, but prioritizing high risk adults spares the most infections. When considering booster doses in the face of increasingly more transmissible and virulent variants, like the Omicron wave of late 2021 and early 2022, the optimal strategy is highly sensitive to initial vaccine efficacy.

---

[1] See the following studies, among many others, for examples: Bansal, Grenfell, and Meyers 2007; Funk, Salathé, and Jansen 2010; Pitzer et al. 2009; Berkman and Syme 1979; Braveman and Gottlieb 2014; House, Kessler, and Herzog 1990; Link and Phelan 1995; Weiss and McMichael 2004; Phelan, Link, and Tehranifar 2010; Shevitz et al. 1996

Next, in Chapter 2, I continue the analysis of SARS-CoV-2 and COVID-19 into the ensuing decade, with an Agent-Based Network simulation of SARS-CoV-2 outbreaks assuming waning immunity over time. Agent-Based Models differ from compartmental models, like the ones used in Chapter 1, by simulating a population of individuals with explicit rules governing behavior rather than differential equations describing flows between partitions of the population. While the literature on ABMs in infectious disease modeling is smaller and more recent than compartmental models, the increased flexibility—especially when including network dynamics, stochasticity, and heterogeneity—allows for a more flexible way of representing infectious disease transfer. In this model, we draw on BICS data to simulate random draws of networks using the Configuration model (Albert-László Barabási 2021) alongside household and school contacts. As well, we include vaccine distribution and booster doses with a flexible distribution schedule, isolation of clinically infectious cases, and demography (fertility and non-COVID mortality). Our model is a free-standing application written in C++ and is made publicly available for download, and can be extended to a variety of disease modeling applications. We find that booster doses, when distributed annually at certain times of the year, have the power to reduce or eliminate a large wintertime outbreak depending on the epidemiology of the pathogen. In the absence of transmission seasonality, distributing vaccines in the first half of the year can eliminate a large outbreak; in addition when seasonality is included, vaccines distributed in early fall—like vaccines for influenza—can reduce the winter-time outbreak.

Finally, in Chapter 3, I look at Chagas disease, a rare parasitic infection in rural Brazil. Substantial progress has been made in non-pharmaceutical interventions to limit the number of infections in the last 30 years, but Chagas continues to affect many people in northeastern Brazil. When treated in the acute phase of infection, Chagas disease has a high recovery rate; without treatment, life-threatening complications including cardiomypoathy and megacolon can occur in a substantial portion of people. Acute infection is generally nonspecific; symptoms include fever and eyelid edema that can be easily missed. Known as a 'disease of poverty,' most cases of Chagas disease—both acute and chronic—occur in relatively impoverished areas among people who are disenfranchised from the health care system. Overall, we find that an increasing population in areas of Chagas endemicity may result an increase in new cases over the years to come, mediated by the changing bioclimate.

For this project, I use a Knorr-Held-type spatio-temporal statistical model adapted to suit the requirements of a rare disease like Acute Chagas Disease. Geostatistical models of this sort are justified when the location of observations in two-dimensional space may be related to the outcome of interest (Haining 2004; Wachter 2005; Cressie and Wikle 2011; Anselin, Luc and Bera, A 1998), such as the distribution of disease vectors. Spatial non-independence can be summarized by Tobler's first law of geography (Tobler, 1970): "Everything is related, but close things are more related than far things." In many geostatistical applications—especially for use with disease surveillance data—space is used as a means of recovering a latent or unobserved process related to risk of contracting disease. Traditional statistical modeling techniques, especially when used within a regression framework, often assume that data are 'independently and identically distributed' (*iid*). The *iid* assumption holds that each observation is not conditional on other observations and that the data generating process for all observations is the same. For many geostatistical questions that seek to investigate if location is a predictor of an outcome, this assumption may not hold both from a mechanistic and statistical point of view. Statistically, this implies that our data are not *iid*, and in fact assuming independence may be problematic both statistically and inferrentially. Should the data-generating process have a strong spatial pattern, using an aspatial model that assumes an independent error process—say, like in OLS regression—may chalk all spatial variation up to statistical error, when the reality may be a mixture of spatially-structured heterogeneity and aspatial random variation. Using a model that cannot detect spatially-structured hetergeneity will overestimate the variance of the model, increasing the type-2 (false negative) error rate and decreasing the statistcal power of the analysis. In the worst case, using an aspatial regression model for a spatial process will likely fail to capture our process of interest and have poor predictive accuracy. While using the proper spatial specification is ultimately a question of getting the error process right, the consequences can be extreme for any causal conclusions of an improperly-specified model.

Together, these three studies each approach intervention to fight disease and inequality in disease risk. In Chapter 1, I find that age and employment drive differential risk for COVID-19, and how to distribute vaccines in a way to minimize deaths given that differential risk. In Chapter 2, I continue to analyze SARS-CoV-2, finding that risk of infection is not uniform throughout the year due to waning immunity and seasonal transmission forcing. Both of these studies show that the distribution of vaccines has a significant effect on the number of infections and deaths due to SARS-CoV-2, and that strategies for prioritization and timing should be done in a way that minimizes the consequences of the disease. Finally, in Chapter 3, I turn to Chagas Disease in Brazil, finding that despite progress towards elimination new diagnoses are likely to increase over the ensuing decade. However, this is not necessarily even throughout the country, driven by population growth and potentially climate. Although interventions have considerably reduced the frequency of Chagas infection since the 1970s and 80s, I find that under the present level of interventions Chagas disease can be expected to persist, and more elaborate interventions may be necessary for elimination in the future. This dissertation contributes to our understanding of risk and control strategies SARS-CoV-2 and Chagas Disease in the future. Both diseases are highly complex in etiology and epidemiology; however, both present opportunities for control and intervention as explored in this dissertation.

Chapter 1

Evaluating primary and booster vaccination prioritization strategies for COVID-19 by age and high-contact employment status using data from contact surveys

Abstract

The debate around vaccine prioritization for COVID-19 has revolved around balancing the benefits from: (1) the direct protection conferred by the vaccine amongst those at highest risk of severe disease outcomes, and (2) the indirect protection through vaccinating those that are at highest risk of being infected and of transmitting the virus. While adults aged 65+ are at highest risk for severe disease and death from COVID-19, essential service and other in-person workers with greater rates of contact may be at higher risk of acquiring and transmitting SARS-CoV-2. Unfortunately, there have been relatively little data available to understand heterogeneity in contact rates and risk across these demographic groups. Here, we retrospectively analyze and evaluate vaccination prioritization strategies by age and worker status. We use a mathematical model of SARS-CoV-2 transmission and uniquely detailed contact data collected as part of the Berkeley Interpersonal Contact Survey to evaluate five vaccination prioritization strategies: (1) prioritizing only adults over age 65, (2) prioritizing only high-contact workers, (3) splitting prioritization between adults 65+ and high-contact workers, (4) tiered prioritization of adults over age 65 followed by high-contact workers, and (5) tiered prioritization of high-contact workers followed by adults. We find that for the primary two-dose vaccination schedule, assuming 70% uptake, a tiered roll-out that first prioritizes adults 65+ averts the most deaths (31% fewer deaths compared to a no-vaccination scenario) while a tiered roll-out that prioritizes high contact workers averts the most number of clinical infections (14% fewer clinical infections compared to a no-vaccination scenario). We also consider prioritization strategies for booster doses during a subsequent outbreak of a hypothetical new SARS-CoV-2 variant. We find that a tiered roll-out that prioritizes adults 65+ for booster doses consistently averts the most deaths, and it may also avert the most number of clinical cases depending on the epidemiology of the SARS-CoV-2 variant and the vaccine efficacy.

COVID-19[1] vaccines have been shown to be highly effective at preventing severe illness and death (Baden et al. 2021; Polack et al. 2020). Following the introduction of vaccination in the U.S. in December 2020, infection rates decreased dramatically through the first quarter of 2021 as increasing shares of the population were protected via vaccine-derived immunity (Gupta et al. 2021). Due to a limited vaccine supply initially, there has been a complicated debate around the trade offs of vaccine prioritization strategies for adults over age 65, healthcare workers, other essential frontline workers, and the general public (Schaffer DeRoo, Pudalov, and Fu 2020; Persad, Peek, and Emanuel 2020; Persad et al. 2021; Giubilini, Savulescu, and Wilkinson 2021).

Control interventions for infectious diseases can have different public health objectives: while the highest priority is often to limit total deaths, secondary priorities can include limiting total infections or reducing the number of infected persons at a given time to below a critical care capacity threshold (Bubar et al. 2021; Buckner, Chowell, and Springborn 2021). Prioritization of vaccination towards a specific group has the direct benefit of reducing infections and deaths in that group. However, for vaccines that prevent transmission, there is also an indirect benefit of limiting secondary infections. Modeling studies that evaluate the total effect of vaccination—the sum of the direct and indirect effects on incidence and deaths—find that benefits of vaccination may extend beyond those conferred to recipients of vaccination themselves (Bubar et al. 2021; Buckner, Chowell, and Springborn 2021).

The risk of severe disease, hospitalization, and death from COVID-19 increases sharply with age (Levin et al. 2020; O'Driscoll et al. 2021), indicating that vaccinating adults 65+ may be most effective at reducing total hospitalizations and deaths due to COVID-19. On the other hand, in-person workers with higher rates of person-to-person contacts are at an increased risk of being infected with and transmitting SARS-CoV-2.

Thus, vaccine prioritization strategies need to balance: (1) the direct protection conferred by the vaccine amongst those at highest risk of severe disease outcomes, and (2) the benefits of indirect protection and potentially achieving herd immunity more quickly through vaccinating those that are at highest risk of being infected and of transmitting the virus. The indirect benefits of preventing COVID-19 by prioritizing high-contact workers could outweigh the direct benefit of vaccinating adults 65+, depending on the public health objective, the prevalence of non-pharmaceutical interventions, and the epidemiology of the virus (Buckner, Chowell, and Springborn 2021).

Vaccinating high-contact workers also has important implications for social and economic equity. Historically-disadvantaged groups, especially Blacks and Hispanics, are overrepresented in essential and front-line occupations and have younger age-distributions compared to Whites (Nelson et al. 2022). Prioritizing high-contact workers, therefore, delivers proportionally more doses to racial and ethnic groups that have also been hardest hit by

---

[1] Chapter 1 has been adapted for publication in *Epidemics* as Roubenoff, Feehan, and Mahmud 2023.

the pandemic (Andrasfay and Goldman 2021; Wrigley-Field et al. 2020; Wrigley-Field et al. 2021) compared to a purely age-based prioritization.

In the US, distribution of the primary vaccine doses prioritized a combination of adults 65+ and essential workers in progressive phases, beginning with those living in long-term care settings and healthcare workers. Eligibility was first opened to adults 65+ and to certain occupational groups before opening to the general public (Dooling 2021). The debate on prioritization is, however, still relevant in many other parts of the world, as well as in the US for future booster doses. It is also important to retrospectively evaluate prioritization strategies to inform response to future pandemics. Mathematical models that account for both the direct and indirect effects of vaccination can help guide policy decisions on prioritization. However, there is little data on contact rates by worker status and age available for the U.S., making these models hard to parameterize. The relatively few contact surveys conducted during 2020-2021 indicated that total contacts had substantially reduced compared to pre-pandemic measures (Liu et al. 2021a, Feehan and Mahmud 2021). However, previously available data has reported contact rates disaggregated by age, but not by both age and occupational status. In a survey of workers in the US, Kiti et al. 2021 found median number of reported contacts was low (around two) and that household structure—rather than age or race—was responsible for the variation in contacts between respondents. However, the study was limited to only three US companies, where many workers reported being able to work remotely offsite. A nationally representative study, covering a more diverse range of occupations, reported that most contacts during this period did happen at work, and that nonwhite and workers in essential occupations had amongst the highest total contact rates (Nelson et al. 2022). Here, we retrospectively analyze and evaluate vaccination prioritization strategies by age and worker status using detailed contact data from surveys to parameterize a mathematical transmission model for SARS-CoV-2.

We compare total effects of prioritizing adults 65+ for vaccination versus prioritizing workers who potentially have a higher risk of contracting COVID-19 due to their in-person work status (Baker, Peckham, and Seixas 2020; Hawkins 2020; Selden and Berdahl 2020). The model accounts for contact patterns between age and occupation groups using contact survey data collected as part of the Berkeley Interpersonal Contact Survey (BICS; Feehan and Mahmud 2021), which has been collecting detailed information about a respondent's daily behavior and their disease-relevant interpersonal contact since March 2020.

We find that prioritizing adults 65+ for primary vaccination averts 25% more deaths than prioritizing high contact workers and 11% more than a split strategy. However, the most clinical infections are averted by strategies that prioritize high contact workers first.

We extend the model to consider the prioritization of booster doses during a hypothetical future outbreak of a new SARS-CoV-2 variant that is able to partially evade vaccine-derived immunity from primary vaccination. When considering booster doses, the reduction in deaths is greatest when prioritizing adults 65+, but all three strategies have similar effects on clinical infections. These results highlight the impact of various vaccination prioritization strategies and can help guide policies during future outbreaks.

1.1 Methods

1.1.1 Data

The BICS survey, collected in several waves beginning in March 2020, is an online survey aimed at capturing the frequency and nature of respondents' physical and conversational contacts over a 24-hour period. Respondents (egos) in the BICS survey are asked about members of their household and their total number of non-household contacts, as well as detailed information on up to 3 of their previous day's contacts (alters). Here we use data from wave 4 of the BICS survey, collected between November 30th and December 8th, 2020, where $2,993$ respondents provided detailed information on $10,001$ contacts and were asked additional questions on their work status and work contacts. Respondents who reported being employed were asked to indicate the number of close contacts they had while performing duties for their job.

Respondents are divided into 3 categories: low contact (LC) adults (age 18-64) not reporting having any close work contacts, high contact (HC) adults (age 18-64) reporting having work contacts, and adults 65+. Although heterogeneity in contact rates and mortality may exist within groups, these age groups were chosen to maintain sufficient sample size especially in the oldest age group. 34% of respondents aged 18-64 reported having interpersonal contacts at work and are labeled as high contact (HC). Children (under age 18) were not included in the survey, although adult respondents could report their contact with children. The procedure for estimating contacts for the 0-18 age group is described below. Alters are divided into children aged 0-18, adults 18-64, and adults 65+ and weighted by the number of total close contacts reported by each ego (see Feehan and Mahmud 2021 for more detail on the weighting procedure). Working-age alters aged 18-64 are categorized as being work contacts if the reported relationship was a coworker or client or if the reported contact happened at work or a store. This consisted of 22% of reported contacts for adults aged 18-64. For some working-age alters we were unable to determine work status from the survey data. This was due to missing data for non-household alters on the purpose of the reported contact and for all reported household contacts (since their work status was not collected). For the alters whose in-person work status is indeterminable from the survey data provided, alters are randomly re-labeled as having in-person work such that the proportion of high contact alters matched the surveyed ego proportion. We also perform a sensitivity test to understand the impact of this random reallocation.

The survey responses are used to construct an age and work-status structured contact matrix between children, low-contact adults (18-64), high-contact adults (18-64), and adults 65+ (Figure 1.1 and supplementary sections A.1.1-A.1.2) using standard methods described in more detail by Feehan and Mahmud 2021 and Jarvis et al. 2020. Briefly, the raw contact matrix $\boldsymbol{M}$ has entries m_{ij} corresponding to the average number of daily contacts between respondents (ego) in group i with their reported contacts in group j, adjusted for survey weights. Total contacts in a population must be reciprocal, but may not be in the survey data due to sampling and differences in survey reporting. We impose reciprocity using previously described methods (Feehan and Mahmud 2021). To adjust for reciprocity, population data for each group is taken from the American Community Survey (ACS) 2019 5-year estimates (US Census Bureau 2019). Adults aged 18-64 in the ACS are categorized as working in-person (34%) and not working in-person (66%) using the survey proportions since equivalent data on in-person work status is unavailable from the ACS. As children are not included in the BICS survey, contacts between children are derived from the POLYMOD survey as described in the supplementary materials section A.1.3. The reciprocal contact matrix, $\boldsymbol{C}$, is the reciprocity-enforced average daily contact matrix. We note that respondents in our survey in the 65+ and Adult LC categories have comparable rates of contact to the survey of remote workers conducted by Kiti et al. 2021.

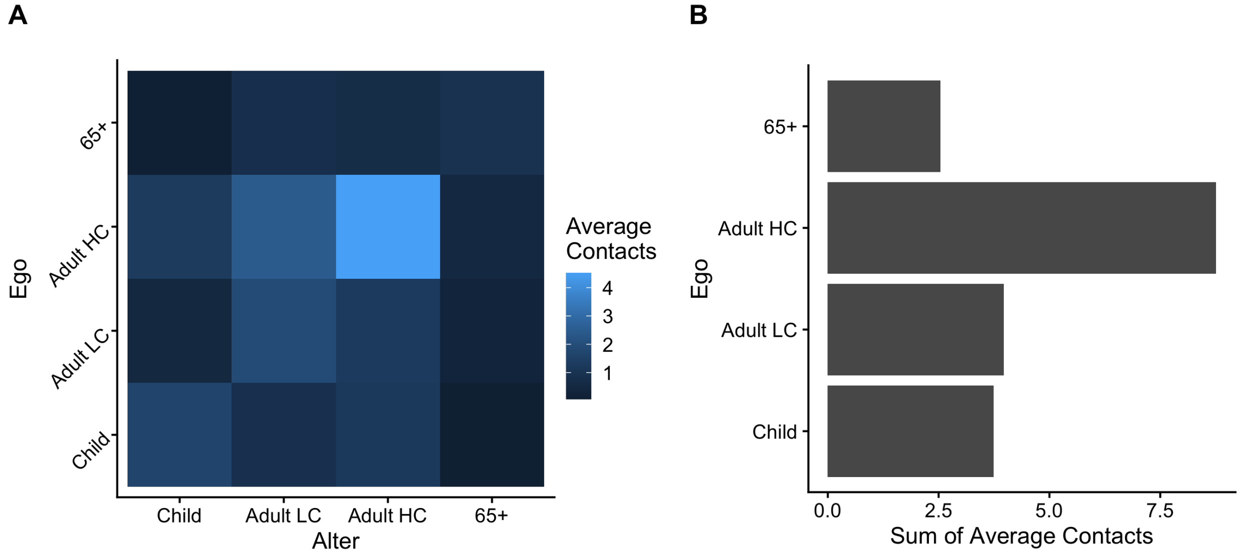


Figure 1.1: (A) Age and work-status structured contact matrix showing daily average number of reported contacts, after adjusting for reciprocity in total contacts and survey weights. (B) Total number of daily contacts for each group, calculated as the sum of each row of the matrix in panel A (total contacts across all groups that they have contact with). For both figures, “Adult LC” and “Adult HC” correspond to adults without and with in-person work contacts (Low Contact and High Contact, respectively). Within-group contacts for children (0-18) are derived from the POLYMOD survey (Mossong et al. 2008).

1.1.2 SARS-CoV-2 transmission model

To model SARS-CoV-2 transmission dynamics and COVID-19 incidence and mortality, we use a deterministic, continuous time compartmental model (outlined in the supplementary section A.1.4, similar to Bubar et al. 2021 and Buckner, Chowell, and Springborn 2021). The model allows for heterogenous mixing between age and in-person employment status groups as specified by the contact matrix derived from the BICS survey data. Susceptibles[2] (compartments S and S_x, described below) who are exposed to SARS-CoV-2 enter an exposed (latent) phase (E). Depending on their age group, exposed individuals proceed to have clinical (symptomatic) infection (I_c) with probability $[0.35, 0.4, 0.75]$ for children, adults 18-64, and adults 65+ respectively; remaining cases experience subclinical (asymptomatic) infection (I_{sc}; adapted from Davies et al. 2020). We allow for subclinical transmission, at a reduced probability (50%; Davies et al. 2020) relative to clinical infections. Probability of death after clinical infection (μ) varies according to age; subclinical mortality is assumed to be zero. Fatal cases proceed to compartment D and non-fatal cases recover to compartment R.

We model a two-dose primary vaccination schedule, distributed 25 days apart, to match the two-dose Pfizer and Moderna vaccines that account for the majority of the vaccination doses delivered in the United States (CDC 2020). Susceptibles awaiting the vaccine in compartment S proceed to compartment V_a after the first dose and then to V_b after the second dose; individuals who have contracted SARS-CoV-2 are ineligible for the vaccine. We incorporate 'leaky' vaccine efficacy and vaccine hesitancy using methods similar to Bubar et al. 2021 and assume an 80% reduction in infections after the first dosage and 90% after the second, consistent with efficacy estimates during the initial roll-out of the vaccines. Leaky vaccines are incorporated by proportionally reducing the force of infection by the corresponding vaccine efficacy for all vaccinated individuals after one or two vaccine doses. In our implementation, vaccination reduces the probability of becoming infected, but does not reduce the probability of a vaccinated yet infectious individual of transmitting the disease (Tenforde 2021; Thompson 2021). Breakthrough infections occurring among those who have received either the first or second dose proceed to the exposed compartment and then on to the infected compartments (with the same probabilities as unvaccinated exposed individuals). Vaccine hesitancy is incorporated by imposing a 70% uptake of the primary vaccination doses, derived from the National Immunization Survey's May 2021 primary uptake for seniors (National Center for Immunization and Respiratory Diseases (NCIRD) 2022b; National Center for Immunization and Respiratory Diseases (NCIRD) 2022a). At the start of the simulation, 70% of susceptibles are in the S compartment and awaiting vaccination; the remaining 30% of susceptibles who refuse or are otherwise ineligible for the vaccine are placed in the compartment S_x, and are otherwise identical to individuals in the S compartment. At its peak, nearly 5 million Americans were being vaccinated per day (CDC 2020); however, as initially vaccine rollout was slower, we assume an average 2 million vaccinations per day distributed equally between first and second shots. First doses are distributed until the S compartment is depleted, either through vaccination or infection. Over the course of the simulation, vaccine uptake is less than the 70% of the target population as those who have become infected while susceptible and awaiting vaccination are ineligible for vaccination. After distribution of vaccines to the priority group, vaccines (including surplus doses intended for members of the priority group who became infected and subsequently ineligible) are distributed to any remaining adults 65+, adults 18-64, and children proportional

---

[2] All model compartments are indexed by group i. Here, we drop group-specific subscript i, indicating the count of individuals in each compartment in group i, for readability.

to the remaining susceptible population size in each group. Those who have recovered from infection are not eligible for vaccination in our simulations.

The next generation matrix (NGM; Bansal, Grenfell, and Meyers 2007; Diekmann, Heesterbeek, and Roberts 2010; Bubar et al. 2021) for the model is:

$$NGM_{ij} = \frac{u_i C_{ij}}{\gamma}\left[\rho_j + (1-\rho_j)\alpha\right] \tag{1.1}$$

Where γ is the recovery rate, ρ_j is the probability that an exposed individual in group j is clinically infectious, α is the relative infectiousness of clinical versus sub-clinical cases, u_i is the age-dependent susceptibility to infection after contact with an infectious individual, and C_{ij} is the entry in contact matrix $\boldsymbol{C}$ corresponding to the average number of daily contacts a respondent in group i has with an individual in group j. The dominant eigenvalue of the NGM is the basic reproduction number R, and when the population is fully susceptible this is equal to the basic reproduction number, R_0. We scaled the values of u_i to calibrate to a specific value of R_0 by optimizing a scaling factor for u_i such that the largest eigenvalue of the NGM matches an assumed R_0 value (Davies et al. 2020; Bubar et al. 2021).

The starting population for the simulations is the United States population on January 1st, 2021, with approximately 73.4 million children, 200.5 million adults 18-64, and 50.8 million adults 65+ (US Census Bureau 2019). Working-age adults were split into working in person and not working in person (including unemployed) using the survey proportions of 34% and 66%, respectively, as derived from BICS data. On January 1st, there were $20,166,028$ confirmed cases, $352,390$ deaths, and $11,426,602$ known active cases. The starting conditions for deaths, subclinical cases, recovered cases, and exposed individuals are detailed in section A.1.

1.1.3 Model parameters

Estimates for u_i and ρ_i are taken from literature (Davies et al. 2020; Bubar et al. 2021). To account for uncertainty in other model parameters, we performed 1000 simulations with transmission and mortality parameters drawn from their assumed distributions (described below) using latin hypercube sampling. R_0 was assumed to be distributed normally with mean 2.5 and standard deviation 0.54 (following Feehan and Mahmud 2021). Child, adult 18-64, and adult 65+ mortality were drawn from uniform distributions $[0.003\%, 0.005\%]$, $[0.20\%, 0.26\%]$, $[6.9\%, 10.4\%]$, respectively (bounds for uniform distribution are the 95% confidence intervals from Levin et al. 2020 for age groups 0-34, 45-54, and 75-84). Average latent period is assumed to be distributed uniformly between 2 and 4 days and average duration of infectiousness is assumed to be distributed uniformly between 4 and 6 days, such that median draws follow values assumed by the literature (see supplemental table A.1.6 for sources on all parameters). Additional parameters are outlined in table A.1.6. For all simulations, we estimated the percent reduction in total clinical infections and total deaths from January 1, 2021 until December 31st, 2021 compared to a no-vaccination scenario for five vaccine prioritization strategies: (1) prioritizing only adults over age 65, (2) prioritizing only adults 18-64 with in-person work contacts, (3) splitting priority vaccines evenly between adults 65+ and adults working in person, (4) a 'tiered' strategy that prioritizes adults 65+ before high contact workers, and (5) a 'tiered' strategy that prioritizes high contact workers before adults 65+ (further details are provided in the supplementary material A.1.7). The 'tiered' strategies are intended to replicate the CDC's decision to progressively distribute the vaccine in a series of decreasing priorities. In our simulations, tiered roll-outs differ from the single priority strategies by allowing a second-priority group to have access to the vaccine before general distribution (when doses are distributed proportionally to eligible group size). For example, for the single priority 65+ strategy, after all of those eligible in the oldest age group have been vaccinated, remaining vaccines are distributed to other groups proportional to remaining eligible group size; however, during the tiered 65+ strategy, after eligible adults 65+ have been vaccinated doses are distributed to eligible HC adults before being distributed to other groups. For clarity, we present the simulation results using the median draw for each parameter for discussion of effect sizes ($R_0 = 2.5$, $\mu = [0.00004, 0.0023, 0.08]$, latent period = 3 days, and infectious period = 5 days), but show the full range of simulation results across the 1000 parameter combinations.

1.1.4 Booster dose model for a subsequent outbreak of a SARS-CoV-2 variant

We also extend the model (outlined in supplementary material section A.1.8) to consider a subsequent outbreak caused by a new, more transmissible SARS-CoV-2 variant, such as the Omicron variant, where vaccines may be less effective (Collie et al. 2022; Hogan et al. 2021; UK Health Security Agency 2021; Thompson 2022). In this situation, distribution of a third "booster" dose is necessary to increase protection against clinical infection, hospitalization, and death. To maintain consistency across results, the starting population size is the same as in the previous simulation (the US population on January 1st 2021). In the booster simulation, we take the January 1st 2022 estimate of 34%, 78%, and 95% of children, adults 18-64, and adults 65+ as having received a primary course of vaccination (2 doses, represented by compartment V_b; National Center for Immunization and Respiratory Diseases (NCIRD) 2022a; National Center for Immunization and Respiratory Diseases (NCIRD) 2022b). Remaining susceptibles are assumed to refuse the vaccine and are placed in compartment S_x. We assume that 70% of individuals in V_b who have received the two-dose primary course of vaccination will receive booster doses; hesitancy or ineligibility for booster doses is incorporated by moving 30% of these individuals in V_b to V_{bx}, indicating that they decline the booster dose. One million booster doses are distributed daily among individuals in V_b until that compartment reaches zero individuals either through infection or vaccination. In this booster model, no individuals will be present in compartment S (awaiting first dose) or V_a (awaiting awaiting second dose). We consider the same five strategies as before for priority distribution of booster doses. Since the vaccine's effectiveness in reducing transmission is unknown for primary and boosted individuals for new variants, we conduct 1000 simulations with randomly drawn values for vaccine efficacy from an assumed distribution; we also draw

1000 values of R_0 from an assumed distribution. The vaccine efficacy is randomly drawn from distributions derived from the CDC's estimates of vaccine efficacy against the Omicron variant (Thompson 2022). In these simulations, primary (2-dose) vaccine efficacy is drawn from a uniform distribution between 32% and 43%; vaccine efficacy of the booster dose is drawn from a uniform distribution between 79% and 84%. (Thompson 2022). Breakthrough infections among individuals who have received 2 or 3 doses of the vaccine proceed through the Exposed and Infectious compartments. A meta-analysis of the Omicron variant's R_0 estimates shows considerable variation from 5.5 to 24, with a median of 10 (IQR: 7.25, 11.88; Liu and Rocklöv 2022). To account for a wide range of transmissibility of potential new variants, we draw R_0 uniformly between 2 and 12. All other parameters are held at their median values as above. Similar to the primary simulation, we present results from a simulation with the median draw of all parameters.

1.2 Results

1.2.1 Prioritization strategies for primary vaccine doses

For a given combination of simulation parameters, we identify the vaccine prioritization strategies that results in the fewest number of deaths due to COVID-19 and the fewest number of clinical infections. In the simulation with median parameter values, tiered 65+ roll out reduces deaths by 31.32% (723,866 deaths averted) compared to a no-vaccination scenario. This strategy saves 25.13% (532,567) more lives than prioritizing high contact workers, and 11.49% (206,087) more lives than splitting prioritization between workers and adults 65+. We note that this strategy is only marginally more effective at reducing deaths (0.34%) than prioritizing only adults 65+ before general distribution, indicating that any distribution strategy that gives initial priority to older adults will limit the most deaths. For a tiered 65+ roll out, there is also a modest reduction in clinical infections— 13% fewer (10.9 million infections averted) compared to no vaccination. However, we find that the most effective strategy for limiting clinical infections is through a tiered roll-out that first prioritizes high contact workers. This strategy reduces infections by 13.9% compared to no vaccination (11.5 million clinical infections averted), although we note that the reduction in clinical infections is similar among all prioritization schemes. Overall, we find that strategies that prioritize high contact workers, even when split with adults 65+, do limit clinical infections but fail to confer the lifesaving benefit of strategies prioritizing adults over age 65.

Figure 1.2a shows the percent reduction in deaths and clinical infections relative to no vaccination, across a range of values for the mortality and transmission parameters. Our results are consistent across a wide range of parameter combinations. In 100% of the simulations, the tiered 65+ roll-out was the most effective strategy for limiting deaths due to COVID-19. All prioritization strategies performed remarkably similarly for reducing the most number of clinical infections. In 62.5% of simulations the tiered HC roll-out was the most effective at limiting clinical infections; in the remaining 37.5% of simulations, the most effective strategy appears to be the tiered 65+ roll-out. Through sensitivity analysis, we show below that these differences are driven by variations in R_0.

Figures 1.2b and 1.2c show the trajectories of cumulative clinical infections and deaths averted compared to null for the five prioritization strategies assuming median parameter values. For each vaccine strategy, we define deaths or clinical infections averted as the difference between deaths or clinical infections in the null scenario versus the vaccination scenario.

These results indicate a marked departure from no vaccination between mid January and February. Vaccines reach their peak lifesaving power very quickly within the first month and the relative benefit increases through February. This demonstrates that the timing of vaccines is critical: prioritizing seniors limits the most deaths because they are able to develop vaccine-derived immunity before the peak of the outbreak. Further, tiered roll out strategies show an extended benefit over their single-prioritization counterparts through February and March after distribution of vaccines to the first priority group. Trajectories for each demographic group are shown in the model appendix. Interestingly, when deaths are broken down by age and worker status, the tiered 65+ strategy averts the most deaths only in the adults 65+ group. For all other groups (children 0-18 and adults 18-64), the tiered HC strategy averts the most deaths. However, since deaths are relatively much higher in the adults 65+ group, the tiered 65+ strategy averts the most deaths in the population overall (see Supp Fig A.4).

We conduct additional analyses to test the sensitivity of our results to the choice of simulation parameters. For each simulation, we separately vary R_0 between 1 and 5, μ_{65+} between 1% and 10%, and the proportion of priority vaccinations split between high contact workers and adults 65+ between 0 and 1. All other parameters are kept the same as the baseline. We see that for all values of R_0, the deaths are lowest by prioritizing adults 65+ for vaccination (figure 1.3). However, the strategy for averting the most clinical infections is sensitive to R_0, and a cross-over in the most effective strategy to limit clinical infections occurs around $R_0 = 2.6$. When R_0 is less than 2.6, prioritizing high contact workers results in the fewest clinical infections. With higher values of R_0, prioritizing adults 65+ leads to the fewest clinical infections. With larger values of R_0, the peak of the outbreak happens earlier in simulation time, resulting in more cases before most vaccines are distributed. Since high-contact individuals in a population will be infected earlier in an outbreak (Mossong et al. 2008) and, therefore, removed from the eligible pool of vaccine recipients, the strategies prioritizing high-contact workers are no longer the most effective for reducing clinical infections for high values of R_0. A strategy prioritizing high-contact workers is, therefore, less effective in reducing clinical infections (compared to a low transmission scenario) since susceptibles in that group will be depleted faster and fewer doses will be distributed to them overall. In a higher transmission setting, the most effective strategy for reducing clinical infections becomes the strategy that targets those most at risk of having a clinical infection: in our model, seniors are almost twice as likely to develop clinical symptoms than adults aged 18-64 (69% vs. 36%). Instead, for high values of R_0 prioritizing adults 65+ is most effective both for reducing deaths as well as clinical infections. However, these results are likely dependent on the timing of vaccine introduction and the practicalities of reaching a large enough population for vaccination before the peak of the outbreak.

As expected, the reduction in the number of deaths is highly sensitive to μ (supplemental figure A.7); however, the most effective strategy consistently remains prioritizing adults 65+ for vaccination due to the strong age gradient in mortality. Figure 1.4 shows the effect on total deaths and clinical infections as the proportion of vaccinations given to adults 65+ is varied under the split vaccination scenario. Deaths are always minimized when giving 100% of priority vaccines to adults 65+, and total infections are minimized when priority is given to high contact workers. However, we do find a minimum number of clinical infections occurring when 55% of vaccinations are given to seniors. This non-linearity in clinical infections is a result of the fact that the burden of clinical infections by subgroup is jointly determined by the age-dependent susceptibility to infection (higher amongst HC workers) and probability of symptomatic illness (highest amongst adults 65+).

1.2.2 Prioritization of Booster Doses during a subsequent outbreak

In the simulation conducted with the median parameter draws, prioritization has a smaller effect on relative outcomes for booster doses than the primary doses. A tiered 65+ roll out for boosters not only continues to save the most lives, but in a departure from the primary simulation, also reduces clinical infections the most—although like the primary simulation, the reduction in infections is nearly equal between the three strategies. With median parameter values (R_0: 7; primary vaccine efficacy: 37.5%; booster vaccine efficacy: 81.5%), a tiered 65+ roll-out reduces deaths by 14.23% (393,562) compared to no booster doses. This strategy reduces deaths by 11.5% more than prioritizing only HC workers and 7% more than a tiered HC roll out. Tiered 65+ roll-out reduces infections by 5.8% (6.5 million infections averted); however, all three strategies reduce infections by nearly the same amount (within 1%). We note that in this subsequent outbreak, the population of eligible HC workers awaiting vaccination is considerably smaller than the primary outbreak; while the eligible 65+ population between the two scenarios is similar (33 million adults 65+ are awaiting primary vaccination in the primary outbreak compared to 31 million for booster doses), the population of eligible high contact workers decreases from about 45 million to about 35 million. This is driven by lower uptake rates of primary vaccination among those below age 65 compared to adults 65+. These results are replicated across the 1000 sets of simulated parameters, shown in figure 1.5. Across all 1000 simulations the median percentage reduction in deaths compared to no boosters is greatest during a tiered 65+ roll out. Clinical infections are reduced the most during a tiered 65+ roll out in 49.1% of simulated parameters; in the other 50.9% of simulations, tiered HC roll out was most effective. Similarly to the primary scenario, this variation in the number of clinical infections averted is driven by R_0.

When R_0 is low, HC or Tiered HC prioritization can limit the most clinical infections (figure 1.6). For both deaths and clinical infections the difference between all distribution strategies is smaller when R_0 is higher (figure 1.6).

1.3 Discussion

While previous modeling studies have evaluated COVID-19 vaccine distribution strategies Matrajt et al. 2021, Foy et al. 2021; see (Saadi et al. 2021 for a review for 43 such studies), very few had considered heterogeneity in contact rates by both age and worker status. We show, using uniquely detailed contact data, that even when we account for high rates of contact among a sub-group, prioritizing vaccination of the oldest age group averts the most deaths from COVID-19. However, we also find that prioritizing high-transmission groups can limit the spread of disease.

We find that the most effective strategy, given our model parameters, for reducing deaths due to COVID-19 via vaccination was a tiered roll out that first prioritized adults over age 65 followed by high contact workers, with remaining doses split among low contact adults 18-64 and children (Tiered 65+ strategy). However, the most clinical infections are averted through a tiered roll out that prioritized high contact workers followed by adults over age 65 before general distribution (Tiered HC strategy).

The most effective strategy for distribution of booster doses in our model to limit deaths is similar to the primary scenario—it is most effective to prioritize seniors for reducing mortality—and this strategy may also reduce the number of clinical infections the most depending on the transmissibility of the novel variant strain. Without an inflow of susceptibles from births or waning immunity in our model, this difference for booster dose prioritization strategies is driven almost entirely by the lower uptake rate for primary and booster doses by 18-64 year olds compare to adults 65+. However, we note that all three distribution strategies for booster doses result in nearly equal reductions in clinical infections.

These simulations were designed to test the total effect of vaccination—both the direct benefits of vaccination on the prioritized group and the indirect effects of reducing community spread through social contact patterns—using empirical estimates of contact rates across groups. When distributing initial doses, the indirect effects of vaccinating high-risk workers or splitting vaccines between workers and seniors did not confer a greater reduction in deaths than the direct effects of prioritizing all seniors but did reduce the spread of COVID-19 overall.

While limiting deaths is often the primary public health objective, other priorities can include limiting total infections or reducing the number of infected persons at a given time. A healthcare system that is stressed beyond the critical care capacity of ER/ICU admissions may see both increased COVID-19 case fatality and excess secondary deaths (Phua et al. 2020, Miura et al. 2021, and Wood et al. 2020. Prioritizing high contact workers with in-person employment may alleviate strain on the healthcare system and reduce the number of people living with chronic COVID-19 symptoms. In a real-world setting, there is undoubtedly a benefit to prioritizing some essential workers in healthcare or eldercare settings who have more exposure to high-mortality populations; substantial heterogeneity exists even within our surveyed high risk workers. Results of our 'split' analysis show that when some vaccines for adults 65+ are diverted to high risk workers, fewer deaths are averted, but the total number of clinical infections drops sharply. The most lives are saved when adults 65+ are prioritized, which is consistent with early CDC guidance that prioritized the elderly and a limited number of healthcare workers first.

Public health objectives and the most effective strategy may also depend on the timing of vaccine distribution (Han et al. 2021). For example,Mylius et al. 2008 find that for pandemic influenza, prioritizing the oldest age group is most effective at limiting deaths, but only if distribution begins near the peak of the outbreak; otherwise, prioritizing young adults and schoolchildren for early doses is most effective. Substantial differences in the age-specific contact, mortality, and vaccine uptake indicate that the most beneficial early-target groups for vaccination are considerably different between COVID-19 and influenza (Fitzpatrick and Galvani 2021). Understanding how the effectiveness of strategies may change over the course of the COVID-19 pandemic is a key direction of future research.

Our results also highlight the significant impact of imperfect vaccines and imperfect vaccine uptake on clinical infections and deaths. We show that the potential disease-averting power of vaccines is lost due to delays in vaccine availability for the general population, as many susceptibles awaiting the vaccine are infected before they are able to be vaccinated (figure A.8). High-contact groups are more likely to be infected earlier in the outbreak, thereby reducing the size of the population eligible for vaccination. This, in combination with low vaccine uptake, reduces the effectiveness of vaccination strategies in high transmission scenarios. This is especially true for the high contact group, who are more likely to be infected earlier on in the outbreak before vaccination is fully rolled out.

Our study has several limitations. First, although our population is faceted by age and employment, the model does not consider other non-pharmaceutical interventions, such as the use of face masks, that can reduce transmission during interpersonal contact. Brief contact between high-risk workers and customers that are masked and distanced may be ultimately insignificant compared to adults lacking such contacts but who are engaging in risky behavior during personal time. Quantifying overall risk along other dimensions of contact, such as whether the contact took place indoors out outdoors, the duration of contact, whether a mask was worn, etc., is an important direction for future research. Second, we are limited — by the survey sample size — to using only coarse age and contact status categories in our contact matrix. By including all respondents and their contacts over age 65 in our oldest age group and using the average infection fatality ratio, our analysis may mask important heterogeneity by age in mortality (Levin et al. 2020). Further, our analysis utilizes POLYMOD data, which were collected in the United Kingdom before the onset of the COVID-19 pandemic, for inferring contact patterns for the youngest age group. POLYMOD respondents are not necessarily representative of contact patterns in the United States during our period of interest. Given the relatively low rates of clinical infections and deaths in the youngest age group, any bias in the analyses is likely to be small.

Third, for simplicity and due to lack of data for parameterization, our model does not include waning of immunity from natural infection or from vaccination, although this presents a substantial challenge to the control of COVID-19. Additionally, our implementation of 'leaky' vaccines does not account for reduced symptom severity or reduced probability of transmitting COVID-19. The natural history and epidemiology of future variants may change the landscape of mortality, virulence, and interaction with vaccines. While our analysis is limited in scope to the conditions from the first quarter of 2021, extending this model to the future will need to consider how waning and partial immunity may affect the most effective vaccination strategies for reducing deaths or clinical infections.

Finally, this analysis considers prioritization only by age and worker status and does not include heterogeneity within or between groups. Structural barriers in health care, including financial cost, health complications, and discrimination within the healthcare system—if related to essential worker status—may mean that essential workers face increased health risks in addition to increased contact network size.

There are also likely to be benefits to prioritizing people for vaccination along other socio-demographic axes of risk that are not considered here—especially race, ethnicity, and geographic location—by identifying social inequality itself as a driver of outbreaks (Wrigley-Field et al. 2021; Link and Phelan 1995). The BICS survey population reflects the composition of the essential workforce in the United States — Blacks and Hispanics comprise of a larger share of the high-contact population (11.12% and 22.49% respectively) relative to their shares in 65+ population (9.12% and 5.04% respectively). Nonetheless, given the strong age gradient in mortality due to COVID-19, the tiered 65+ prioritization strategy averts the most deaths among respondents of all ethnicities (see supplement A.1.10), assuming that transmission and mortality are equal by ethnicity and gender within our defined demographic groups (children, HC adults, LC adults, adults 65+). Unfortunately, due to limitations in the sample size we are unable to construct a contact matrix that is disaggregated by age, employment, and race or ethnicity, and thus unable to fully account for heterogeneities across these groups. Although the results of our simulations indicate that relying on indirect effects of prioritizing high-contact workers for vaccination does not ultimately save more lives, high-contact workers may be at increased risk of death compared to other adults because of socio-demographic disadvantage within with the healthcare system not captured within our model. Our analysis supports prioritization of vaccination for the highest risk members of society—which we identify through advanced age, but may instead be a complex combination of socio-demographic disadvantage. This is an important direction for future research.

Overall our results highlight the impact of two dimensions of risk — contact behavior and age-dependent susceptibility and risk of severe disease— on the effect of vaccination strategies. This work also shows the utility of combining mathematical models with detailed contact data by socio-demographic groups, particularly for understanding sensitivity of the effectiveness of different prioritization strategies to the epidemiology of the circulating virus strain. As novel strains are likely to continue to emerge (Brüssow 2022), these data and models can be especially useful for tailoring prioritization strategies for subsequent outbreaks.

Replication Code

All analyses was conducted using R software (version R version 4.0.2). Replication code is publicly available at `https://github.com/eroubenoff/BICS_employment_replication_code` (Roubenoff, Feehan, and Mahmud 2022a).

Data availability

We have deposited our data in the Harvard Dataverse, `https://doi.org/10.7910/DVN/K8YPVZ` (Roubenoff, Feehan, and Mahmud 2022b).

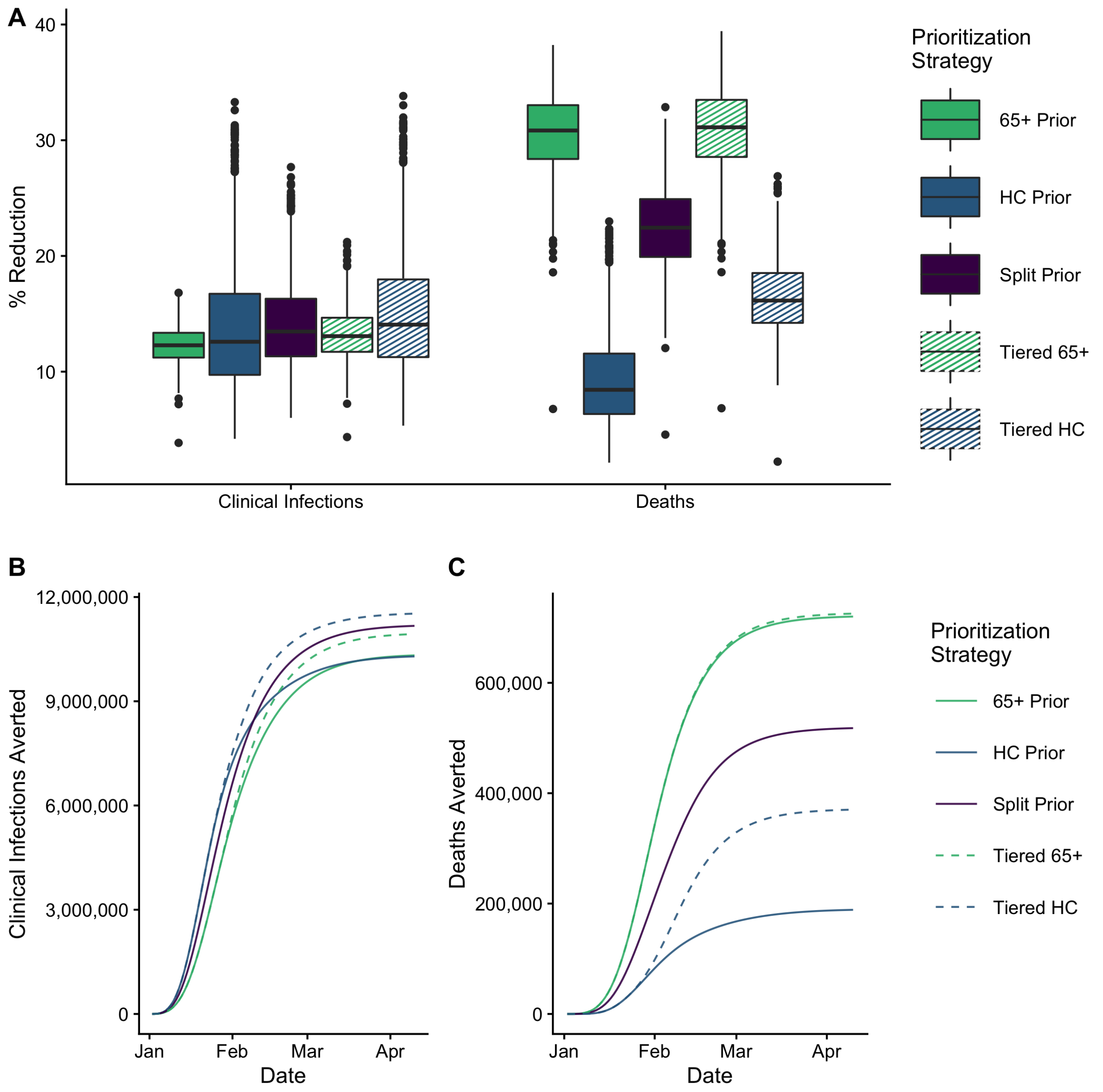


Figure 1.2: (A) For primary vaccination, percent reduction in clinical infections and deaths when compared to no vaccination for randomly drawn transmission parameters. The median percent reduction in deaths was highest in a tiered strategy that prioritizing seniors and lowest when only prioritizing contact risk workers; clinical infections are reduced the most by a tiered system that prioritizes workers first, although all strategies produce similar results. (B) and (C): For baseline parameters, trajectories of daily cumulative clinical infections (B) and deaths (C) averted relative to a no vaccination scenario, calculated as the cumulative difference between each strategy and null through each date. For each vaccine strategy, we define deaths or clinical infections averted as the difference between deaths or clinical infections in the null scenario versus the vaccination scenario. When prioritizing seniors the reduction in deaths begins nearly immediately, whereas the indirect benefit from prioritizing HC workers begins later and is lower in magnitude. The opposite is observed for clinical infections.

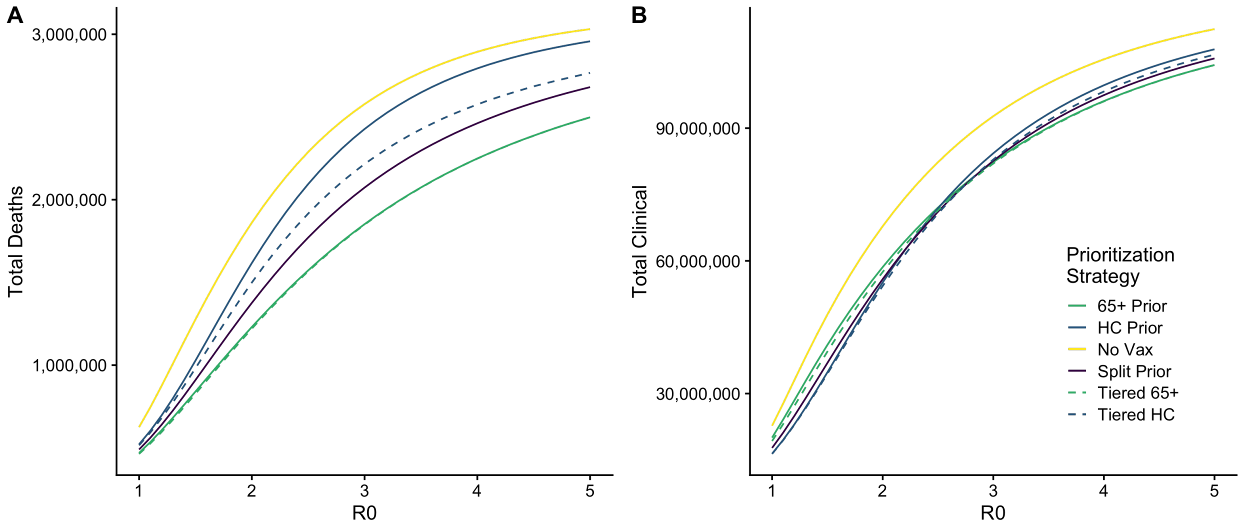


Figure 1.3: Results of the sensitivity analysis of R_0 (varied linearly between 1 and 5 in 0.1 increments) and the resultant count of total deaths (A) and clinical infections (B). Across all values, a strategy that prioritizes adults 65+ either directly or in a tiered roll out limits the most deaths. At low values of R0 (<2.5) the reduction in clinical infections is greatest in a 'Tiered HC' roll out; however, the most effective strategy with high R0 is through a Tiered 65+ strategy.

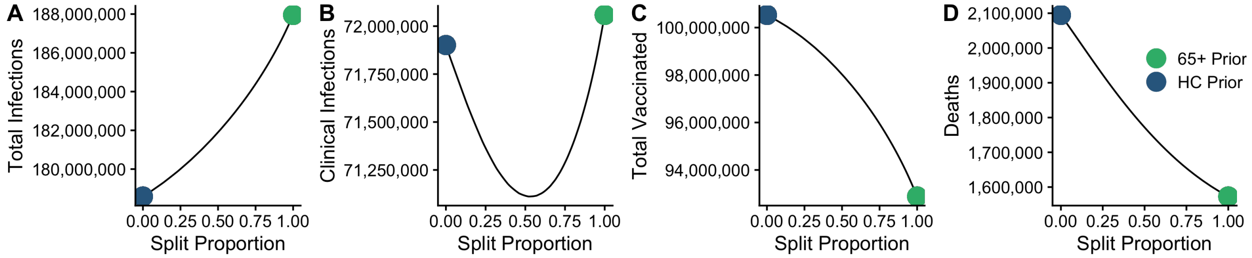


Figure 1.4: The proportion of vaccines split between seniors and HC workers is varied from 0% to 100% and shown with counts of (A) total infections, (B) clinical infections, (C) total vaccinated, and (D) total deaths. Extremes (corresponding to the HC Prior and 65+ Prior strategies) are shown. When priority vaccines are given more to adults 65+, deaths are minimized but total infections are maximized, indicating that while this strategy limits deaths it fails to limit transmission effectively. Additionally, more susceptibles are eligible for vaccination under this strategy. However, the minimum number of clinical infections is minimized when 59% of vaccines are distributed to 65+. This effect is mediated by increased susceptibility to infection and increased probability of symptomatic infection among seniors, and the increased priority group size among HC workers.

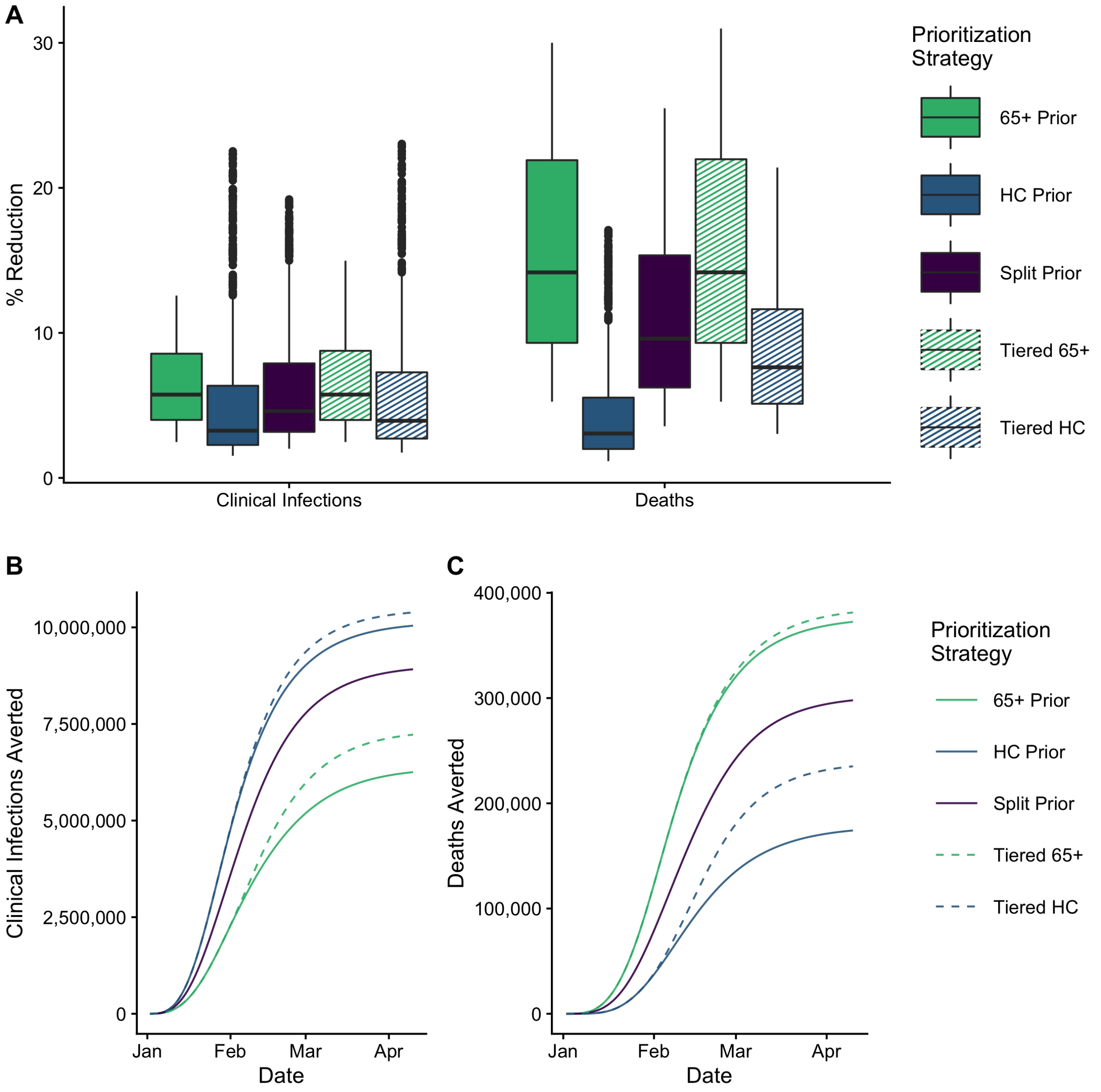


Figure 1.5: For distribution of booster doses: (A) percent reduction in clinical infections and deaths when compared to no vaccination for randomly drawn transmission parameters. (B) and (C): For baseline parameters, counts of clinical infections (B) and deaths (C) averted relative to a no vaccination scenario. For each vaccine strategy, we define deaths or clinical infections averted as the difference between deaths or clinical infections in the null scenario versus the vaccination scenario.

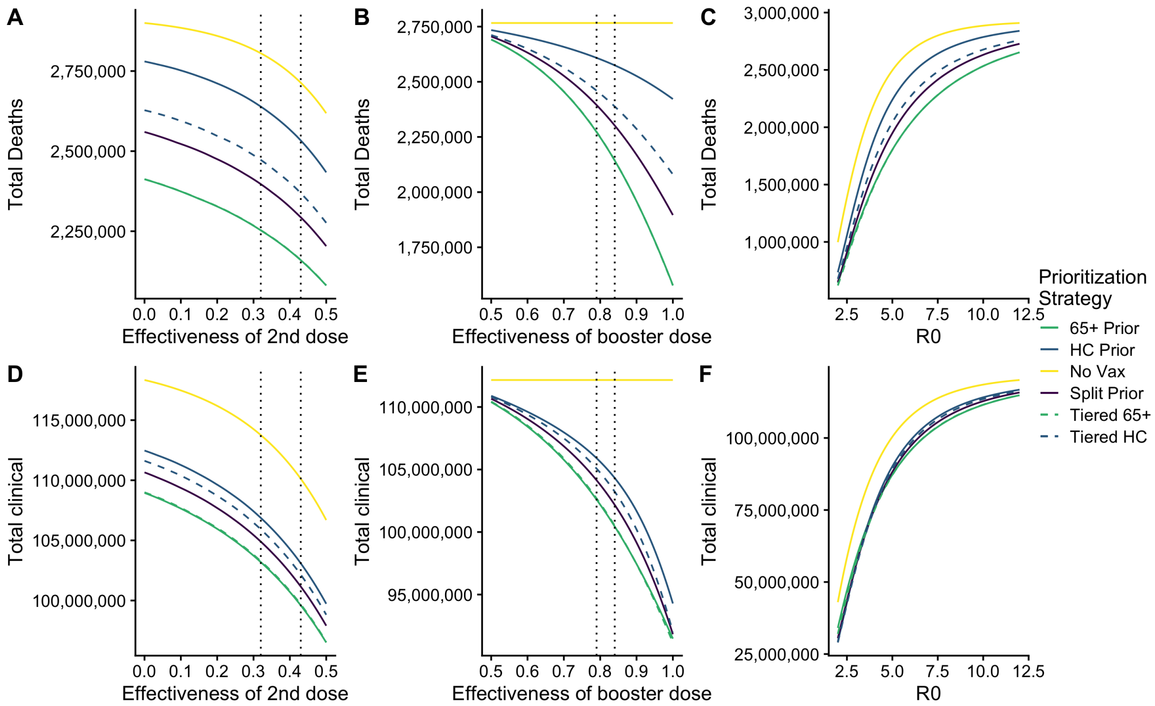


Figure 1.6: Relationship of stochastically drawn parameters in assessing the effect of booster dose prioritization.

Chapter 2

How will COVID-19 persist in the future? Simulating future dynamics of COVID-19 using an agent-based network model

Abstract

Despite the United States Center for Disease Control (CDC)'s May 2023 expiration of the declared public health emergency pertaining to the COVID-19 pandemic (Silk 2023), approximately 3 years after the first cases of SARS-CoV-2 appeared in the United Sates, thousands of new cases persist daily. Many questions persist about the future dynamics of SARS-CoV-2's in the United States, including: will COVID continue to circulate as a seasonal disease like influenza, and will annual vaccinations be required to prevent outbreaks? In response, we present an Agent Based Networked Simulation of COVID-19 transmission to evaluate recurrent future outbreaks of the disease, accounting for contact heterogeneity and waning vaccine-derived and natural immunity. Our model is parameterized with data collected as part of the Berkeley Interpersonal Contact Survey (BICS; Feehan and Mahmud 2021) and is used to simulate time series of confirmed cases of and deaths due to SARS-CoV-2, paying special attention to seasonal forces and waning immunity (Kronfeld-Schor et al. 2021; Liu et al. 2021b; Nichols et al. 2021). From the BICS ABM model we simulate SARS-CoV-2 dynamics over the 10-year period beginning in 2021 with waning immunity and inclusion of annual booster doses under a variety of transmission scenarios. We find that SARS-CoV-2 outbreaks are likely to occur frequently, and that distribution of booster doses during certain times of the year—notably in the late winter/early spring—may reduce the severity of a wintertime outbreak depending on the seasonal epidemiology of the pathogen.

2.1 Introduction

Three years after the first cases of SARS-CoV-2—the pathogen responsible for the COVID-19 pandemic—appeared in the United States, many control measures put in place during the early phase of the pandemic have been eliminated (Silk 2023) in favor of a desire to return to 'business as usual'. This includes mask mandates, shelter-in-place and work from home ordinances, physical distancing guidelines, and recommendations to isolate symptomatic cases. While vaccines for COVID-19 were a source of optimism through early 2021, it became clear over the subsequent months that waning natural and vaccine-derived immunity and the pathogen's immunity-evading mutations rendered the vaccines less effective at ending the pandemic than initially hoped (Levin et al. 2021,Centers for Disease Control and Prevention 2021). Additionally, uptake of booster doses has lagged far behind targets (National Center for Immunization and Respiratory Diseases (NCIRD) 2022a; National Center for Immunization and Respiratory Diseases (NCIRD) 2022b). Over time, 'pandemic fatigue' has set in as compliance with disease-preventing behavioral mandates, especially mask usage and contact limitation, has slipped and such ordinances have been lifted (Reicher and Drury 2021). The current phase of the pandemic is substantially different from the initial phase, characterized by a highly transmissible but less severe form of the illness owing in part to many factors acting in different directions, including: highly transmissible but less severe later variants (Davies et al. 2021; Strasser et al. 2022; Yang et al. 2022), higher levels of partial or full immunity from vaccine or prior infection (Clarke 2022), higher levels of social contacts approaching pre-pandemic levels, low or absent rates of mask usage and physical distancing (Crane et al. 2021), and better treatments reducing the probability of death or severe illness after infection (National Institutes of Health 2022). While better treatments, milder variants, and prior immunity has resulted in a far lower case fatality ratio than the early days of the pandemic, avoiding the illness is still a persistent challenge for those who remain at elevated risk of severe illness and death due to COVID-19, including the elderly and people with chronic health conditions.

Understanding the various drivers of future SARS-CoV-2 outbreaks can help to plan for future interventional strategies, including vaccine distribution, school or work closure, and other non-pharmaceutical interventions. Infectious disease models can help to plan for these future outbreak scenarios by helping understand how SARS-CoV-2 will exist in our medium to long-term future. Here, we consider the effect of recurrent outbreaks, annual distribution of booster doses, and seasonal change in transmission of SARS-CoV-2. It is presently unknown if SARS-CoV-2 will demonstrate a strong seasonal pattern; however, it is hypothesized to follow the seasonal patterns of influzena and other coronaviruses (Kronfeld-Schor et al. 2021).

To answer these questions, we use a stochastic Agent-Based Network Model paramaterized with contact data from the Berkeley Interpersonal Contact Survey (BICS; Feehan and Mahmud 2021). Using BICS data allows us to consider how contact heterogeneity, household structure, and other network dynamics play into the periodicity and size of future outbreaks. Our model also includes seasonal forcing of transmission parameters, waning immunity from vaccines and prior infection, and variable-rate case importation to capture interaction with counter-seasonal populations (i.e., travel between the hemispheres experiencing opposite seasons).

Agent-Based Models are alternative to compartmental models and allow for more flexible and dynamic transmission dynamics, including network structure (Ajelli et al. 2010; Bansal et al. 2010). Roubenoff, Feehan, and Mahmud 2023, included in the first chapter of this dissertation, utilized a compartmental model for analyzing SARS-CoV-2 transmission. While these types of models are heavily utilized for analyzing disease transmission, one particular limitation of these models with contact data is their requirement for relatively few and well-defined categorizations. ABMs are more flexible and contrast with compartmental models by keeping track of the disease status for each individual in the simulation, rather than the tally of individuals in a particular compartment. In place of differential equations describing flows between compartments, ABMs use highly explicit (either deterministic or stochastic) rules governing interaction (Bonabeau 2002; Bruch and Atwell 2015; Baker et al. 2013; He, Ionides, and King 2010). Interactions are governed by some aspect of the simulated population contained within an objective function, such that interactions between nodes of certain values are more likely that others. As a group, ABMs are free of many of the analytic requirements of compartmental models—especially the need for of explicit transition properties between states, only an objective function for optimization—earning them the descriptor 'plug and play' (He, Ionides, and King 2010). Importantly for our purposes, ABMs allow for flexibility in how the population mixes, allowing for contact inequality between simulated agents through either a spatial or network component. Network models are a particular type of agent-based model that assume an explicit network structure for disease-transmitting contacts. In network models, instead of a homogenous or even matrix-structured contact pattern employed by compartmental models, disease transmission is simulated as occurring over a graph representing network connections (Bansal, Grenfell, and Meyers 2007; Danon et al. 2011; Keeling and Eames 2005).

A plethora of agent-based and network simulations of the COVID-19 pandemic have been published and tools released. The flexible yet highly specific ways that ABMs can be used to model social interactions is ideal for testing network and behavioral interventions for COVID-19. These models include a wide set of techniques, including models that explicitly account for how individuals navigate geographic space (Cuevas 2020) and social networks (Hunter and Kelleher 2021). Agent-based models have used to estimate parameters for the COVID-19 outbreak in France (Hoertel et al. 2020), Ireland (Hunter and Kelleher 2021), and Colombia (Gomez et al. 2021). Models can utilize existing contact data, such as Moghadas et al. 2021 and Sah et al. 2021, who use data from the POLYMOD survey (Mossong et al. 2008) as inputs to an ABM to evaluate willingness to vaccinate. Holmdahl et al. 2020 test a series of behavioral interventions in nursing homes using a two-cohort ABM (patients and caregivers), finding that testing frequency and isolation are the most effective ways to limit the spread of disease. We draw on a number of general ABMs developed for COVID-19, including Covasim (Kerr et al. 2021), an aspatial model that combines Erdos-Reyni Poisson random networks with SynthPops networks that are generated from empirical contact data, and OpenABM (Hinch et al. 2021), that simulates social network interaction through stochastic network simulation at the household, occupational, and random connectivity additively.

Although it remains to be seen, research has suggested that SARS-CoV-2 may exhibit seasonality similarly to influenza and other coronaviruses, which exhibit higher incidence in the colder months (Nichols et al. 2021). Indeed, in the United States, the highest number of cases were observed in winters 2020-2021 and 2021-2022. The periodicity and severity of future SARS-CoV-2 outbreaks is currently unknown, largely since the rate of mutations and long-term vaccine-derived and natural immunity is unknown, but many mechansims are theorized (Kronfeld-Schor et al. 2021). Seasonal forcing of respiratory diseases involves a consideration of multiple temporal factors relevant to modeling the transmission of SARS-CoV-2 including seasonal changes in host behavior and immune function (Altizer et al. 2006; Grassly and Fraser 2006). Although modeling studies suggest that climate may mediate the timing and peak incidence of SARS-CoV-2 outbreaks, susceptible supply driven by population immunity is the primary driver of such dynamics (Baker et al. 2020). Many childhood diseases, namely measles, exhibit seasonal cycles driven by the birth rate and increased contact between non-immune children during the school year (Metcalf et al. 2009). In some situations, it can be assumed that the pathogen is circulating at a very low level locally in between outbreaks; in others, like influenza, seasonal human-animal interactions and case importation between populations in opposite hemispheres experiencing counter-cyclical temperature-forced outbreaks may drive timing (Lofgren et al. 2007; Lowen and Steel 2014).

We use a stochastic agent-based network simulation of SARS-CoV-2 transmission parameterized with data from the Berkeley Interpersonal Contact Survey (Feehan and Mahmud 2021) and include seasonality, annual vaccination, waning immunity, and demography. To our knowledge, this represents the first agent-based model for examining COVID-19 endemic outbreak cycles and seasonality. We find that outbreaks are likely to occur regularly and that annually-distributed booster doses can be an effective tool to eliminate regular outbreaks. Depending on seasonal epidemiology of the pathogen, booster doses are most effective when distributed at certain times of year; in the absence of seasonality, booster doses are most effective when distributed in the first half of the year, but in a seasonally-forced transmission scenario distributing vaccines in early fall is more successful at eliminating major annual outbreaks.

2.2 Methods

2.2.1 Data

Like Roubenoff, Feehan, and Mahmud 2023 (included in Chapter 1 of this dissertation), we also utilize here contact survey data collected by Feehan and Mahmud 2021 as part of the ongoing Berkeley Interdisciplinary Contact Survey (BICS), which captures disease-transmitting behavior during the COVID-19 pandemic. The BICS survey, collected in several waves beginning in March 2020, is an online survey aimed at capturing the frequency and nature of respondents' face-to-face contact over a 24-hour period. Respondents to the BICS survey are recruited through a quota sample using an online survey panel provider, Lucid.

Respondents are asked to report the total number of close, face-to-face contacts they had over the previous 24 hours, and to elaborate on three such contacts in detail. Respondents are also asked to report information regarding their demographic information, household structure, and other questions regarding their behavior. In this chapter, we utilize responses from Wave 6 (n = 5418, 12 May - 25 May 2021) of the BICS national (U.S.) survey to capture post vaccination contact patterns.

2.2.2 Model: BICS ABM

Using Hunter, Mac Namee, and Kelleher 2017's taxonomy for categorizing Agent Based Models, the simulation model used here is disease-specific to COVID-19 and society-specific to the behaviours captured from respondents in the BICS sample frame. Behavior is modeled on networks and is without transportation and without environment. The BICS ABM simulation population is constructed of individuals (also referred to as agents, nodes, or vertices) within households. We simulate interaction and disease spread among a population of 1000 households (approx. 3200 individuals) representative of the U.S. according to the procedure described below and in the model supplement. Each agent in the simulation directly corresponds to a respondent in the BICS or POLYMOD surveys sampled with survey weights to match the distribution of age and sex of the US population, and the agents' demographic and behavioral data is derived from the corresponding survey respondent. The simulation includes three types of social contacts: household contacts with household members, school contacts for children below the age of 18, and non-household 'random' contacts. As employment data are not available for this wave of the survey, 'random' contacts are designed to include employment contacts for adults. Household contacts and school contacts are drawn randomly at the start of the simulation according to the procedure described below and are maintained throughout the simulation; random draws of graphs representing random non-household contacts are taken during each daytime time step. In this way, the total network is *dynamic* as it changes throughout the course of the simulation.

The graph of household contacts is drawn according to the procedure described in the model supplement, which is similar to the SynthPops procedure utilized in COVASIM (Kerr et al. 2021; Mistry et al. 2021). Briefly, a supplied number of households are created with the following two-step procedure. First, BICS survey respondents are repeatedly sampled with replacement and adjustment for survey weights to be heads of household. Heads of household are chosen to match the age- and sex-distribution of adults in the United States using 2021 American Community Survey estimates (US Census Bureau 2022). Then, households are filled by sampling BICS respondents (again, with replacement and adjustment for survey weights) who match the household head's reported household members' age and gender, until each household is the proper size. Respondents under the age of 18 were not ascertained in the BICS survey; instead, they are sampled uniformly from the POLYMOD UK survey (Mossong et al. 2008). Throughout the simulation, each node's behavior is derived from the corresponding BICS survey respondent's responses; nodes derived from POLYMOD respondents are derived from the corresponding fields in the POLYMOD survey[1].

The model progresses in hourly steps through simulation time. During morning and evening hours (6pm-8am), agents have contact with all members of their household. During daytime hours, random connections occur between members of the simulation population governed by the degree (number of non-household contacts) supplied by each survey respondent. A daily contact network is drawn using the Network Configuration Model, described below. Each such contact is designated to begin at a random time during the day chosen uniformly between 8am and 6pm. Each contact has a randomly chosen duration sampled according to the following probabilities: respondents to wave 6 of the BICS survey indicated that 17.1% of contacts were less than one minute, 45.2% were between 1 minute and 15 minutes, 18.7% were less than one hour, and 18.9% were more than 1 hour. Here, we choose to use the marginal distribution rather than individual-level responses due to computational limitations. During the duration of each contact, respondents are disconnected from other members of their household and reconnected after the conclusion of the random contact. If a node is clinically infectious, they may enter isolation for the duration of symptoms; if asymptomatic, they continue to mix as before. Isolation is incorporated by a parameter representing the multiplicative reduction in random daily contacts: for example, an isolation parameter of 0.1 means that a node normally with 10 daily non-household contacts with would have 1 such contact while in isolation.

Random contacts are drawn using the Network Configuration Model, which generates a random graph of contacts that preserves each node's degree—here, the number of daily non-household contacts. The network configuration model creates random networks using only a provided degree distribution as an input. The configuration model works through a two-step procedure (Albert-László Barabási 2021):

1. First, assign a degree to each node in the network such that the distribution of degrees matches the desired distribution. For each node, assign a number of 'stubs' or half-edges equal to the degree of that node.

2. Second, randomly and uniformly join stubs to create edges until there are no stubs remaining in the network.

Without alteration, the model may produce self-edges (a node connected to itself), or multi-edges (multiple edges connecting a pair of nodes). Sampling 'simple' graphs that lack self- or multi-edges is a computationally intensive procedure and non-uniform in its graph-generating process. In our application, we continue to sample graphs uniformly and remove self-edges, but maintain multi-edges. Realizations of this type are likely to have total degree less than would be implied by the supplied degree distribution.

---

[1] Some fields collected as part of the BICS survey are not available in the POLYMOD data and are imputed. For the present application, only mask/NPI usage is unavailable, and is taken to be the False.

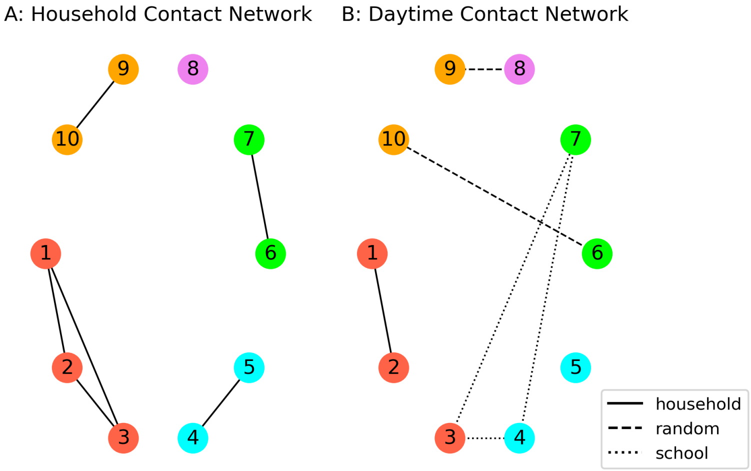


Figure 2.1: Illustration of network structures used during simulation. (A): household contact network representing evening and morning contacts, and (B): daytime contact network, consisting of school contacts and randomly drawn contacts. While school contacts are maintained throughout the simulation (with the exception of summertime school closures), random contacts are re-drawn hourly.

2.2.3 Transmission of SARS-CoV-2

All agents begin susceptible and vaccine roll-out begins at the beginning of simulation time. At a given time $T0$ a supplied number of index cases are chosen randomly to be exposed to SARS-CoV-2. In addition, a vector representing the number of cases imported daily is provided as input to the simulation—in the present application, one case weekly is imported to ensure that SARS-CoV-2 is constantly circulating at a low level. At exposure, each agent is assigned a randomly drawn number of hours spent as exposed and infectious; they then proceed to either symptomatic or asymptomatic infection with a supplied probability. Baseline probability of transmission—before considering vaccine efficacy, contact duration, non-pharmaceutical interventions, and asymptomatic reduction in transmisison probability—from an infected node to a susceptible node occurs with probability $\beta(t)$, where t is the simulation's tth day in the year. The value of $\beta(t)$ thus represents the probabiltiy of transmission during an hour-long contact between one clinically symptomatic node and one susceptible, unvaccinated node without mask usage.

Various factors multiplicatively reduce the probability of transmission. First, transmission is reduced proportional to the duration of contact in fractions of an hour; a 15-minute contact is $1/4$-th as likely to result in transmission as an hour-long contact. As well, transmission from an asymptomatic node to a susceptible node occurs at a reduced probability α relative to symptomatic nodes. The susceptible node's vaccination status reduces the probability of transmission by the corresponding vaccine efficacy; the infectious node's vaccination status is not assumed to affect transmission probability. As detailed further below, we allow for reinfection after a set amount of time; previous infection offers protection for the recovered node as a reduction in the transmission probability. Finally, the model includes a single Non-Pharmaceutical Intervention (NPI), designed to capture the combined disease-blocking effectiveness of masks, physical distancing, and other preventative measures. If BICS respondents corresponding to both nodes in a random contact report any mask usage, the probability of transmission is proportionally reduced by a supplied effectiveness. If NPI effectiveness is set to 0, the simulation is effectively in the absence of NPI usage. We assume that NPIs are not used among household contacts. At the conclusion of the infectious period, asymptomatic nodes all recover; symptomatic nodes die with a supplied age-based probability.

Unlike compartmental models that often hinge on parameter R_0—the basic reproduction ratio, representing the average number of secondary cases caused by an index case in a fully susceptible population, estimated as the product of the transmission probability, contact rate, and duration of infectiousness—no such closed form solution for R_0 in an ABM necessarily exists. Although the ABM's ability to model contact and transmission through separate processes and objective functions allows for for increased flexibility, including time-variable and stochastic transmission probability, heterogeneous contact rates and network structure, variable duration of contact, isolation of infectious cases, and uneven NPI and vaccination usage, the parameters that govern the overall transmission dynamics are not well-defined in closed form. Instead, the R_0 must be estimated from the the contact rate or incidence curve generated by the simulation itself (Hunter, Namee, and Kelleher 2018; Venkatramanan et al. 2018; Hunter and Kelleher 2021; Hoertel et al. 2020). We estimate R_0 by observing the hourly contact rates $\hat{c}_h$ in the initial 7 days of the simulation, using the expression:

$$\hat{c}_h = \sum_{e \in E_h} (Duration_e \cdot NPI_e) \tag{2.1}$$

which estimates the average hourly number of edges weighted by the duration and effectiveness of non-pharmaceutical interventions, if used during each contact. Then, the average contact rate $\bar{c}$ is taken to be the average of the hourly contact rates $\hat{c}_h$. Finally, R_0 is taken to be:

$$R_0 = \bar{\beta} \cdot \bar{d} \cdot \frac{\bar{c}}{N} \cdot [\rho + (1-\rho)\alpha] \tag{2.2}$$

where parameter $\bar{\beta}$ is the average transmission probability over the course of the simulation period, $\bar{d}$ is the average duration of infectiousness, N is the size of the population, and the expression $\rho + (1-\rho)\alpha$ represents the reduction in infectiousness among subclinical cases. Since R_0 is not necessarily known until the conclusion of the simulation, we instead determine the overall transmission rate by varying the transmission probably β, then estimating R_0 post-facto.

2.2.4 Vaccine effectiveness, waned immunity, and reinfection

An important feature of the model is waning natural and vaccine-derived immunity. Vaccination occurs uniformly in the population assuming that all eligible members of the population

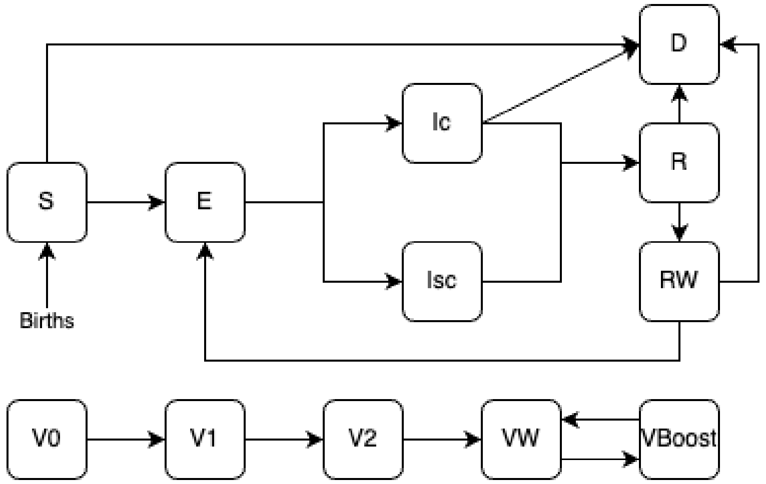


Figure 2.2: Schematic for disease states in the Agent-Based Model used in the present simulation, including disease status (top) and vaccination status (bottom). Disease states include susceptible (S), exposed/pre-infectious (E), clinically infectious (Ic), subclinically infectious (Isc), recovered (R), recovered with waned immunity (RW), and deceased (D). Vaccination states include unvaccinated (V0), first and second primary doses (V1, V2), waned immunity (VW), and boosted (VBoost).

are equally able to access the vaccine (differing from the prioritization procedure taken in Chapter 1), at a baseline 70% uptake rate. A set number of vaccine doses are available daily, and are distributed at 8am each day according to optional priority rules. Nodes are eligible for a second dose of the vaccine after 25 days. After a supplied amount of time, immunity wanes, and nodes are eligible for booster doses; booster doses are made available at a set date every year and are distributed with the same rate and priority schedule as primary doses. Only nodes with waned immunity are eligible for booster doses. All vaccination statuses (1st dose, 2nd dose, waned, and booster) have fixed proportional reductions in transmission probability.

As well, we include reinfection in the model with a similar procedure. Nodes that have recovered from infection remain completely protected from infection for a fixed amount of time; after this, nodes are assigned disease status 'Recovered-Waned' indicating that they may be re-infected, yet have some protection against future infection. Waned immunity is assumed to be the same for clinical and subclinical infections, and the protection offered does not depend on the number of previous infections.

At present, the duration of immunity after infection and vaccination is not known, but is estimated to be approximately 6 months of near-complete immunity followed by a steady decrease over time (Centers for Disease Control and Prevention 2021). This is implemented in our model with a pair of parameters: first, the duration of complete immunity, governing the time after infection or vaccination that an individual experiences the full effect of vaccine-derived or natural immunity; second, the wanted immunity effectiveness. For simplicity, these parameters are held to be the same for both vaccine-derived and post-infectious immunity.

2.2.5 Incorporation of Seasonality

Seasonal environmental changes are known to affect infectious disease transmission in predictable, annually-recurrent cycles. Although seasonality is well documented in many infectious diseases, the underlying mechanisms are frequently poorly understood or difficult to tease out from other compounding effects (Fisman 2012). For respiratory infections transmitted between humans via the airborne pathway such as SARS-CoV-2 and influenza, seasonal effects can be grouped in three broad areas: environmentally-driven changes in host behavior, such as summertime school closings or increased wintertime indoor gatherings; the pathogen's ability to survive outside of the human host adapted to certain climatic conditions, in turn affecting fitness for transmission; and seasonal changes in the host's immune response, possibly due to changes in temperature or sunlight exposure (Altizer et al. 2006; Grassly and Fraser 2006; Dowell 2001; Kronfeld-Schor et al. 2021; Buonomo, Chitnis, and d'Onofrio 2018; Held and Paul 2012). Additionally, seasonal migration—especially temporary domestic migration with an annual cyclical pattern—may fundamentally change the landscape of interactions and population at risk (Buckee, Tatem, and Metcalf 2017).

Incorporating seasonality into a compartmental model is typically done by adding sinusoidal temporal forcing to the transmission parameter β as $\beta(t) = \beta_0(1+\beta_1 cos(2\pi t))$, through a binary indicator in the case of seasonal school closings as $\beta(t) = \beta_0(1+\beta_1 \text{term}(t))$, or other time-dependent functional form (Grassly and Fraser 2006; Keeling, Rohani, and Grenfell 2001). Here, the basic reproductive number R_0 represents the average number of secondary cases from a single index case introduced at a random time of the year, and is defined as $\hat{R}_0 = D \int \beta(t) dt$ where D is the average duration of infection. Forcing functions can be easily extended to include age-structured contact (Bolker and Grenfell 1993), and time-series methods allow for modeling of outbreak dynamics without fitting a functional form to the transmission parameter (Metcalf et al. 2009; Finkenstädt and Grenfell 2000). Extending beyond compartmental models, seasonality can be incorporated into modeling of incidence data; Held and Paul (2012) demonstrate how seasonal incidence can be decomposed into an endemic and epidemic component with independent temporal structure.

The long-term seasonal patterns and drivers of COVID-19 are still unknown and not necessarily possible to disentangle from control efforts, especially noting non-pharmaceutical interventions like shelter-in-place ordinances and mask usage. Weaver et al. 2022 note in a review that COVID-19 may be more stable and more transmissible in cooler environments, consistent with influenza[2], although both stability in high humidity and low humidity have been observed (Morris et al. 2021; Marr et al. 2019; Matson et al. 2020; Dabisch et al. 2021). SARS-CoV-2's preference for colder, drier conditions is consistent with climatic effects observed with influenza (Lofgren et al. 2007; Lowen and Steel 2014; Shaman and Kohn 2009). Another review article by Mecenas et al. 2020 finds a similar conclusion. While climate may affect SARS-CoV-2's transmissibility directly, the indirect effect of climate's effect on human behaviors has been demonstrated to be a much stronger effect (Susswein, Rest, and Bansal

---

[2] See Weaver et al's review of the following studies, among others: Chin et al. 2020; Liu et al. 2020; Ma et al. 2021; Matson et al. 2020; Morris et al. 2021; Nottmeyer and Sera 2021; Raiteux et al. 2021; Riddell et al. 2020; Sera et al. 2021; Smith et al. 2021; Wu et al. 2020

2023; Damette, Mathonnat, and Goutte 2021; Weaver et al. 2022). Indeed, research on contact patterns that relate to COVID-19 have are known to be a substantial driver of outbreak dynamics (Feehan and Mahmud 2021).

While many applications of ABMs to infectious disease focus on investigating the interaction-level, network, or transportation aspects of infectious disease, few have focused directly on seasonality. Arduin et al. 2017 incorporate seasonal forcing of pneumococcal infections linked to influenza infection using a fixed multiplication of the transmission probability during the flu season, similar to the school-term forcing of the transmission probability used in compartmental models by Keeling, Rohani, and Grenfell 2001. Similarly, Williams et al. 2022 incorporate seasonal forcing to an ABM used to study influenza by adding a multiplicative effect to the transmission parameter related to the temporal distance of each time period from the winter solstice, which is 'intended to account for factors that may influence transmissibility across a range of seasons due to variability in factors such as temperature, humidity, and changes in contact rates.' In an application of ABMs to COVID-19, Krivorotko et al. 2022 use an time-series model to decompose incidence counts into a time series effect, a seasonal effect, and a noise component; the seasonal effect and the time-series effects are specified to have a temporally autocorrelative function. ABMs have also been used to study seasonality in non-human diseases (Dawson et al. 2018; Oraby et al. 2014).

We incorporate seasonality in two ways: in the transmission probability $\beta(t)$ and in the number of nonhousehold contacts. We allow for seasonal forcing of the transmission probability to capture how the transmissibility of the pathogen may change with weather, modeled as:

$$\beta(t) = \beta_0(1 + \beta_1 * cos(2\pi/365 * t)) \tag{2.3}$$

where β_0 represents the average transmission probability and β_1 represents the amplitude of seasonal forcing (Liu et al. 2021b; Grassly and Fraser 2006). Second, we include school contacts between children under 18, which are derived from the POLYMOD survey (Mossong et al. 2008). School contacts are drawn once at the start of the simulation and maintained through simulation time. Schoolchildren are taken to have school contacts during the same business hours as random contacts between September 1st and June 1st annually; children who are clinically infectious with SARS-CoV-2 are kept home from school until they recover.

2.2.6 Incorporation of Demography

Optional in the model is the inclusion of basic demographic vital rates in the form of age-specific fertility and mortality data. At baseline, we draw from the CDC's 2021 estimates (Martin, Hamilton, and Osterman 2022; Xu et al. 2022) summarized to the age categories used in the model (0-18, 18-25, then in 10 year increments through age 85). The rates used at baseline are shown in figure B.2. These rates are used to randomly introduce new susceptibles into the population and randomly remove members of the population, representing deaths due to non-SARS-CoV-2 causes. Each birth represents a new, fully susceptible and unvaccinated child in the population in the household of the birthing parent; they are assumed to have the number of non-household and school contacts equal to the population average. Demography is incorporated into the model once monthly as a Bernoulli random draw for each member of the population with rate equal to the supplied rate, divided by 12 (for males, fertility rate is 0).

2.2.7 Simulation Procedure and Parameters

We identify a set of baseline transmission parameters in line with those used by Roubenoff, Feehan, and Mahmud 2023, adapted to fit the parametric needs of our Agent Based Model. For all simulations, the number of households is fixed at $1,000$, producing approximately $3,200$ individuals at the start of the simulation. Simulations are run for 10 years and are seeded with 5 index cases at time $t = 0$, intended to mimic the wintertime outbreak of early January 2021. During the course of the simulation, one case is imported weekly to ensure that all SARS-CoV-2 is constantly circulating at a low level. To account for a wide range of transmission scenarios, we consider three levels of transmissibility: low, with $\beta_0 = 0.01$ (implying $R_0 \approx 1.3$); moderate, with $\beta_0 = 0.025$ ($R_0 \approx 3.4$); and high, with $\beta_0 = 0.05$ ($R_0 \approx 6.5$). Baseline simulations are without seasonal forcing of transmission or contact but with school contacts included during school-year weekdays (Monday-Friday, 9am-3pm). Seasonal forcing of both transmission and contact parameters is introduced with low (10%), moderate (25%), or extreme (50%) seasonal amplitude as described above. To introduce stochasticity into the model, each infected case in the simulation contains a randomly drawn duration of latent period of between 48 and 96 hours after transmission; this is followed by a randomly drawn infectious period of between 72 and 120 hours. These distributions are held constant across all simulations. We assume that each case has a 20% chance of being subclinical—fewer than used by Roubenoff, Feehan, and Mahmud 2023 (derived from Johansson et al. 2021), but in line with recent meta-review estimates by Buitrago-Garcia et al. 2022. That same analysis identified seven studies comparing the secondary attack rate of asymptomatic cases and symptomatic cases with an average ratio of 32%. At baseline we assume that symptomatic cases have all of their normal random contacts and do not isolate, for a 'business as usual' scenario; we test the effect of isolation in sensitivity analysis. However, children are assumed to always be kept at home when ill. NPI effectiveness is set to zero, equivalent to the absence of NPIs or masks, but is varied during sensitivity analysis.

Vaccine effectiveness is taken to be 80% after one, dose, 90% after two doses, and 80% after three doses, consistent with estimates published in 2021 and 2022 (Tenforde 2021,Thompson 2021; Thompson 2022) and values used by Roubenoff, Feehan, and Mahmud 2023. Unclear at the present moment is the duration of immunity after infection and vaccination and the effectiveness of waned immunity; a 2021 CDC brief (updated 2022) estimates 6 months of nearly complete immunity that diminishes over time (Centers for Disease Control and Prevention 2021). We take 6 months to be the baseline assumption but vary this duration in monthly increments from 6 months to two years in a sensitivity test. We assume a pessimistic 25% efficacy of waned natural- and vaccine-derived immunity.

All main simulations are run for 10 years in replicates of 10, and we report a number of summary values for all simulations. These include the total number of clinical and subclinical cases, deaths due to SARS-CoV-2, and the timing and size of all outbreaks after year 5.

All estimates are standardized to the population size to account for populations that vary randomly in size. Outbreaks are found using the `Python` library `scipy`'s `find_peaks` function on the daily sequence of clinical cases, for a minimum incidence threshold of 5% of the population infected over a 30-day window.

2.3 Results

To elucidate future outbreaks of SARS-CoV-2, we simulate outbreaks at various levels of transmissibility, and test the distribution of annual booster doses in the absence of and presence of seasonality. We find that the optimal date of booster dose distribution for reducing the number of clinical infections is different for the simulations with and without seasonality; in the absence of seasonality booster doses in the first half of the year are most effective at eliminating a large annual outbreak, but with seasonality booster doses are most effective when distributed in early fall.

Infectiousness of COVID-19 and the duration of immunity after infection and vaccination have a strong effect on the dynamics of outbreaks. In a moderate transmission scenario, where the base probability of transmission for an hour-long contact the absence of NPIs or vaccination $\beta_0 = 0.025$ (corresponding to an R_0 of approximately 3.2), an average of 6.16 clinical infections occur per capita over a 10-year period. Over this period, an average of 33.6 outbreaks occur, each infecting an average of 17.7% of the population. However, when β_0 is raised to 0.05 (corresponding $R_0 \approx 6.5$), outbreaks are fewer (an average of 16.0 over the 10-year period) but more severe, with an average outbreak size of 59.63% and about 9.75 infections per capita—nearly one per person per year. These simulations are summarized in figure 2.3 and trajectories are shown in figure 2.4. We observe this dynamic throughout many simulations: when SARS-CoV-2 is more transmissible or less mitigated, outbreaks are fewer but more severe. Mortality in the high-transmission scenario is higher, proportional to the number of clinical infections—2.63% on average compared to 1.67% in the moderate transmission scenario and 1.1% in the low-transmission scenario. These simulations suggest that the public health planning and response for future variants may differ based on their epidemiology. More transmissible variants are likely to have fewer, larger outbreaks that may overwhelm the healthcare system capacity, but show few cases outside of the season; less transmissible variants may require a year-round response, but with less severe outbreaks.

Distributing annual booster shots consistently lowers the rates of SARS-CoV-2 for all simulations, but may independently induce a seasonal pattern in outbreaks. The ability of booster doses to successfully limit SARS-CoV-2 outbreaks is dependent on the timing of their distribution and whether or not seasonal forcing of transmission is included in the model. In the absence of seasonal forcing of transmission and contact (figures 2.5 and 2.6), distributing booster doses earlier in the year is most effective at reducing the size of outbreaks: assuming a 75% update of vaccinations distributed annually between January and May, about $1.68-2.35$ clinical infections occur per capita (average outbreak size ranging from 4.25% to 6.13% of the population); when distributed after July 1st, this rises to $4.34 - 4.76$ clinical infections per capita (average outbreak size ranging from 13.21% to 15.61%). In a

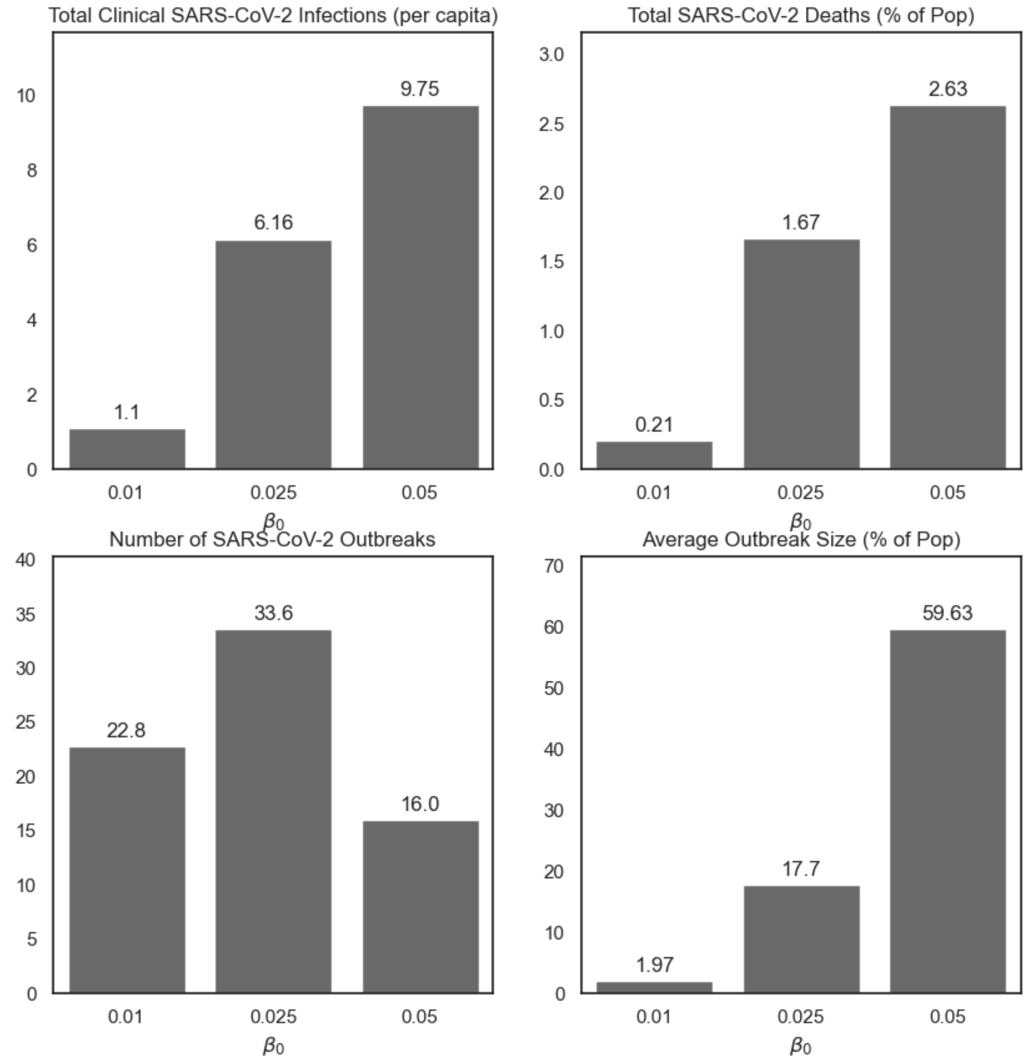


Figure 2.3: Summary of SARS-CoV-2 clinical infections, deaths, number of outbreaks, and average outbreak size for different values of β_0, the average baseline transmission probability, in the absence of seasonal forcing, isolation, or vaccine distribution. Simulations are run for 10 years in replicates of 10. Approximate corresponding values of R_0 are: $\beta_0 = 0.01$, $R_0 \approx 1.3$; $\beta_0 = 0.025$, $R_0 \approx 3.4$; $\beta_0 = 0.05$, $R_0 \approx 6.5$.

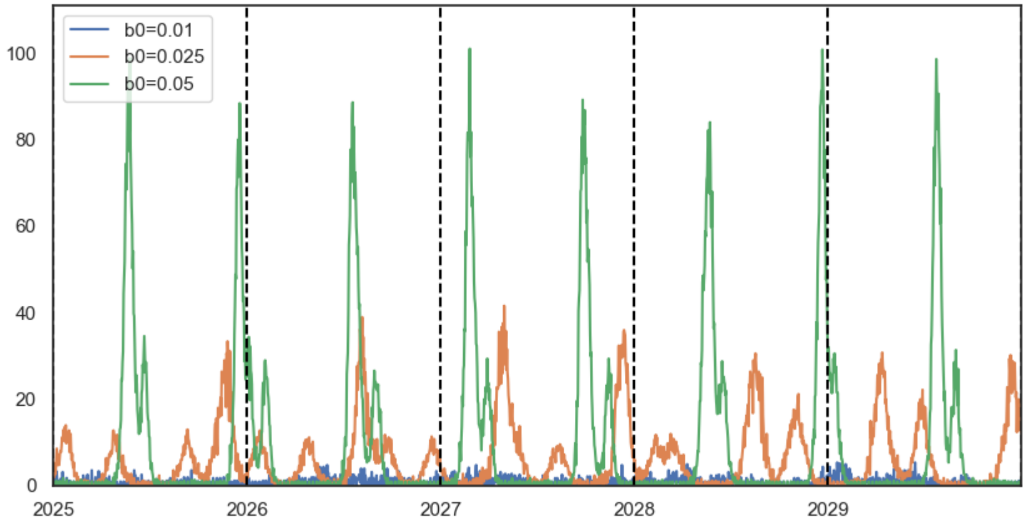


Figure 2.4: 5-year trajectory of SARS-CoV-2 clinical infections for different values of β_0, the average baseline transmission probability, averaged across 10 simulations each. When β_0 is high, there are generally 1-2 large outbreaks per year; when lower, outbreaks are smaller and more frequent. Approximate corresponding values of R_0 are: $\beta_0 = 0.01$, $R_0 \approx 1.3$; $\beta_0 = 0.025$, $R_0 \approx 3.4$; $\beta_0 = 0.05$, $R_0 \approx 6.5$.

high-vaccine uptake scenario (90% uptake, shown in the model supplement), the overall rates of SARS-CoV-2 are lower only when vaccines are distributed in the first half of the year; when vaccines distributed in the latter half of the year, the total cases and severity of outbreaks is comparable to the regular-uptake scenario. Outbreaks are observed to generally occur around 6 months—which is also the duration of full immunity after vaccination—after the date that booster doses are made available, with the strongest seasonal patterns observed with June-September distribution inducing a strong wintertime and October - December inducing an early spring outbreak. Indeed, as observed in figure 2.6, the peak out the outbreak appears to be 'shifted' approximately 6 months after the date of booster dose distribution, the duration of of full immunity after vaccination. This phenomenon—whereby January-May boosters nearly eliminate annual outbreaks in the steady state but later-year boosters fail to do so, albeit shift the outbreak timing—is likely driven by the initial outbreak (set to occur in January of 2021 to capture the large winter outbreak in the United States), which establishes the clock for a sufficient number of individuals with waned immunity to appear for an outbreak to occur predictably after. These simulations indicate that, in the absence of seasonality, the timing of booster dose distribution may have the power to govern the timing of the primary annual outbreak.

We also tested the distribution of vaccines in the presence of seasonal forcing of the transmission parameters ($\beta_1 = 0.5$), shown in figure 2.7 and 2.8. Unlike when distributing booster doses in the absence of seasonal forcing, in which the timing of booster dose distribution shifts the timing of the main outbreak, in these simulations with seasonal forcing a substantial wintertime outbreak occurs at nearly the same time every year. However, the size of this outbreak, as well as the presence of secondary outbreaks throughout the year, depends on the timing of booster doses. When boosters are distributed in the first half of the year—January 1st through May 1st—a moderate-sized fall outbreak occurs. When boosters are distributed by July 1st, this fall outbreak nearly doubles in size; a September 1st distribution day results in a less predictable situation, with multiple (2-3) smaller outbreaks throughout the year. However, distributing doses too late (November 1st) results in a large summertime outbreak, despite the relatively lowered transmission rate during the summer months. These dynamics are similar when seasonal forcing is present in the transmission parameters and the contact rate, shown in the supplementary material. Overall, simulations where vaccines are distributed between January and June resulted in $2.76 - 3.36$ infections per capita—1-1.5 fewer than when vaccines are distributed in the highest months of July or December ($4.82, 5.19$ respectively). However, distributing booster doses in September or October results in fewer infections, comparable with simulations when vaccines distributed earlier in the year, and smaller outbreaks ($8.69\% - 9.59\%$ of the population infected on average per outbreak).

As a sensitivity test, we varied the duration of immunity after infection and vaccination, and find that this parameter has a substantial effect on the timing and size of outbreaks. As this parameter governs the rate and which susceptibles are effectively re-introduced into the population, our results are onsistent with Baker, Peckham, and Seixas 2020. Recurrent outbreaks of SARS-CoV-2 are driven by the seasonality included in the model but also by

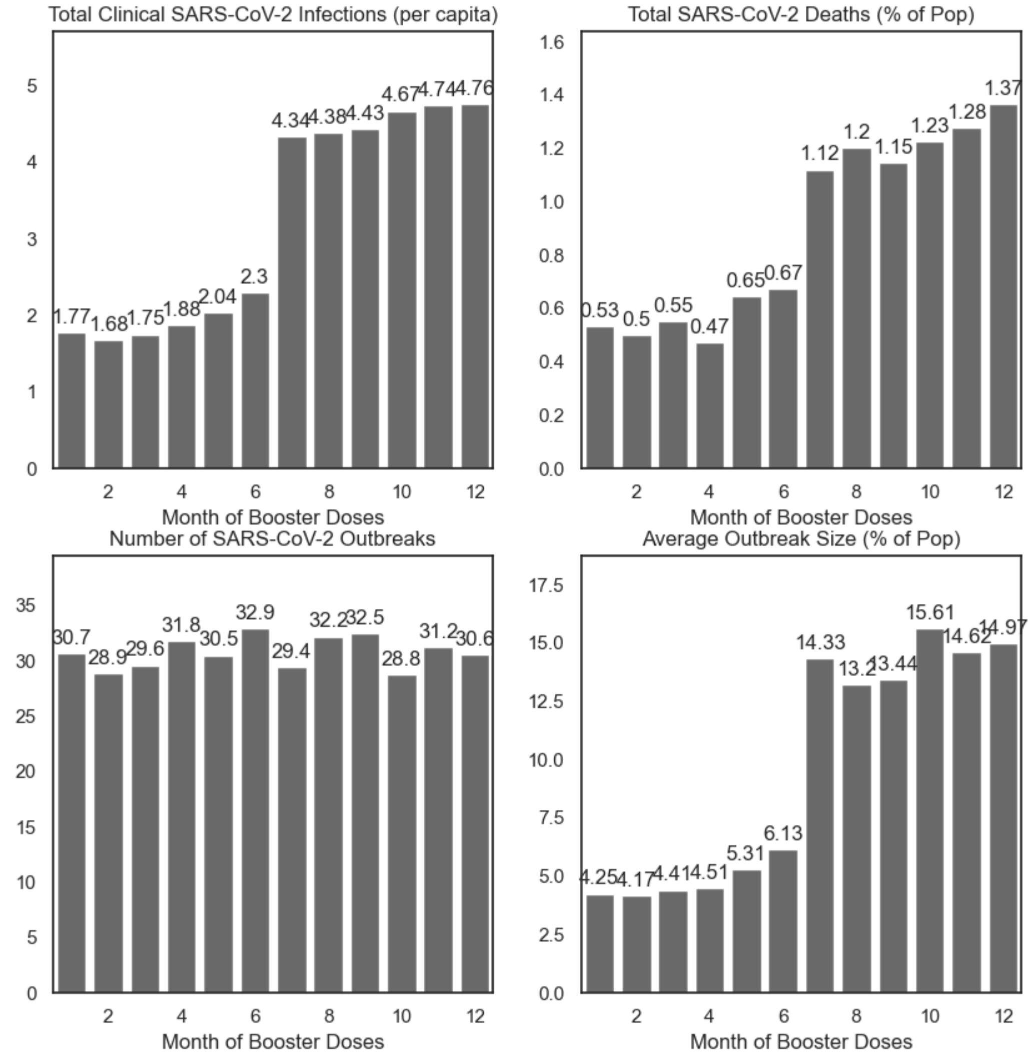


Figure 2.5: Summary of simulations by day of booster dose distribution, varied as the first of each month, in the absence of seasonal forcing of the transmission parameter β. Simulations are run for 10 years in replicates of 10.

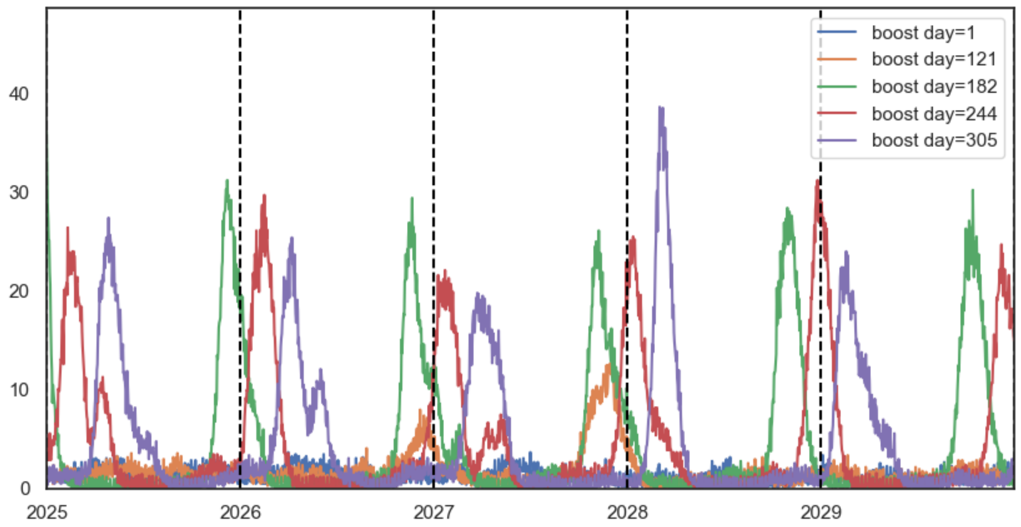


Figure 2.6: 5-year trajectory of simulations by day of booster dose distribution, in the absence of seasonal forcing, for selected distribution days: Jan 1st (day 1), May 1st (day 121), July 1st (day 182), Sept 1st (day 244), and Nov 1st (day 305), averaged across 10 replications. When doses are distributed earlier in the year—Jan 1st-May 1st—the major outbreaks are largely averted, but persist when doses are distributed too late in the year.

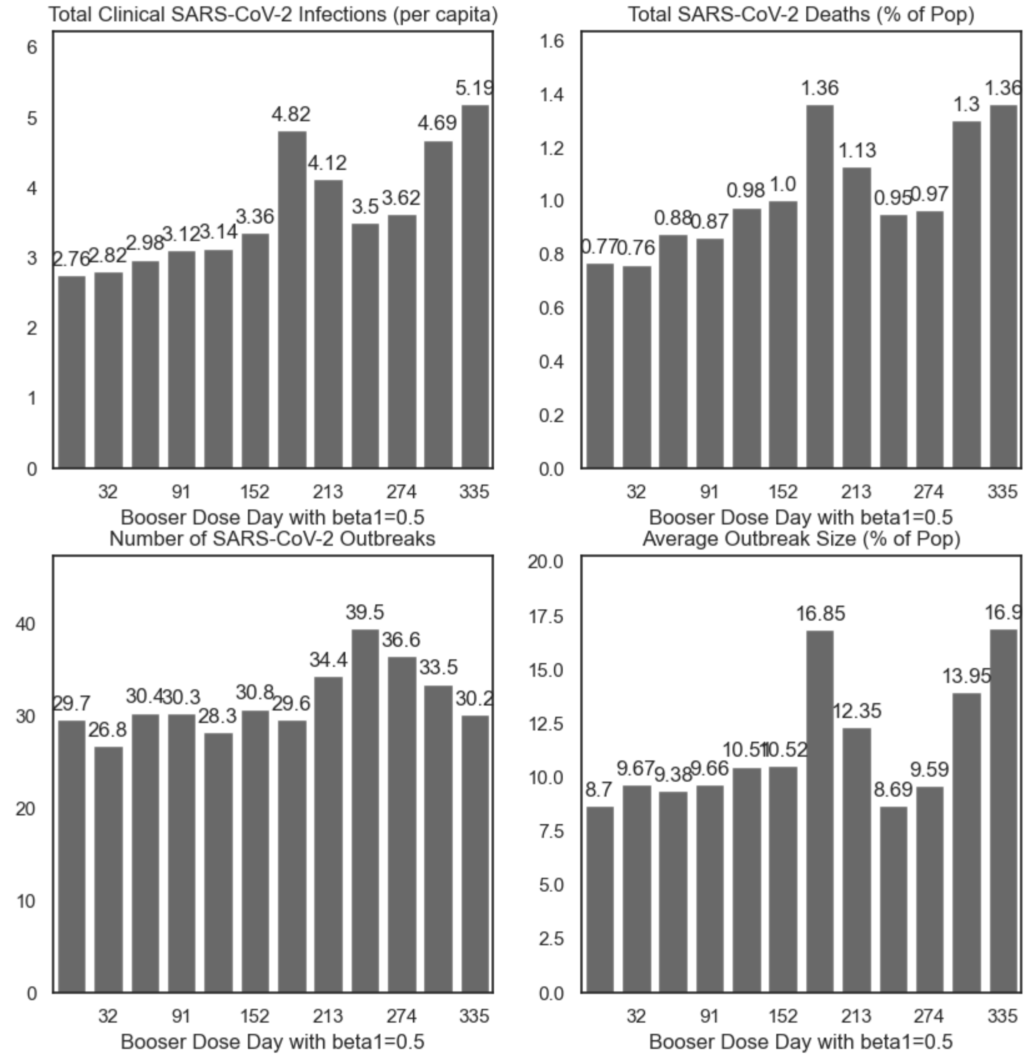


Figure 2.7: Summary of simulations by day of booster dose distribution, varied as the first of each month, in the presence of seasonal forcing of the transmission parameter β. Simulations are run for 10 years in replicates of 10.

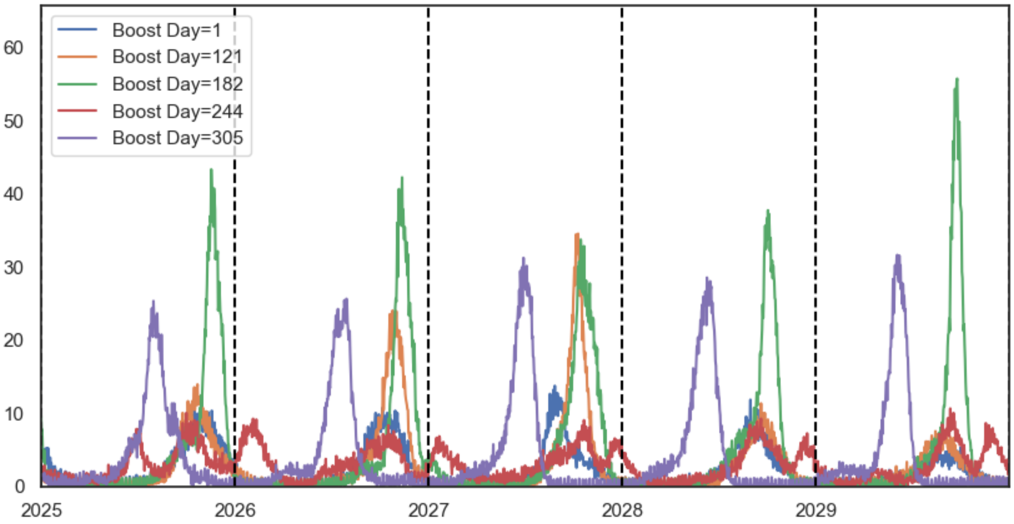


Figure 2.8: 5-year trajectory of simulations by day of booster dose distribution, for selected distribution days: Jan 1st (day 1), May 1st (day 121), July 1st (day 182), Sept 1st (day 244), and Nov 1st (day 305), in the presence of seasonal forcing of the transmission parameter β, averaged across 10 replications.

the effect of waning natural and vaccine-derived immunity, such that even models without seasonal forcing and booster dose distribution may exhibit predictable outbreaks (figure 2.9). Rather than these outbreaks occurring at specific times throughout the year, they occur a certain amount of time after the previous outbreak—generally equal to the duration of complete immunity. At 6 month immunity, generally two outbreaks occur per year, spaced slightly more than 6 months apart, with periodic secondary outbreaks between; at one year of full immunity, large outbreaks occur slightly more than one year apart, without secondary outbreaks. These simulations indicate that preparations for outbreaks should include evaluation of the previous major outbreak.

We also explored the possibility of isolation as a means to control SARS-CoV-2, despite it being unlikely to be used as a general control strategy in the future. Isolation remains an effective way to limit the spread of SARS-CoV-2 in the steady state, however only the higher degrees of isolation— reducing non-household contacts among clinically infectious individuals to less than 50% of their normal amounts as compared to a 'business as usual' scenario— has a substantial effect on transmission dynamics (figure 2.10). At 50% isolation, an average of 2.72 clinical infections occur per capita, down from 4.84 when clinically infectious nodes have 75% of their normal random contacts and 6.22 in the complete absence of isolation. With 50% isolation, the outbreak size drops dramatically to 6.05% of the population infected during an average of 40.8 outbreaks, down from from 18.7% of population infected in an average of 32.2 outbreaks in the absence of isolation. More extreme isolation reduces the

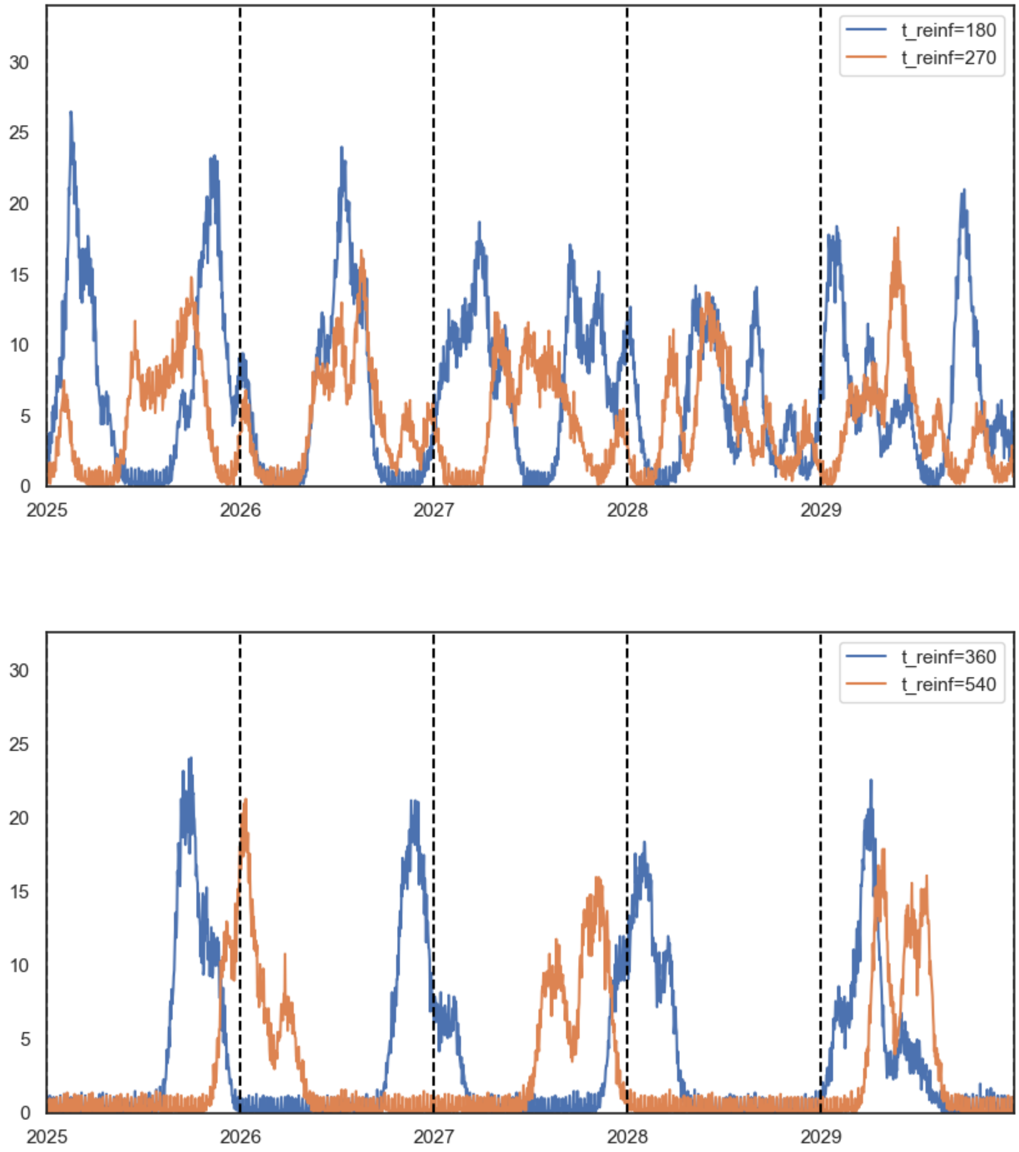


Figure 2.9: Trajectory of SARS-CoV-2 clinical infections with varied duration of complete immunity after infection and vaccination, illustrating how outbreak timing can be affected by immunity. Top: full immunity lasting for 180 days (6 months) and 270 days (9 months); bottom: full immunity lasting for 360 days (12 months) and 540 days (18 months). Averaged across 10 replications.

severity of outbreaks even further: at 25% of normal contacts, clinical infections average one per capita; with perfect isolation, clinical infections are less than 0.5 per capita.

2.4 Discussion

Across all simulations, we observed frequent and predictable SARS-CoV-2 outbreaks over a 10-year period, even with the annual distribution of booster doses as the primary disease-averting intervention. Depending on the epidemiology of the pathogen—namely, should SARS-CoV-2 exhibit seasonality in transmission probability and contact—outbreaks may occur at predictable times of the year, and distribution of booster doses may be able to mitigate the worst of seasonal outbreaks. Our results are consistent with Baker et al. 2020, who find that outbreak cycles are primarily determined by the levels of susceptibility in the population, although seasonality is an important moderator in outbreak dynamics. Different vaccination campaigns may be needed in areas that exhibit stronger transmission seasonality. We find that distributing booster doses in the first half of the year—January through May—may be an effective strategy at limiting recurrent outbreaks depending on the seasonality exhibited by the pathogen. We find that in simulations without seasonal forcing, distributing booster doses in the first half of the year is most effective at limiting outbreaks; however, with the inclusion of seasonal forcing of transmission and contact, distributing booster doses in September or October may limit the average outbreak size the most. In these simulations, distributing booster doses in the late fall 'shifts' the outbreak to the summer, when transmissibility is lower. Since influenza vaccines are distributed in the fall, including booster doses for SARS-CoV-2 at the same time may be easiest in implementation, but less successful than Springtime distribution in limiting outbreaks should SARS-CoV-2 fail to exhibit transmission seasonality.

In addition to illustrating how vaccination interventions can avert the burden of illness and death due to SARS-CoV-2, our simulations further understanding of how to prepare for future SARS-CoV-2 outbreaks. Although distributing booster doses in early fall—like annual vaccinations for influenza—avoids a large wintertime outbreak in a seasonally forced environment, multiple smaller outbreaks may occur throughout the year. This may be preferable to avoid exceeding the treatment capacity of the health system. However, this strategy may not be effective in a less seasonally-variable climate, where booster dose timing merely shifts the main outbreak to later in the year after immunity wanes. Overall, we generally observe variance in mortality proportional to the number of clinical infections that occur in the simulations, indicating both that the most effective strategies for limiting clinical infections also limit deaths. Although we focus on the number of infections as the primary outcome, reducing the number of deaths due to SARS-CoV-2 may be possible with the interventional strategies outlined here.

While agent-based models have been used in a variety of applications for COVID-19, ours represents one of the first to examine long-term dynamics using real population contact data. Our results hint at a rather bleak outlook for the future of SARS-CoV-2: that outbreaks are extremely likely to continue given that natural and vaccine-derived immunity wanes

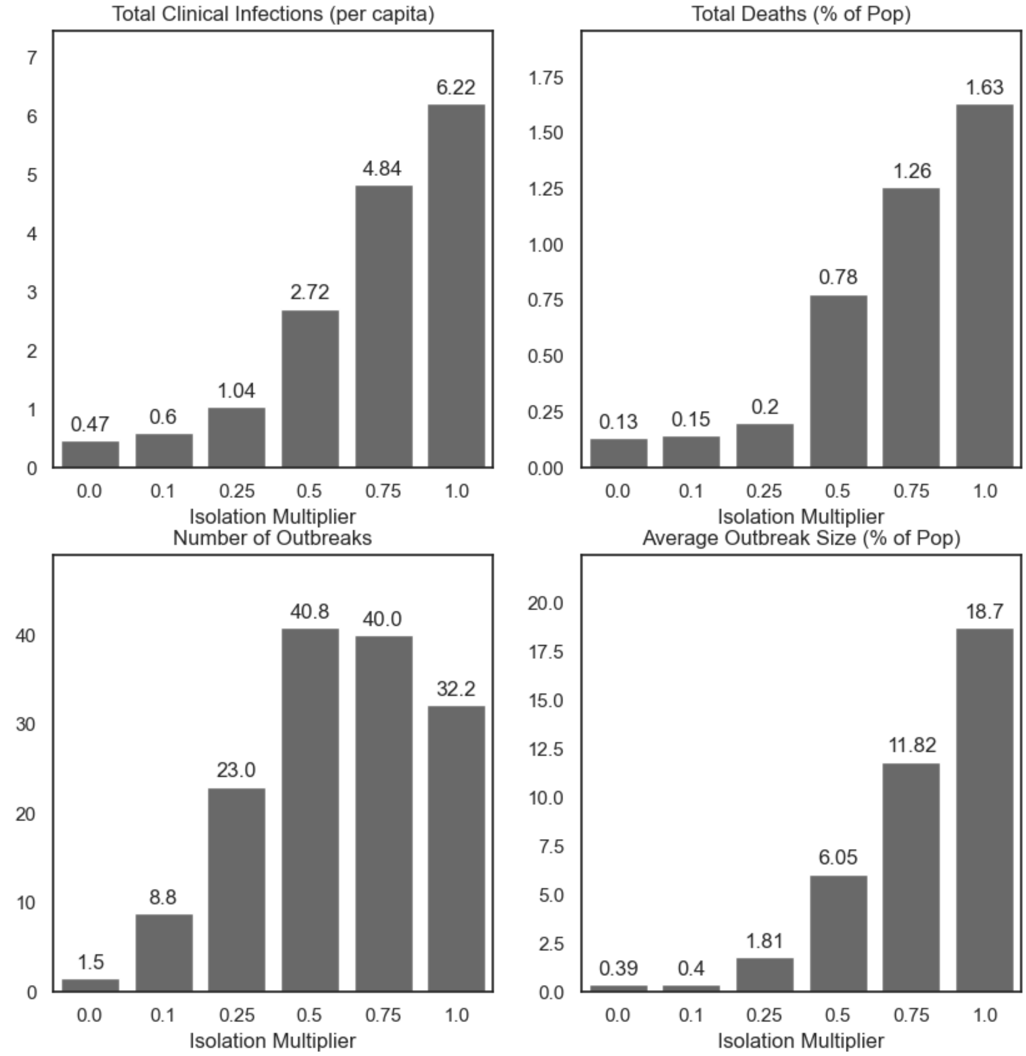


Figure 2.10: Summary of SARS-CoV-2 clinical infections, deaths, number of outbreaks, and average outbreak size for different levels of isolation, in the absence of seasonal forcing or vaccine distribution. Isolation multiplier is a factor used to scale a clinically infectious node's random, non-household contacts; 1.0 indicates business as usual and 0 indicates perfect isolation.

over time. However, annual booster doses—especially when timed properly—and isolation of infectious cases may be effective control strategies. The investigation above tells us that outbreaks can be expected to occur more frequently than annually, and the epidemiology of the pathogen—namely, the base transmission probability and the duration of immunity after infection and vaccination— determines the frequency and severity of outbreaks more than vaccination timing and within a reasonable range of effectiveness. As our approach to SARS-CoV-2 shifts from eradication to management, the strategies presented here show how we can time the distribution of doses to minimize strain on the healthcare system and limit chronic complications from SARS-CoV-2 infections.

Our model has a number of shortcomings that limit its generalizability. First, we are limited by computing resources to a population of 1000 households (approximately $3,200$ individuals). During development, we found that using too few households resulted in simulations that were highly unstable; with a larger population, simulation results were more consistent between runs but took much longer to complete. As a result, we chose the number of households to balance numerical stability while maintaining a reasonable run time of approximately 4 minutes per simulation. Larger populations may exhibit different dynamics as the spread of infection may take longer; as such, it is not known presently if this chosen population size is representative of a larger population or is limited in generalizability to smaller communities. As well, like in Chapter 1, we borrow contact patterns for the youngest age group from the POLYMOD UK survey (Mossong et al. 2008), which may not be representative of that age group in the US during the COVID-19 pandemic. However, since our application here is in consideration of the long-term patterns of SARS-CoV-2,' we believe that these contacts represent a return to 'business as usual' for school children.

A key feature of our model is in the random network generation, that produces daily draws of random contacts according to the network configuration model parameterized with degree and duration from the BICS survey. Like any random network generation, the network structure of contacts produced by this model may affect transmission dynamics. Future work should draw inspiration from the COVASIM agent based model (Kerr et al. 2021), which allows for comparison of outbreak dynamics between simulations with Poisson random networks, Hybrid networks, or SynthPops networks depending on the needs of the simulation and data available. As well, we do not include any assortative mixing as emphasized by the model used by Roubenoff, Feehan, and Mahmud 2023 or serial (repeat) contacts, instead choosing a network generation function that maintains the degree of each node. Employment contacts, like school contacts for children under the age of 18, provide a consistent set of individuals having contact most days; inclusion of these such contacts may affect outbreak in unpredictable ways. Future development of the model should include associativity by age and employment status; however, doing so was outside of the scope of the present analysis.

Implementation Overview

The BICS ABM model is implemented in the `C++` language and with use of the `iGraph`-C library (Csárdi and Nepusz, T. 2006; Csárdi et al. 2023), and a Python 3.8 API. The core

C++ implementation was chosen over other languages, like `Python` or `R`, to maximize speed. Full implementation details are given in the supplementary material. Each simulation of 1000 households for 10 years completes in approximately 4 minutes on an Apple MacBook Air M1, and the entire suite of simulations completes in approximately 8 hours. Replication code is publicly available at `https://github.com/eroubenoff/BICS_ABM`.

Chapter 3

Bayesian Spatiotemporal Projection of Chagas Disease Incidence in Brazil

Abstract

Chagas Disease is a parasitic infection caused by the *T. Cruzi* parasite endemic to Central and South America and transmitted through contact with *Triatomine* insects, commonly known as "kissing bugs." Although the symptoms of Acute Chagas Disease (ACD) are nonspecific, untreated chronic infection can lead to heart disease, enlarged esophagus and colon, and stroke. Chagas disease has become increasingly rare owing to a series of public health interventions, including insect eradication campaigns in Brazil through the 1980's that considerably reduced the number of new acute cases. However, hundreds of new acute cases still are diagnosed annually, primarily in the states of Pará, Amapá, and Acre. Moreover, the population in areas of high Chagas endemicity are changing: many areas are growing and becoming increasingly urban, whereas others are decreasing in population. We estimate the Incidence Rate (IR) for Acute Chagas disease over the period 2001-2019 in Brazil at the municipal level and investigate the variation of these rates with climatic factors. These estimates are used to project forward incidence of Acute Chagas Disease over the following decade 2020-2029. Modeling ACD presents numerous methodological challenges since incidence is rare, with extreme overdispersion of zero-case counts, and vectors exhibit a highly spatially- and temporally-clustered pattern. We use a spatially- and temporally-autoregressive small-area smoothing models to estimate the true latent risk in developing Acute Chagas Disease. The Bayesian model presented here involves spatio-temporal smoothing via a Zero-Inflated (Lambert 1992), Knorr-Held (2000)-Type spatio-temporal model with a BYM2 (Morris, 2019) spatial convolution to predict smoothed incidence rates of Chagas disease. As well, we include estimates of Brazil's growing population and projected bioclimate to evaluate how climate and population change may affect ACD rates. We estimate that cases will continue to increase in the absence of control efforts, primarily driven by a growing peri-urban population in regions of Chagas endemicity.

3.1 Introduction

Chagas disease is a vector-borne parasitic infection in humans caused by the *T. Cruzi* parasite, and is transmitted to humans primarily through contact with infected *Triatomine* insects commonly known as "kissing bugs" (WHO Expert Committee on the Control of Chagas Disease 2002; Pérez-Molina and Molina 2018; Canals et al. 2017). Transmission of *T. Cruzi* to humans can occur following bites from infected kissing bugs or contact with their feces; human-human transmission can occur via blood transfusion and congenitally from mothers to children via the vertical pathway. Over 80% of transmission occurs from human-vector contact, with congenital transmission responsible for nearly all other new infections; screening of blood donation has nearly eliminated all transmission from transfusions (World Health Organization 2015). Acute symptoms of Chagas disease include fever, inflammation of the infection site, eyelid edema, and swollen lymph nodes and tonsils. Acute symptoms resolve spontaneously over 4-8 weeks and treatment during the acute phase with antiparasitic medication is highly effective at curing infection (Bern et al. 2007). However, since acute symptoms are generally nonspecific and the burden of infection affects many communities lacking affordable access to high-quality healthcare, many acute cases go undiagnosed and untreated. Untreated chronic Chagas infection can cause cardiomypoathy, megacolon, stroke, and megaesophagus in 30-40% of patients in the 10-30 years following acute infection (Pérez-Molina and Molina 2018; Sosa-Estani and Segura 2015). Although Chagas is fairly rare—acute infection incidence is on the order of a few hundred infections annually in Brazil for a population of around 200 million—it is this possibility of chronic complications, especially cardiomyopathy, that makes early identification and control of Chagas an important public health concern.

Chagas disease is also known as a disease of poverty affecting mostly poor and indigenous rural communities in South America (Fernández, Gaspe, and Gürtler 2019; Sosa-Estani and Segura 2015; Dias 1987; Tarleton et al. 2007). Many people are at increased risk of infection due to the use of certain residential construction materials hospitable to *Triatomine* infestation, especially untreated wood. People working certain jobs that entail contact with *Triatomine* habitats—including forestry or agriculture—may be at increased risk of exposure during employment. A number of non-pharmaceutical interventions can alleviate much of the probability of contact, including insecticide usage, bed netting, and removal of certain residential construction materials known to be *Triatomine* habitats. Intervention campaigns through the 1980's focused directly on the class dynamics of Chagas risk by implementing control efforts across the sociodemographic ladder (Dias 1987), yet many persisting Chagas hotspots occur in poor and rural parts of Brazil. Controlling the incidence of Chagas remains an important issue of equity.

Since then, Public health efforts to eliminate new Chagas infections and pharmaceutical developments to treat latent chronic infections and complications have been successful at reducing the burden of Chagas disease in Brazil (World Health Organization 2015; Sosa-Estani and Segura 2015). The two primary methods of transmission—human-vector contact, either through bites or contact with vector feces, and the vertical pathway from mothers to infants—have have required vastly different interventions, both with success. Early studies dating to the late 1940s found that continuous use of residential insecticides was highly effective at eliminating transmission, indicating that residential contact with *Triatomines* may have been responsible for the majority of transmission risk (Dias 1987). Eradication programs by Brazil's SUCAM (Superintendencia de Campanhas de Saude Publica) in the 1980s involved identifying areas of risk by sampling insects in homes and generating maps of high risk locations. All homes within more than 700 high-risk municipalities, regardless of known infestation, were sprayed with insecticide every 3-6 months until under 5% of homes were found to be infested with any insects and no *Triatomines* were found in any homes. Overall, more than 5 million homes were sprayed with insecticide, resulting in a 73% reduction in the number of infested homes by 1986 (*ibid.*) and the total elimination of transmission by *T. infestans*—previously the vector responsible for most transmission—resulting in a 94% reduction in new acute cases (Gurgel-Gonçalves et al. 2012).

Congenitally-transmitted Chagas disease via the vertical pathway is less frequent than transmission via contact with *Triatomines* (World Health Organization 2015). Screening for Chagas Disease among potential mothers living in high-risk areas and initiating treatment in advance of pregnancy is ideal for reducing the probability of successful vertical transmission, although commencing treatment after pregnancy appears to be well tolerated by both the mother and the fetus (Cevallos and Hernández 2014). While it is not currently possible to entirely eliminate vertical transmission, treatment of infants with suspected Chagas infection within the first year of life is very successful at eliminating the disease from children (Carlier et al. 2011; Moya, Basso, and Moretti 2005). Transmission may occur at any time during pregnancy, but is theorized to be more likely to occur during the second and third trimesters (*ibid*). Most congenital transmission occurs from mothers who are in the chronic phase of disease, however vertical transmission has been documented from mothers who are acutely infected at conception or become infected during the course of pregnancy. It has been proposed that the level of parasitemia of the mother may affect the probability of vertical transmission and the severity of infection at birth (Carlier et al. 2011).

Antiparasitic medications benzniazole and nifurtimox have proven efficacy against Chagas disease, and the former is usually recommended for treatment (Bern et al. 2007). If treated in the acute phase, complete parasitological cure can occur in 60-85% of vector-transmitted infections and more than 90% of congenital infections when treatment is administered within the first year of life (Altcheh et al. 2011; Carlier et al. 2011; Cevallos and Hernández 2014; Moya, Basso, and Moretti 2005). If Chagas disease is left untreated until the chronic phase, treatment is less effective— only 60% of participants achieved negative serology within 3-4 years. Even if not resulting in a complete cure, treatment may slow the development of Chagas cardiomyopathy and other potentially lethal complications, and treatment is recommended for all patients presenting with positive serology (*ibid.*).

Despite progress towards elimination, there are still an estimated 1-4.6 million people currently infected with chronic Chagas disease and approximately 6,000 deaths per year (Simões et al. 2018). Pérez-Molina and Molina 2018 estimate that in 2010, over 70 million people were at risk of contracting Chagas disease across 21 countries in Latin America, and 38,593 new infections were reported that year. This count of acute infections down considerably from 55,585 in 2005 and from more than 700,000 between 1980-1985. Most individuals living with Chagas disease are located in three countries—Argentina, Brazil, and Mexico—and the most new infections were reported in Bolivia (World Health Organization 2015). While preventative interventions have brought the new infection rate down considerably in Brazil, the rate of decrease has not necessarily been equal across the country. The WHO estimates that as of 2015, the incidence of new Chagas infections in Brazil via human-vector contact was 0.084 per 100,000 population and via congential transmission 0.020 per 100 live births (*ibid.*). Our analysis aides in identifying areas where future interventions can further alleviate risk of the disease.

The distribution of *triatomines* is a highly spatial process within endemic areas, and as a result risk of contracting Chagas disease is a complex interaction of local vector population, local human population, and interaction between the two. Despite the elimination of *T. infestans*, previously the vector responsible for the most new cases of Chagas Disease, 62 known species of *Triatomines* in Brazil are responsible for transmission; some, including *Panstrongylus geniculatus* and *P. megistus*, are widespread over the country, whereas others are more localized spatially (Gurgel-Gonçalves et al. 2012). Certain biomes, including the Cerrado tropical savannah in the central region and Caatinga shrub forest in the northeast, have a higher diversity of species. Not all species of *Triatomines* are equally likely to transmit Chagas disease to humans; for example, while the most epidemiologically relevant species may be *P. megistus*, which is common in domiciles and a frequent carrier of *T. cruzi*, the behavior and habitat of *T. sordida* is more likely to result in contact from agricultural activities but unlikely to result in residential contact. Gurgel-Gonçalves et al. 2012 remark that nearly all areas of Brazil have some risk of Chagas disease, but certain regions, especially the Cerrado and Caatinga, present more risk.

Climate change presents an ambiguous threat to incidence of Chagas disease. Tamayo et al. 2018 find that *Triatomine* vectors of Chagas disease may exhibit increased fecundity and egg viability in warmer temperatures. *T. Cruzi* as well may exhibit increased viability at warmer temperatures, suggesting that incidence of Chagas disease will likely increase with climate change. Medone et al. 2015 find that the changing climate will likely create more geographic areas that are suitable habitats for Chagas vectors correlates with the force of infection of acute Chagas disease in Argentina and Venezuela. They find that warmer temperatures are unfavorable to vectors; although current Chagas hotspots may see decreases with increased temperatures, the overall geographic distribution of vectors may shift as previously too-cold areas warm.

Since Chagas disease is primarily found in rural areas in Brazil, multi-decadal trends in the urban and rural population may be a mediator in the future trajectory and control of Chagas disease (Delazeri, Da Cunha, and Oliveira 2022; Perz 2000; Randell and VanWey 2014). Brazil has become increasingly urbanized since the 1960s; internal migrants from rural, Chagas-endemic areas have resulted in identification of both acute and chronic cases of Chagas disease in places where Chagas is not historically found (Coura and Borges-Pereira 2010; Martins-Melo et al. 2014; Moncayo and Silveira 2009). Like all urban areas in Brazil, larger municipalities in Chagas-endemic areas have been increasing in size faster than the nearby rural areas, many of which have even seen population declines. The identification and control strategies of the 1980s that targeted known areas of Chagas endemicity may not be as effective for identifying latently infected internal migrants who have moved to the larger cities ourside of endemic strategies. As well, declines in the rural population are fundamentally changing the spatial distribution of new cases of Chagas disease. Although the rural population is declining overall, declines are not uniform across all areas of Chagas endemicity, and projections of future cases must include population change as well.

To identify areas of persisting and future Chagas endemicity, we borrow Bayesian disease mapping methods for modeling higher-incident spatially-clustered diseases, such as cancer (Best, Richardson, and Thomson 2005; Napier et al. 2019; Riebler et al. 2016; Wakefield 2007; Wikle, Berliner, and Cressie, Noel 1998; Knorr-Held 2000; Knorr-Held and Best 2001) and adapt these methods to suit the highly rare nature of Chagas disease. Since the rates of ACD are low—on the order of hundreds of cases annually for a population of over 200 million—and present in a highly clustered pattern in certain regions of Brazil, a specialized modeling approach is needed to highlight the spatio-temporal structure of Chagas disease. We use a Knorr-Held (2000) spatio-temporal model adapted with a rare count, zero-inflated model (Lambert 1992; Lee et al. 2016; Rathbun and Fei 2006; Ver Hoef and Jansen 2007) and include climate covariates and population growth to analyze how patterns of Chagas disease might play out for the ensuing decade. We find that with an increasing population and climate trends, it is likely that cases of Chagas will continue to increase in the absence of additional intervention, potentially as high as doubling in incidence.

3.2 Methods

3.2.1 Data

Counts of Acute Chagas Disease (ACD), aggregated by municipality [1] of residence and year between 2001 and 2019, are collected by the Ministry of Health's Departamento de Informática do Sistema Único de Saúde (DATASUS; Department of Informatics of the Unified Health System) and retrieved from the agency's TABNET database (Ministério da Saúde, Brasil 2023). We chose to use municipalities—the finest level of geographic aggregation available—in order to analyze spatial variability that may be lost at larger levels of geographic aggregation, like state or region. Over the period 2001-2019, 5568 cases of Acute Chagas Disease were reported among residents of 826 municipalities, with the highest counts in the northern states[2] of Pará and Amapá. The municipality-specific total counts of Acute Chagas Disease reports are displayed in figure 3.1. Official population estimates at the municipality level are taken from Brazil Instituto Brasileiro de Geografia e Estatística (IBGE)'s annual population estimates for 2001-2006, 2008-2009, and 2011-2019 and the 2010 census

---

[1] *Município* is translated to English as municipality, but are functionally closer to US Counties by population size, geographic size, and governance.

[2] Brazil has 27 Federative Units (*unidades federativas*, abbreviated as UFs), consisting of 26 states and one federal district (Brasília). Here, we refer to all 27 UFs as states.

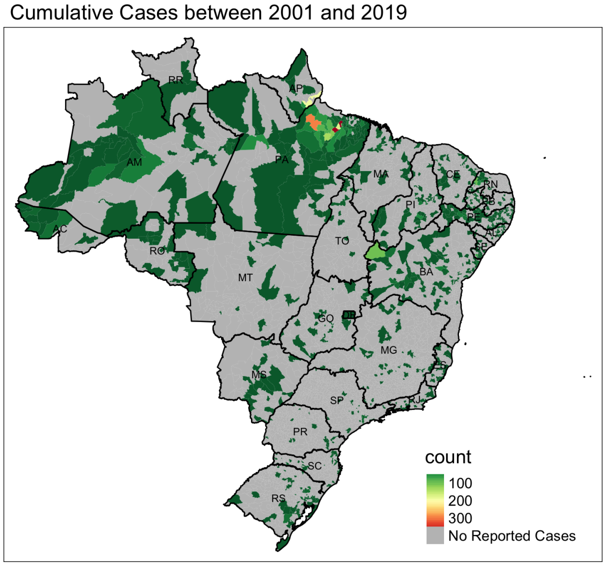


Figure 3.1: Counts of Chagas disease between 2001 and 2019 at the municipality level. Of the 5568 municipalities in Brazil, 4472 municipalities reported no cases of ACD during this period.

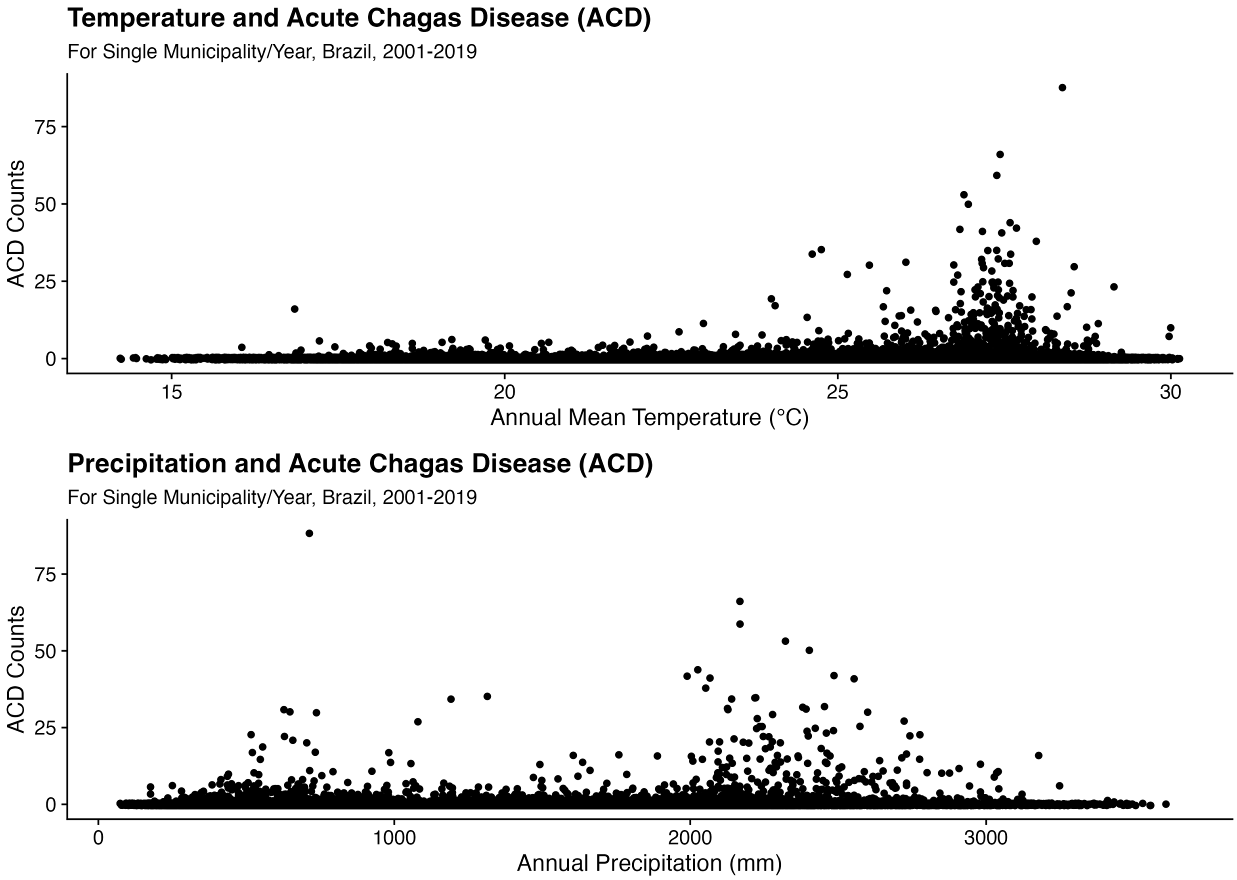
counts (Instituto Brasileiro de Geografia e Estatística 2023). No data are present for 2007; population for this year is taken as the linear interpolation between 2006 and 2008. Climate data are retrieved from the European Union's Copernicus Climate Change Service (C3S) Climate Data Store (CDS) Global Bioclimatic Indicators from 1950-2100 Derived from Climate Projections (Wouters et al. 2021), which contains a suite of 19 bioclimatic variables averaged annually. These variables are the same as those in the WorldClim (Fick and Hijmans 2017) dataset[3], and are listed in supplementary table C.2; a selection are displayed in figures 3.2 and 3.3. Broadly, we see in figure 3.2 that most reports of ACD occur in areas that are warmer (annual mean temp $> 25°C$) and wetter (annual precipitation above 2000 mm), although many cases do occur in drier climates. In Pará and Amapá, the UFs where most cases of ACD occur, there is a slight trend towards warmer and wetter weather, although substantial year-to-year variations present (figure 3.3). Projections are performed using the GFDL-ESM2M (NOAA, USA) algorithm.

3.2.2 Conditionally AutoRegressive (CAR) statistical models for disease incidence data

Many classes of geospatial models for areal data (polygons, like municipíos) exist, a few of which are discussed here. Distributional models used in Bayesian modeling can be divided into two groups: Conditionally AutoRegressive (CAR) models, that describe probability for observations as conditional on their neighbors, and Simultaneous AutoRegressive (SAR)

---

[3] `http://www.worldclim.org/`


Temperature and Acute Chagas Disease (ACD)

For Single Municipality/Year, Brazil, 2001-2019

Precipitation and Acute Chagas Disease (ACD)

For Single Municipality/Year, Brazil, 2001-2019

Figure 3.2: Counts of Acute Chagas Disease (ACD) by mean annual temperature and total precipitation, two of the 19 bioclimatic variables used in the analysis, displayed over the period 2001-2019.

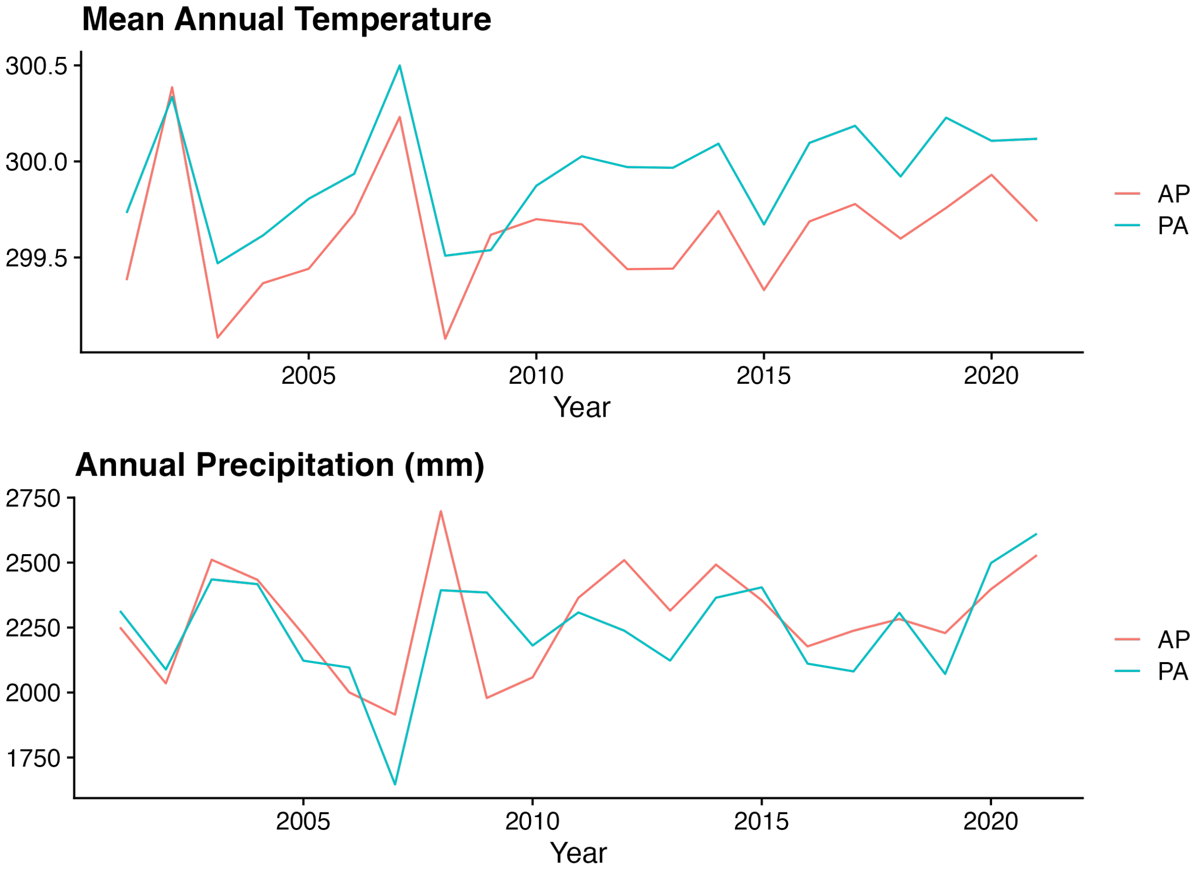


Figure 3.3: Mean annual temperature and total precipitation, two of the 19 bioclimatic variables used in the analysis, displayed over the period 2001-2030 for the two states with the highest incidence of ACD, Pará (PA) and Amapá (AP).

models, that use a matrix of adjacency weights to adjust for spatial dependence. At a high level, these two models differ in the scale of spatial dependence: the CAR model involves local smoothing, where the SAR model involves global smoothing. Here, we use a CAR model in our application to rates of Chagas disease, which is a highly local process found only in certain relatively isolated regions of Brazil and where global correlative structure is likely to over-smooth small-area variation. More commentary is provided in the supplementary material. As well, CAR models have a computational advantage over SAR models: they do not require matrix inversion, which can be computationally expensive or impossible when modeling thousands of small-area samples[4].

The simplest implementation of the CAR model is the Intrinsically AutoRegressive (IAR or ICAR) model, also called the BYM model after authors Besag, York, and Mollie (1991). For a general Gaussian spatial process $\boldsymbol{\phi}$, the CAR model is conditionally specified for each geographical unit as a normal distribution with expectation equal to the average of its neighbors and variance τ:

$$\phi_i | \boldsymbol{\phi_{j \sim i}} \sim N\left(\frac{1}{n_i}\sum_{j\sim i}\phi_j, \frac{\tau^2}{n_i}\right) \tag{3.1}$$

Where ϕ_i is an observation at the ith spatial unit, $\boldsymbol{\phi_{j \sim i}}$ indicates the set of observations among the neighbors j of i, and n_i is the number of neighbors of i. Throughout, we refer to equation 3.1 as the CAR and IAR models interchangably. The IAR distribution can also be extended to Poisson, Binomial, and Logistic distributions as well (Besag, York, and Mollie, 1991; Haining, 2004).

To utilize the CAR distribution in a disease modeling context, Banerjee, Carlin, and Gelfand (2015) recommend using a pair of *random effects* for the standardized incidence rates of disease:

$$Y_i | \psi_i \overset{iid}{\sim}, \tau Poisson(Pop_i e^{\psi_i}) \tag{3.2}$$

$$\psi_i = \boldsymbol{x_i'\beta} + \theta_i + \phi_i \tag{3.3}$$

where $\boldsymbol{x}$ and β are vectors of spatially-varying covariates, θ_i captures heterogeneity with an *iid* normal prior $N(0, 1/\tau_h)$ and ϕ_i captures spatial clustering with prior $CAR(\tau_c)$ as in equation 3.1. Here, parameters τ_h and τ_c represent precision. Dividing extra-Poisson variability into 'heterogeneity' and 'clustering' poses a problem: should τ_h and τ_c be too large, the model will be unable to identify the two random effects. Indeed, priors on variance must be carefully chosen in order to allow for identifiability of θ and ϕ, which poses an existential question as to the utility of these models in the first place. Instead, other specifications including Leroux, Lei, and Breslow 2000 and the very closely related BYM2 model (Riebler et al. 2016) as implemented by Morris et al. 2019, which introduce a convolution of the spatial and aspatial error terms which allows for identifiability. The BYM2 model is a similar

---

[4] Related to the SAR models is the field of Spatial Econometrics (see Anselin 2003 for an overview). Spatial Econometrics uses the SAR model in a maximum likelihood regression framework. Many related specifications form the suite of Spatial Econometric models, each with their own implication about spatial correlative structure and dependence (Golgher and Voss 2016)

Poisson-GLM framework to the original BYM model, but replacing the pair of random effects with a convolved error term:

$$\phi + \theta = \left((\sqrt{\rho/s})\phi^* + (\sqrt{1-\rho})\theta^* \right) \sigma \tag{3.4}$$

where $\rho \in [0, 1]$ represents the proportion of variance that comes from the spatial clustering random effect and how much comes from the heterogeneity random effect; ϕ^* is the ICAR distribution; $\theta^* \sim N(0, n)$ where n is the number of connected subgraphs (in our application $n = 1$), s is the scaling factor such that $Var(\phi) \approx 1$ (Riebler et al. 2016); and $\sigma > 0$ is the overall standard deviation for the combined error terms (Morris et al. 2019[5]). The BYM2 model improves upon the original form by allowing for independent definitions of the two prior distributions without involving ρ in the sampling process as is done for the proper CAR model. In doing so, this model involves the identification of only a single set of random effects rather than a pair of independent random effects, which improves identification by separating the dependency structure. Further, this avoids the need for informative priors in the BYM model as emphasized by Banerjee, Carlin, and Gelfand 2015. As well, in this context ρ has an informative interpretation, although it still does not map onto other indicators like Moran's I. Morris et al. report that Stan's Hamiltonian Monte Carlo (HMC) and No U-Turn Sampler (NUTS) provide faster and more precise inference with the BYM2 model than other samplers like WINBUGS and JAGS. The related Leroux (2000) model, which is similar to the BYM2 model but specifies the neighborhood matrix differently, has been shown through simulation to be superior to the original BYM model, and is employed by many in disease mapping studies (Lee, 2011).

Extending the BYM model to include temporal effects

We follow the Knorr-Held (Knorr-Held 2000) framework for Bayesian spatio-temporal modeling. The Knorr-Held model adds time structure in a way that mirrors the BYM model (3.3) by adding temporally autoregressive effects α, temporal random effects γ, and a spatio-temporal interaction term δ:

$$\begin{aligned} Y_i|\psi_i &\overset{iid}{\sim} Poisson(E_i e^{\psi_i}) \\ \psi_i &= \mu + \gamma_t + \alpha_t + \theta_i + \phi_i + \delta_{it} \end{aligned} \tag{3.5}$$

Where μ is the overall intercept, γ is an unstructured temporal component distributed $N(0, \sigma_\gamma)$, α is a structured temporal component that can be specified as an AR(1) or AR(2) process, and δ is a spatio-temporal interaction term. ϕ and θ are as described above. Effectively, the Knorr-Held model decomposes the overall pattern into a global temporal trend,

---

[5] In their 2019 paper, Morris et al use a logit-normal prior for ρ, which has mass around either extreme, indicating that the value of ρ should be close to 0 or 1 and is less likely to be in the middle. However, in a 2018 case study predating the publication, the same authors use a $Beta(1/2, 1/2)$ prior, which has a similar U-shape.

a global spatial trend, and an interaction term between the two, in a procedure similar to ANOVA. Prior choice of δ is not straightforward, and requires careful thought about the relationship of space and time in the model. Knorr-Held (2000) lays out four types of priors, depending on the hypothesized interaction of the spatial and temporal dimensions: (I) where all interaction terms are *a priori* independent; (II) where interactions are autoregressive in time but independent in space; (III) where interactions are autoregressive in space but independent in time; and (IV) where interactions are totally dependent in both space and time. Further description of these interaction types is given in appendix section C.4. Knorr-Held's (2000) evaluation includes specification of the same disease model with each of the four interaction types, and evaluation of the resultant model by DIC.

Zero-Inflated Poisson models for rare counts

Since Chagas disease is very rare, most entries in our matrix of counts by municipality and year are zero. While a low-rate Poisson may be able to capture this overdispersion of zeros, a more appropriate specification involves the zero-inflated Poisson model (ZIP; Lambert 1992). The zero-inflated Poisson is a mixture model that includes a Bernoulli process generating zeros and a Poisson process that generates counts (but may also generate some zeros). In this way, the zero counts in the data are effectively split into 'structural' zeros, which are generated from the presence or absence of the process of interest, and 'sampling' zeros, which are true random zero-counts in the presence of the Poisson process. Lambert (1992) note that in simulation, Poisson-only models are sufficient for a dataset that contains at most 68% zeros and 3.4% counts greater than 9, and that the ZIP model may be justified on datasets with higher rates of zeros. In our application to Chagas disease, over 99% of municipality-years have a zero count; nonzero entries have an average of 1.65 (95% CI: 1-7) infections. The ZIP distribution is parameterized by Bernoulli probability π and Poisson rate λ:

$$P(y_i|\pi,\lambda) = \begin{cases} \pi + (1-\pi)\cdot Poisson(0|\lambda) & \text{if } y_i = 0 \\ (1-\pi)\cdot Poisson(y_n|\lambda) & \text{if } y_i > 0 \end{cases} \tag{3.6}$$

The ZIP distribution is appropriate in a GLM framework, where parameters π and λ are specific to each observation y_i and are estimated with logit and log link functions, respectively. When writing the probability statement, we can also take advantage of the fact that $Poisson(0|\lambda) = \lambda^0 e^{-\lambda}/0!$ simplifies to $e^{-\lambda}$, clarifying the condition where $y = 0$:

$$P(y_i = 0) = \pi_i + (1-\pi_i)e^{-\lambda_i} \tag{3.7}$$

$$P(y_i = k) = (1-\pi_i)e^{-\lambda_i}\lambda_i^k/k! \tag{3.8}$$

In turn, the central moments of the ZIP distribution are mean $(1-\pi)\lambda$ and variance $\lambda(1-\pi)(1+\pi\lambda)$ (Lambert, 1992).

The ZIP distribution then prompts an additional modeling decision. When used in a GLM framework, covariates can be added to both the Poisson process and the Bernoulli process. Prior studies have done both: Agarwal, Gelfand, and Citron-Pousty 2002 include a spatially-autocorrelated Poisson process, Rathbun and Fei 2006 include a spatially-dependent Bernoulli process, and Ver Hoef and Jansen 2007 include spatial (and temporal) autocorrelation in both parts. For our application to Chagas disease, it is not clear if the spatial process should be included in either the Bernoulli or Poisson process, or in both parts. Is the probability of any appearance of Chagas disease spatially correlated? Undoubtedly, as the primary determinant in risk of Chagas disease is the highly localized distribution of *T. Cruzi*. However, conditional on the presence of *T. Cruzi*, it is less clear *a priori* if the risk of contracting Chagas also a spatially dependent process.

3.2.3 Model: Estimating rates of ACD with zero-inflation and spatial and temporal autoregression

Here, we integrate the components discussed above into a single Bayesian model that allows for spatial and temporal autoregression as well as overdispersion of zeros. Specifically, we use a ZIP likelihood within a spatio-temporal decomposition framework like the one proposed by Knorr-Held (2000). As well, we innovate by introducing the BYM2-type convolution of the unstructured error and spatially-structured heterogeneity to improve identification on the model posed by Knorr-Held (2000). We encountered convergence issues when including the spatial convolution in the Poisson proccess; as a result; this process is defined in the Bernoulli parameters only. In the Poisson process, we include a grand mean, temporally AR(1) time trend, and a spatial fixed effect with a Knorr-Held Type 1 interaction. We chose the probabilistic programming language and software suite `Stan` to estimate the yearly incidence risk of Chagas disease across all municipalities in Brazil between 2000 and 2019. The model is evaluated in `Stan` 2.20 using the `cmdstanr` (Stan Development Team 2023) interface for the `R` programming language, version 4.20. `Stan` was chosen for its speed relative to other probabilistic programming languages, like GeoBUGS (Lunn, Arnold, and Spiegelhalter 2004) or JAGS (Plummer 2003), especially for its ability to evaluate vectorized probability statements. Although Stan lacks the built-in support for spatial models present in BUGS, the computational gains from vectorization and adaptive sampling allow for quick evaluation and convergence of complicated posteriors, with full implementation details elaborated in the supplementary material. To sample Bayesian posteriors, by default Stan uses the No U-Turn Sampler (NUTS), a variant of Hamiltonian Monte Carlo, in contrast to Gibbs sampling used by BUGS and JAGS.

We run two formulations of the model: first, a non-covariate smoothing model used purely to recover latent rates of Chagas disease unadjusted for other causal factors besides population at risk; and second, a model that includes climatological covariates. The two models differ only in the inclusion of the set of covariates.

Beginning with ZIP-distributed likelihood:

$$Y_{ti}|\pi_{ti}, \lambda_{ti} \sim ZIP(\pi_{ti}, Pop_{ti} \cdot \lambda_{ti}) \tag{3.9}$$

Where indices t and i refer to year t between 2000 and 2018 and municipality i between 1 and 5561, the number of municipalities in Brazil. Assuming a logit link for Bernoulli parameter π and Poisson parameter λ, we take the follow GLM equations for π and λ:

$$\text{logit}^{-1}(\pi_{ti}) = \mu_\pi + \alpha_{\pi,t} + \beta_\pi x_{t,i} + \left((\sqrt{\rho_\pi / s})\phi_{\pi,i} + (\sqrt{1-\rho_\pi})\theta_{\pi,i}\right)\sigma_\pi + \delta_{\pi,ti} \tag{3.10}$$

$$\log^{-1}(\lambda_{ti}) = \mu_\lambda + \alpha_{\lambda,t} + \beta_\lambda x_{t,i} + \theta_{\lambda,i}\sigma_\lambda + \delta_{\lambda,ti} \tag{3.11}$$

Where μ indicates the global mean with uninformative, $U(-20, 20)$ prior; α is an AR(1) structured time effect; ϕ is a structured spatial process with an $IAR(1)$ prior; θ is an unstructured spatial error with an independent $N(0,1)$ prior; ρ indicates the proportion of variance that comes from the spatially structured process, with prior $Beta(1/2, 1/2)$ (Mitzi Morris 2018); σ is the variance of the convolved spatial term; and δ is a spatio-temporal interaction with a normal prior at mean 0. Finally, $\beta_. x_{t,i}$ indicates a set of coefficients and covariates, which are absent in the main smoothing model but include a set of environmental covariates in the climate model. For simplicity, the effect of these covariates is assumed to be constant throughout space and time. We follow Knorr-Held's recommendation to drop the unstructured temporal component γ to improve identification of the model and the parameters in equation 3.10 are otherwise same as described in equation 3.5. Finally, we specify uninformative Gamma-distributed hyperpriors for variance as recommended by Knorr-Held 2000:

$$\sigma, \sigma_\alpha, \sigma_u, \sigma_\delta \sim Gamma(2, 1) \tag{3.12}$$

The quantities of interest include the expected number of acute Chagas Disease cases in year t in municipality i, which is given by:

$$E[Y_{t,i}] = (1 - \pi_{t,i})\lambda_{t,i} \cdot Pop_{t,i} \tag{3.13}$$

And the incidence rate:

$$IR_{t,i} = \frac{E[Y_{t,i}]}{Pop_{t,i}} = (1 - \pi_{t,i})\lambda_{t,i} \tag{3.14}$$

Altogether, this model was evaluated in Stan using the CmdStanR interface for the R-language. Four chains were in parallel run for 2000 warmup iterations and 1000 posterior draws, and evaluated in approximately 12 hours on an Apple MacbookAir M1. The corresponding Stan code is included in the supplementary material.

Climate Model

To investigate the relationship of climate with Chagas incidence, we include covariates to the model related to temperature, precipitation, and vegetation. Determining covariates relevant to the incidence of Chagas disease is not straightforward as the climate processes that affect *Triatomines* and the *T. Cruzi* parasite may not necessarily be the same as those governing transmission to humans. Further, interventional strategies limiting transmission have shown an overall decrease in incidence of Chagas disease, which may confound identification of climatic factors influencing transmission. Nonetheless, we have chosen to include the following covariates in our model.

In laboratory settings, it was found that *Triatomines* incubated at warmer temperatures (30C vs. 28C and 26C) mature faster and had higher levels of *T. Cruzi* parasites in stool, although insect mortality did increase slightly (Tamayo et al. 2018). Further, *Triatomines* may be able to adapt to changes in temperature in complex ways (Clavijo-Baquet et al. 2021). Ecological modeling of Chagas Disease in North America indicates that as temperatures rise, the distribution of *Triatomines* may shift towards the north and northeastern part of the region (Garza et al. 2014).

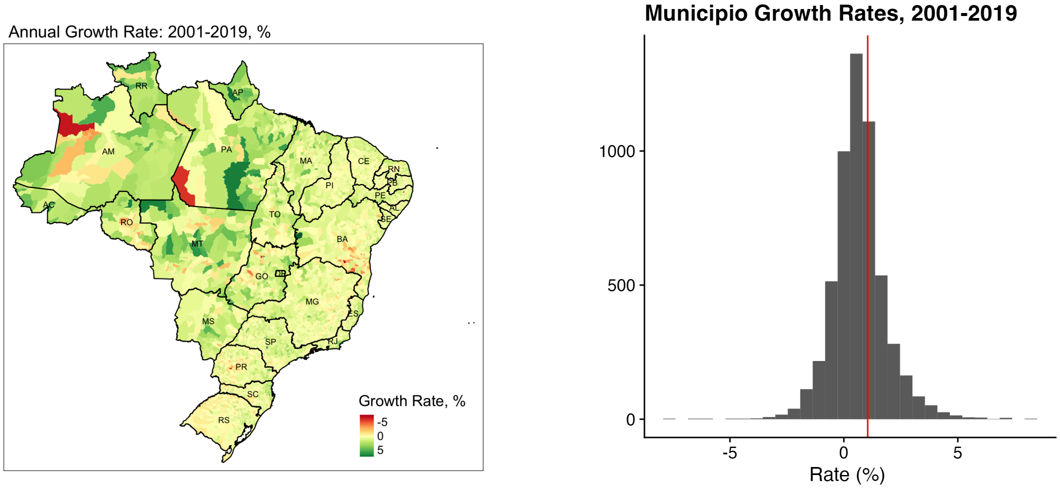
To avoid multi-collinearity among our climatological factors, like Medone et al. 2015 we use Principal Component Analysis (PCA) on the 19 WorldClim Bioclimatic Indicators (Fick and Hijmans 2017), which we retrieved from the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) dataset, "Global Bioclimatic Indicators from 1950-2100 Derived from Climate Projections" (Wouters et al. 2021). PCA is a dimensionality-reducing procedure that decomposes the matrix of covariates by municipality-year into an ordered set of orthogonal vectors, or principal components. Each principal component represents a 'trend' or pattern in the data with the first component representing the most dominant pattern by proportion of variance explained, and each subsequent component representing less of the variance. Each observation can then be described as a linear combination of principal components and coefficients. PCA can be used for Principal Component Regression; rather than using the covariates directly, each observation's location in principal component-space is used as a covariate. Since all principal components are orthogonal, this avoids any potential multi-collinearity in the data.

We find that the first six principal components explained 95% of the variation in the data. Values for these principal components are displayed in supplementary table C.3 and the variance explained in supplementary figure C.3. The largest principal component, responsible for just over 50% of the variance in the dataset, is related to warmer, drier weather year-round. In turn, the second principal component (17% of total variance) is related to cooler temperatures with more seasonal fluctuation but less precipitation year-round. Third and subsequent principal components are less clear in their interpretation and are responsible for decreasing amounts of variance in the dataset.

Projection of future incidence

The resulting quantities estimated from the main smoothing model and the covariate model are used to estimate the incidence of Chagas disease over the 10 year period from 2020 to 2030 using projected population counts and projected climate variables. To project this data, we calculate the average annual exponential growth rate over the period 2001-2019 for each municipality as:

$$r_i = exp\left(\frac{\log Pop_i(2019) - \log Pop_i(2001)}{18}\right) - 1 \tag{3.15}$$



(a) (b)

Figure 3.4: (A): Average annual percentage change in population for all municipalities in Brazil between 2001 and 2019. (B): Average annual growth rate, percent, for all municipalities in Brazil between 2001 and 2019. The overall growth rate for Brazil, indicated in red, is 1.04%, higher than the average municipality growth rate of 0.6%.

Growth rates at the municipality level are displayed in figure 3.4. The overall growth rate of Brazil is 1.04% per year between 2001 and 2019, and the average municipality grew by 0.6%. The 2019 population for each municipality is projected forward each year for 10 years as:

$$Pop_i(t \in [2020 : 2030]) = Pop_i(2019)(1 + r_i)^{t-2019} \tag{3.16}$$

This projected population is used as an input to the model to predict future incidence of Chagas disease. We report two sets of projected rates of Chagas disease: one from the main smoothing model and one from the covariate model. In the covariate model, each municipality-year's location in prinicipal component space for the predicted bioclimatic variables is used as an additional input. The incidence in municipality i at future year $t\star$ is estimated as:

$$E[Y_{t\star,i}] = (1 - \pi_{t\star,i})\lambda_{t\star,i} \cdot Pop_{t\star,i} \tag{3.17}$$

$$\text{logit}^{-1}(\pi_{t\star,i}) = \hat{\mu}_\pi + \alpha_{\pi,t\star} + \hat{\beta}_\pi x_{t\star,i} + \left((\sqrt{\hat{\rho}_\pi / s})\hat{\phi}_{\pi,i} + (\sqrt{1 - \hat{\rho}_\pi})\hat{\theta}_{\pi,i} \right) \hat{\sigma}_\pi + \delta_{\pi,t\star,i} \tag{3.18}$$

$$\log^{-1}(\lambda_{t\star,i}) = \hat{\mu}_\lambda + \alpha_{\lambda,t\star} + \hat{\beta}_\lambda x_{t\star,i} + \hat{\theta}_{\lambda,i}\hat{\sigma}_\lambda + \delta_{\lambda,t\star,i} \tag{3.19}$$

Where all quantities are the median value of those estimated in the main model except time-trend quantity $\alpha_.$ for both α_π and α_λ, which is taken as an AR(1) random walk given the distribution of estimated time trend terms:

$$\alpha_{.,2020} \sim N(\hat{\alpha}_{.,2019}, \hat{\sigma}_{.,\alpha}) \tag{3.20}$$

$$\alpha_{.,2021} \sim N(\alpha_{.,2020}, \hat{\sigma}_{.,\alpha}) \tag{3.21}$$

$$... \tag{3.22}$$

$$\alpha_{.,t\star} \sim N(\alpha_{.,t\star-1}, \hat{\sigma}_{.,\alpha}) \tag{3.23}$$

We conduct 1000 simulated random draws of future incidence and report the predicted rates of Chagas disease.

3.3 Results

3.3.1 Results of Main Smoothing Model

The overall incidence rate of Acute Chagas Disease in Brazil between 2001 and 2019 is estimated to be 0.121 per 100k person-years of life (PYL), although substantial heterogeneity in risk exists between and within regions. Figure 3.5 shows the municipality level 18-year incidence rate of acute Chagas Disease. Estimated incidence of Chagas disease is highly spatially variable with a strong regional trend, with two major areas of vulnerability: first, the northern Amazon states of Amapá (AP) and Pará (PA), which have the highest smoothed incidence rates in the country at 1.80 and 1.69 per 100k PYL—almost an order of magnitude higher than the national average—as well as Acre (AC; 0.317 per 100kPYL) and Amazonas (AM; 0.188 per 100kPYL). These states are highly rural and have a smaller population than the coastal states, but contain the majority of Acute Chagas Disease risk. The second main region of transmission includes the northeastern, Caatinga states of Rio Grande do Norte (RN; 0.334 per 100kPYL), Sergipe (SE; 0.247 per 100kPYL), Piauí (PI; 0.197 per 100kPYL) and Pernambuco (PE; 0.316 per 100kPYL). We do not observe increase transmission rates in the Cerrado, which includes the state of Goiânia and Mato Grosso do Sul, as reported by Gurgel-Gonçalves et al. 2012 besides a slight elevation in Tocantins. The states with the lowest estimated rates of Chagas Disease are the federal district of Brasilia and Sao Paulo.

Within the Amazon states of Pará and Amapá, which have the highest overall rates of new Acute Chagas Disease diagnoses, 31 of 160 municipalities had 18-year incidence rates higher than 1 per 100k PYL; the highest rates of ACD were found in Breves (population 93,000) and Limoeiro do Ajuru (25,000), with 15.9 and 15.8 cases per 100k PYL respectively. Six more municipalities had rates above 10 per 100kPYL: Curralinho, Abaetetuba, Bagre, Muaná, Anajás, and São Sebastião Da Boa Vista. However, the most cases were predicted to be found in Belém, the capital and largest city in Pará, at 386 over the 18 year period for a population of approximately 1.4 million.

A zero-inflated models represent the observed data as being a mixture of two processes: here, the probability of never being exposed to Chagas disease represented through the Bernoulli process, and the incidence rate given exposure represented through the Poisson process. Lambert 1992 refers to the over-dispersion of zeros generated through these pro-

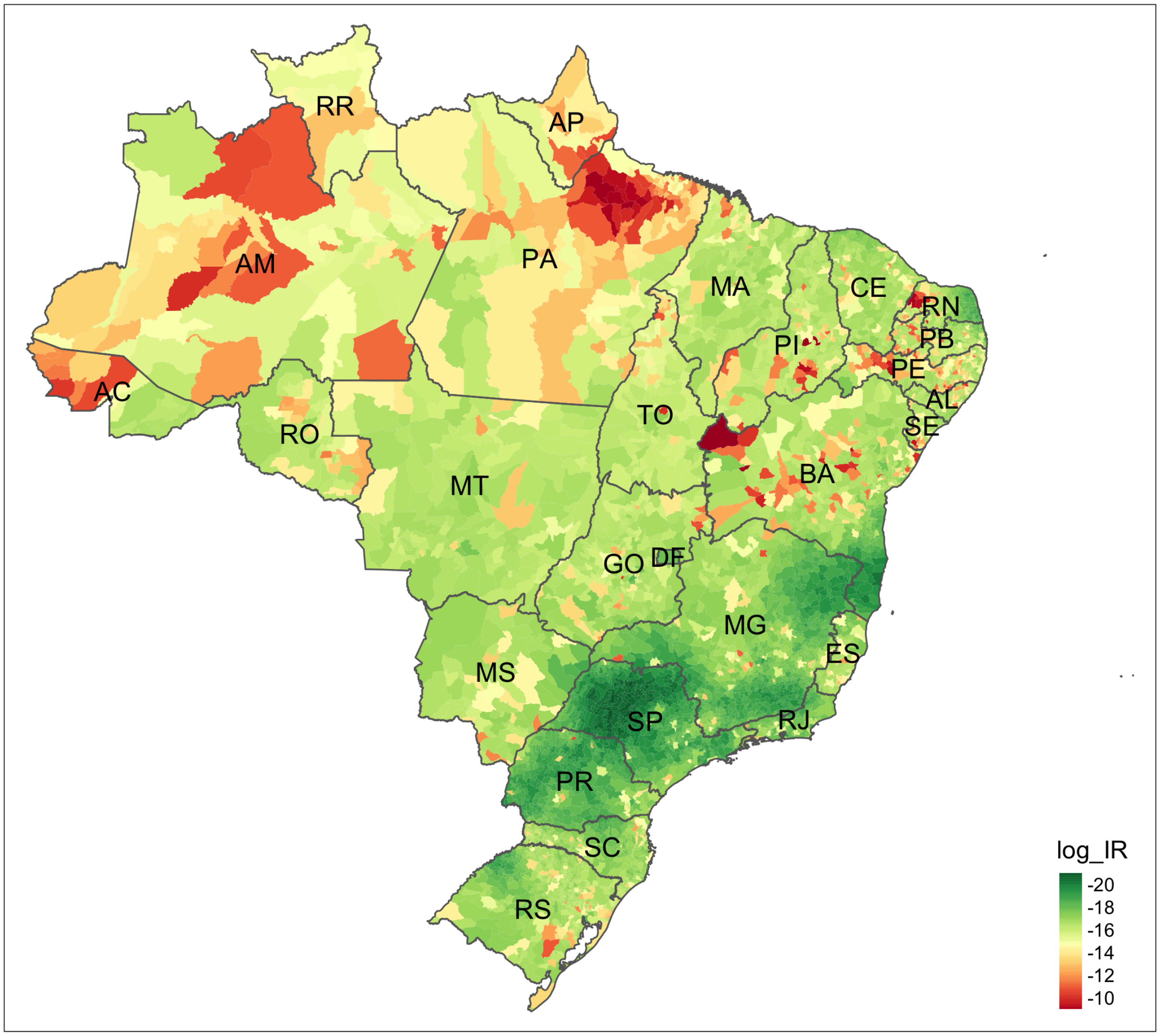


Figure 3.5: Overall log incidence rate at the municipality level over 2001-2019. Red indicates higher rates of ACD and green indicates lower rates of ACD. Incidence rate is calculated as $\log \sum_t \left((1 - \hat{\pi}_{m,t})\hat{\lambda}_{m,t}/Pop_{m,t} \right)$.

cesses as these as 'structural' and 'non-structural'/'sampling' zeros, respectively. Although we found a very strong spatial process governing the rate of 'structural' zeros—probability of never being exposed to Chagas disease (shown in figure 3.6)—we did not find a strong spatial process in the rate given exposure. Since Chagas disease is transmitted to humans given contact with disease-transmitting vectors with a particular habitat, we interpret this to mean that Chagas-carrying *Triatomines* are more likely to live in certain locales, the rate of contact and transmission within those locales is more spatially constant.

Overall, the Northern and Amazon states were found of have a high probability of exposure to Chagas Disease and the coastal and southern states were less likely to have exposure. The total spatial term for π is shown in figure 3.6. Parameter ρ, indicating the proportion of spatial variance derived from the ICAR term, evaluated to 0.985, indicating that the spatially-clustered process was responsible for most of the Chagas incidence and the random process θ contributed very little to the overall distribution, indicating further that the location of *Triatomines* may be driving the location of cases.

Where the probability of exposure to Chagas disease shows a strong spatial pattern, the rate of Chagas—after normalization for population—does not show nearly the degree of spatial autocorrelation as the Bernoulli process. The spatial structure for the Poisson process λ that estimates the rate of disease is shown in supplementary figure C.1. Instead, even though the Poisson process parameter λ is normalized to municipal population, the Poisson process instead appears to be highlighing population locations rather than a spatially-autocorrelated process. However, Moran's I test for spatial autocorrelation did find that the spatial heterogeneity term θ was statistically significantly spatially autocorrelated, albeit weakly ($E[I_0] = -0.0001; I_a = 0.12; p(I_0 < I_a) < 2e - 16$). Future evaluations of this model will need to carefully consider how to incorporate autocorrelation into the Poisson process while maintaining model identification.

The overall time trend parameters α_π and α_λ, which are specified as AR(1) processes, both show a difference from 0 on the linear scale, indicating that there is a global temporal component in both the Bernoulli and Poisson processes (figure C.2A). However, after adding in mean terms μ and applying the logit and log transforms as shown in figure C.2B, the overall time trend tells a different story: the Bernoulli probability π, indicating probability of non-exposure to acute Chagas Disease, decreases from 55% in 2000 to a maximum of 23.6% by 2006 only to increase to nearly 100% for the remainder of the study period. This indicates that over the course of the study period, country-wide exposure to Chagas disease increased before decreasing to nearly 0 after 2007, at which point the distriubtion of Chagas cases ceased to be a country-wide phenomenon and instead became more spatially localized. The Poisson rate is stable around 2e-5 per capita over the course of the study period.

Overall, the model converged well and showed good mixing between the chains for the main parameters π and λ. The Root Mean Squared Error (RMSE) of the model, evaluated as:

$$RMSE = \sqrt{\frac{1}{|T| \cdot |M|} \sum_{t \in T, i \in M} (\hat{y}_{t,i} - y_{t,i})^2}$$

is 0.175. Convergence is evaluated using statistic $\hat{R}$, which evaluates the agreement of

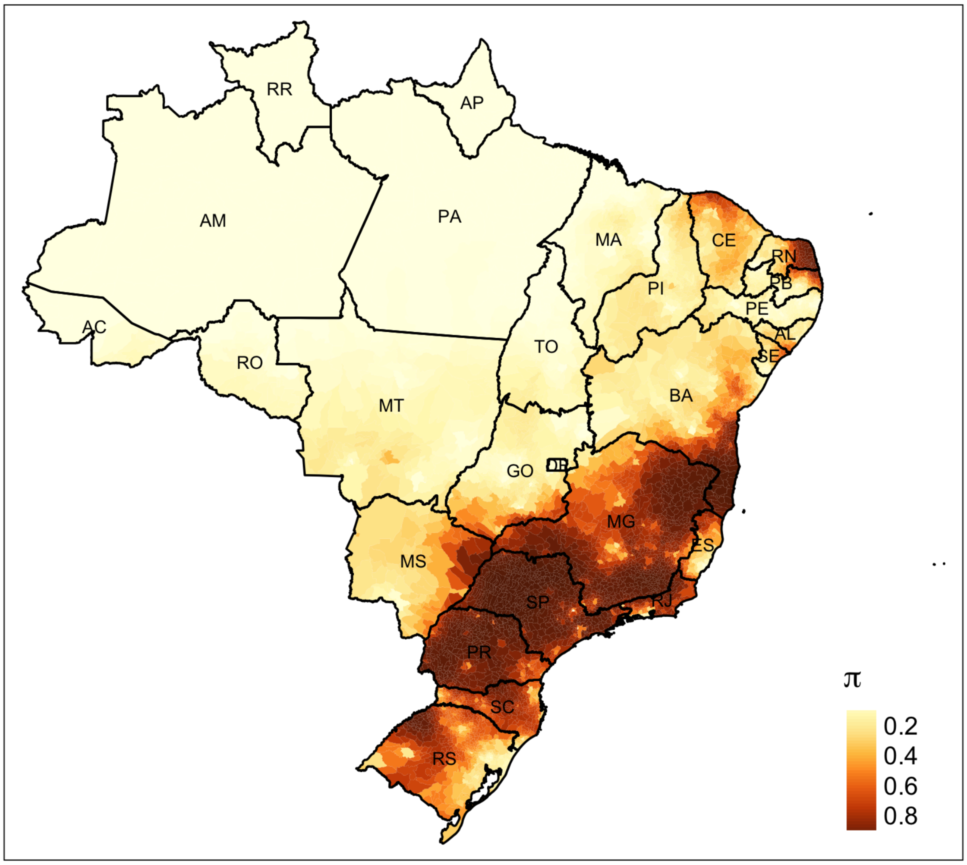


Figure 3.6: Estimated spatial process π governing the probability of an individual never being exposed to Chagas Disease over the study period. This process is an inverse-logit transformation of a linear combination of Conditionally AutoRegressive (CAR) term ϕ_π capturing risk that is spatially clustered, possibly due to Triatomine habitat or contact rates.

between-chain estimates, and Effective Sample Size (ESS), which evaluates the number of samples correcting for autocorrelation. Supplementary table C.1 shows the distribution of $\hat{R}$ and ESS for all parameters $\pi, \lambda, \phi, \theta, \alpha$, and δ.

3.3.2 Results of Climate Model

The climate model includes the specification as the main model above with the inclusion of each municipality-year's location in principal component space among the 19 WorldClim Bioclimatic variables. We include the first six principal component dimensions as covariates β in both the Bernoulli process governing overdispersion of zeros and the Poisson process governing rate of Chagas disease. The RMSE of the climate model is 0.183, which is slightly higher than that of the main model, indicating that controlling for climate produces a worse fit and may reduce the accuracy of the model, possibly due to overfitting. Posterior densities of β are included in supplementary figure C.4. In the Poisson process, posterior estimates of coefficients for the first three principal components were found to be statistically significantly different from zero, whereas in the Bernoulli process, only the second principal component was found to be significantly different from zero. The values of parameters in the climate model are similar to the values in the main model (shown in figure C.5), except for ϕ_π, the spatial clustering term in the Bernoulli process.

To better analyze these climatic factors, we transformed these estimated coefficients from principal component space back to the scale of the original variables before applying the inverse-link function and intercept terms to show the values visualized in figure 3.7. These values are calculated as $g^{-1}(\mu + \hat{\beta}^\star)$, where g is the link function *log* for the Poisson process λ and *logit* for the Bernoulli process π, and $\hat{\beta}^\star$ represents the estimated coefficients transformed from principal component space to the original coefficients. Ultimately, β_π coefficients represent a one-unit change in each variable on the probability of non-exposure to Chagas disease, and β_λ represents the effect of a one-unit change in each variable on the predicted rate of Chagas disease, per million person years, conditional on exposure. Overall, we see that these variables do not affect the rate of Chagas disease, only the probability of non-exposure. Non-exposure to Chagas disease is more likely in climates that are highly seasonal, and less likely in wetter wetter climates.

3.3.3 Projected Rates 2020-2030

We used the main smoothing and climate covariate models to estimate counts of Chagas disease over the decade 2020-2030. As elaborated above, the projection procedure utilizes the estimated intercept and spatial parameters, and using randomly-drawn temporal structure. The main model estimates a median of 4461 cases of Acute Chagas Disease over the decade 2020 (IQR: 1,653 - 13,859), almost double the number of cases in the previous decade (2,612). Predicted incidence is similar when including the bioclimate covariates, estimating a median of 4461 cases (IQR: 1619 - 13,270). Figure 3.8 shows the median annual predicted counts of Chagas disease across Brazil and interquartile range between 2020 and 2030. A map of projected incidence and a comparison of observed and projected rates are shown in figure

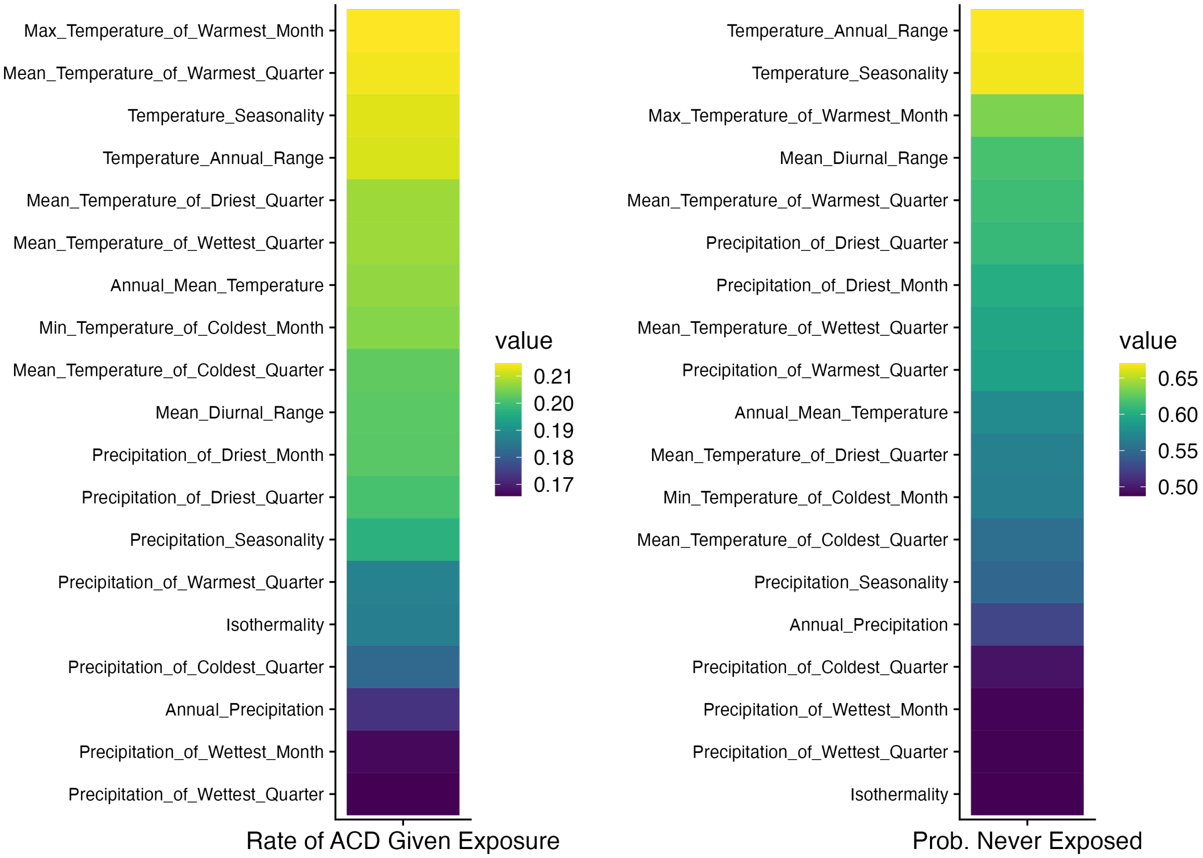


Figure 3.7: Coefficients for 19 WorldClim Bioclimatic Variables used in the climate model estimated for both the Poisson process (λ, left), governing the incidence rate of ACD given exposure, and the Bernoulli process (π, right), governing the probability of never being exposed to ACD. Lighter colors on the left figure indicate that a higher value of the coefficient corresponds to a higher rate of ACD in the population where exposure is present, and on the right figure indicate that probability of never being exposed is higher. Coefficients are estimated in principal component space and transformed to the natural scale, and applied the corresponding log^{-1} link function for the Poisson process and $logit^{-1}$ for the Bernoulli process.

3.9 and for selected municipalities in figure 3.10 and 3.11. Most of 'hot spots' for new cases are predicted to be in the same locations as 2001-2019, including Abaetetuba, Belém, and Breves in Pará, and Macapá in Amapá. However, the largest increases are projected to be in smaller, rural municipalities with high growth rates in the states of Amazonas—especially municipalities Apuí and Tefé—and Piauí. The climate covariate model implies slightly lower rates than the main model, implying that projected bioclimatic conditions may result in fewer infections, although the overall difference is likely small. The trend observed in figure 3.8 shows a highly variable trajectory year to year—much more than the main model predicts—indicating that annual climate fluctuations may have substantial effects on predicted rates.

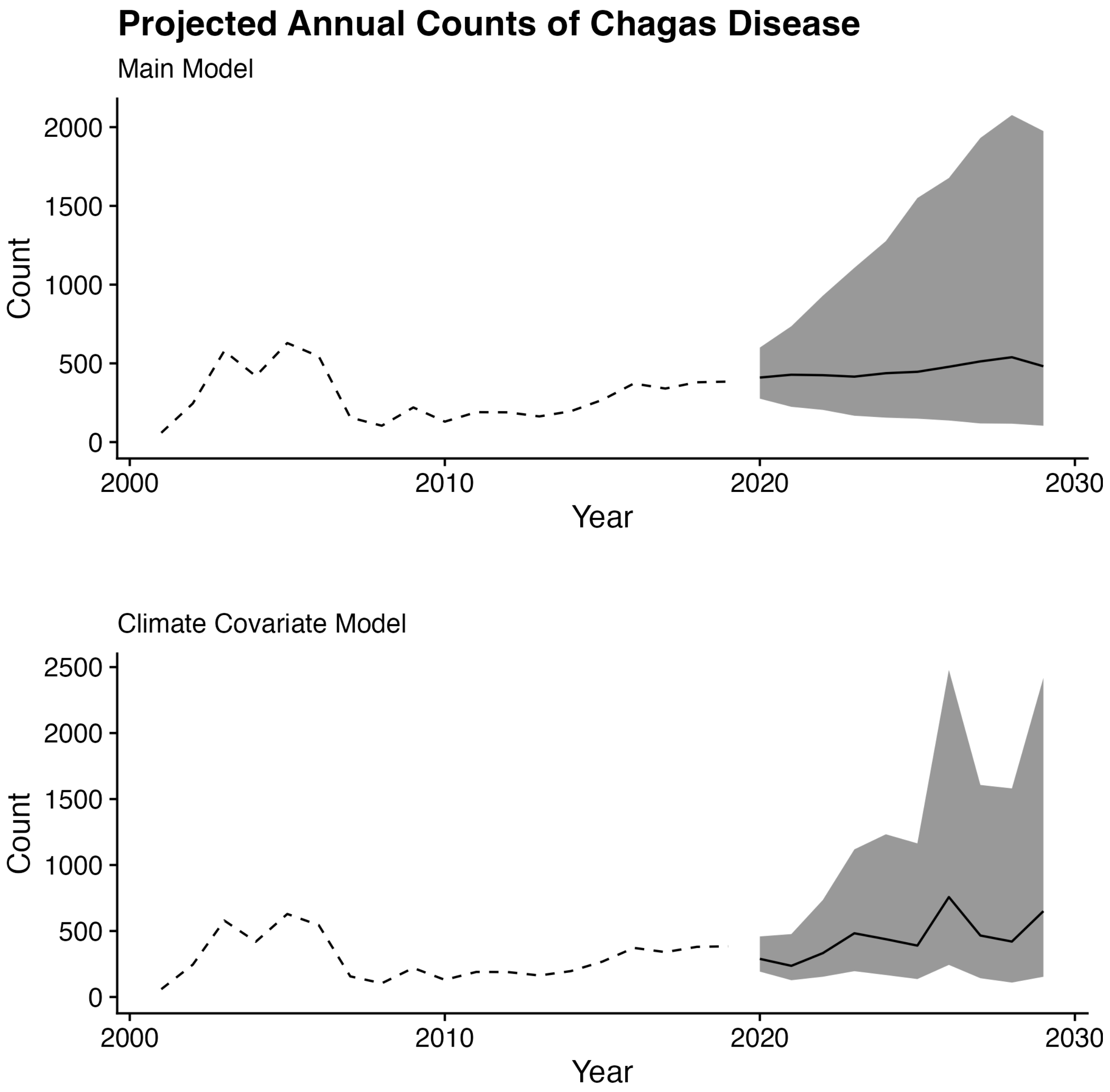
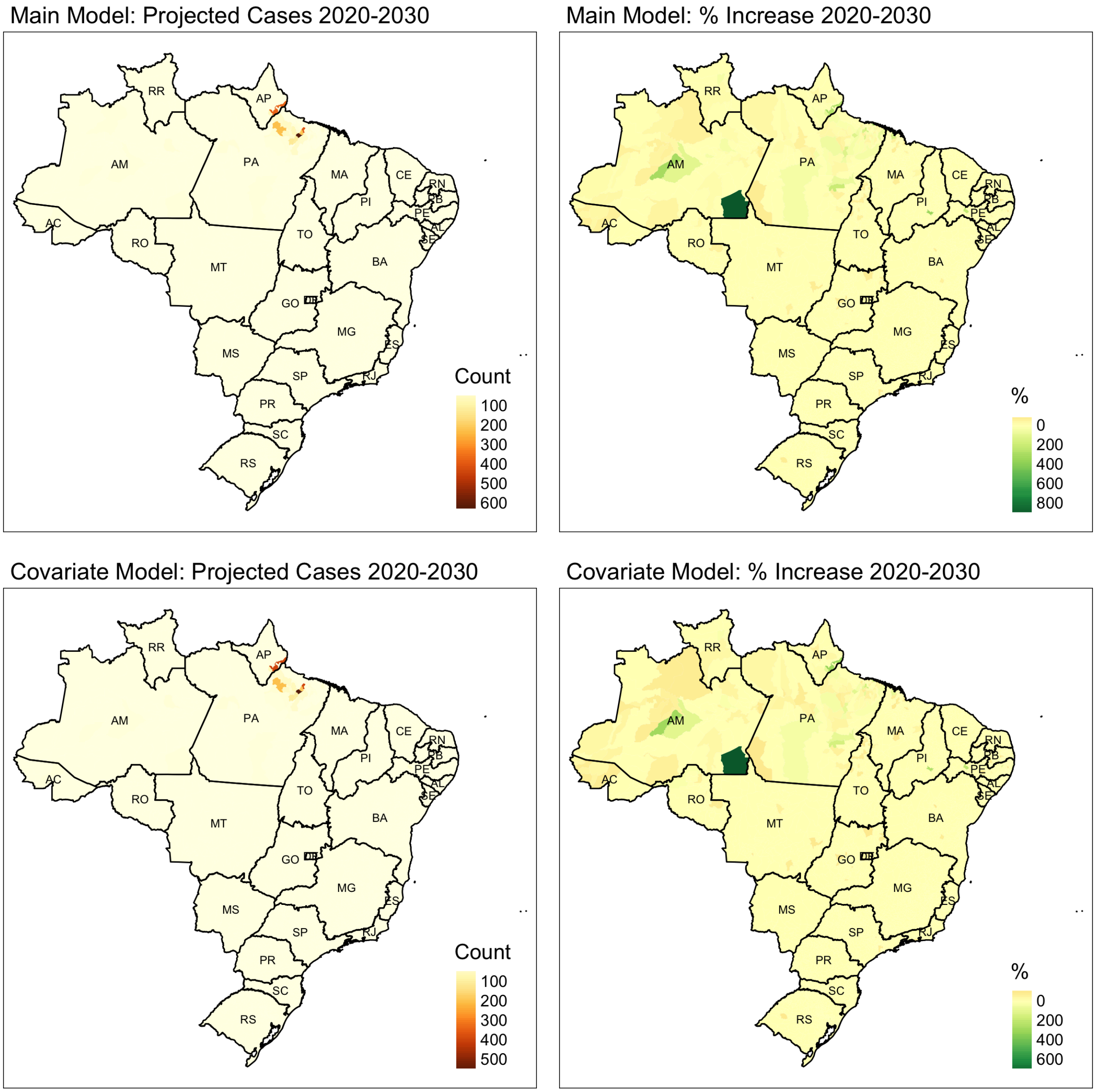


Figure 3.8: Observed counts across Brazil, 2001-2019, summary of 1000 projected counts, 2020-2030, from the main smoothing model (top) and climate covariate model (bottom). Median simulated counts are shown in black and interquartile range, representing 50% of simulations, is shown in grey.



Main Model: Projected Cases 2020-2030

Main Model: % Increase 2020-2030

Covariate Model: Projected Cases 2020-2030

Covariate Model: % Increase 2020-2030

Figure 3.9: Projected incidence and percent increase compared to the previous decade.

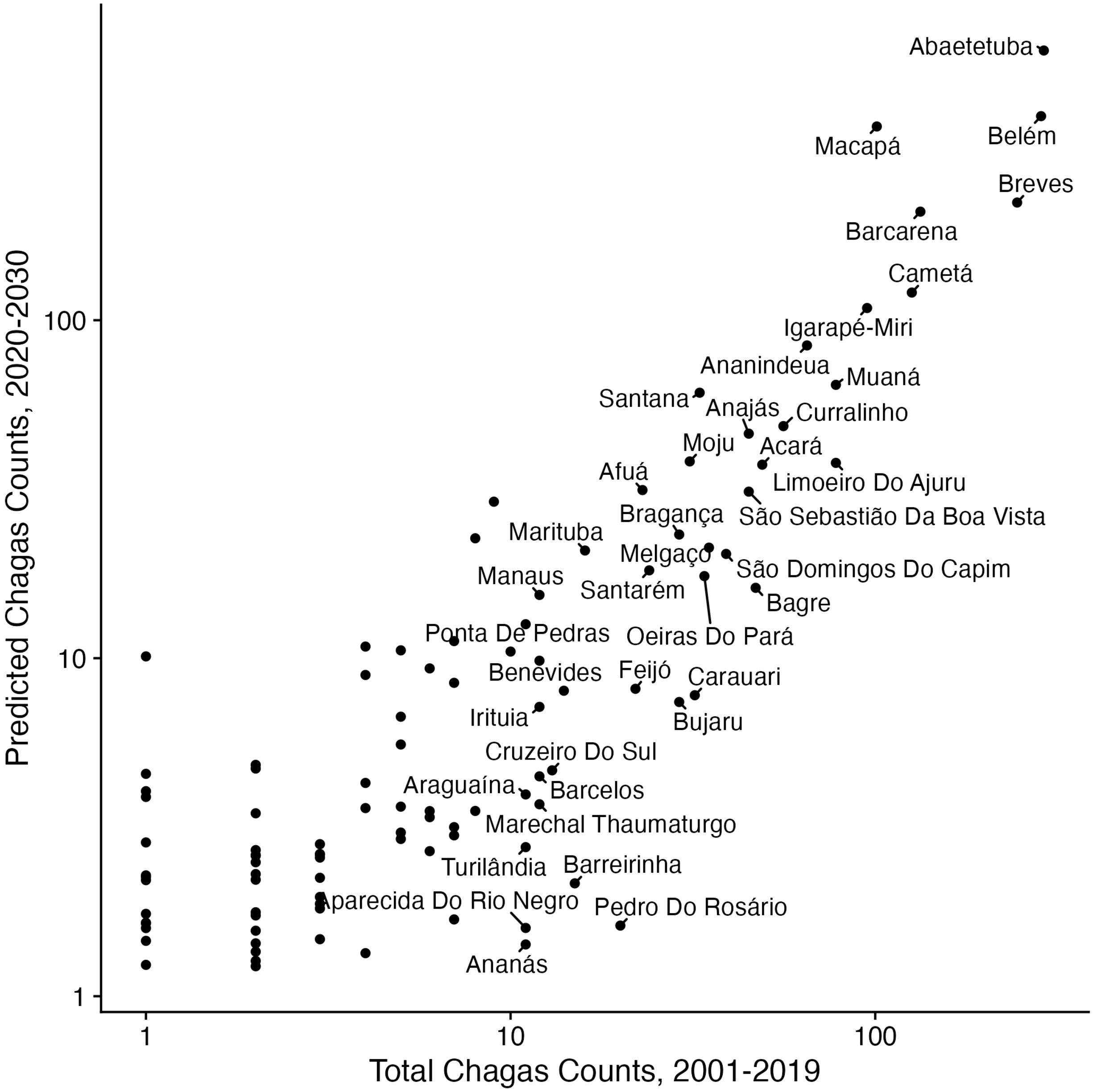


Figure 3.10: Observed and Projected rates, main model

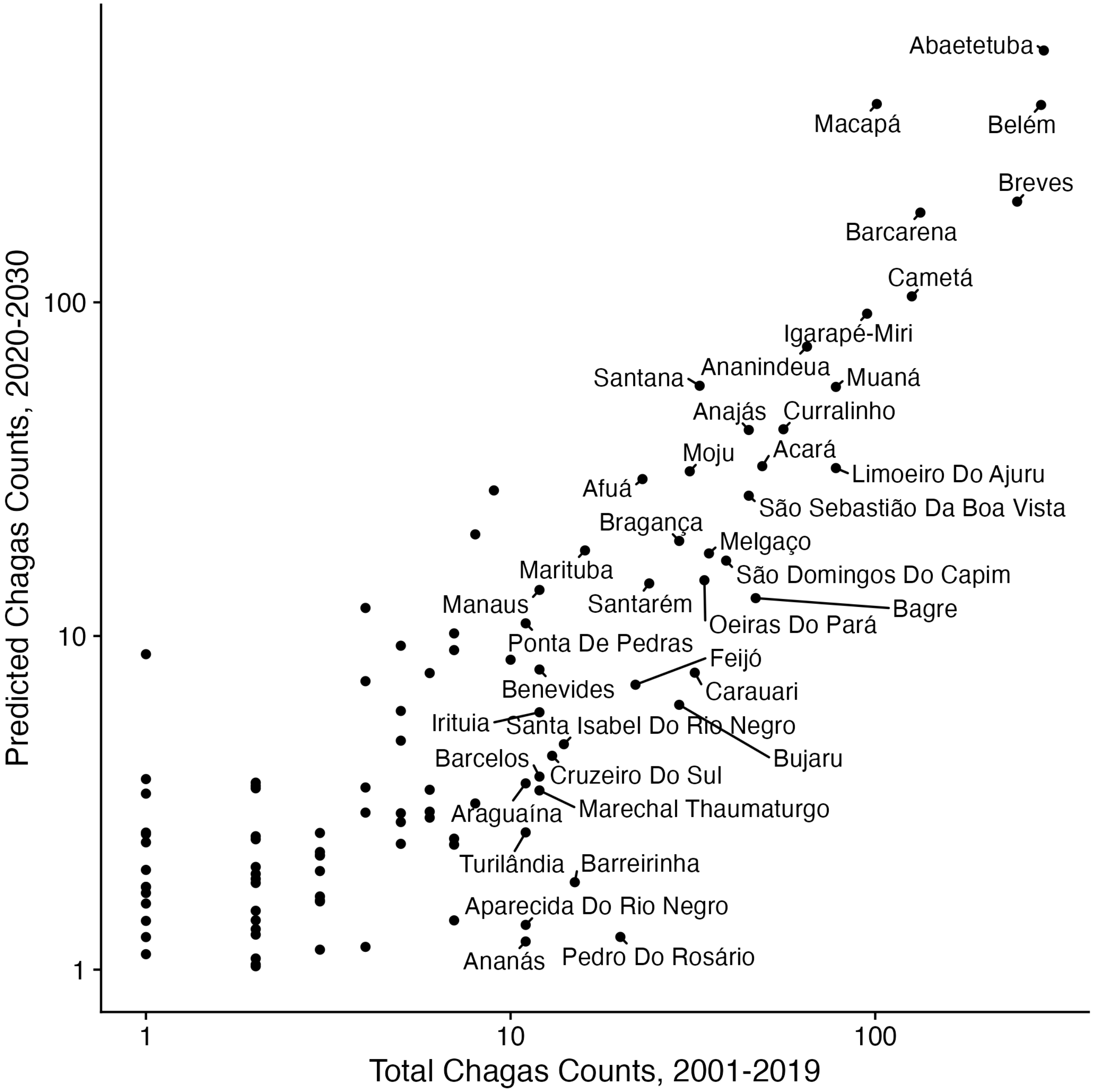


Figure 3.11: Observed and Projected rates, climate model

3.4 Discussion

Despite progress towards elimination, Chagas disease remains a significant threat to public health, and additional intervention will be needed to further reduce rates of new ACD cases in Brazil. The results of our modeling imply that rates of ACD will increase consistently over the ensuing decade—potentially as much as doubling compared to the previous decade—driven primarily by an increasing population in high-risk areas. Climate change may result in the exacerbation of this trend. However, this appears to be at odds with the country-wide decrease in ACD rates as a result of *Triatomine* eradication campaigns since the 1980's. While these campaigns have been enormously successful, risk of ACD is likely to persist without additional intervention. As a result, our work here serves as a call to continue the campaigns Nonetheless, we predict that risk for ACD will persist, and further intervention will be necessary to continue the decrease observed in previous decades.

This model has a number of limitations, some of which may be addressable in future modeling studies. First, our data include official reports of Acute Chagas Disease as submitted to SINAN. We were unable to find literature estimates of under-reporting rates of ACD not submitted to SINAN. It is possible that there are many annual cases not captured in the dataset, implying that our estimates are not only incomplete, but subject to a 'survivorship paradox' where the locations of highest epidemiological interest are not captured in the dataset. Since the treatment success rate is high for acutely diagnosed cases of Chagas disease, it may be reasonable to estimate the number of 'missed' acute cases from backwards-projection of chronic cases. However, due to the substantial lag time (at least 10-30 years) between exposure and chronic symptoms, this was not possible with the data available from SINAN. However, should this topic be revisited in the following decades, a back-projection approach may be useful for retrospectively estimating the underreporting rate.

Our model uses Knorr-Held Type I spatio-temporal interactions δ, which are the least sophisticated of the structures outlined by Knorr-Held (2000). This strategy essentially estimates uncorrelated space-time fixed effects, which we found to be ideal for precise internal estimation; however, since these fixed effects lack a temporally autoregressive definition, it is not possible to use these fixed effects to project the model to predict future incidence without introducing meaningless statistical noise. In our prediction we simply dropped the interaction term from the model; in effect, only the un-interacted spatial and simulated random walk temporal terms were used for future projection. We believe that using the basic Type I interactions allowed for better estimation of the independent spatial and temporal components as we found that the Type IV term is only weakly identified in our model; so this decision was not without benefit. Choice of a spatio-temporal interaction that includes a temporally autoregressive term, such as Type II or Type IV (see supplementary material for more elaboration), would allow for projection in the form of a random walk through interaction space-time. For example, should a Type IV prior be used that assumes total spatial and temporal dependence in the interaction, future space-time interaction terms can be simulated as a random draw from a Multivariate Normal distribution with a variance-covariance matrix derived from the previously-estimated interaction terms, similar to the simulation procedure for spatially autocorrelated data (Banerjee, Carlin, and Gelfand 2015).

Special care will likely be needed to assure propriety of the Type IV distribution, which is outside of the scope of the present study.

Our population projection would benefit from a more precise estimation methodology than the crude exponential growth model used here. Future analysis should consider municipality, year, and age-specific rates of fertility, mortality, and migration—especially internal migration—to inform population projections. We were unable to obtain these quantities at the level of spatial and temporal granularity required. We believe that in the short term—namely, the single decade between 2020 and 2030—this crude methodology allows for understanding how heterogeneity in growth rates may relate to future incidence of Acute Chagas Disease. Nonetheless, it is not suitable for long-term projections. Model-based estimates that utilize the readily-available state-level rates to determine small-area estimates, possibly similar to the Lee-Carter (1992) procedure or the one used by Alexander, Zagheni, and Barbieri 2017, may allow for more precise estimation of future municipality-level population. As well, an extension of the model to predict age-specific rates of Chagas disese may aide researchers in planning for interventions.

Finally, we make the critical assumption in both models that population and climate change are the sole drivers of future incidence. Other factors, such as housing construction materials, poverty, habitat destruction, and residential or industrial development encroaching on *Triatomine* habitats may increase affect rates of Chagas disease even in the absence of population or climate change. Further, the inherent assumption of linearity in our model assumes that as climate and population change, predicted incidence of Chagas disease will respond. This may not be the case: while climate may partially determine the geographic distribution of *Triatomines*, which may in turn affect incidence, it is likely that the relationship of climate with *Triatomine* populations is too complex to be captured by a linear model of the sort used here. Since there are many species of *Triatomines*, each with different habitats, behaviors, and virulence, the ultimate effect of climate on Chagas incidence is undoubtedly complex and nonlinear. Further, our climate model has a slightly higher error than the main model despite the inclusion of additional covariates, indicating that the model may be suffering from overfitting. As well, many exogenous factors could affect the distribution of *Triatomines* under future climate conditions, including interventional strategies to limit *Triatomine* habitats like the residential insecticidal campaigns of the 1980s, development of urbanization and infrastructure, and climate adaptations, environmental destruction and conservation practices that may affect *Triatomine habitats.* Should future climate conditions create new habitats for *Triatomines*, it is not clear at present if the insects are mobile enough to find these habitats, or if accidental importation by humans—such as improper handling of lumber—may catalyze a shift in *Triatomine* distribution.

Replication Code

Replication code is publicly available at `https://github.com/eroubenoff/chagas_modeling`.

Bibliography

Agarwal, Deepak K., Alan E. Gelfand, and Steven Citron-Pousty (2002). “Zero-inflated models with application to spatial count data”. In: *Environmental and Ecological Statistics* 9.4, pp. 341–355. ISSN: 1573-3009. DOI: `10.1023/A:1020910605990`. URL: `https://doi.org/10.1023/A:1020910605990` (visited on 04/11/2023).

Ajelli, Marco et al. (2010). “Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models”. In: *BMC Infectious Diseases* 10.1, p. 190. ISSN: 1471-2334. DOI: `10.1186/1471-2334-10-190`. URL: `https://doi.org/10.1186/1471-2334-10-190` (visited on 09/01/2021).

Albert-László Barabási (2021). *Network Science by Albert-László Barabási.* URL: `http://networksciencebook.com/` (visited on 10/06/2021).

Alexander, Monica, Emilio Zagheni, and Magali Barbieri (2017). “A Flexible Bayesian Model for Estimating Subnational Mortality”. In: *Demography* 54.6, pp. 2025–2041. ISSN: 1533-7790. DOI: `10.1007/s13524-017-0618-7`. URL: `https://doi.org/10.1007/s13524-017-0618-7` (visited on 12/15/2019).

Altcheh, Jaime et al. (2011). “Adverse events after the use of benznidazole in infants and children with Chagas disease”. In: *Pediatrics* 127.1, e212–218. ISSN: 1098-4275. DOI: `10.1542/peds.2010-1172`.

Altizer, Sonia et al. (2006). “Seasonality and the dynamics of infectious diseases”. In: *Ecology Letters* 9.4, pp. 467–484. ISSN: 1461-0248. DOI: `10.1111/j.1461-0248.2005.00879.x`.

Andrasfay, Theresa and Noreen Goldman (2021). “Reductions in 2020 US life expectancy due to COVID-19 and the disproportionate impact on the Black and Latino populations”. In: *Proceedings of the National Academy of Sciences* 118.5. ISSN: 0027-8424, 1091-6490. DOI: `10.1073/pnas.2014746118`. URL: `https://www.pnas.org/content/118/5/e2014746118` (visited on 10/25/2021).

Anselin, Luc (2003). “Spatial Externalities, Spatial Multipliers, And Spatial Econometrics”. In: *International Regional Science Review* 26.2, pp. 153–166. ISSN: 0160-0176. DOI: `10.1177/0160017602250972`. URL: `https://doi.org/10.1177/0160017602250972` (visited on 08/04/2020).

Anselin, Luc and Bera, A (1998). “Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics”. In: *Handbook of Applied Economic Statistics*. Ed. by Ullah, A and Giles, D. New York: Marcel Dekker, p. 53.

Arduin, Hélène et al. (2017). “An agent-based model simulation of influenza interactions at the host level: insight into the influenza-related burden of pneumococcal infections”.

In: *BMC Infectious Diseases* 17.1, p. 382. ISSN: 1471-2334. DOI: `10.1186/s12879-017-2464-z`. URL: `https://doi.org/10.1186/s12879-017-2464-z` (visited on 06/05/2023).

Baden, Lindsey R. et al. (2021). "Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine". In: *The New England Journal of Medicine* 384.5, pp. 403–416. ISSN: 1533-4406. DOI: `10.1056/NEJMoa2035389`.

Baker, Jack D. et al. (2013). "An evaluation of the accuracy of small-area demographic estimates of population at risk and its effect on prevalence statistics". In: *Population Health Metrics* 11.1, p. 24. ISSN: 1478-7954. DOI: `10.1186/1478-7954-11-24`. URL: `https://doi.org/10.1186/1478-7954-11-24` (visited on 04/01/2020).

Baker, Marissa G., Trevor K. Peckham, and Noah S. Seixas (2020). "Estimating the burden of United States workers exposed to infection or disease: A key factor in containing risk of COVID-19 infection". In: *PloS One* 15.4, e0232452. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0232452`.

Baker, Rachel E. et al. (2020). "Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic". In: *Science* 369.6501, pp. 315–319. DOI: `10.1126/science.abc2535`. URL: `https://www.science.org/doi/10.1126/science.abc2535` (visited on 08/01/2023).

Banerjee, Sudipto, Bradley P. Carlin, and Alan E. Gelfand (2015). *Hierarchical modeling and analysis for spatial data*. Second edition. Monographs on statistics and applied probability 135. Boca Raton: CRC Press, Taylor & Francis Group. 562 pp. ISBN: 978-1-4398-1917-3.

Bansal, Shweta, Bryan T Grenfell, and Lauren Ancel Meyers (2007). "When individual behaviour matters: homogeneous and network models in epidemiology". In: *Journal of the Royal Society Interface* 4.16, pp. 879–891. ISSN: 1742-5689. DOI: `10.1098/rsif.2007.1100`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2394553/` (visited on 04/01/2020).

Bansal, Shweta et al. (2010). "The dynamic nature of contact networks in infectious disease epidemiology". In: *Journal of Biological Dynamics* 4.5, pp. 478–489. ISSN: 1751-3758. DOI: `10.1080/17513758.2010.503376`. URL: `https://doi.org/10.1080/17513758.2010.503376` (visited on 08/01/2023).

Berkman, Lisa and Krishna, Aditi (2014). "Social Network Epidemiology". In: *Social epidemiology*. Ed. by Lisa F. Berkman, Ichirō Kawachi, and M. Maria Glymour. Second edition. Oxford: Oxford University Press. ISBN: 978-0-19-939533-0.

Berkman, Lisa F and S Leonard Syme (1979). "Social Networks, Host Resistance, and Mortality: A Nine-Year Follow-Up Study of Alameda County Residents". In: *SOCIAL NETWORKS*, p. 19.

Bern, Caryn et al. (2007). "Evaluation and Treatment of Chagas Disease in the United StatesA Systematic Review". In: *JAMA* 298.18, pp. 2171–2181. ISSN: 0098-7484. DOI: `10.1001/jama.298.18.2171`. URL: `https://doi.org/10.1001/jama.298.18.2171` (visited on 04/25/2023).

Besag, Julian, Jeremy York, and Annie Mollie (1991). "Bayesian image restoration, with two applications in spatial statistics". In: *Annals of the Institute of Statistical Mathematics*

43.1, pp. 1–20. ISSN: 0020-3157, 1572-9052. DOI: `10.1007/BF00116466`. URL: `http://link.springer.com/10.1007/BF00116466` (visited on 03/24/2020).

Best, Nicky, Sylvia Richardson, and Andrew Thomson (2005). “A comparison of Bayesian spatial models for disease mapping”. In: *Statistical Methods in Medical Research* 14.1, pp. 35–59. ISSN: 0962-2802. DOI: `10.1191/0962280205sm388oa`. URL: `https://doi.org/10.1191/0962280205sm388oa` (visited on 06/23/2020).

Blackwood, Julie and Lauren Childs (2018). “An Introduction to Compartmental Modeling for the Budding Infectious Disease Modeler”. In: *Letters in Biomathematics* 5.1. ISSN: 23737867. DOI: `10.30707/LiB5.1Blackwood`. URL: `https://lettersinbiomath.journals.publicknowledgeproject.org/index.php/lib/article/view/81` (visited on 06/15/2020).

Bolker, BM and BT Grenfell (1993). “Chaos and biological complexity in measles dynamics”. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 251.1330, pp. 75–81. ISSN: 0962-8452, 1471-2954. DOI: `10.1098/rspb.1993.0011`. URL: `https://royalsocietypublishing.org/doi/10.1098/rspb.1993.0011` (visited on 06/05/2023).

Bonabeau, Eric (2002). “Agent-based modeling: Methods and techniques for simulating human systems”. In: *Proceedings of the National Academy of Sciences* 99 (suppl 3), pp. 7280–7287. ISSN: 0027-8424, 1091-6490. DOI: `10.1073/pnas.082080899`. URL: `https://www.pnas.org/content/99/suppl_3/7280` (visited on 09/01/2021).

Braveman, Paula and Laura Gottlieb (2014). “The Social Determinants of Health: It’s Time to Consider the Causes of the Causes”. In: *Public Health Reports* 129.1, pp. 19–31. ISSN: 0033-3549. DOI: `10.1177/00333549141291S206`. URL: `https://doi.org/10.1177/00333549141291S206` (visited on 06/14/2020).

Bruch, Elizabeth and Jon Atwell (2015). “Agent-Based Models in Empirical Social Research”. In: *Sociological Methods & Research* 44.2, pp. 186–221. ISSN: 0049-1241. DOI: `10.1177/0049124113506405`. URL: `https://doi.org/10.1177/0049124113506405` (visited on 09/01/2021).

Brüssow, Harald (2022). “COVID-19: Omicron – the latest, the least virulent, but probably not the last variant of concern of SARS-CoV-2”. In: *Microbial Biotechnology*. ISSN: 1751-7915. DOI: `10.1111/1751-7915.14064`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/1751-7915.14064` (visited on 05/18/2022).

Bubar, Kate M. et al. (2021). “Model-informed COVID-19 vaccine prioritization strategies by age and serostatus”. In: *Science*. URL: `https://www.science.org/doi/abs/10.1126/science.abe6959` (visited on 09/07/2021).

Buckee, Caroline O., Andrew J. Tatem, and C. Jessica E. Metcalf (2017). “Seasonal Population Movements and the Surveillance and Control of Infectious Diseases”. In: *Trends in Parasitology* 33.1, pp. 10–20. ISSN: 1471-4922. DOI: `10.1016/j.pt.2016.10.006`. URL: `https://www.sciencedirect.com/science/article/pii/S1471492216301891` (visited on 06/05/2023).

Buckner, Jack H., Gerardo Chowell, and Michael R. Springborn (2021). “Dynamic prioritization of COVID-19 vaccines when social distancing is limited for essential workers”. In: *Proceedings of the National Academy of Sciences* 118.16, e2025786118. ISSN: 0027-8424, 1091-6490. DOI: `10.1073/pnas.2025786118`. URL: `http://www.pnas.org/lookup/doi/10.1073/pnas.2025786118` (visited on 02/01/2022).

Buitrago-Garcia, Diana et al. (2022). "Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: Update of a living systematic review and meta-analysis". In: *PLOS Medicine* 19.5, e1003987. ISSN: 1549-1676. DOI: `10.1371/journal.pmed.1003987`. URL: `https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1003987` (visited on 07/04/2023).

Buonomo, B., N. Chitnis, and A. d'Onofrio (2018). "Seasonality in epidemic models: a literature review". In: *Ricerche di Matematica* 67.1, pp. 7–25. ISSN: 1827-3491. DOI: `10.1007/s11587-017-0348-6`. URL: `https://doi.org/10.1007/s11587-017-0348-6` (visited on 05/22/2023).

Canals, Mauricio et al. (2017). "Modeling Chagas disease in Chile: From vector to congenital transmission". In: *Biosystems* 156-157, pp. 63–71. ISSN: 0303-2647. DOI: `10.1016/j.biosystems.2017.04.004`. URL: `http://www.sciencedirect.com/science/article/pii/S0303264716302283` (visited on 02/03/2021).

Carlier, Yves et al. (2011). "Congenital Chagas disease: recommendations for diagnosis, treatment and control of newborns, siblings and pregnant women". In: *PLoS neglected tropical diseases* 5.10, e1250. ISSN: 1935-2735. DOI: `10.1371/journal.pntd.0001250`.

Cassel, John (1976). "The Contribution of the Social Environment to Host Resistance". In: *American Journal of Epidemiology* 104.2, pp. 107–123. ISSN: 1476-6256, 0002-9262. DOI: `10.1093/oxfordjournals.aje.a112281`. URL: `https://academic.oup.com/aje/article-lookup/doi/10.1093/oxfordjournals.aje.a112281` (visited on 01/14/2022).

CDC (2020). *COVID Data Tracker*. Centers for Disease Control and Prevention. URL: `https://covid.cdc.gov/covid-data-tracker` (visited on 09/09/2021).

Centers for Disease Control and Prevention (2021). *Science Brief: SARS-CoV-2 Infection-induced and Vaccine-induced Immunity*. Science Brief: SARS-CoV-2 Infection-induced and Vaccine-induced Immunity. URL: `https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/vaccine-induced-immunity.html` (visited on 07/04/2023).

Cevallos, Ana María and Roberto Hernández (2014). "Chagas' Disease: Pregnancy and Congenital Transmission". In: *BioMed Research International* 2014, p. 401864. ISSN: 2314-6133. DOI: `10.1155/2014/401864`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4052072/` (visited on 04/25/2023).

Chang, Serina et al. (2021). "Mobility network models of COVID-19 explain inequities and inform reopening". In: *Nature* 589.7840, pp. 82–87. ISSN: 1476-4687. DOI: `10.1038/s41586-020-2923-3`. URL: `https://www.nature.com/articles/s41586-020-2923-3` (visited on 01/19/2022).

Chin, Alex W H et al. (2020). "Stability of SARS-CoV-2 in different environmental conditions". In: *The Lancet Microbe* 1.1, e10. ISSN: 2666-5247. DOI: `10.1016/S2666-5247(20)30003-3`. URL: `https://www.sciencedirect.com/science/article/pii/S2666524720300033` (visited on 06/05/2023).

Choi, Joyce et al. (2020). "An exploration of land use and poverty as an integrative model for mitigating Chagas disease in El Salvador". In: *BIOS* 91.2, pp. 125–132. ISSN: 0005-3155,

1943-6289. DOI: `10.1893/BIOS-D-19-00006`. URL: `https://bioone.org/journals/bios/volume-91/issue-2/BIOS-D-19-00006/An-exploration-of-land-use-and-poverty-as-an-integrative/10.1893/BIOS-D-19-00006.full` (visited on 02/08/2021).

Clarke, Kristie E. N. (2022). "Seroprevalence of Infection-Induced SARS-CoV-2 Antibodies — United States, September 2021–February 2022". In: *MMWR. Morbidity and Mortality Weekly Report* 71. ISSN: 0149-21951545-861X. DOI: `10.15585/mmwr.mm7117e3`. URL: `https://www.cdc.gov/mmwr/volumes/71/wr/mm7117e3.htm` (visited on 02/27/2023).

Clavijo-Baquet, Sabrina et al. (2021). "Thermal performance of the Chagas disease vector, Triatoma infestans, under thermal variability". In: *PLOS Neglected Tropical Diseases* 15.2, e0009148. ISSN: 1935-2735. DOI: `10.1371/journal.pntd.0009148`. URL: `https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0009148` (visited on 05/08/2023).

Collie, Shirley et al. (2022). "Effectiveness of BNT162b2 Vaccine against Omicron Variant in South Africa". In: *New England Journal of Medicine* 386.5, pp. 494–496. ISSN: 0028-4793. DOI: `10.1056/NEJMc2119270`. URL: `https://doi.org/10.1056/NEJMc2119270` (visited on 02/03/2022).

Coura, José Rodrigues and José Borges-Pereira (2010). "Chagas disease: 100 years after its discovery. A systemic review". In: *Acta Tropica*. Chagas disease: 100 years of the discovery and beyond 115.1, pp. 5–13. ISSN: 0001-706X. DOI: `10.1016/j.actatropica.2010.03.008`. URL: `https://www.sciencedirect.com/science/article/pii/S0001706X10000884` (visited on 05/07/2023).

Crane, Matthew A. et al. (2021). "Change in Reported Adherence to Nonpharmaceutical Interventions During the COVID-19 Pandemic, April-November 2020". In: *JAMA* 325.9, pp. 883–885. ISSN: 0098-7484. DOI: `10.1001/jama.2021.0286`. URL: `https://doi.org/10.1001/jama.2021.0286` (visited on 02/27/2023).

Cressie, Noel A. C. and Christopher K. Wikle (2011). *Statistics for spatio-temporal data*. Wiley series in probability and statistics. Hoboken, N.J: Wiley. 588 pp. ISBN: 978-0-471-69274-4.

Csárdi, G. and Nepusz, T. (2006). "The igraph software package for complex network research". In: *InterJournal Complex Systems*, p. 1695. URL: `https://cir.nii.ac.jp/crid/1370294643763138711` (visited on 07/06/2023).

Csárdi, Gábor et al. (2023). *igraph*. Version 0.10.5. DOI: `10.5281/ZENODO.3630268`. URL: `https://zenodo.org/record/3630268` (visited on 07/06/2023).

Cuevas, Erik (2020). "An agent-based model to evaluate the COVID-19 transmission risks in facilities". In: *Computers in Biology and Medicine* 121, p. 103827. ISSN: 0010-4825. DOI: `10.1016/j.compbiomed.2020.103827`. URL: `https://www.sciencedirect.com/science/article/pii/S001048252030192X` (visited on 11/16/2022).

Currie, Janet (2011). "Inequality at Birth: Some Causes and Consequences". In: *American Economic Review* 101.3, pp. 1–22. ISSN: 0002-8282. DOI: `10.1257/aer.101.3.1`. URL: `https://pubs.aeaweb.org/doi/10.1257/aer.101.3.1` (visited on 07/15/2020).

Dabisch, Paul et al. (2021). “The influence of temperature, humidity, and simulated sunlight on the infectivity of SARS-CoV-2 in aerosols”. In: *Aerosol Science and Technology* 55.2, pp. 142–153. ISSN: 0278-6826. DOI: `10.1080/02786826.2020.1829536`. URL: `https://doi.org/10.1080/02786826.2020.1829536` (visited on 06/05/2023).

Damette, Olivier, Clément Mathonnat, and Stéphane Goutte (2021). “Meteorological factors against COVID-19 and the role of human mobility”. In: *PLOS ONE* 16.6, e0252405. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0252405`. URL: `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0252405` (visited on 06/05/2023).

Danon, Leon et al. (2011). “Networks and the Epidemiology of Infectious Disease”. In: *Interdisciplinary Perspectives on Infectious Diseases* 2011, e284909. ISSN: 1687-708X. DOI: `10.1155/2011/284909`. URL: `https://www.hindawi.com/journals/ipid/2011/284909/` (visited on 09/01/2021).

Davies, Nicholas G. et al. (2020). “Age-dependent effects in the transmission and control of COVID-19 epidemics”. In: *Nature Medicine* 26.8, pp. 1205–1211. ISSN: 1546-170X. DOI: `10.1038/s41591-020-0962-9`. URL: `https://www.nature.com/articles/s41591-020-0962-9` (visited on 08/28/2020).

Davies, Nicholas G. et al. (2021). “Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England”. In: *Science* 372.6538. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.abg3055`. URL: `https://science.sciencemag.org/content/372/6538/eabg3055` (visited on 05/05/2021).

Dawson, Daniel E. et al. (2018). “Investigating behavioral drivers of seasonal Shiga-Toxigenic Escherichia Coli (STEC) patterns in grazing cattle using an agent-based model”. In: *PLOS ONE* 13.10, e0205418. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0205418`. URL: `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0205418` (visited on 06/05/2023).

Delazeri, Linda Márcia Mendes, Dênis Antônio Da Cunha, and Lais Rosa Oliveira (2022). “Climate change and rural–urban migration in the Brazilian Northeast region”. In: *GeoJournal* 87.3, pp. 2159–2179. ISSN: 1572-9893. DOI: `10.1007/s10708-020-10349-3`. URL: `https://doi.org/10.1007/s10708-020-10349-3` (visited on 05/07/2023).

Dias, J.C.P. (1987). “Control of chagas disease in Brazil”. In: *Parasitology Today* 3.11, pp. 336–341. ISSN: 01694758. DOI: `10.1016/0169-4758(87)90117-7`. URL: `https://linkinghub.elsevier.com/retrieve/pii/0169475887901177` (visited on 02/25/2021).

Diekmann, O., J. A. P. Heesterbeek, and M. G. Roberts (2010). “The construction of next-generation matrices for compartmental epidemic models”. In: *Journal of the Royal Society Interface* 7.47, pp. 873–885. ISSN: 1742-5689. DOI: `10.1098/rsif.2009.0386`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2871801/` (visited on 09/01/2021).

Dong, Ensheng, Hongru Du, and Lauren Gardner (2020). “An interactive web-based dashboard to track COVID-19 in real time”. In: *The Lancet Infectious Diseases* 20.5, pp. 533–534. ISSN: 14733099. DOI: `10.1016/S1473-3099(20)30120-1`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S1473309920301201` (visited on 02/01/2022).

Dooling, Kathleen (2021). “The Advisory Committee on Immunization Practices’ Updated Interim Recommendation for Allocation of COVID-19 Vaccine — United States, December 2020". In: *MMWR. Morbidity and Mortality Weekly Report* 69. ISSN: 0149-21951545-861X. DOI: `10.15585/mmwr.mm695152e2`. URL: `https://www.cdc.gov/mmwr/volumes/69/wr/mm695152e2.htm` (visited on 09/09/2021).

Dowell, S F (2001). "Seasonal variation in host susceptibility and cycles of certain infectious diseases." In: *Emerging Infectious Diseases* 7.3, pp. 369–374. ISSN: 1080-6040. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2631809/` (visited on 05/22/2023).

Feehan, Dennis M. and Ayesha S. Mahmud (2021). "Quantifying population contact patterns in the United States during the COVID-19 pandemic". In: *Nature Communications* 12.1, p. 893. ISSN: 2041-1723. DOI: `10.1038/s41467-021-20990-2`. URL: `https://www.nature.com/articles/s41467-021-20990-2` (visited on 05/06/2021).

Fernández, María del Pilar, María Sol Gaspe, and Ricardo E. Gürtler (2019). "Inequalities in the social determinants of health and Chagas disease transmission risk in indigenous and creole households in the Argentine Chaco". In: *Parasites & Vectors* 12.1, p. 184. ISSN: 1756-3305. DOI: `10.1186/s13071-019-3444-5`. URL: `https://doi.org/10.1186/s13071-019-3444-5` (visited on 02/03/2021).

Fick, Stephen E. and Robert J. Hijmans (2017). "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas". In: *International Journal of Climatology* 37.12, pp. 4302–4315. ISSN: 1097-0088. DOI: `10.1002/joc.5086`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/joc.5086` (visited on 07/14/2023).

Finkenstädt, B. F. and B. T. Grenfell (2000). "Time series modelling of childhood diseases: a dynamical systems approach". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 49.2, pp. 187–205. ISSN: 1467-9876. DOI: `10.1111/1467-9876.00187`. URL: `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9876.00187` (visited on 06/15/2020).

Fisman, D. (2012). "Seasonality of viral infections: mechanisms and unknowns". In: *Clinical Microbiology and Infection* 18.10, pp. 946–954. ISSN: 1198-743X. DOI: `10.1111/j.1469-0691.2012.03968.x`. URL: `https://www.sciencedirect.com/science/article/pii/S1198743X14610910` (visited on 06/28/2023).

Fitzpatrick, Meagan C. and Alison P. Galvani (2021). "Optimizing age-specific vaccination". In: *Science* 371.6532, pp. 890–891. DOI: `10.1126/science.abg2334`. URL: `https://www.science.org/doi/full/10.1126/science.abg2334` (visited on 02/23/2023).

Foy, Brody H. et al. (2021). "Comparing COVID-19 vaccine allocation strategies in India: A mathematical modelling study". In: *International Journal of Infectious Diseases* 103, pp. 431–438. ISSN: 1201-9712. DOI: `10.1016/j.ijid.2020.12.075`. URL: `https://www.sciencedirect.com/science/article/pii/S1201971220325996` (visited on 02/23/2023).

Funk, Sebastian, Marcel Salathé, and Vincent A. A. Jansen (2010). "Modelling the influence of human behaviour on the spread of infectious diseases: a review". In: *Journal of The Royal Society Interface* 7.50, pp. 1247–1256. ISSN: 1742-5689, 1742-5662. DOI: `10.1098/rsif.2010.0142`. URL: `https://royalsocietypublishing.org/doi/10.1098/rsif.2010.0142` (visited on 06/15/2020).

Galea, Sandro et al. (2011). “Estimated Deaths Attributable to Social Factors in the United States”. In: *American Journal of Public Health* 101.8, pp. 1456–1465. ISSN: 0090-0036. DOI: `10.2105/AJPH.2010.300086`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3134519/` (visited on 07/23/2020).

Garza, Miroslava et al. (2014). “Projected Future Distributions of Vectors of Trypanosoma cruzi in North America under Climate Change Scenarios”. In: *PLOS Neglected Tropical Diseases* 8.5, e2818. ISSN: 1935-2735. DOI: `10.1371/journal.pntd.0002818`. URL: `https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0002818` (visited on 05/08/2023).

Giubilini, Alberto, Julian Savulescu, and Dominic Wilkinson (2021). “Queue questions: Ethics of COVID-19 vaccine prioritization”. In: *Bioethics* 35.4, pp. 348–355. ISSN: 1467-8519. DOI: `10.1111/bioe.12858`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/bioe.12858` (visited on 09/07/2021).

Golgher, André Braz and Paul R. Voss (2016). “How to Interpret the Coefficients of Spatial Models: Spillovers, Direct and Indirect Effects”. In: *Spatial Demography* 4.3, pp. 175–205. ISSN: 2164-7070. DOI: `10.1007/s40980-015-0016-y`. URL: `https://doi.org/10.1007/s40980-015-0016-y` (visited on 06/12/2020).

Gomez, Jonatan et al. (2021). “INFEKTA—An agent-based model for transmission of infectious diseases: The COVID-19 case in Bogotá, Colombia”. In: *PLOS ONE* 16.2, e0245787. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0245787`. URL: `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245787` (visited on 09/01/2021).

Grassly, Nicholas C and Christophe Fraser (2006). “Seasonal infectious disease epidemiology”. In: *Proceedings of the Royal Society B: Biological Sciences* 273.1600, pp. 2541–2550. ISSN: 0962-8452. DOI: `10.1098/rspb.2006.3604`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1634916/` (visited on 06/15/2020).

Gupta, Sumedha et al. (2021). “Vaccinations Against COVID-19 May Have Averted Up To 140,000 Deaths In The United States”. In: *Health Affairs (Project Hope)*. ISSN: 1544-5208. DOI: `10.1377/hlthaff.2021.00619`.

Gurgel-Gonçalves, Rodrigo et al. (2012). *Geographic Distribution of Chagas Disease Vectors in Brazil Based on Ecological Niche Modeling*. Journal of Tropical Medicine. DOI: `https://doi.org/10.1155/2012/705326`. URL: `https://www.hindawi.com/journals/jtm/2012/705326/` (visited on 02/03/2021).

Haining, Robert P (2004). “Part E: Modelling Spatial Data”. In: *Spatial Data Analysis*, pp. 289–378.

Han, Shasha et al. (2021). “Time-varying optimization of COVID-19 vaccine prioritization in the context of limited vaccination capacity”. In: *Nature Communications* 12.1, p. 4673. ISSN: 2041-1723. DOI: `10.1038/s41467-021-24872-5`. URL: `https://www.nature.com/articles/s41467-021-24872-5` (visited on 02/23/2023).

Hawkins, Devan (2020). “Social Determinants of COVID-19 in Massachusetts, United States: An Ecological Study”. In: *Journal of Preventive Medicine and Public Health = Yebang Uihakhoe Chi* 53.4, pp. 220–227. ISSN: 2233-4521. DOI: `10.3961/jpmph.20.256`.

He, Daihai, Edward L. Ionides, and Aaron A. King (2010). "Plug-and-play inference for disease dynamics: measles in large and small populations as a case study". In: *Journal of the Royal Society Interface* 7.43, pp. 271–283. ISSN: 1742-5689. DOI: `10.1098/rsif.2009.0151`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2842609/` (visited on 08/16/2020).

Held, Leonhard and Michaela Paul (2012). "Modeling seasonality in space-time infectious disease surveillance data". In: *Biometrical Journal* 54.6, pp. 824–843. ISSN: 1521-4036. DOI: `10.1002/bimj.201200037`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201200037` (visited on 05/23/2023).

Hinch, Robert et al. (2021). "OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing". In: *PLOS Computational Biology* 17.7, e1009146. ISSN: 1553-7358. DOI: `10.1371/journal.pcbi.1009146`. URL: `https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009146` (visited on 09/01/2021).

Hoertel, Nicolas et al. (2020). "A stochastic agent-based model of the SARS-CoV-2 epidemic in France". In: *Nature Medicine* 26.9, pp. 1417–1421. ISSN: 1546-170X. DOI: `10.1038/s41591-020-1001-6`. URL: `https://www.nature.com/articles/s41591-020-1001-6` (visited on 09/01/2021).

Hogan, A et al. (2021). *Report 48: The value of vaccine booster doses to mitigate the global impact of the Omicron SARS-CoV-2 variant*. Imperial College London. DOI: `10.25561/93034`. URL: `http://spiral.imperial.ac.uk/handle/10044/1/93034` (visited on 02/03/2022).

Holmdahl, Inga et al. (2020). *Frequent testing and immunity-based staffing will help mitigate outbreaks in nursing home settings*. medRxiv, p. 2020.11.04.20224758. DOI: `10.1101/2020.11.04.20224758`. URL: `https://www.medrxiv.org/content/10.1101/2020.11.04.20224758v2` (visited on 03/04/2022).

House, James S., Ronald C. Kessler, and A. Regula Herzog (1990). "Age, Socioeconomic Status, and Health". In: *The Milbank Quarterly* 68.3, p. 383. ISSN: 0887378X. DOI: `10.2307/3350111`. URL: `https://www.jstor.org/stable/3350111?origin=crossref` (visited on 06/14/2020).

Hunter, Elizabeth and John D. Kelleher (2021). "Adapting an Agent-Based Model of Infectious Disease Spread in an Irish County to COVID-19". In: *Systems* 9.2, p. 41. ISSN: 2079-8954. DOI: `10.3390/systems9020041`. URL: `https://www.mdpi.com/2079-8954/9/2/41` (visited on 09/01/2021).

Hunter, Elizabeth, Brian Mac Namee, and John D. Kelleher (2017). "A Taxonomy for Agent-Based Models in Human Infectious Disease Epidemiology". In: *Journal of Artificial Societies and Social Simulation* 20.3, p. 2. ISSN: 1460-7425.

Hunter, Elizabeth, Brian Mac Namee, and John Kelleher (2018). "An open-data-driven agent-based model to simulate infectious disease outbreaks". In: *PLOS ONE* 13.12, e0208775. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0208775`. URL: `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0208775` (visited on 09/01/2021).

Instituto Brasileiro de Geografia e Estatística (2023). *Tabela 6579 - População residente estimada*. URL: `https://sidra.ibge.gov.br/tabela/6579` (visited on 05/05/2023).

Jarvis, Christopher I. et al. (2020). "Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK". In: *BMC Medicine* 18.1, p. 124. ISSN: 1741-7015. DOI: `10.1186/s12916-020-01597-8`. URL: `https://doi.org/10.1186/s12916-020-01597-8` (visited on 11/03/2021).

Johansson, Michael A. et al. (2021). "SARS-CoV-2 Transmission From People Without COVID-19 Symptoms". In: *JAMA Network Open* 4.1, e2035057. ISSN: 2574-3805. DOI: `10.1001/jamanetworkopen.2020.35057`. URL: `https://doi.org/10.1001/jamanetworkopen.2020.35057` (visited on 11/28/2022).

Keeling, Matt J and Ken T.D Eames (2005). "Networks and epidemic models". In: *Journal of The Royal Society Interface* 2.4, pp. 295–307. DOI: `10.1098/rsif.2005.0051`. URL: `https://royalsocietypublishing.org/doi/10.1098/rsif.2005.0051` (visited on 09/01/2021).

Keeling, Matt J., Pejman Rohani, and Bryan T. Grenfell (2001). "Seasonally forced disease dynamics explored as switching between attractors". In: *Physica D: Nonlinear Phenomena* 148.3, pp. 317–335. ISSN: 01672789. DOI: `10.1016/S0167-2789(00)00187-1`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0167278900001871` (visited on 06/05/2023).

Kerr, Cliff C. et al. (2021). "Covasim: an agent-based model of COVID-19 dynamics and interventions". In: p. 2020.05.10.20097469. DOI: `10.1101/2020.05.10.20097469`. URL: `https://www.medrxiv.org/content/10.1101/2020.05.10.20097469v3` (visited on 09/01/2021).

Kiti, Moses C. et al. (2021). "Social contact patterns among employees in 3 U.S. companies during early phases of the COVID-19 pandemic, April to June 2020". In: *Epidemics* 36, p. 100481. ISSN: 1755-4365. DOI: `10.1016/j.epidem.2021.100481`. URL: `https://www.sciencedirect.com/science/article/pii/S1755436521000347` (visited on 02/23/2023).

Knorr-Held, Leonhard (2000). "Bayesian modelling of inseparable space-time variation in disease risk". In: *Statistics in Medicine* 19.17, pp. 2555–2567. ISSN: 1097-0258. DOI: `10.1002/1097-0258(20000915/30)19:17/18<2555::AID-SIM587>3.0.CO;2-#`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/1097-0258%2820000915/30%2919%3A17/18%3C2555%3A%3AAID-SIM587%3E3.0.CO%3B2-%23` (visited on 04/17/2023).

Knorr-Held, Leonhard and Nicola G. Best (2001). "A Shared Component Model for Detecting Joint and Selective Clustering of Two Diseases". In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 164.1, pp. 73–85. ISSN: 0964-1998. URL: `https://www.jstor.org/stable/2680535` (visited on 04/17/2023).

Krieger, Nancy (2001). "Theories for social epidemiology in the 21st century: an ecosocial perspective". In: *International Journal of Epidemiology* 30.4, pp. 668–677. ISSN: 1464-3685, 0300-5771. DOI: `10.1093/ije/30.4.668`. URL: `https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/30.4.668` (visited on 01/14/2022).

Krivorotko, Olga et al. (2022). “Agent-based modeling of COVID-19 outbreaks for New York state and UK: Parameter identification algorithm”. In: *Infectious Disease Modelling* 7.1, pp. 30–44. ISSN: 2468-0427. DOI: `10.1016/j.idm.2021.11.004`. URL: `https://www.sciencedirect.com/science/article/pii/S2468042721000798` (visited on 06/05/2023).

Kronfeld-Schor, N. et al. (2021). “Drivers of Infectious Disease Seasonality: Potential Implications for COVID-19”. In: *Journal of Biological Rhythms* 36.1, pp. 35–54. ISSN: 0748-7304. DOI: `10.1177/0748730420987322`. URL: `https://doi.org/10.1177/0748730420987322` (visited on 05/22/2023).

Lambert, Diane (1992). “Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing”. In: *Technometrics* 34.1, pp. 1–14. ISSN: 0040-1706. DOI: `10.2307/1269547`. URL: `https://www.jstor.org/stable/1269547` (visited on 11/16/2022).

Lee, Ronald D. and Lawrence R. Carter (1992). “Modeling and Forecasting U.S. Mortality”. In: *Journal of the American Statistical Association* 87.419, pp. 659–671. ISSN: 0162-1459. DOI: `10.1080/01621459.1992.10475265`. URL: `https://doi.org/10.1080/01621459.1992.10475265` (visited on 04/01/2020).

Lee, Youngjo et al. (2016). “Spatial modeling of data with excessive zeros applied to reindeer pellet-group counts”. In: *Ecology and Evolution* 6.19, pp. 7047–7056. ISSN: 2045-7758. DOI: `10.1002/ece3.2449`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5513232/` (visited on 04/11/2023).

Leroux, Brian G., Xingye Lei, and Norman Breslow (2000). “Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence”. In: *Statistical Models in Epidemiology, the Environment, and Clinical Trials* 116, pp. 179–191. DOI: `10.1007/978-1-4612-1284-3_4`. URL: `http://link.springer.com/10.1007/978-1-4612-1284-3_4` (visited on 06/23/2020).

Levin, Andrew T. et al. (2020). “Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications”. In: *European Journal of Epidemiology* 35.12, pp. 1123–1138. ISSN: 1573-7284. DOI: `10.1007/s10654-020-00698-1`. URL: `https://doi.org/10.1007/s10654-020-00698-1` (visited on 12/06/2022).

Levin, Einav G. et al. (2021). “Waning Immune Humoral Response to BNT162b2 Covid-19 Vaccine over 6 Months”. In: *New England Journal of Medicine* 385.24, e84. ISSN: 0028-4793. DOI: `10.1056/NEJMoa2114583`. URL: `https://doi.org/10.1056/NEJMoa2114583` (visited on 07/06/2023).

Link, Bruce G. and Jo Phelan (1995). “Social Conditions As Fundamental Causes of Disease”. In: *Journal of Health and Social Behavior*, pp. 80–94. ISSN: 0022-1465. DOI: `10.2307/2626958`. URL: `https://www.jstor.org/stable/2626958` (visited on 06/14/2020).

Liu, Carol Y. et al. (2021a). “Rapid Review of Social Contact Patterns During the COVID-19 Pandemic”. In: *Epidemiology (Cambridge, Mass.)* 32.6, pp. 781–791. ISSN: 1044-3983. DOI: `10.1097/EDE.0000000000001412`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8478104/` (visited on 11/14/2022).

Liu, Jiangtao et al. (2020). "Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China". In: *Science of The Total Environment* 726, p. 138513. ISSN: 0048-9697. DOI: `10.1016/j.scitotenv.2020.138513`. URL: `https://www.sciencedirect.com/science/article/pii/S004896972032026X` (visited on 06/05/2023).

Liu, Xiaoyue et al. (2021b). "The role of seasonality in the spread of COVID-19 pandemic". In: *Environmental Research* 195, p. 110874. ISSN: 0013-9351. DOI: `10.1016/j.envres.2021.110874`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7892320/` (visited on 03/03/2023).

Liu, Ying and Joacim Rocklöv (2022). "The effective reproductive number of the Omicron variant of SARS-CoV-2 is several times relative to Delta". In: *Journal of Travel Medicine*, taac037. ISSN: 1195-1982. DOI: `10.1093/jtm/taac037`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8992231/` (visited on 04/28/2022).

Lofgren, Eric et al. (2007). "Influenza Seasonality: Underlying Causes and Modeling Theories". In: *Journal of Virology* 81.11, pp. 5429–5436. DOI: `10.1128/JVI.01680-06`. URL: `https://journals.asm.org/doi/10.1128/JVI.01680-06` (visited on 02/01/2023).

Lowen, Anice C. and John Steel (2014). "Roles of Humidity and Temperature in Shaping Influenza Seasonality". In: *Journal of Virology* 88.14, pp. 7692–7695. ISSN: 0022-538X. DOI: `10.1128/JVI.03544-13`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4097773/` (visited on 02/01/2023).

Lunn, Dave, Richard Arnold, and David Spiegelhalter (2004). "GeoBUGS User Manual". In: p. 45.

Ma, Yiqun et al. (2021). "Role of meteorological factors in the transmission of SARS-CoV-2 in the United States". In: *Nature Communications* 12.1, p. 3602. ISSN: 2041-1723. DOI: `10.1038/s41467-021-23866-7`. URL: `https://www.nature.com/articles/s41467-021-23866-7` (visited on 06/05/2023).

Marr, Linsey C. et al. (2019). "Mechanistic insights into the effect of humidity on airborne influenza virus survival, transmission and incidence". In: *Journal of The Royal Society Interface* 16.150, p. 20180298. DOI: `10.1098/rsif.2018.0298`. URL: `https://royalsocietypublishing.org/doi/10.1098/rsif.2018.0298` (visited on 06/05/2023).

Martin, Joyce, Brady Hamilton, and Michelle Osterman (2022). *Births in the United States, 2021.* National Center for Health Statistics (U.S.) DOI: `10.15620/cdc:119632`. URL: `https://stacks.cdc.gov/view/cdc/119632` (visited on 07/01/2023).

Martins-Melo, Francisco Rogerlândio et al. (2014). "Prevalence of Chagas disease in Brazil: A systematic review and meta-analysis". In: *Acta Tropica* 130, pp. 167–174. ISSN: 0001-706X. DOI: `10.1016/j.actatropica.2013.10.002`. URL: `https://www.sciencedirect.com/science/article/pii/S0001706X13002763` (visited on 05/07/2023).

Matrajt, Laura et al. (2021). "Vaccine optimization for COVID-19: Who to vaccinate first?" In: *Science Advances* 7.6, eabf1374. DOI: `10.1126/sciadv.abf1374`. URL: `https://www.science.org/doi/full/10.1126/sciadv.abf1374` (visited on 02/23/2023).

Matson, M. Jeremiah et al. (2020). "Effect of Environmental Conditions on SARS-CoV-2 Stability in Human Nasal Mucus and Sputum - Volume 26, Number 9—September 2020 - Emerging Infectious Diseases journal - CDC". In: DOI: `10.3201/eid2609.202267`.

URL: `https://wwwnc.cdc.gov/eid/article/26/9/20-2267_article` (visited on 06/05/2023).

McEvoy, David et al. (2020). *The relative infectiousness of asymptomatic SARS-CoV-2 infected persons compared with symptomatic individuals: A rapid scoping review*. medRxiv, p. 2020.07.30.20165084. DOI: `10.1101/2020.07.30.20165084`. URL: `https://www.medrxiv.org/content/10.1101/2020.07.30.20165084v1` (visited on 03/04/2022).

McMichael, A J (1999). "Prisoners of the Proximate: Loosening the Constraints on Epidemiology in an Age of Change". In: 149.10, p. 11.

Mecenas, Paulo et al. (2020). "Effects of temperature and humidity on the spread of COVID-19: A systematic review". In: *PLOS ONE* 15.9, e0238339. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0238339`. URL: `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0238339` (visited on 06/05/2023).

Medone, Paula et al. (2015). "The impact of climate change on the geographical distribution of two vectors of Chagas disease: implications for the force of infection". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1665. ISSN: 0962-8436. DOI: `10.1098/rstb.2013.0560`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4342964/` (visited on 03/01/2021).

Metcalf, C. Jessica E. et al. (2009). "Seasonality and comparative dynamics of six childhood infections in pre-vaccination Copenhagen". In: *Proceedings of the Royal Society B: Biological Sciences* 276.1676, pp. 4111–4118. ISSN: 0962-8452. DOI: `10.1098/rspb.2009.1058`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2821338/` (visited on 06/15/2020).

Ministério da Saúde, Brasil (2023). *Informações de Saúde (TABNET) – DATASUS*. URL: `https://datasus.saude.gov.br/informacoes-de-saude-tabnet/` (visited on 05/05/2023).

Mistry, D et al. (2021). *SynthPops: a generative model of human contact networks.* URL: `https://github.com/InstituteforDiseaseModeling/synthpops` (visited on 11/30/2022).

Mitzi Morris (2018). *Spatial Models in Stan: Intrinsic Auto-Regressive Models for Areal Data*. Stan. URL: `https://mc-stan.org/users/documentation/case-studies/icar_stan.html` (visited on 04/27/2023).

Miura, Fuminari et al. (2021). "Optimal vaccine allocation for COVID-19 in the Netherlands: A data-driven prioritization". In: *PLOS Computational Biology* 17.12, e1009697. ISSN: 1553-7358. DOI: `10.1371/journal.pcbi.1009697`. URL: `https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009697` (visited on 02/23/2023).

Moghadas, Seyed M et al. (2021). "The Impact of Vaccination on Coronavirus Disease 2019 (COVID-19) Outbreaks in the United States". In: *Clinical Infectious Diseases* 73.12, pp. 2257–2264. ISSN: 1058-4838. DOI: `10.1093/cid/ciab079`. URL: `https://doi.org/10.1093/cid/ciab079` (visited on 11/16/2022).

Moncayo, Álvaro and Antonio Carlos Silveira (2009). "Current epidemiological trends for Chagas disease in Latin America and future challenges in epidemiology, surveillance and health policy". In: *Memórias do Instituto Oswaldo Cruz* 104, pp. 17–30. ISSN: 0074-0276, 1678-8060. DOI: `10.1590/S0074-02762009000900005`. URL: `http://www.scielo.br/j/mioc/a/H5tnZvvMbmchMdXHDsZBBxv/?lang=en` (visited on 05/07/2023).

Morris, Dylan H et al. (2021). “Mechanistic theory predicts the effects of temperature and humidity on inactivation of SARS-CoV-2 and other enveloped viruses”. In: *eLife* 10. Ed. by Wendy S Garrett, C Brandon Ogbunugafor, and Andreas Handel, e65902. ISSN: 2050-084X. DOI: `10.7554/eLife.65902`. URL: `https://doi.org/10.7554/eLife.65902` (visited on 06/05/2023).

Morris, Mitzi et al. (2019). “Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan”. In: *Spatial and Spatio-temporal Epidemiology* 31, p. 100301. ISSN: 1877-5845. DOI: `10.1016/j.sste.2019.100301`. URL: `https://www.sciencedirect.com/science/article/pii/S1877584518301175` (visited on 04/04/2023).

Mossong, Joël et al. (2008). “Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases”. In: *PLOS Medicine* 5.3, e74. ISSN: 1549-1676. DOI: `10.1371/journal.pmed.0050074`. URL: `https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0050074` (visited on 04/01/2020).

Moya, Pedro, Beatriz Basso, and Edgardo Moretti (2005). “[Congenital Chagas disease in Córdoba, Argentina: epidemiological, clinical, diagnostic, and therapeutic aspects. Experience of 30 years of follow up]”. In: *Revista Da Sociedade Brasileira De Medicina Tropical* 38 Suppl 2, pp. 33–40. ISSN: 0037-8682.

Mylius, Sido D. et al. (2008). “Optimal allocation of pandemic influenza vaccine depends on age, risk and timing”. In: *Vaccine* 26.29, pp. 3742–3749. ISSN: 0264-410X. DOI: `10.1016/j.vaccine.2008.04.043`. URL: `https://www.sciencedirect.com/science/article/pii/S0264410X08005057` (visited on 02/23/2023).

Napier, Gary et al. (2019). “A Bayesian space–time model for clustering areal units based on their disease trends”. In: *Biostatistics* 20.4, pp. 681–697. ISSN: 1465-4644. DOI: `10.1093/biostatistics/kxy024`. URL: `https://academic.oup.com/biostatistics/article/20/4/681/5039880` (visited on 06/17/2020).

National Center for Immunization and Respiratory Diseases (NCIRD) (2022a). *National Immunization Survey Adult COVID Module (NIS-ACM)*. COVIDVaxViews| Data | Centers for Disease Control and Prevention (cdc.gov) | Data | Centers for Disease Control and Prevention. URL: `https://data.cdc.gov/Vaccinations/National-Immunization-Survey-Adult-COVID-Module-NI/udsf-9v7b` (visited on 02/02/2022).

— (2022b). *National Immunization Survey Child COVID Module (NIS-CCM)*. COVIDVaxViews| Data | Centers for Disease Control and Prevention (cdc.gov) | Data | Centers for Disease Control and Prevention. URL: `https://data.cdc.gov/Vaccinations/National-Immunization-Survey-Child-COVID-Module-NI/uny6-e3dx` (visited on 02/02/2022).

National Institutes of Health (2022). *COVID-19 Treatment Guidelines Panel: Clinical Spectrum of SARS-CoV-2 Infection. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines.* COVID-19 Treatment Guidelines. URL: `https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/` (visited on 02/27/2023).

Nelson, Kristin N. et al. (2022). "Nationally representative social contact patterns among U.S. adults, August 2020-April 2021". In: *Epidemics* 40, p. 100605. ISSN: 1755-4365. DOI: `10.1016/j.epidem.2022.100605`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9242729/` (visited on 11/14/2022).

Nichols, G. L. et al. (2021). "Coronavirus seasonality, respiratory infections and weather". In: *BMC Infectious Diseases* 21.1, p. 1101. ISSN: 1471-2334. DOI: `10.1186/s12879-021-06785-2`. URL: `https://doi.org/10.1186/s12879-021-06785-2` (visited on 02/01/2023).

Nottmeyer, Luise N. and Francesco Sera (2021). "Influence of temperature, and of relative and absolute humidity on COVID-19 incidence in England - A multi-city time-series study". In: *Environmental Research* 196, p. 110977. ISSN: 0013-9351. DOI: `10.1016/j.envres.2021.110977`. URL: `https://www.sciencedirect.com/science/article/pii/S0013935121002711` (visited on 06/05/2023).

O'Driscoll, Megan et al. (2021). "Age-specific mortality and immunity patterns of SARS-CoV-2". In: *Nature* 590.7844, pp. 140–145. ISSN: 1476-4687. DOI: `10.1038/s41586-020-2918-0`. URL: `https://www.nature.com/articles/s41586-020-2918-0` (visited on 09/07/2021).

Oraby, Tamer et al. (2014). "Modeling seasonal behavior changes and disease transmission with application to chronic wasting disease". In: *Journal of Theoretical Biology* 340, pp. 50–59. ISSN: 0022-5193. DOI: `10.1016/j.jtbi.2013.09.003`. URL: `https://www.sciencedirect.com/science/article/pii/S0022519313004244` (visited on 06/05/2023).

Pérez-Molina, José A and Israel Molina (2018). "Chagas disease". In: *The Lancet* 391.10115, pp. 82–94. ISSN: 01406736. DOI: `10.1016/S0140-6736(17)31612-4`. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0140673617316124` (visited on 02/26/2021).

Persad, Govind, Monica E. Peek, and Ezekiel J. Emanuel (2020). "Fairly Prioritizing Groups for Access to COVID-19 Vaccines". In: *JAMA* 324.16, pp. 1601–1602. ISSN: 0098-7484. DOI: `10.1001/jama.2020.18513`. URL: `https://doi.org/10.1001/jama.2020.18513` (visited on 09/07/2021).

Persad, Govind et al. (2021). "Public Perspectives on COVID-19 Vaccine Prioritization". In: *JAMA Network Open* 4.4, e217943–e217943. ISSN: 2574-3805. DOI: `10.1001/jamanetworkopen.2021.7943`. URL: `https://doi.org/10.1001/jamanetworkopen.2021.7943` (visited on 09/09/2021).

Perz, Stephen G. (2000). "The Rural Exodus in the Context of Economic Crisis, Globalization and Reform in Brazil". In: *The International Migration Review* 34.3, pp. 842–881. ISSN: 0197-9183. DOI: `10.2307/2675947`. URL: `https://www.jstor.org/stable/2675947` (visited on 05/07/2023).

Phelan, Jo C., Bruce G. Link, and Parisa Tehranifar (2010). "Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications". In: *Journal of Health and Social Behavior* 51.1, S28–S40. ISSN: 0022-1465. DOI: `10.1177/0022146510383498`. URL: `https://doi.org/10.1177/0022146510383498` (visited on 06/14/2020).

Phua, Jason et al. (2020). “Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations”. In: *The Lancet Respiratory Medicine* 8.5, pp. 506–517. ISSN: 2213-2600, 2213-2619. DOI: `10.1016/S2213-2600(20)30161-2`. URL: `https://www.thelancet.com/journals/lanres/article/PIIS2213-2600(20)30161-2/fulltext?LinkSource=PassleApp` (visited on 02/23/2023).

Pitzer, Virginia E. et al. (2009). “Demographic variability, vaccination, and the spatiotemporal dynamics of rotavirus epidemics”. In: *Science (New York, N.Y.)* 325.5938, pp. 290–294. ISSN: 1095-9203. DOI: `10.1126/science.1172330`.

Plummer, Martyn (2003). “JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling”. In: p. 10.

Polack, Fernando P. et al. (2020). “Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine”. In: *The New England Journal of Medicine* 383.27, pp. 2603–2615. ISSN: 1533-4406. DOI: `10.1056/NEJMoa2034577`.

Raiteux, Jérémy et al. (2021). “Inactivation of SARS-CoV-2 by Simulated Sunlight on Contaminated Surfaces”. In: *Microbiology Spectrum* 9.1, 10.1128/spectrum.00333–21. DOI: `10.1128/spectrum.00333-21`. URL: `https://journals.asm.org/doi/10.1128/Spectrum.00333-21` (visited on 06/05/2023).

Randell, Heather F. and Leah K. VanWey (2014). “Networks Versus Need: Drivers of Urban Out-Migration in the Brazilian Amazon”. In: *Population Research and Policy Review* 33.6, pp. 915–936. ISSN: 1573-7829. DOI: `10.1007/s11113-014-9336-7`. URL: `https://doi.org/10.1007/s11113-014-9336-7` (visited on 05/07/2023).

Rathbun, Stephen L. and Songlin Fei (2006). “A spatial zero-inflated poisson regression model for oak regeneration”. In: *Environmental and Ecological Statistics* 13.4, pp. 409–426. ISSN: 1573-3009. DOI: `10.1007/s10651-006-0020-x`. URL: `https://doi.org/10.1007/s10651-006-0020-x` (visited on 04/11/2023).

Reicher, Stephen and John Drury (2021). “Pandemic fatigue? How adherence to covid-19 regulations has been misrepresented and why it matters”. In: *BMJ* 372, n137. ISSN: 1756-1833. DOI: `10.1136/bmj.n137`. URL: `https://www.bmj.com/content/372/bmj.n137` (visited on 02/27/2023).

Riddell, Shane et al. (2020). “The effect of temperature on persistence of SARS-CoV-2 on common surfaces”. In: *Virology Journal* 17.1, p. 145. ISSN: 1743-422X. DOI: `10.1186/s12985-020-01418-7`. URL: `https://doi.org/10.1186/s12985-020-01418-7` (visited on 06/05/2023).

Riebler, Andrea et al. (2016). “An intuitive Bayesian spatial model for disease mapping that accounts for scaling”. In: *Statistical Methods in Medical Research* 25.4, pp. 1145–1165. ISSN: 0962-2802. DOI: `10.1177/0962280216660421`. URL: `https://doi.org/10.1177/0962280216660421` (visited on 06/12/2020).

Rose, Geoffrey (1985). “Sick individuals and sick populations”. In: *International Journal of Epidemiology* 14, pp. 32–38.

Roubenoff, Ethan, Dennis M. Feehan, and Ayesha S. Mahmud (2022a). “Replication code for Roubenoff, Feehan, and Mahmud, " Evaluating primary and booster vaccination prioritization strategies for COVID-19 by age and high-contact employment status using data from contact surveys"". In: URL: https://github.com/eroubenoff/BICS_employment_replication_code (visited on 11/14/2022).

Roubenoff, Ethan, Dennis M. Feehan, and Ayesha S. Mahmud (2022b). *Replication Data for: Evaluating primary and booster vaccination prioritization strategies for COVID-19 by age and high-contact employment status using data from contact surveys*. DOI: 10.7910/DVN/K8YPVZ. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/K8YPVZ (visited on 11/14/2022).

— (2023). "Evaluating primary and booster vaccination prioritization strategies for COVID-19 by age and high-contact employment status using data from contact surveys". In: *Epidemics* 43, p. 100686. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2023.100686. URL: https://www.sciencedirect.com/science/article/pii/S1755436523000221 (visited on 06/27/2023).

Saadi, Nuru et al. (2021). "Models of COVID-19 vaccine prioritisation: a systematic literature search and narrative review". In: *BMC Medicine* 19.1, p. 318. ISSN: 1741-7015. DOI: 10.1186/s12916-021-02190-3. URL: https://doi.org/10.1186/s12916-021-02190-3 (visited on 02/23/2023).

Sah, Pratha et al. (2021). "Accelerated vaccine rollout is imperative to mitigate highly transmissible COVID-19 variants". In: *EClinicalMedicine* 35, p. 100865. ISSN: 2589-5370. DOI: 10.1016/j.eclinm.2021.100865.

Schaffer DeRoo, Sarah, Natalie J. Pudalov, and Linda Y. Fu (2020). "Planning for a COVID-19 Vaccination Program". In: *JAMA* 323.24, pp. 2458–2459. ISSN: 0098-7484. DOI: 10.1001/jama.2020.8711. URL: https://doi.org/10.1001/jama.2020.8711 (visited on 09/09/2021).

Segerstrom, Suzanne C. and Gregory E. Miller (2004). "Psychological Stress and the Human Immune System: A Meta-Analytic Study of 30 Years of Inquiry". In: *Psychological bulletin* 130.4, pp. 601–630. ISSN: 0033-2909. DOI: 10.1037/0033-2909.130.4.601. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1361287/ (visited on 08/27/2020).

Selden, Thomas M. and Terceira A. Berdahl (2020). "COVID-19 And Racial/Ethnic Disparities In Health Risk, Employment, And Household Composition". In: *Health Affairs (Project Hope)* 39.9, pp. 1624–1632. ISSN: 1544-5208. DOI: 10.1377/hlthaff.2020.00897.

Sera, Francesco et al. (2021). "A cross-sectional analysis of meteorological factors and SARS-CoV-2 transmission in 409 cities across 26 countries". In: *Nature Communications* 12.1, p. 5968. ISSN: 2041-1723. DOI: 10.1038/s41467-021-25914-8. URL: https://www.nature.com/articles/s41467-021-25914-8 (visited on 06/05/2023).

Shaman, Jeffrey and Melvin Kohn (2009). "Absolute humidity modulates influenza survival, transmission, and seasonality". In: *Proceedings of the National Academy of Sciences* 106.9, pp. 3243–3248. DOI: 10.1073/pnas.0806852106. URL: https://www.pnas.org/doi/full/10.1073/pnas.0806852106 (visited on 06/05/2023).

Shevitz, Abby et al. (1996). "The Association Between Youth, Women, and Acquired Immunodeficiency Syndrome". In: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 13.5, pp. 427–433. ISSN: 1525-4135. URL: https://journals.lww.com/jaids/Fulltext/

`1996/12150/The_Association_Between_Youth,_Women,_and_Acquired.5.aspx` (visited on 09/04/2020).

Silk, Benjamin J. (2023). "COVID-19 Surveillance After Expiration of the Public Health Emergency Declaration — United States, May 11, 2023". In: *MMWR. Morbidity and Mortality Weekly Report* 72. ISSN: 0149-21951545-861X. DOI: `10.15585/mmwr.mm7219e1`. URL: `https://www.cdc.gov/mmwr/volumes/72/wr/mm7219e1.htm` (visited on 06/30/2023).

Simões, Taynãna César et al. (2018). "Chagas disease mortality in Brazil: A Bayesian analysis of age-period-cohort effects and forecasts for two decades". In: *PLOS Neglected Tropical Diseases* 12.9, e0006798. ISSN: 1935-2735. DOI: `10.1371/journal.pntd.0006798`. URL: `https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0006798` (visited on 10/27/2021).

Smith, Thomas P. et al. (2021). "Temperature and population density influence SARS-CoV-2 transmission in the absence of nonpharmaceutical interventions". In: *Proceedings of the National Academy of Sciences* 118.25, e2019284118. DOI: `10.1073/pnas.2019284118`. URL: `https://www.pnas.org/doi/full/10.1073/pnas.2019284118` (visited on 06/05/2023).

Sosa-Estani, Sergio and Elsa Leonor Segura (2015). "Integrated control of Chagas disease for its elimination as public health problem - A Review". In: *Memórias do Instituto Oswaldo Cruz* 110.3, pp. 289–298. ISSN: 0074-0276. DOI: `10.1590/0074-02760140408`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4489466/` (visited on 02/08/2021).

Stan Development Team (2023). *Stan Modeling Language Users Guide and Reference Manual, version 2.31.* URL: `https://mc-stan.org`.

Strasser, Zachary H. et al. (2022). "Estimates of SARS-CoV-2 Omicron BA.2 Subvariant Severity in New England". In: *JAMA Network Open* 5.10, e2238354. ISSN: 2574-3805. DOI: `10.1001/jamanetworkopen.2022.38354`. URL: `https://doi.org/10.1001/jamanetworkopen.2022.38354` (visited on 02/27/2023).

Susswein, Zachary, Eva C Rest, and Shweta Bansal (2023). "Disentangling the rhythms of human activity in the built environment for airborne transmission risk: An analysis of large-scale mobility data". In: *eLife* 12. Ed. by Niel Hens, Diane M Harper, and Guillaume Béraud, e80466. ISSN: 2050-084X. DOI: `10.7554/eLife.80466`. URL: `https://doi.org/10.7554/eLife.80466` (visited on 08/01/2023).

Tamayo, Laura D. et al. (2018). "The effect of temperature increase on the development of Rhodnius prolixus and the course of Trypanosoma cruzi metacyclogenesis". In: *PLoS Neglected Tropical Diseases* 12.8, e0006735. ISSN: 1935-2727. DOI: `10.1371/journal.pntd.0006735`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6110519/` (visited on 03/30/2023).

Tarleton, Rick L et al. (2007). "The Challenges of Chagas Disease— Grim Outlook or Glimmer of Hope?" In: *PLoS Medicine* 4.12. ISSN: 1549-1277. DOI: `10.1371/journal.pmed.0040332`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2222930/` (visited on 02/08/2021).

Tenforde, Mark W. (2021). “Effectiveness of Pfizer-BioNTech and Moderna Vaccines Against COVID-19 Among Hospitalized Adults Aged ≥65 Years — United States, January–March 2021”. In: *MMWR. Morbidity and Mortality Weekly Report* 70. ISSN: 0149-21951545-861X. DOI: `10.15585/mmwr.mm7018e1`. URL: `https://www.cdc.gov/mmwr/volumes/70/wr/mm7018e1.htm` (visited on 09/07/2021).

Thompson, Mark G. (2021). “Interim Estimates of Vaccine Effectiveness of BNT162b2 and mRNA-1273 COVID-19 Vaccines in Preventing SARS-CoV-2 Infection Among Health Care Personnel, First Responders, and Other Essential and Frontline Workers — Eight U.S. Locations, December 2020–March 2021”. In: *MMWR. Morbidity and Mortality Weekly Report* 70. ISSN: 0149-21951545-861X. DOI: `10.15585/mmwr.mm7013e3`. URL: `https://www.cdc.gov/mmwr/volumes/70/wr/mm7013e3.htm` (visited on 09/07/2021).

— (2022). “Effectiveness of a Third Dose of mRNA Vaccines Against COVID-19–Associated Emergency Department and Urgent Care Encounters and Hospitalizations Among Adults During Periods of Delta and Omicron Variant Predominance — VISION Network, 10 States, August 2021–January 2022”. In: *MMWR. Morbidity and Mortality Weekly Report* 71. ISSN: 0149-21951545-861X. DOI: `10.15585/mmwr.mm7104e3`. URL: `https://www.cdc.gov/mmwr/volumes/71/wr/mm7104e3.htm` (visited on 03/02/2022).

UK Health Security Agency (2021). *Technical briefing: Update on hospitalisation and vaccine effectiveness for Omicron VOC-21NOV-01 (B.1.1.529).*

US Census Bureau (2019). “Table B01001: Sex by Age, 2015-2019. American Community Survey 5-year Estimates”. In: URL: `data.census.gov`.

— (2022). *Table B01001: Sex by Age, 2021. American Community Survey 1-year Estimates.* URL: `https://data.census.gov/table?q=B01001:+SEX+BY+AGE&tid=ACSDT1Y2021.B01001` (visited on 07/06/2023).

Venkatramanan, Srinivasan et al. (2018). “Using data-driven agent-based models for forecasting emerging infectious diseases”. In: *Epidemics.* The RAPIDD Ebola Forecasting Challenge 22, pp. 43–49. ISSN: 1755-4365. DOI: `10.1016/j.epidem.2017.02.010`. URL: `https://www.sciencedirect.com/science/article/pii/S1755436517300221` (visited on 09/01/2021).

Ver Hoef, Jay M. and John K. Jansen (2007). “Space—time zero-inflated count models of Harbor seals”. In: *Environmetrics* 18.7, pp. 697–712. ISSN: 1099-095X. DOI: `10.1002/env.873`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/env.873` (visited on 04/11/2023).

Wachter, Kenneth W. (2005). “Spatial demography”. In: *Proceedings of the National Academy of Sciences* 102.43, pp. 15299–15300. ISSN: 0027-8424, 1091-6490. DOI: `10.1073/pnas.0508155102`. URL: `https://www.pnas.org/content/102/43/15299` (visited on 06/18/2020).

Wakefield, Jon (2007). “Disease mapping and spatial regression with count data”. In: *Biostatistics* 8.2, pp. 158–183. ISSN: 1465-4644. DOI: `10.1093/biostatistics/kxl008`. URL: `https://academic.oup.com/biostatistics/article/8/2/158/230741` (visited on 06/23/2020).

Weaver, Amanda K. et al. (2022). "Environmental Factors Influencing COVID-19 Incidence and Severity". In: *Annual Review of Public Health* 43.1, pp. 271–291. DOI: `10.1146/annurev-publhealth-052120-101420`. URL: `https://doi.org/10.1146/annurev-publhealth-052120-101420` (visited on 06/05/2023).

Weiss, Robin A and Anthony J McMichael (2004). "Social and environmental risk factors in the emergence of infectious diseases". In: *Nature Medicine* 10 (S12), S70–S76. ISSN: 1078-8956, 1546-170X. DOI: `10.1038/nm1150`. URL: `http://www.nature.com/articles/nm1150` (visited on 06/15/2020).

WHO Expert Committee on the Control of Chagas Disease (2002). *Control of chagas disease: second report of the WHO Expert Committee*. Brasilia, Brazil: WHO.

Wikle, Christopher K, L Mark Berliner, and Cressie, Noel (1998). "Hierarchical Bayesian space-time models". In: *Evironmental and Ecological Statistics*, p. 38.

Williams, Katherine V. et al. (2022). "Can a Two-Dose Influenza Vaccine Regimen Better Protect Older Adults? An Agent-Based Modeling Study". In: *Vaccines* 10.11, p. 1799. ISSN: 2076-393X. DOI: `10.3390/vaccines10111799`. URL: `https://www.mdpi.com/2076-393X/10/11/1799` (visited on 06/05/2023).

Wood, Richard M. et al. (2020). "COVID-19 scenario modelling for the mitigation of capacity-dependent deaths in intensive care". In: *Health Care Management Science* 23.3, pp. 315–324. ISSN: 1572-9389. DOI: `10.1007/s10729-020-09511-7`. URL: `https://doi.org/10.1007/s10729-020-09511-7` (visited on 02/23/2023).

World Health Organization (2015). "Chagas disease in Latin America: an epidemiological update based on 2010 estimates". In: *Weekly Epidemiological Record* 6. URL: `https://apps.who.int/iris/bitstream/handle/10665/242316/WER9006_33-44.PDF?sequence=1&isAllowed=y` (visited on 03/29/2023).

Wouters, H et al. (2021). *Global bioclimatic indicators from 1950 to 2100 derived from climate projections*. DOI: `10.24381/CDS.A37FECB7`. URL: `https://cds.climate.copernicus.eu/doi/10.24381/cds.a37fecb7` (visited on 05/08/2023).

Wrigley-Field, Elizabeth et al. (2020). "Racial Disparities in COVID-19 and Excess Mortality in Minnesota". In: *Socius* 6, p. 2378023120980918. ISSN: 2378-0231. DOI: `10.1177/2378023120980918`. URL: `https://doi.org/10.1177/2378023120980918` (visited on 10/25/2021).

Wrigley-Field, Elizabeth et al. (2021). "Geographically targeted COVID-19 vaccination is more equitable and averts more deaths than age-based thresholds alone". In: *Science Advances* 7.40, eabj2099. ISSN: 2375-2548. DOI: `10.1126/sciadv.abj2099`. URL: `https://www.science.org/doi/10.1126/sciadv.abj2099` (visited on 11/22/2021).

Wu, Yu et al. (2020). "Effects of temperature and humidity on the daily new cases and new deaths of COVID-19 in 166 countries". In: *Science of The Total Environment* 729, p. 139051. ISSN: 0048-9697. DOI: `10.1016/j.scitotenv.2020.139051`. URL: `https://www.sciencedirect.com/science/article/pii/S0048969720325687` (visited on 06/05/2023).

Xu, Jiaquan et al. (2022). *Mortality in the United States, 2021*. National Center for Health Statistics (U.S.) DOI: `10.15620/cdc:122516`. URL: `https://stacks.cdc.gov/view/cdc/122516` (visited on 07/01/2023).

Yang, Zhen et al. (2022). "Clinical Characteristics, Transmissibility, Pathogenicity, Susceptible Populations, and Re-infectivity of Prominent COVID-19 Variants". In: *Aging and Disease* 13.2, pp. 402–422. ISSN: 2152-5250. DOI: `10.14336/AD.2021.1210`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8947836/` (visited on 02/27/2023).

Appendix A

Supplementary Material for Chapter 1

A.1 Epidemiological Model

A.1.1 Contact Matrix

Table A.1: Baseline Contact Matrix $\boldsymbol{C}$ used in simulations

	Child	Adult LC	Adult HC	65+
Child	1.637	0.816	1.197	0.062
Adult LC	0.455	1.918	1.283	0.311
Adult HC	1.285	2.472	4.550	0.495
65+	0.093	0.834	0.689	0.908

A.1.2 Alternate Specification of the Contact Matrix

Two alternate specifications of the contact matrix are used for sensitivity tests. First, unlike in the primary simulation where high/low contact status was randomly assigned (in equal proportions to adult respondents surveyed) to alters whose contact status was indeterminable, a matrix is derived without the above described process for unknown alter redistribution. As a number of reported alters are of unknown risk status (no occupation or purpose), in the baseline matrix we assume that the distribution of alters' contact status at the aggregate level is equal to egos' contact status. Instead of randomly re-labeling unknown adult alters such that the reported alters have the same survey-weighted proportion of high/low contact status as egos, in this alternate specification we label all unknown alters as having Low Contact status. This represents a conservative estimate for alter contact patterns. Second, we derive the contact matrix otherwise as normal, but with an alternate specification for the POLYMOD matrix $\boldsymbol{P}$ used to derive child-child contacts. This departs from the baseline simulation, where we use POLYMOD contacts with school contacts removed to represent school closures at the time of writing. For this alternate specification, we derive the matrix that includes these POLYMOD contacts.

The primary simulation is run with these two alternate contact matrix specifications contact matrix (shown in figure A.1). All other parameters are as baseline. Both total deaths and clinical infections are consistently higher with the two alternate specifications compared to baseline. In neither set of simulations did the the most effective strategies change.

Table A.2: Contact Matrix without random unknown alter reallocation

	Child	Adult LC	Adult HC	65+
Child	1.637	0.858	1.207	0.062
Adult LC	0.478	2.009	1.260	0.319
Adult HC	1.296	2.427	4.332	0.486
65+	0.093	0.855	0.677	0.922

Table A.3: Contact Matrix derived without POLYMOD school contacts removed

	Child	Adult LC	Adult HC	65+
Child	3.271	0.847	1.193	0.060
Adult LC	0.472	1.930	1.277	0.304
Adult HC	1.281	2.459	4.491	0.524
65+	0.090	0.814	0.730	0.874

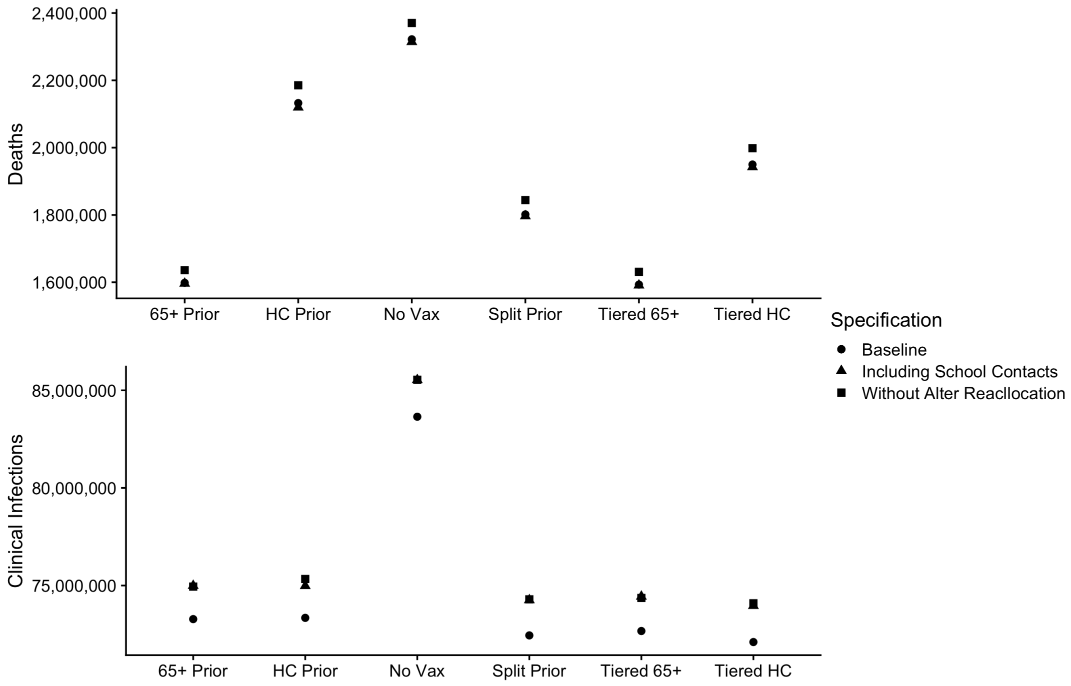


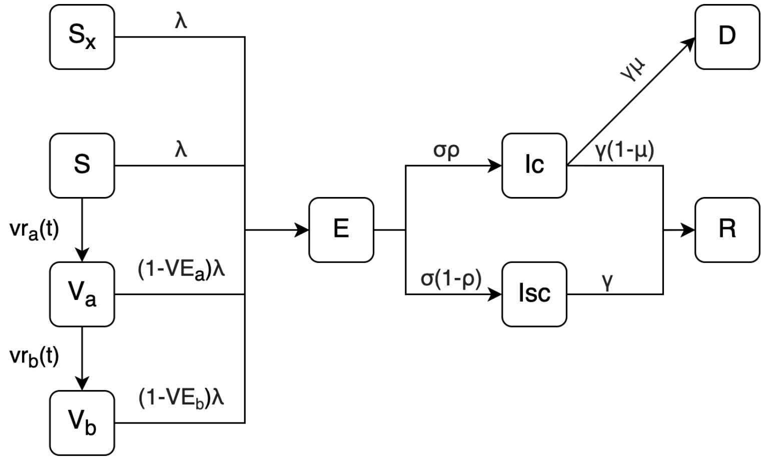
Figure A.1: Deaths and infections for the 5 intervention strategies compared to null for alternate contact matrix specifications, without random unknown alter reallocation and without POLYMOD school contacts removed.

A.1.3 Derivation of contacts for the youngest age group

Since the youngest age group (below 18 years of age) were not included in the survey, we use data from the POLYMOD survey (Mossong et al. 2008) in the United Kingdom. First, BICS respondents are re-grouped without employment disaggregation into three age categories—$0 - 18,14 - 65$, and 65+—and used to derive reciprocal weighted average contact matrix $\boldsymbol{B}$. Then, data from the POLYMOD survey are grouped into the same three age categories and likewise used to create reciprocal weighted average contact matrix $\boldsymbol{P}$. As in Feehan and Mahmud 2021, we calculate the ratio of dominant eigenvalues (ρ) of $\boldsymbol{B}$ and $\boldsymbol{P}$. The entry for the youngest age group in $\boldsymbol{C}$ is then taken as the entry for the youngest age group in $\boldsymbol{P}$ scaled by this ratio:

$$\boldsymbol{C}_{<18,<18} = \frac{\rho(\boldsymbol{B})}{\rho(\boldsymbol{P})} \cdot \boldsymbol{P}_{<18,<18} \tag{A.1}$$

A.1.4 Primary (2-dose) vaccination model



$$\lambda = u \sum_j C_j \frac{Ic_j + \alpha Isc_j}{N_j} \tag{A.2}$$

$$\frac{dS_x}{dt} = -\lambda S_x \tag{A.3}$$

$$\frac{dS}{dt} = -\lambda S - VR_a(t) \tag{A.4}$$

$$\frac{dV_a}{dt} = -\lambda(1 - VE_a)Va + VR_a(t) - VR_b(t) \tag{A.5}$$

$$\frac{dV_b}{dt} = -\lambda(1 - VE_b)Vb + VR_b(t) \tag{A.6}$$

$$\frac{dE}{dt} = -\sigma E + \lambda\Big[S + S_x + (1 - VE_a)V_a + (1 - VE_b)V_b\Big] \tag{A.7}$$

$$\frac{dI_c}{dt} = -\gamma I_c + \rho\sigma E \tag{A.8}$$

$$\frac{dI_{sc}}{dt} = -\gamma I_{sc} + (1 - \rho)\sigma E \tag{A.9}$$

$$\frac{dR}{dt} = (1 - \mu)\gamma I_c + \gamma I_{sc} \tag{A.10}$$

$$\frac{dD}{dt} = \mu\gamma Ic \tag{A.11}$$

Figure A.2: Primary vaccination model. (Caption continued on following page)

A.1.5 Derivation of the Starting Population

The starting population is derived from the Johns Hopkins COVID-19 dashboard data repository (JHU; Dong, Du, and Gardner 2020) and American Community Survey 2019 estimates

Figure A.2: (Continued from previous page) Compartments S, S_x, E, I_c, I_{sc}, R, D, V_a, V_b correspond to susceptible and waiting to be vaccinated, susceptible but not waiting to be vaccinated, exposed, clinically infectious, subclinically infectious, recovered, deceased, vaccinated 1st dose, and vaccinated 2nd dose, respectfully. Parameter u is the group-specific probability of infection upon contact with an infected person; C_{ij} is the number of contacts a person in group i has with a person in group j; α is the relative infectiousness of subclinical cases; $VE_{a,b}$ are reduction in infection after 1 and 2 doses of the vaccine; ρ_i is probability of symptomatic illness after exposure and first and second shots; μ, σ, γ are the mortality, rate of progressing from E to I, and recovery rate. $VR_{ai,bi}(t)$ are the time-specific vaccination rates for 1st and 2nd vaccines. *NB: Group index i is dropped to reduce visual clutter.*

(ACS; US Census Bureau 2019) using parameters from the baseline simulation listed in table A.8. The JHU data contain the state level counts of confirmed COVID-19 cases and deaths; in addition, 45 of 51 states (including DC) list counts of active cases. To obtain the initial size of the clinically infectious compartment, we scaled the national count of total confirmed cases by the proportion of cases currently active for states with active cases listed. For these 45 states with listed active cases, on 1 January 2021 there were $14,405,003$ total confirmed cases, of which 40.5% ($5,829,986$) were currently active. This proportion is used to scale the total confirmed cases ($20,278,813$) to yield the total initial size of the clinically infectious compartment: $8,207,366$.

The BICS wave 4 survey indicated that, adjusted for survey weights, 34% of adults were working in person (HR) and 66% were not (LR; including unemployed). These survey data were used to scale the ACS age-disaggregated data to yield the total group sizes N:

Child	LC Adult	HC Adult	65+
73,429,392	132,319,841	68,164,766	50,783,796

Table A.4: Group sizes N derived from the ACS

We assume that the total count of Active cases is distributed proportionally to group size (N) and the age-dependent probability of infection (u, table A.8) to produce infectious and symptomatic compartment $Ic(0)$:

$$Ic_i(0) = Active \cdot \frac{u_i \cdot N_i}{\sum_j u_j \cdot N_j} \tag{A.12}$$

Individuals in the exposed (pre-infectious) compartment are taken to be proportional I_c. The constant of proportionality is taken as the ratio of average time spent in the E compartment to average time spent in the I_c compartment. This assumes that the size of the outbreak is stationary at the start of the simulation.

$$E_i(0) = Ic_i(0) * \frac{latentPeriod}{infectiousPeriod} \tag{A.13}$$

Remaining members of the population are all assumed to be susceptible. Susceptibles are then divided into S and S_x, indicating if they are awaiting or declining the vaccine. We assume that all groups will have 70% uptake. For baseline parameters of μ and ρ, the initial population state is:

Table A.5: Primary Simulation Starting Population

	Child	Adult LC	Adult HC	65+
N	$73,553,240$	$132,073,398$	$68,037,811$	$49,238,581$
S	$50,346,846$	$88,093,322$	$45,381,408$	$33,018,446$
S_x	$21,577,220$	$37,754,281$	$19,449,175$	$14,150,763$
E	$610,941$	$2,334,673$	$1,202,710$	$776,015$
I_c	$1,018,235$	$3,891,122$	$2,004,517$	$1,293,358$
I_{sc}	0	0	0	0
R	0	0	0	0
D	0	0	0	0
V_a	0	0	0	0
V_b	0	0	0	0
V_{bx}	0	0	0	0
V_{boost}	0	0	0	0

In the booster simulation, we utilize uptake data derived from the CDC to initialize a proportion of the susceptible population in compartment V_b, indicating that they have received a primary course of vaccination. Individuals who decline the primary vaccine remain in compartment S_x until infected. We assume that 70% of individuals who have received a primary course of vaccination will receive boosters; the remaining 30% are placed in compartment V_{bx}, indicating that they decline the vaccine.

Finally, we conducted a series of simulations assuming high uptake of the primary vaccine among both low and high contact adults to match older adult uptake (95%). These simulations closely mirrored those for the main booster simulations—prioritizing older adults, either solely or in a tiered roll-out limited the most deaths, and the number of infections limited was nearly equal across all strategies.

Table A.6: Booster Simulation Starting Population

	Child	Adult LC	Adult HC	65+
N	$73,553,240$	$132,073,398$	$68,037,811$	$49,238,581$
S	0	0	0	0
S_x	$47,462,186$	$27,092,608$	$13,956,798$	$2,394,698$
E	$615,314$	$2,351,386$	$1,211,320$	$781,570$
I_c	$1,025,523$	$3,918,976$	$2,018,866$	$1,302,616$
I_{sc}	0	0	0	0
R	0	0	0	0
D	0	0	0	0
V_a	0	0	0	0
V_b	$17,115,152$	$69,097,300$	$35,595,579$	$31,331,788$
V_{bx}	$7,335,065$	$29,613,129$	$15,255,248$	$13,427,909$
V_{boost}	0	0	0	0

Table A.7: Booster Simulation Starting Population with High Primary Uptake

	Child	Adult LC	Adult HC	65+
N	$73,553,240$	$132,073,398$	$68,037,811$	$49,238,581$
S	0	0	0	0
S_x	$47,462,186$	$6,388,806$	$3,291,203$	$2,394,698$
E	$615,314$	$2,351,386$	$1,211,320$	$781,570$
I_c	$1,025,523$	$3,918,976$	$2,018,866$	$1,302,616$
I_{sc}	0	0	0	0
R	0	0	0	0
D	0	0	0	0
V_a	0	0	0	0
V_b	$17,115,152$	$83,589,962$	$43,061,495$	$31,331,788$
V_{bx}	$7,335,065$	$35,824,269$	$18,454,927$	$13,427,909$
V_{boost}	0	0	0	0

A.1.6 Model Parameters

Parameter	Description	Value	Source
C_{ij}	Contact rate between individuals in group i and j	Estimated	Berkeley Interpersonal Contact Survey (BICS)
α	Relative Infectiousness of clinical versus sub-clinical cases	0.5	Assumption, in line with the range presented in McEvoy et al. 2020
$1/\sigma$	Mean latent period	3 days	Davies et al. 2020; Bubar et al. 2021
$1/\gamma$	Mean infectiousness period	6 days	Davies et al. 2020; Bubar et al. 2021
ρ	Probability of having clinical symptoms per age group	$[0.26, 0.36, 0.69]$	Davies et al. 2020 Extended Data Fig. 7
u	Relative susceptibility to infection per age group	$[0.39, 0.83, 0.74]$	Davies et al. 2020Extended Data Fig. 7; Bubar et al. 2021
μ	Mortality per group	[0.004%, 0.023%, 8%]	Levin et al. 2020
R_0	Basic Reproduction Ratio	2.5	Feehan and Mahmud 2021
VE_a, VE_b	Primary model: Vaccine effectiveness after first dose and second dose	$0.80, 0.90$	Tenforde 2021; Thompson 2021
VE_b, VE_{boost}	Booster model: Vaccine effectiveness after second dose and booster dose	0.37, 0.81	Thompson 2022
v_t	Time between first and second primary doses	25 days	Assumed

Table A.8: Description of parameters for baseline model. Groups specified are for children, all adults, and seniors, respectively.

The force of infection λ_i for a susceptible individual in group i is:

$$\lambda_i = u_i \sum_j C_{ij} \frac{I_{c_j} + \alpha I_{sc_j}}{N_j} \tag{A.14}$$

where u_i is the age-dependent susceptibility to infection after contact with an infectious individual, C_{ij} is the entry in contact matrix $\boldsymbol{C}$ corresponding to the average number of daily contacts a respondent in group i has with an individual in group j; I_{c_j} and and I_{sc_j} are the number of clinically infectious and subclinically infectious individuals in group j, α is the relative infectiousness of clinical versus subclinical infections, and N_j is the total population size (Davies et al. 2020; Bubar et al. 2021).

Strategy	Description
No Vax	Null scenario; no vaccine distribution
65+ Prior	Initial prioritization for seniors over age 65; remaining vaccines distributed proportionally to eligible group size
HC Prior	Initial prioritization for high contact workers; remaining vaccines distributed proportionally to eligible group size
Split Prior	Priority vaccine dose are split evenly between seniors over age 65 and high contact workers; remaining vaccines distributed proportionally to eligible group size
Tiered 65+	First priority vaccines are given to seniors over age 65 and second to high contact workers; remaining doses are distributed to children and low contact adults proportional to eligible group size
Tiered HC	First priority vaccines are given to to high contact workers and second to seniors over age 65; remaining doses are distributed to children and low contact adults proportional to eligible group size

Table A.9: Summary of the 5 vaccination prioritization scenarios for analysis

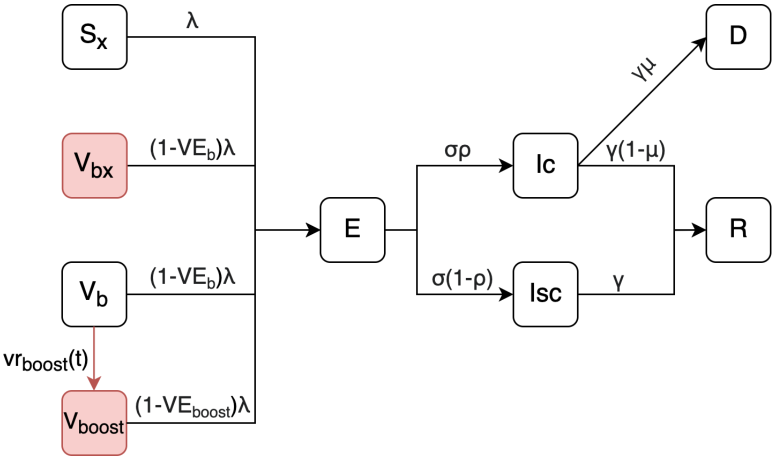
A.1.7 Vaccine distribution

We took a general value of 2 million vaccine doses distributed daily in all primary simulations. There are six strategies tested for a set of parameters:

- No Vax: No vaccination (no priority or general doses distributed)
- 65+ Prior: Prioritizing only seniors before general distribution
- HC Prior: Prioritizing only adults 18-64 with in-person contacts before general distribution
- Split Prior: Splitting priority vaccines evenly (i.e., 500,000 each) between seniors and working adults before general distribution
- Tiered 65+: ‘Tiered’ rollout that prioritizes, in order, seniors before working adults before general distribution
- Tiered HC: ‘Tiered’ rollout that prioritizes, in order, working adults before seniors before general distribution

Here, ‘general distribution’ refers to distributing vaccine doses proportionally to remaining group size. For all strategies besides the no vaccination baseline, general distribution occurs after all priority doses have been distributed. The tiered strategies are intended to replicate the CDC’s guidance, which opened up priority eligibility to the elderly, high risk people, and essential workers in progressive tiers before opening up eligibility to all. In the primary 2-dose vaccination model, the 2 million eligible doses are split evenly between first and second doses, meaning that 1 million susceptible can receive the first dose daily and 1 million people in V_a can receive the second dose daily.

A.1.8 Booster (3rd dose) vaccination Model



$$\lambda_i = u_i \sum_j C_{ij} \frac{Ic_j + \alpha Isc_j}{N_j} \tag{A.15}$$

$$\frac{dS_x}{dt} = -\lambda S_x \tag{A.16}$$

$$\frac{dV_{bx}}{dt} = -\lambda(1 - VE_b)Vb_x \tag{A.17}$$

$$\frac{dV_b}{dt} = -\lambda(1 - VE_b)Vb - VR_{boost}(t) \tag{A.18}$$

$$\frac{dV_{boost}}{dt} = -\lambda(1 - VE_{boost})V_{boost} + VR_{boost}(t) \tag{A.19}$$

$$\frac{dE}{dt} = -\sigma E + \lambda\Big[S_x + (1 - VE_{Vb})(V_b + V_{bx}) + (1 - VE_{boost})V_{boost}\Big] \tag{A.20}$$

$$\frac{dIc}{dt} = -\gamma I_c + \rho\sigma E \tag{A.21}$$

$$\frac{dIsc}{dt} = -\gamma Isc + (1 - \rho)\sigma E \tag{A.22}$$

$$\frac{dR}{dt} = (1 - \mu)\gamma I_c + \gamma I_{sc} \tag{A.23}$$

$$\frac{dD}{dt} = \mu\gamma I_c \tag{A.24}$$

Figure A.3: Booster vaccination model. (Caption continued on following page)

Figure A.3: (Continued from previous page) The model is extended to accommodate the prioritization of additional booster doses for people who have already received two doses of the vaccine (compartment V_b). Additions to the model are indicated in red; compartments S and V_a are removed as no shots are allocated to unvaccinated individuals. Susceptible compartments S_x, V_{bx}, V_b, and Vboost indicate resepectfully to individuals who have received no doses of the vaccine but will not receive boosters, those who have received two doses but will not get booster doses, those waiting for booster doses, and those who have received booster doses. $VRboost_i$ is the rate at which booster doses are given to people in Vb_i. $VE_{b,boost}$ indicate the efficacy of the vaccine at reducing transmission. *NB: Group index i is dropped to reduce visual clutter.*

A.1.9 Booster Model Sensitivity Analysis

We find that the most effective prioritization strategy for reducing deaths is highly sensitive to model parameters (figure 1.6), suggesting that the epidemiology of any subsequent variants will determine the most effective vaccination strategy for booster doses. Holding all other parameters at the baseline median value, we varied the efficacy of the primary course of vaccination from 0 to 50%, efficacy of the booster doses between 50% and 100%, and R_0 between 2 and 12. All other parameters are held at baseline values. We find that the reduction in deaths is nearly always higher under senior prioritization than under high risk adult prioritization. Efficacy of the primary and booster doses plays a substantial role in shaping the outbreak; 100% effective booster doses reduce deaths by nearly 50% under 65+ prioritization, whereas a 50% effective booster only reduces deaths by 10%. As R_0 increases from 2 to 12, we observe a similar cross-over in clinical infections as the primary scenario: at values of R_0 below 3.6, Tiered HC and HC Prioritization limits the most clinical infections; above this value, 65+ and Tiered 65+ Prioritization limit the most clinical infections. Additionally, higher values of R_0 generally correspond to less effective interventions, since the outbreak peak is higher by the time of widespread booster uptake.

Investigation of clinical infections averted during High Contact adult booster prioritization (figure 1.5B) reveals an interesting dynamic where counts of averted infections relative to the no booster scenario appear to stall after mid-January. Further investigation of group-level incidence rates (supplemental figure A.9) reveals that when this group is prioritized in part or in full, after mid January counts of infections among HC adults are actually lower in the null scenario—where no booster doses are distributed—than simulations that included booster distribution. We observe that the HC adult curve is 'flattened,' where the peak of the curve is lowered but infections continue to occur for longer. Ultimately only about 500,000 fewer clinical infections occur during HC Prioritization (of about 23 million in the null scenario), indicating that this strategy is much less effective at high values of R_0.

Uptake of the primary vaccine is considerably lower for adults between ages 18 and 65 and than for seniors; as a result, proportionally fewer working-age adults are eligible for a booster dose than seniors. As a sensitivity test, we ran an additional 1000 simulations assuming that HC and LC adult primary vaccine uptake rate matches the 65+ adult uptake rate of 95%; booster uptake remains at 70% of the eligible primary-vaccinated population.

1000 simulations are conducted with randomly drawn R_0 and vaccine efficacy parameters identical to the above simulation. We find a similar overall story: prioritizing seniors first, either solely or in a tiered roll-out, limits the most deaths, and the reductions in clinical infections is nearly equal between all five strategies. With baseline parameters, tiered 65+ roll-out reduces deaths by 26% compared to no booster doses and tiered HC roll-out reducing clinical infections by 12%.

A.1.10 Gender and Ethnicity

Table A.10 shows the gender and ethnicity of respondents in each category (Adult LC, Adult HC, 65+). High Contact adults are almost 20 percentage points less likely to be female than Low Contact adults. The ethnic breakdown of High Contact adults is similar to that of Low Contact adults. Adults aged 65+ are more about 10 percentage points more likely to be white and 15 percentage points less likely to be Hispanic.

Assuming that contact rates, probability of transmission, and mortality are equal by gender and ethnicity, prioritizing the oldest age group for vaccination averts the most deaths in all demographic categories. Even though Hispanics make up 20% of the adult population but only 5% of the 65+ population, prioritizing the oldest age group still averts nearly 20,000 more Hispanic deaths (about 20%) compared to HC prioritization. Since mortality among the oldest age group is about 350 times higher than for adults ages 18-64, the direct benefit of preventing deaths among this age group through vaccination remains the most effective strategy for vaccine distribution.

In aggregate, Tiered HC vaccine rollout limits the most clinical infections, and this remains true for Blacks and Asians. HC prioritization (without tiered roll-out) averts the most deaths among Hispanics. However, Tiered 65+ roll-out averts the most infections among Women and Whites, whose age distributions skew slightly older than the other groups. Like in the primary analysis, we note that differences in clinical infections between strategies is small.

Table A.10: Survey-weighted estimates of Gender and Ethnicity of BICS respondents, by category.

Category	Female	White	Black	Asian	Hispanic
Adult LC	57.68%	73.75%	13.56%	6.04%	20.56%
Adult HC	38.45%	76.38%	11.12%	4.91%	22.49%
65+	52.49%	85.27%	9.12%	3.29%	5.04%

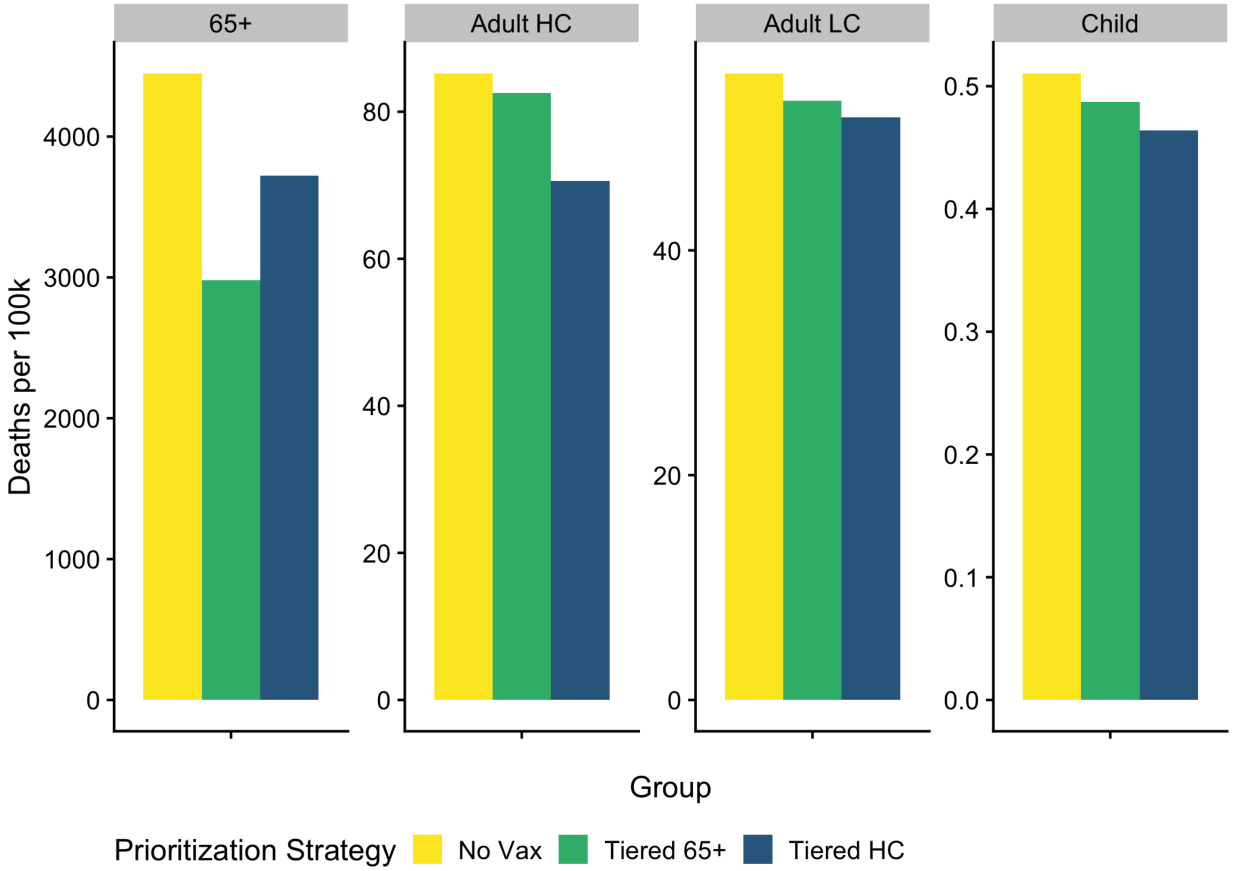


Figure A.4: Deaths by category used in transmission model for selected prioritization strategies (chosen to reduce visual clutter), per 100k in each category.

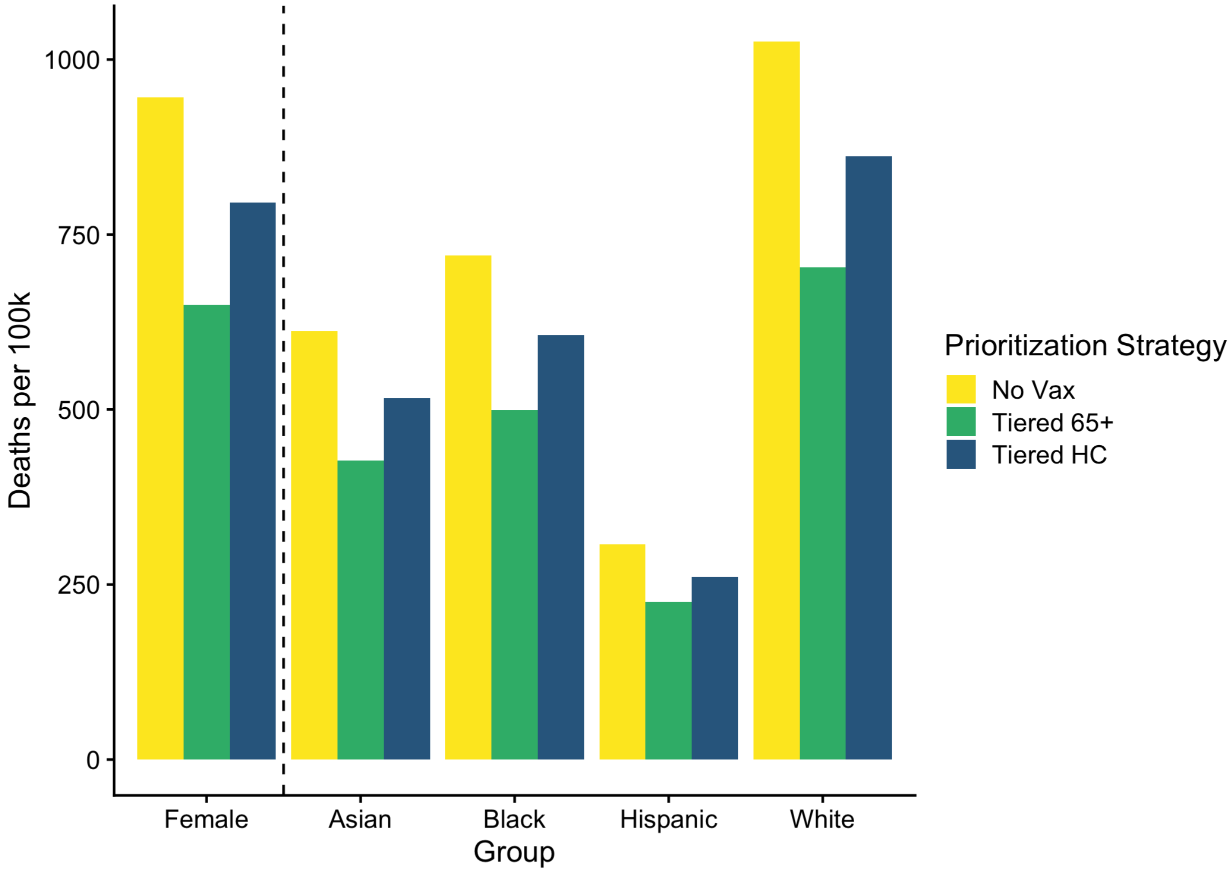


Figure A.5: Deaths per 100k by gender and ethnicity, assuming equal transmission and mortality parameters between groups, for selected prioritization strategies (chosen to reduce visual clutter). Ethnicities are mutually exclusive.

A.2 Additional Figures

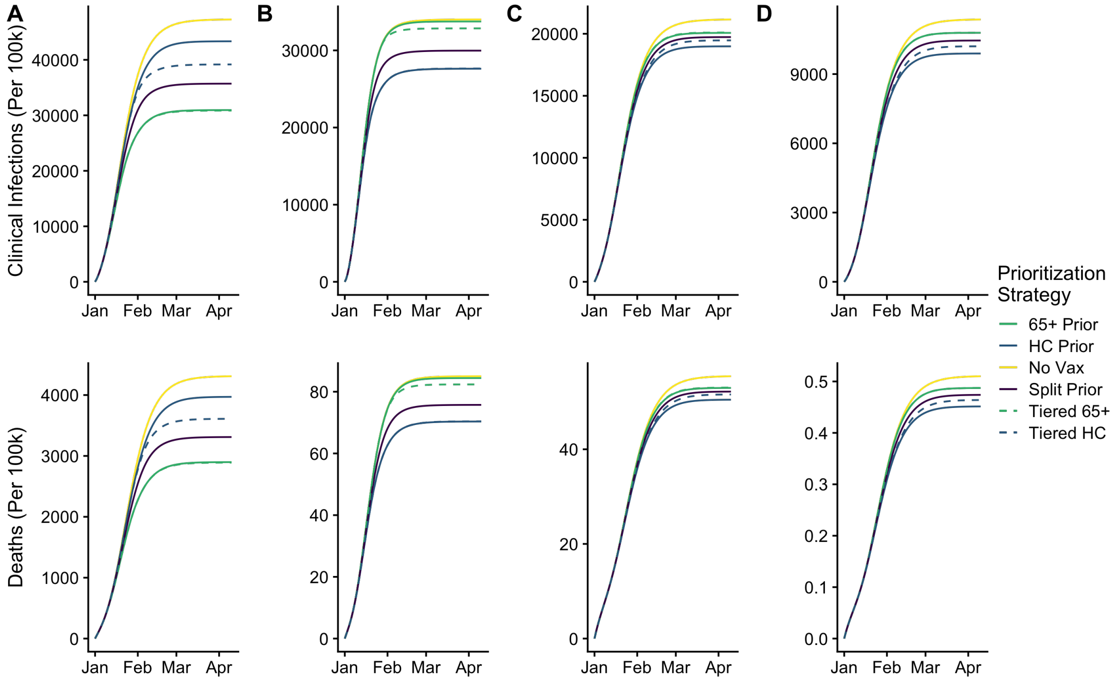


Figure A.6: Cumulative clinical infections and deaths during the baseline simulation for all demographic groups (A-D: Seniors, HC Workers, LC Workers, Children) and prioritization strategies per 100,000 individuals in each group. The largest reduction in infections occur from the direct effects of vaccinating HR workers and seniors, but indirect benefits are observed when seniors or workers receive non-priority access. Mortality reductions are greatest when a seniors or HC workers are directly given priority access.

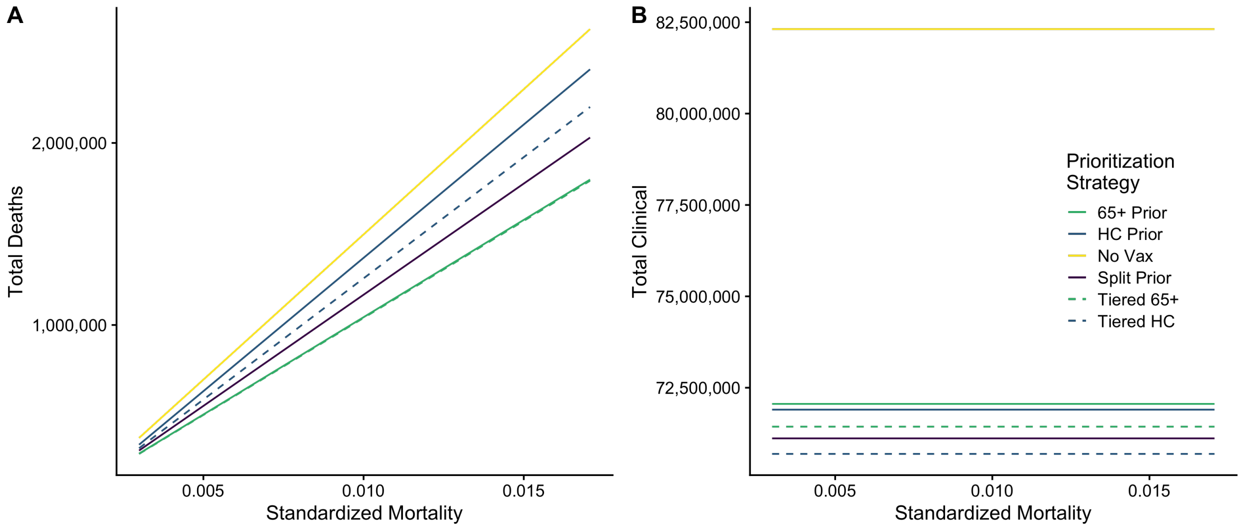


Figure A.7: Results of the sensitivity analysis of 65+ mortality μ_{65+} (varied between 0.01 and 0.1) and the effect on total deaths (A) and clinical infections (B). X-axis indicates total mortality standardized for group size.

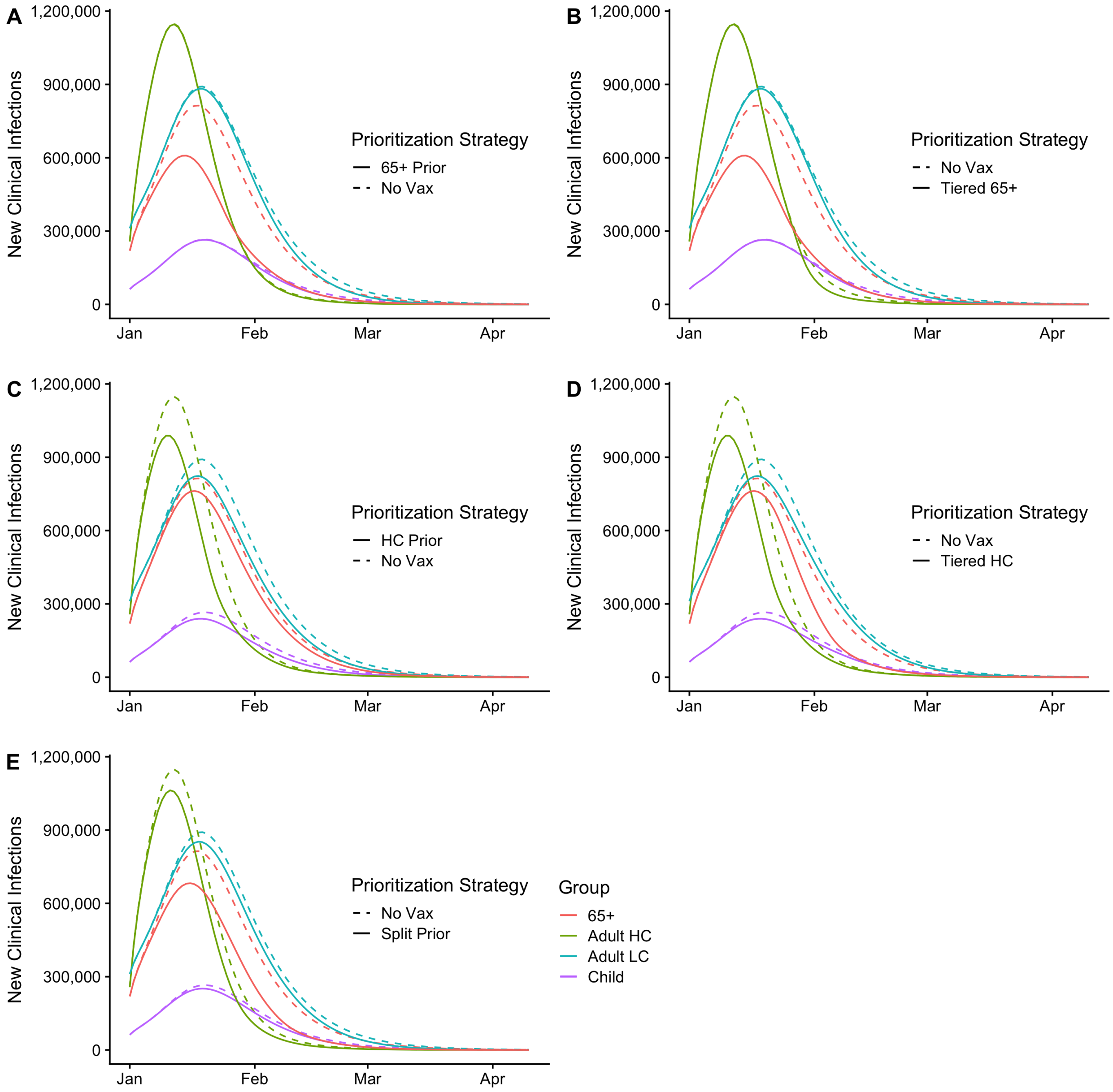


Figure A.8: New clinical infections daily by group and prioritization strategy for the primary (2-dose) vaccine distribution model with baseline parameters, with No Vaccine scenario shown as comparison. (A) 65+ Prioritization; (B) Tiered 65+ Prioritization; (C) High Contact Prioritization; (D) Tiered High Contact Prioritization; (E) Split prioritization.

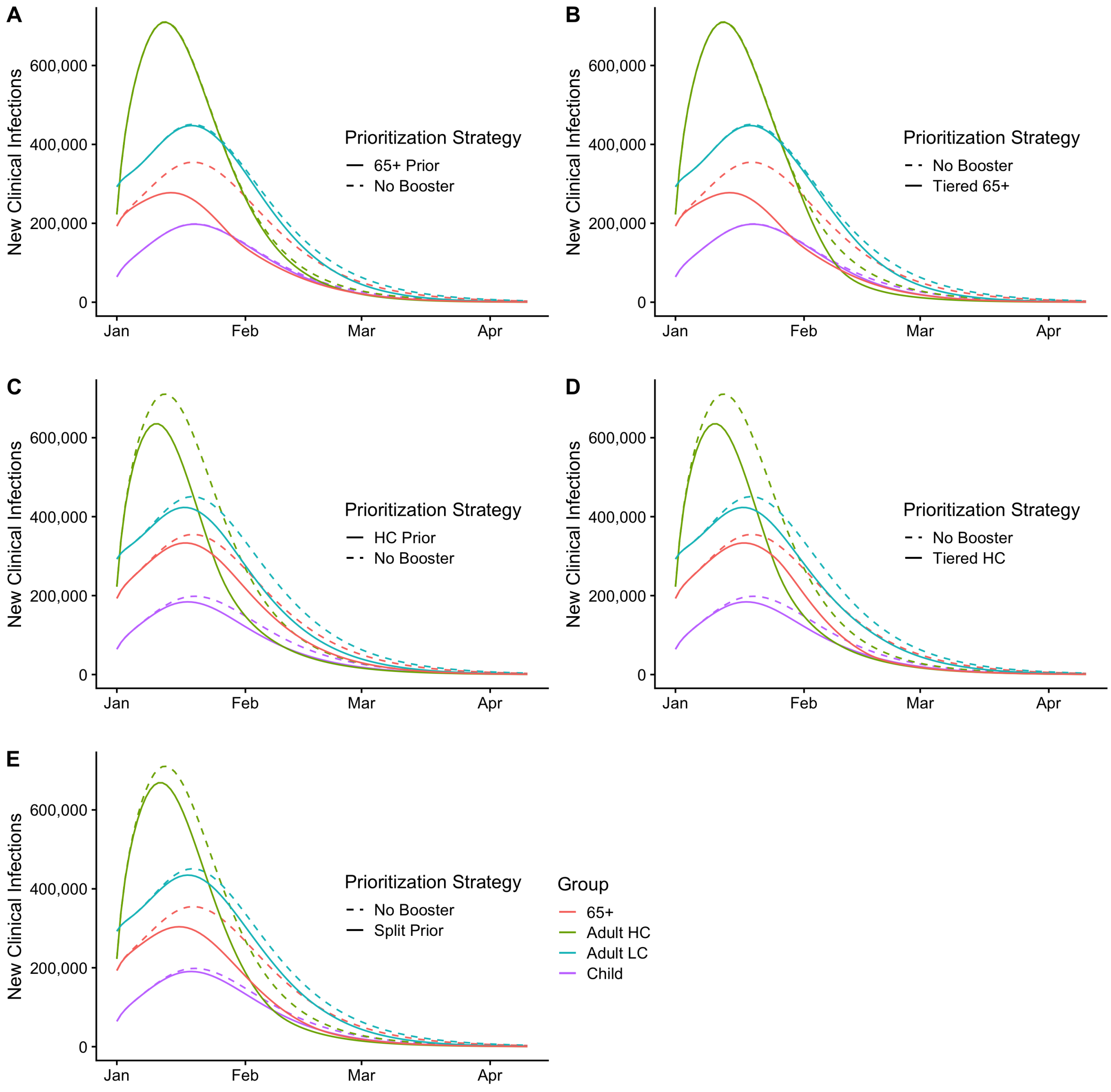


Figure A.9: New clinical infections daily by group and prioritization strategy for the booster vaccine distribution model with baseline parameters, with No Vaccine scenario as comparison. (A) 65+ Prioritization; (B) Tiered 65+ Prioritization; (C) High Contact Prioritization; (D) Tiered High Contact Prioritization; (E) Split prioritization. In scenarios where high contact workers receive priority vaccines, either directly, tiered, or split (panels C-D), incidence among that group is lower after the peak of the outbreak during the No Booster scenario than the booster scenario.

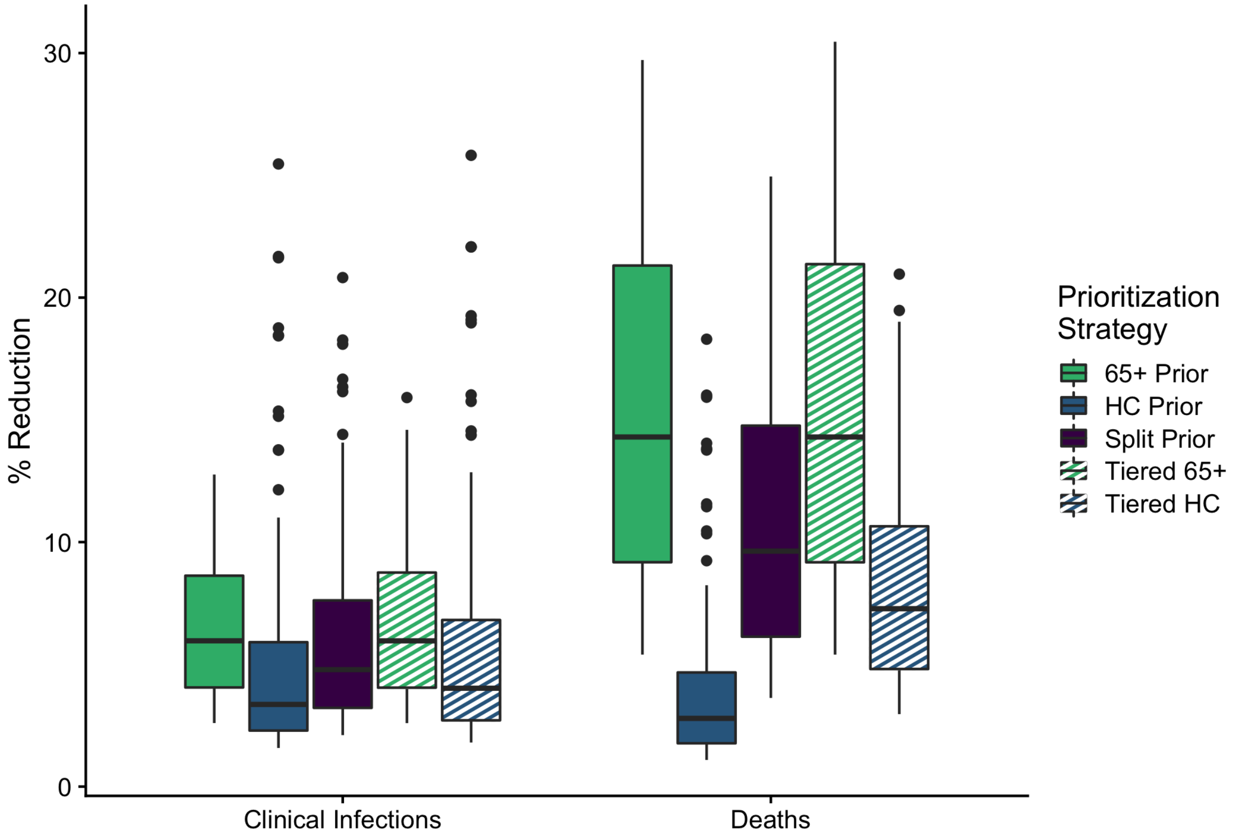


Figure A.10: Results from stochastically drawn parameter for high-uptake booster scenario

Appendix B

Supplementary Material for Chapter 2

B.1 Model Supplement

The model described here is a stochastic Agent-Based Network Simulation for COVID-19 transmission that utilizes data collected as part of the Berkeley Interpersonal Contact Survey to simulation population structure and contact networks. The core algorithm is written in `C++`, compiled using Apple Clang++17, and utilizes CMake 3.16 and utilizes the `igraph 10.4` library (Csárdi and Nepusz, T. 2006; Csárdi et al. 2023), and includes a Python 3.8 user API. The BICS ABM C++ library is compiled to a dynamic library for linkage to the Python API and can be used with an external C++ program.

The BICS ABM C++ library hinges on two input data structures and their Python equivalents: `Params`, a C-struct that contains the input parameters, documented in table B.1; and an array of the simulation population. The input population requires a strict format: each row is an individual and each column represents individual-level data passed to the simulation, flattened to one dimensional array of floats in column-major (Fortran-style) order. Currently, 8 fields are required for each simulated node: the node's household id; age, represented as a categorical index 0-8; gender, where 0 corresponds to male and 1 corresponds to female; number of non-household contacts; number of school contacts, which is taken to be zero for adults; number of times left home (unused presently, but maintained for legacy purposes); vaccine priority (see B.1.1); and a boolean indicating if the node uses NPIs or not. Parameters must be supplied indicating the dimensionality of the dataset. The simulation returns a `trajectory`, a C-struct containing an hourly time-series of all disease states.

The Python API contains a class `BICS_ABM`, which is a wrapper around all of the above utilizing the `ctypes` library for cross-language functionality. Parameters can be passed to the model through the Python API identically as to the C++ library and we recommend interacting with the simulation through the Python API. The class constructor for `BICS_ABM` takes any of the arguments to `Params`, runs the simulation, and saves as fields the components of the resultant trajectory as a `numpy.ndarray`; as such, the trajectory of clinical cases can be accessed as `BICS_ABM.Cc`. The population can be accessed through `BICS_ABM._pop`.

Parameter Name	Description	Default Value
N_HH	Number of simulation households	1000
WAVE	(Python only) BICS survey wave used to derive simulation population	6
GAMMA_MIN,MAX	Bounds for uniformly-sampled duration of latent period	2,4
SIGMA_MIN,MAX	Bounds for uniformly-sampled duration of infectious period	4,6
BETA_VEC	Array of length 365 indicating the daily baseline probability of infection	[0.025, ... 0.025]
BETA0, BETA1	(Python only) Average and amplitude of sinusoidal seasonal forcing of the baseline transmission probability; transformed into BETA_VEC	0.025, 0
CONTACT_MULT_VEC	Array of length 365 indicating daily multiplier for number of random contacts	[1, ... ,1]
C1	(Python only) Amplitude of sinusoidal seasonal forcing of contact multiplier; transformed into CONTACT_MULT_VEC	0
SCHOOL_CONTACTS	Whether to include school contacts for children during weekdays during the school year	True
MU_VEC	Vector of length 9 indicating the age-specific SARS-CoV-2 mortality for each of the 9 age categories	See main text
INDEX_CASES	Number of index cases	5
IMPORT_CASES_VEC	Array of length 365 indicating the number of imported cases each day of the year; only begins after T0	One Weekly
SEED	Random seed	None
N_VAX_DAILY	Number of vaccines distributed daily	100
VE1,2,BOOST	Efficacy of vaccines after first, second, and booster doses	0.8, 0.9, 0.8
VEW	Efficacy of waned immunity for both vaccines and infectious-derived immunity	0.5
ISOLATION_MULTIPLIER	Scaling factor for random contacts by clinically infectious nodes; 0 is full isolation and 1 is business as usual	0.5
T_REINFECTION	Duration (in hours) of full vaccine and infectious-derived immunity before waning	24*180
T0	Date of appearance of index cases	0

ALPHA	Relative infectiousness of subclinical cases	0.32
RHO	Proportion of subclinical cases	0.20
MAX_DAYS	Duration of simulation, in days	10 * 365
BOOSTER_DAY	Day of year that booster doses are made available	244
FERTILITY_VEC	Array of length 9 indicating the age-category specific fertility rate for females, adjusted for age group bin width	See main text
Mortality_VEC	Array of length 9 indicating the age-category specific mortality rate, adjusted for age group bin width	See main text

Table B.1: Parameters used in the ABM model

B.1.1 Model Pseudocode

Psuedocode for the model is given below:

1. In Python API: Establish the following parameters governing the population structure: number of households n_{hh}, survey wave, and vaccine priority. Establish all other transmission parameters and store them in an object `params`.
2. Generate n_{hh} households from the corresponding survey wave using the procedure described below. Assign each household a unique household identifier.
    (a) First, the distribution of adult ages and sex is derived from ACS data, and a set of N_HH households are sampled from this distribution.
    (b) A 'head' of household is randomly chosen from the BICS survey data. A respondent is eligible to be head of household h_i if they are the corresponding age and sex for the sampled household head. Eligible household heads are chosen with replacement and with probability adjusted for survey weights.
    (c) Finally, households are filled by sampling (with replacement and adjustment for survey weights) from the set of BICS respondents who mach each of h_i's reported household members' age, gender, and household size. Children under 18 are not ascertained in the survey; children are instead sampled from the POLYMOD survey. The max size of a household is 6 as as respondents were only asked to report 6 of their household members.
3. Assign vaccine priority to all nodes in the network based off of the rules provided as input, as elaborated in section B.1.1.
4. Create and pass the `params` and population object to the C++ core algorithm.

5. In C++ core: Determine the household contact network assuming that all nodes have contact with all members of their household. Randomly draw a school contact network for children under the age of 18.

6. Repeat the following procedure representing one 'day' of simulation time, where each day contains 24 'hours' of simulation time, until either no nodes are exposed or infectious OR the simulation has occurred for a supplied maximum number of days.

    (a) If hour == 0 (midnight) and an index case is supplied to appear on the current simulation day, transition one node at random into 'Exposed' status.

    (b) Each hour between midnight and 8am: assume that all nodes have contact with all members of their household. Execute `transmission` and `decrement` procedures for all nodes.

    (c) 8am: Distribute vaccine doses to `n_vax` nodes awaiting any dose of the vaccine.

    (d) 8am: generate a random graph of daily contacts using the procedure outlined in section B.1.1; assign a random duration and start time for all contacts.

    (e) Each hour between 8am and 6pm: connect all random contacts; disconnect each node having a random contact from their household nodes; `transmit`; and `decrement`. Reconnect nodes after termination of random contact.

    (f) Each hour between 6pm and midnight: transmit within households assuming that all nodes have contact with all members of their household. Execute `transmit` and `decrement` procedures.

7. At the conclusion of the simulation, return trajectories of each disease and vaccination status.

Update Handlers

The C++ program contains a centralized method for handling and dispatching changes to nodes, edges, and the graph itself. Classes exist for five types of changes can be executed: `UpdateGraphAttribute`, `CreateEdge`, `DeleteEdge`, `UpdateEdgeAttribute`, and `UpdateVertexAttribute`. Vertices cannot be created or destroyed using the update handler. All updates are stored in wrapper class `UpdateList`, which contains an overloaded method `UpdateList.add_update()` to add an update of each type. Updates are then dispatched with the method `UpdateList.add_updates_to_graph(igraph_t*)`, which takes a reference to the graph object as an argument to perform the updates.

The centralized update handlers were developed to streamline the attribute interface. Although the `igraph` API contains methods for updating individual node or edge attributes, it is more computationally efficient to pull all of the attributes, make all changes, then push them back to the graph. Since the `igraph` API involves many dynamically-allocated objects, this meant keeping track of many pointers, being sure to free all used vectors of attributes. With a project of this size, adding a centralized way of dispatching updates helped with debugging many issues with memory management, at the cost of a slight function overhead. As well, the object-oriented interface allows for saving of all, say, household edges in a single object to be connected and disconnected as needed.

Decrement procedure

The `decrement` procedure is among the most important function in the C++ core for tracking the progression of nodes through simulation time. Each 'event' that can occur to a node (infection, development of clinical or subclinical infectiousness, recovery, waning immunity, becoming vaccinated, etc) is accompanied by a duration of the event. The `decrement` procedure decrements the time remaining at each status, and for some events (for example, recovery from infectiousness) will automatically trigger a status change; for other events (like eligibility for vaccination), eligible nodes are placed in a queue for the next event to occur.

Random contacts

Random contacts are drawn for the daytime hours, 8am-6pm, using the network configuration model (as described in the main text; Albert-László Barabási 2021). First, the number of daily-nonhousehold contacts is supplied for each simulation node; this number is first multiplied by the isolation multiplier if the node is clinically infectious and contact multiplier if included in the model, then taken to be a random draw from a Poisson distribution with rate parameter equal to the product of all three terms. This is done to allow for minor stochasticity in the model. Each node is assigned a number of stubs equal to this Poisson random draw; if the total number of stubs subs to an odd number, one stub is randomly deleted until the sum is an even number. Should this sum be zero stubs then the procedure aborts. The configuration model is drawn using the `igraph_degree_sequence_game` from the `igraph` library; the graph is then simplified to remove self-edges but not multi-edges.

Non-Pharmaceutical Interventions

We include a single generic Non-Pharmaceutical Intervention (NPI) intended to capture the combined transmission-preventing power of mask usage, gloves, and physical distancing. In wave 6, about 60% BICS respondents reported the usage of any of these possible NPIs in any of their reported contacts; any simulation nodes representing these respondents are given an NPI status of `True`. During the daytime simulation procedure, if two non-household nodes are connected who both have an NPI status of `True`, then their transmission probability is lowered by a supplied parameter representing the strength of these NPIs (see below).

Vaccine Distribution

Nodes in the population are assigned a discrete priority level for vaccination before the simulation begins. A set number of vaccines are distributed daily among nodes with the highest priority level until no nodes in that priority level remain; remaining doses are distributed

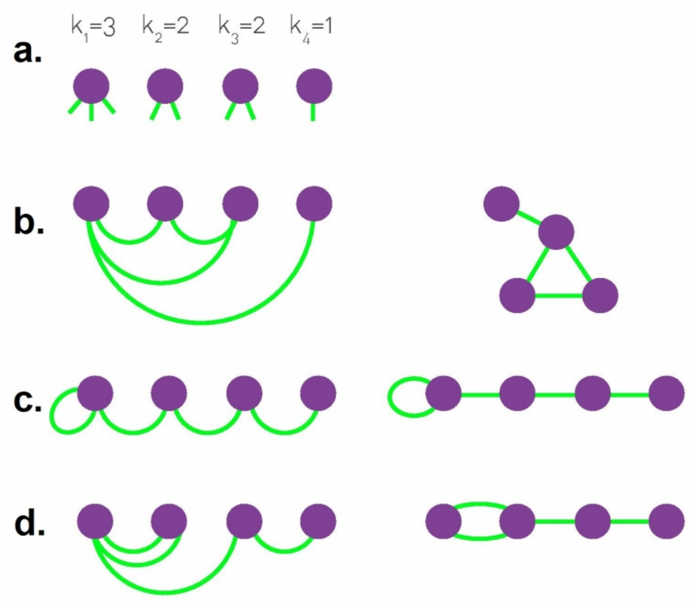


Figure B.1: Albert-László Barabási 2021's diagram representing realizations of the Network Configuration Model showing multiple ways of connecting the four nodes in panel (a) with corresponding degree k. (b): no self- or multi-edges; (c): allowing self-edges but not multi-edges; (d): allowing multi-edges, but not self-edges. Our application would allow for configuration (d) but not (c).

among nodes with the next highest priority level, repeating until the day's number of vaccines are exhausted. A priority level of -1 indicates that the node declines or is ineligible for vaccination. Nodes are eligible to receive the second dose of the vaccine 25 days after they received first dose. After a period of time, vaccine efficacy is assumed to wane; at this point, nodes are eligible to receive an additional booster dose.

Demographic Rates

Fertility and all-cause mortality are incorporated in the simulation according to published age-specific rates, aggregated for each age group in the simulation:

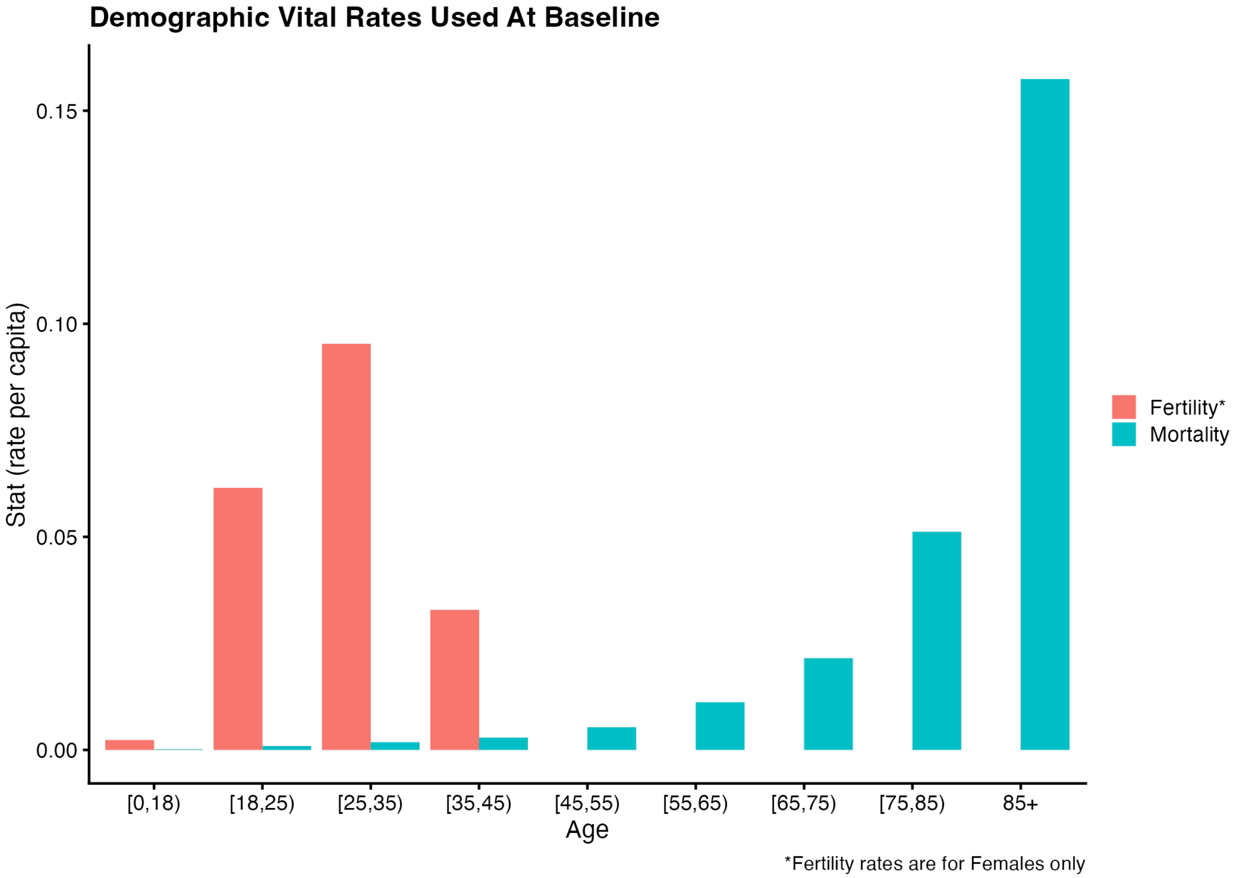


Figure B.2: Baseline demographic vital rates used in the simulation.

B.1.2 Supplementary Figures

Booster Dose Distribution Day, High Uptake

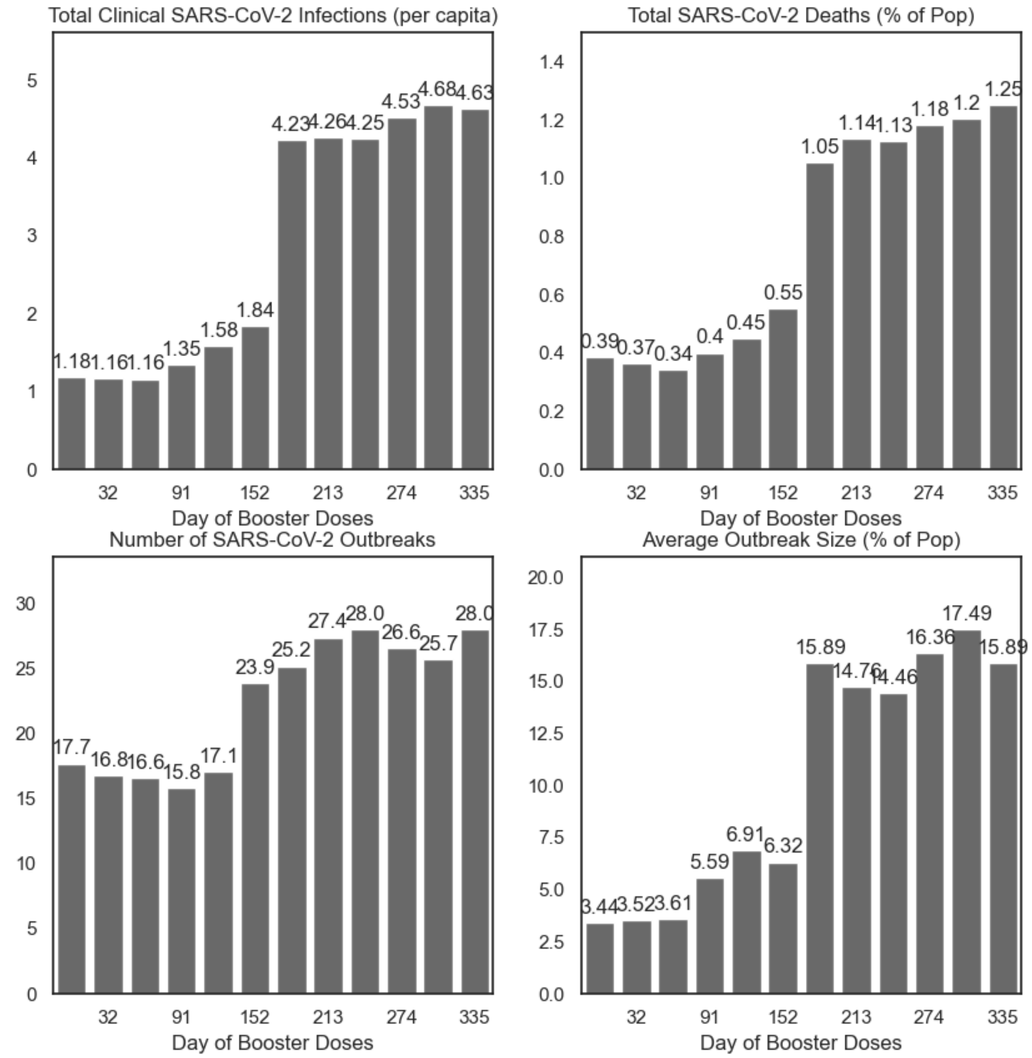


Figure B.3: Summary of simulations by day of booster dose distribution, varied as the first of each month, in the absence of seasonal forcing of the transmission parameter β, with 90% vaccine uptake.

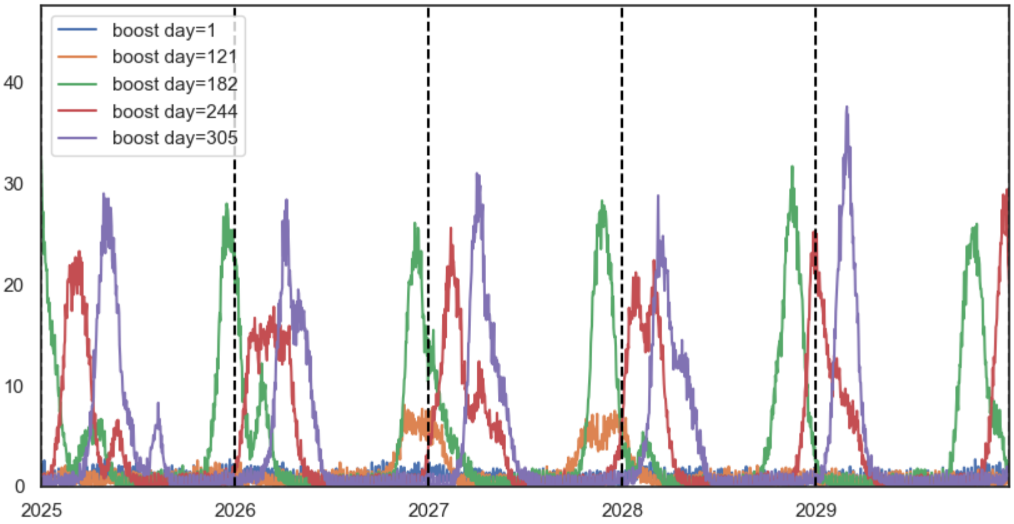


Figure B.4: Trajectory of simulations by day of booster dose distribution, without seasonal forcing, for selected distribution days: Jan 1st (day 1), May 1st (day 121), July 1st (day 182), Sept 1st (day 244), and Nov 1st (day 305), with 90% uptake.

Seasonality in contact rates

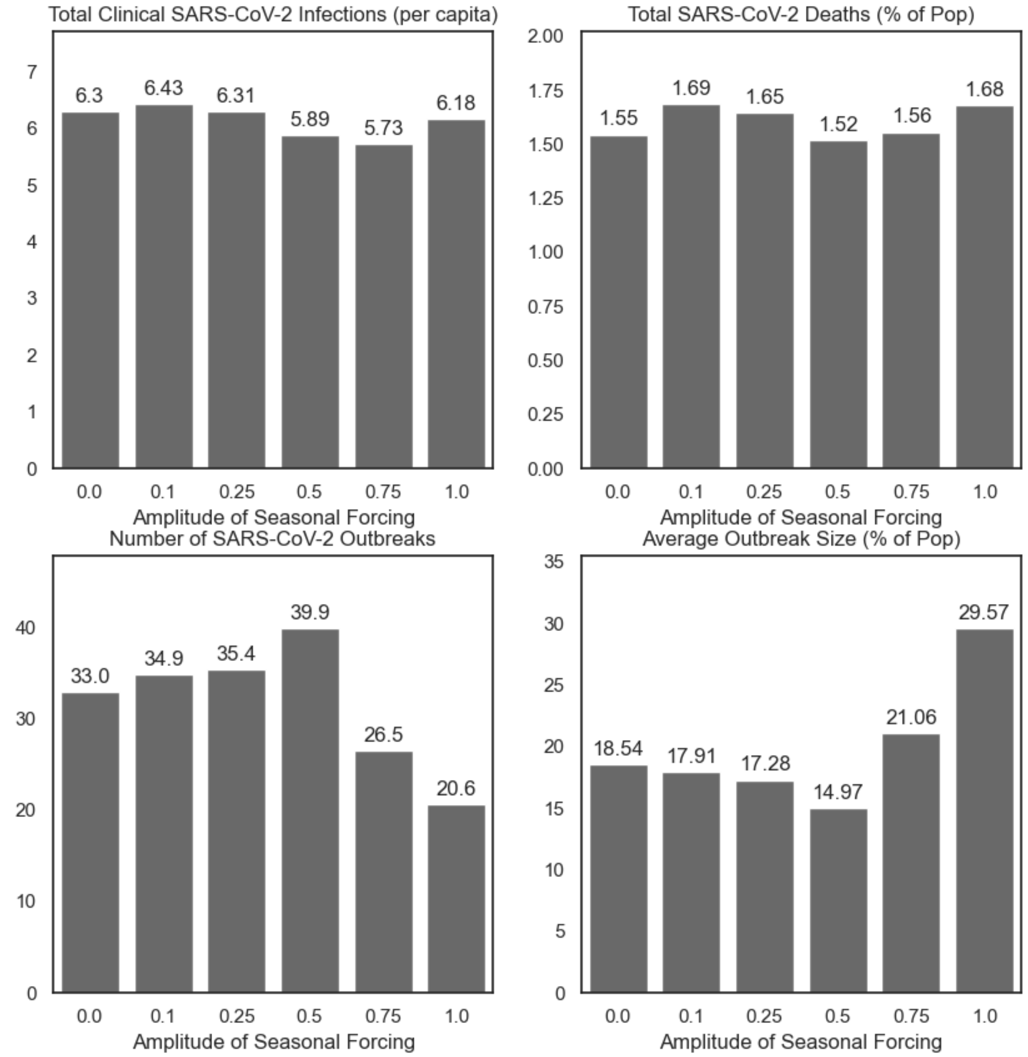


Figure B.5: Summary of simulations at selected levels of c_1, the amplitude of seasonal forcing of the contact parameter.

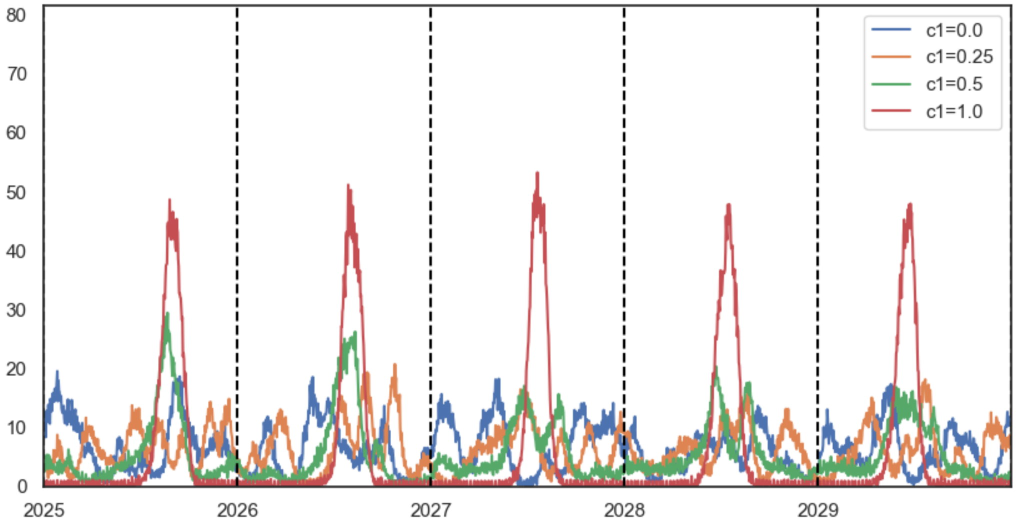


Figure B.6: Trajectories at selected levels of c_1, the amplitude of seasonal forcing of the contact parameter.

Isolation of infectious cases

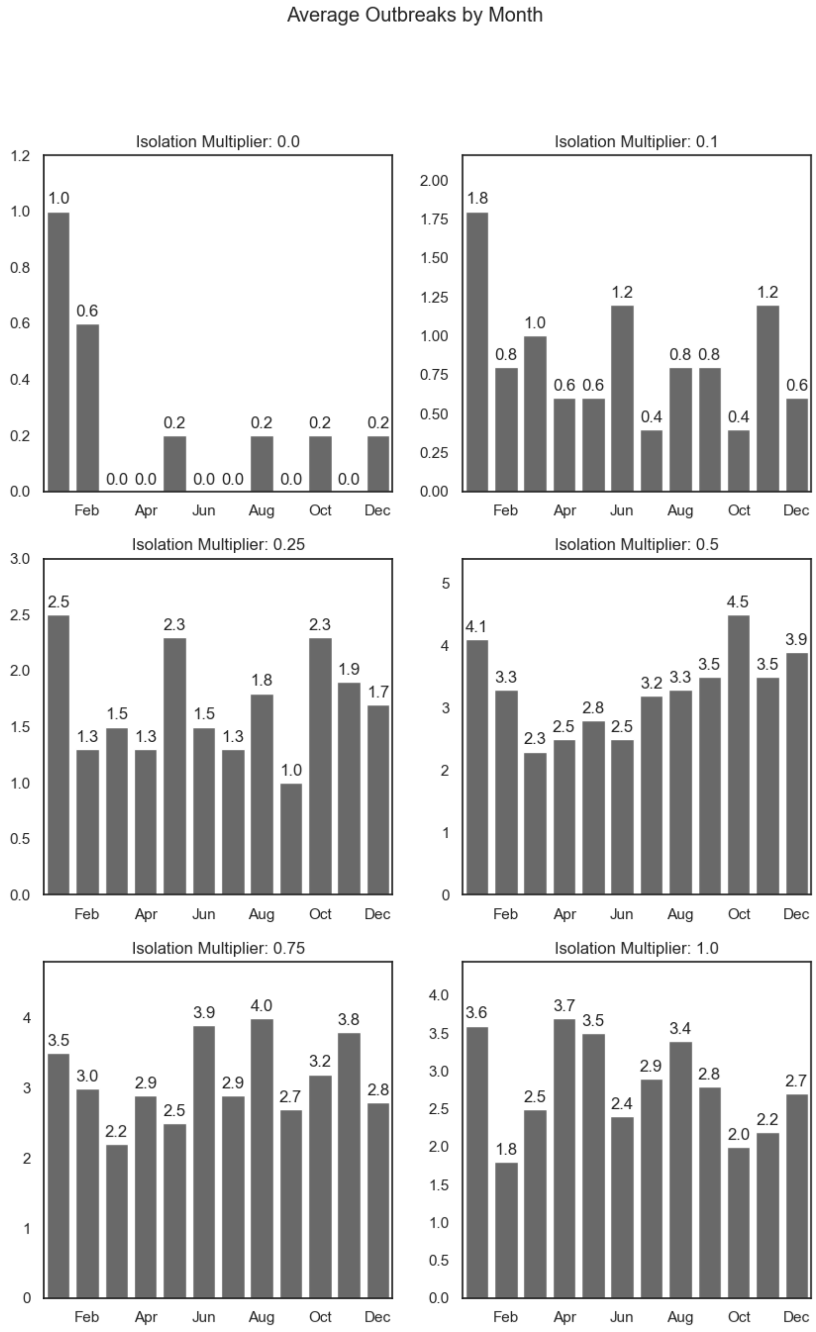


Figure B.7: Outbreak seasonality at selected levels of isolation

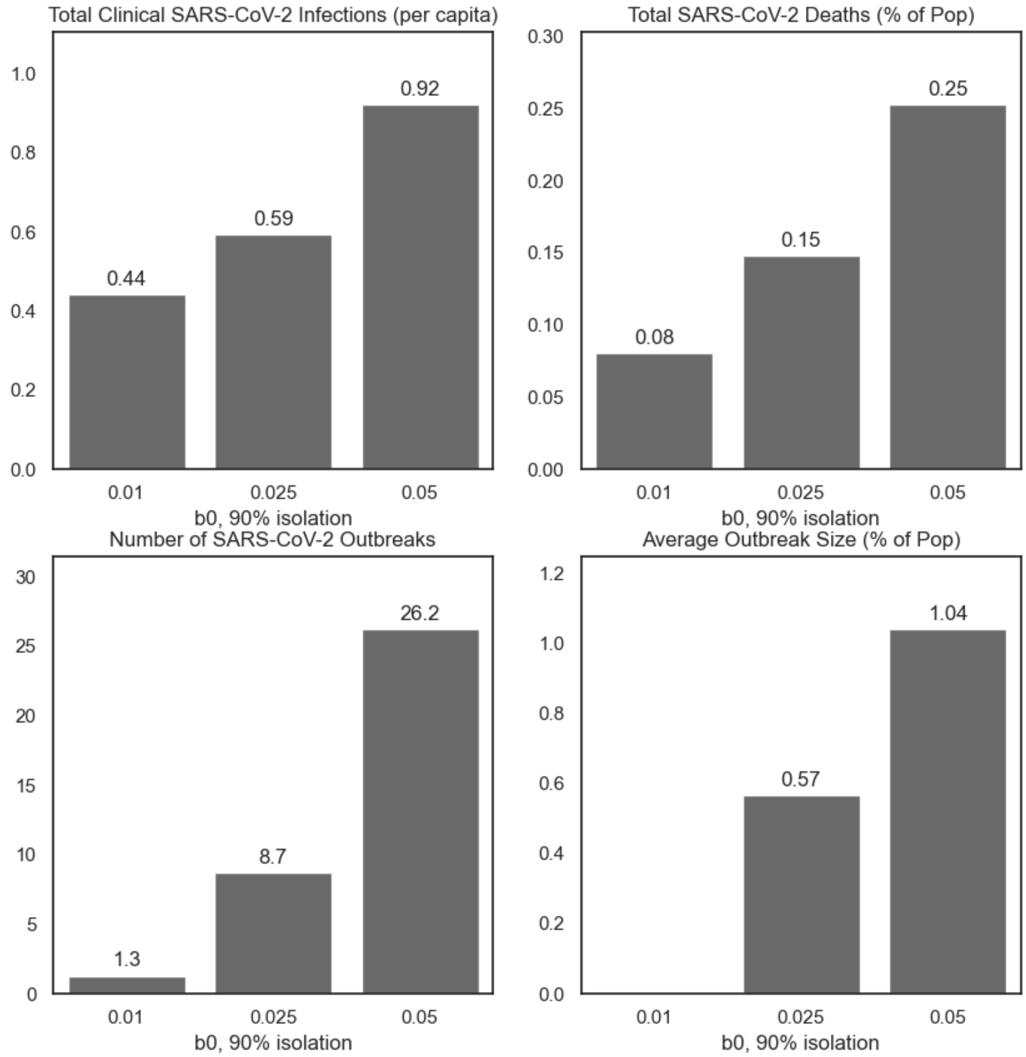


Figure B.8: Summary of simulations at selected levels of β_0 with 90% isolation of clinically infectious cases (corresponding to an isolation multiplier of 10%).

Appendix C

Supplementary Material for Chapter 3

C.1 Adjacency Matrices for Geostatistical Models

For the analysis of Chagas disease, we focus on methods for areal or polygon data, which refer to a region of space which contains a subset of the observations of interest. Polygonal data is a common format for administratively-collected spatial data, often representing a governmentally-defined area—such as a state or province equivalent, city or municipality, or even more specific form such as census tract or block. Areal data exist in contrast to point-referenced data, which instead link each observation with longitude and latitude coordinates. Whereas areal data can be generated from point data using a simple point-in-polygon operation, the reverse process is not possible as the specific coordinates are lost when points are tallied within polygons.

Critical in any spatial statistcal work is the concept of the neighborhood matrix: a mathematical representation of geographic adjacency. For example, this 3x3 grid could be representing by binary neighborhood matrix W:

1	2	3
4	5	6
7	8	9

$$W = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

A symmetric binary matrix like this is most common for representing adjaceny, but can easily be extended to include reciprocal distance weights, higher-order neighbors, or measures of connectivity that are not strict adjacency (for example, consider transit networks or other transportational features that mean the travel time between two locations is not linear with distance). While estimates will change between different matrices W, the following distributional properties remain the same.

In the general case of eq. 3.1 with a non-binary neighborhood matrix $\boldsymbol{W}$, the equivalent, generalized model can be parameterized:

$$\phi_i | \boldsymbol{\phi_{j \neq i}} \sim N\left(\frac{1}{w_{i+}} \sum_j w_{ij}\phi_j, \frac{\tau^2}{w_{i+}}\right) \tag{C.1}$$

where $w_{i+} = \sum_j w_{ij}$, or the sum of matrix $\boldsymbol{W}$ row i.

C.2 Commentary on the BYM- type model

In their 1991 paper, BYM use the IAR distribution (eq. 3.1) to fit a log-linear Poisson GLM of the form:

$$log\mu_i = logE_i + x_i\beta + \phi_i$$

where μ_i is the rate of disease occurrence in unit i, E is expected count unit i, x and β are explanatory variables and regression coefficients, and error term ϕ_i has the prior distribution of eq. 3.1. This model, which utilizes internal standardization, was called by Banerjee, Carlin, and Gelfand "cheating (or at least 'Empirical Bayes')" since the E_i are not fixed but rather itself part of the data, posing a 'null hypothesis' about if there were to be absent a spatial pattern (BCG, p.151)[1]. As well, BCG and Leroux et al (2000) note that this model may have poor performance as using the CAR prior alone as an error term may over-smooth aspatial variation, which may be mechanistically important to the model.

The IAR distribution is conditionally specified for each geographic unit and is improper, meaning that it does not integrate to produce a valid probability distribution, instead only able to show the proportional density between spatial units. This is problematic for stochastic generation and maximum likelihood estimation, but is valid for Bayesian inference as posterior density need only be proportional to the prior density (Besag, York, and Mollie 1991). However, it is possible and mathematically convenient to consider equation 3.1 in its joint, albeit improper, form. Besag (1974) showed that fully conditional distributions of this type can utilize Brook's Lemma (1964) to recover the full conditional form, as a multivariate normal distribution with mean 0 and variance-covariance matrix related to the adjacency matrix. Banerjee, Carlin, and Gelfand demonstrate this concisely, determining the joint distribution of ϕ from a fully specified set of conditionals:

$$p(\phi_1, ..., \phi_n) \propto exp\left\{-\frac{1}{2\tau^2}\sum_{j\sim i}(\phi_i - \phi_j)^2\right\} \tag{C.2}$$

which is also known as the pairwise-differences formula (Banerjee, Carlin, and Gelfand 2015, p.81, eqn. 4.16). Equation C.2 can be utilized to provide convenient estimation in Bayesian

---

[1] This is forgiveable as BYM—who were working in digital image restoration—were the first to demonstrate how this technique could be used in other fields, which has become a foundational technique in Bayesian Disease Mapping.

MCMC, using software such as Stan as pairs of adjacent units can be efficiently stored in program memory, and the proportional density can be quickly computed without the need for matrix inversion. Since Stan estimates the proportional log-density of ϕ up to a constant, Morris et al. 2019 demonstrate that equation C.2 can be quickly evaluated as:

```
phi ~ sigma * -0.5 * dot_self(phi[node1] - phi[node2]);
sum(phi) ~ normal(0, 0.001 * N);
```

where `sigma` is the precision rather than the variance, where `node1` and `node2` are vectors of adjacent pairs, `dot_self` takes the dot product of the vector with itself, and the second line indicates that `phi` is subject to a soft sum-to-0 constraint. Due to this computational efficiency and problematic assumptions needed to make this distribution properly integrate, BCG recommend that the IAR model be used only in the case of a Bayesian prior, and may be frequently the optimal choice for geostatistical inference (BCG, ch. 4 and BCG, ch. 6, p.155)

The expression in C.2 is an improper probability distribution since the joint probability density is only proportional to the derived expression. This is because the variance-covariance matrix implied is singular, meaning the inverse does not have a unique solution and as a result the distribution does not necessarily sum to one, as required for valid probability distributions. For a non-mathematical explaination, consider that each observation is entirely dependent on its neighbors, which allows us to estimate the total distribution only on relative terms without a 'ground truth' or some external source centering the distribution. To demonstrate this impropriety, BCG (p.81) derive equation 3.1 by beginning with adjacency matrix $\boldsymbol{W}$, which has $w_{ij} = 1$ if i and j are neighbors and 0 otherwise; matrix $\boldsymbol{B}$ where $b_{ij} = w_{ij}/w_{i+}$, or a row-standardized version of matrix $\boldsymbol{W}$; and $\boldsymbol{D}$, a diagonal matrix where d_{ii} is equal to the number of neighbors of i and 0 otherwise. Then, equation 3.1 can be written in the conditional form as $\phi_i|\phi_{j\sim i}$ $N(\sum_j b_{ij}y_j, \tau_i^2)$ since $\boldsymbol{B}$ is the row-standardized version of $\boldsymbol{W}$. This would imply that $\boldsymbol{\phi} \sim MVN(\mathbf{0}, [\tau(\boldsymbol{D} - \boldsymbol{W})]^{-1})$. Temporarily disregarding τ, calculating the covariance matrix Σ^{-1} of this distribution involves the calculation $(\boldsymbol{D} - \boldsymbol{W})\mathbf{1} = \mathbf{0}$, which is singular; effectively, too many variables without a constraint to preserve propriety (BCG, p.81). It is possible to make this distribution proper with an additional parameter, often denoted ρ (α in Morris et al. 2019), by defining the inverse covariance matrix $\Sigma^{-1} = \boldsymbol{D} - \rho\boldsymbol{W}$, so long as ρ is chosen to find a singular solution. BCG list the bounds under which ρ will provide a non-singular solution, which is related to the eigenvalues of matrices $\boldsymbol{D}$ and $\boldsymbol{W}$. Then, the full conditional distribution becomes:

$$\phi_i|\boldsymbol{\phi_{j\sim i}} \sim N\left(\rho\frac{1}{n_i}\sum_{j\sim i}\phi_j, \frac{\tau^2}{n_i}\right) \tag{C.3}$$

BCG write that ρ is sometimes reported as being the degree of spatial autocorrelation, but it is clear from equation C.3 that the resultant expression rather expresses some *proportion* of the spatial gaussian process (p.82). As well, ρ does not map clearly onto any other measures of spatial autocorrelation, like Moran's I or Geary's C, and thus its interpretation outside of the model is limited.

Further, the authors remark that the proper CAR model may be attractive in cases where the spatial pattern is weak, and the improper CAR model may over-smooth heterogeneity. However, in simulation, the proper CAR model has been shown to nearly always converge on values of ρ close to 1, as when ρ is less than 1 there presents an identification challenge between the spatial random effects and the non-spatial random effect. BCG remark (p.155) that it appears that the data always *want* ρ to be close to 1. In conclusion, BCG recommend that the improper IAR model be only used as a Bayesian prior, or in the frequentist case, use of a SAR or other proper probability distribution.

As noted above, two issues present with using the IAR model alone as a prior for a spatially autocorrelated error term. First, the IAR model is known to show poor performance when spatial autocorrelation is not very strong, otherwise it will oversmooth random variation in the data. This issue is rectified with a proper CAR model, but as noted above, BCG do not recommend usage of the proper CAR prior. Second, the IAR variance parameter τ has an ambiguous function, and sources differ as to its interpretation. While Leroux (2000) states that this parameter represents both autocorrelation and over-dispersion simultaneously, but Banerjee, Carlin, and Gelfand write that this parameter should not be taken as representing spatial autocorrelation in any mechanistic way.

Ideally, we would have included the BYM2-type spatial convolution term in both the Bernoulli and Poisson processes, however in model development we were unable to reach convergence with the model specified as such: the convolved spatial error for the Poisson parameter λ failed to be identified. Recall that this term is specified in both the Bernoulli and Poisson parameters as $\left((\sqrt{\rho/s})\phi + (\sqrt{1-\rho})\theta\right)\sigma$, where ϕ is the IAR model, θ is $N(0,1)$, ρ represents the proportion of variance having a spatial pattern rather than unstructured error, and σ is the overall variance. The convolved spatial process in the Bernoulli part showed excellent convergence and mixing with posterior estimates of $\rho_\pi \approx 0.75$. As a result, both the spatially clustering term ϕ and the spatial heterogeneity term θ contributed to posterior estimates of π and the likelihood of the model. However, in the Poisson part, median posterior estimates of $\rho_\lambda \approx 0$ such that the convolved error term $\approx \theta\sigma$. As a result, the sampler could not identify values for ϕ_λ, functionally searching the entire parameter space for ϕ_λ without any effect on the likelihood of the model. Ultimately, this caused slow evaluation and divergences, but did not affect the resultant values of other parameters.

C.3 Zero-Inflated Models and their efficient estimation in Stan

To assess the appropriateness of the ZIP distribution for our data, we conducted a naive maximum likelihood estimate of the Poisson parameter λ and the Zero-Inflated Poisson parameters π, λ without any adjustment for spatial structure, annual deviations, or covariates. Likelihood functions for both the Poisson-only specification and ZIP specification were optimized over the municipality-year counts of Chagas incidence using R's multivariate `optim` routine. The maximum likelihood estimate of Poisson-only parameter λ was 0.1053 with log-likelihood $\ell\ell = 57773.01$, and the estimate of ZIP parameters were $\pi = 0.9849$ and $\lambda = 6.9885$ with $\ell\ell = 19330.45$. To test the relative fit of both models, we conduct Wilk's test for likelihood ratios, which assumes that the ratio of two likelihoods is asymptotically distributed as $\chi^2(df = df_{H1} - df_{H0})$. Taking the Poisson-only specification as the null hypothesis and the ZIP specification as the alternative, the probability of observing these data generated by the Poisson-only specification instead of the ZIP specification is $p \approx 0$. Hence, we can comfortably reject the Poisson-specification in favor of the ZIP specification.

Using ZIP models may provide an additional computational advantage over a regular Poisson specification. We found that a model of the type laid out in the previous sections, which models the count of Chagas disease as a Poisson-distributed GLM with terms for fixed effects for spatial and aspatially-clustered errors, showed slow evaluation and poor estimation. While the sheer dimensionality of the model—approximately estimating 6 parameters for 5000 municipalities across 19 years—was undoubtedly responsible for part of the problem, we hypothesized that the complicated posterior geometry caused by the overdispersion of 0s in the dataset was partially to blame. To test this hypothesis, we ran two test cases each with a single UF over the first two years of the study period. We chose Pará (PA), which has the highest number of Chagas cases at 5259 over the 19 year study period in 143 municipalities, and Roraima (RR), with the second lowest number of cases at 10 cases in 15 municipalities[2]. In Stan, the test models were run for 500 warmup iterations and 500 sampling iterations. For PA, the model completed evaluation in 2508 seconds for an average parameter effective sample size (ESS) of 2234.738 (SD = 1413). However, despite RR having one tenth the number of municipalities of PA, the model for RR took more than twice as long to evaluate—5994.4 seconds— for an average parameter ESS of 1416.437 (SD = 451.5). Both models showed convergence ($\hat{R} \leq 1.01$) for more than 99% of parameter estimates. When this test was replicated using the ZIP model with an autoregressive component in the Bernoulli part only, the PA model evaluated in 46.2 seconds with an average parameter effective sample size of 470.96, and the RR model evaluated in 57 seconds with an average ESS of 25.39.

Naive implementation of Spatio-Temporal statistical models involves many pairwise comparisons, which can be prohibitively computationally expensive for MCMC estimation. For example, our spatio-temporal adjacency structure may contain not only neighboring observations between all 5000 municipalities in Brazil, but the temporally-correlated neighbors as well. Assuming that an average municipality has 4 spatial neighbors and 4 temporal neighbors, this will result in inverting a neighborhood matrix of over 1.6 billion elements which is not reasonably evaluated with standard computing resources, let alone for thousands of MCMC iterations. We have taken many steps towards quick and efficient estimation in lieu of this challenge, which is a major contribution of this research in addition to the primary substantive estimation of Acute Chagas Disease incidence.

When sampling a ZIP GLM in Bayesian software such as Stan, we will have to write a custom log-probability mass function (LPMF, in Stan terms) to cover zero-inflation. First,

---

[2] Only the Federal District, DF, had fewer cases, at 4 over the 19 year period, but was not chosen since that UF contains only one municpality.

we assume that Bernoulli parameter π is estimated on a logit scale and Poisson parameter λ is estimated on a log scale, as is common for GLMs. Second, we must evaluate the probability density on the log scale. The following is an unoptimized ZIP estimation, adapted from Stan's documentation for Zero-Inflated Poisson models[3]:

```
for (t in 1:T){
    for (n in 1:N) {
        if (y[t,n] == 0) {
          target += log_sum_exp(bernoulli_logit_lpmf(1 | pi[t,n]),
                                bernoulli_logit_lpmf(0 | pi[t,n])
                                  + poisson_log_lpmf(y[t,n] | pi[t,n]));
        } else {
          += bernoulli_logit_lpmf(0 | pi[t,n])
                       + poisson_log_lpmf(y[t,n] | pi[t,b]);
        }
    }
}
```

where `target` is the log-probability of the model and `log_sum_exp(a,b)` $= \log(\exp(a) + (\exp(b))$. Clearly, this is a highly inefficient way to evaluate the model since the log-probability statement is conditioned on the data `y` which is known and constant through the course of the simulation. In computational efficiency terms, each evaluation of the likelihood will complete in $O(T \cdot N)$ time, meaning that the time to evaluate the log-probability statement is proportional to the number of municipalities times the number of years. For our application to Chagas Disease in Brazil, which contains observations of approximately 5000 municipalities over 19 years, this becomes extraordinarily slow, evaluating 1000 warmup and sampling iterations on the scale of 12-24 hours.

To optimize this Stan modeling statement, the Stan documentation recommends partitioning the data into zero and non-zero elements and evaluating them separately, but does not elaborate on how to do so in a GLM framework, which we have developed for our application. Doing so will allow for separate, efficient vectorized evaluation of the Bernoulli and Poisson GLM statements. Indeed, as explained elsewhere, vectorization is one of the primary benefits of using Stan over other Bayesian MCMC software suites, since vectorized probability statements evaluate much faster and with less overhead than doubly-looped functions. First, consider a matrix of counts Y with dimensions T (number of years) and N (number of municipalities). From Y, we will derive two matrices `zero_idx` and `nonzero_idx` with the same dimensions, containing the indices of zero and nonzero observations, and supported by vectors `zero_max` and `nonzero_max` with dimension T, where each element contains the annual number of zero and nonzero entries. In this way, for matrix row $t \in T$ columns $[1 : \texttt{zero\_max[t]}]$ contains the index of municipalities with zero entries, and $[\texttt{zero\_max[t]} + 1 : N]$ are uninitialized. Beginning with matrix of counts Y:

$$Y = \begin{bmatrix} 0 & 1 & 0 & 3 & 2 \\ 1 & 0 & 0 & 0 & 1 \\ 2 & 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 3 & 1 & 2 \end{bmatrix}$$

---

[3] https://mc-stan.org/docs/stan-users-guide/zero-inflated.html

that yields derived zero and non-zero matrices:

$$\texttt{zero\_idx} = \begin{bmatrix} 1 & 3 & / & / & / \\ 2 & 3 & 4 & / & / \\ 2 & 4 & / & / & / \\ 1 & 2 & 3 & 4 & 5 \\ / & / & / & / & / \end{bmatrix} \quad \texttt{nonzero\_idx} = \begin{bmatrix} 2 & 4 & 5 & / & / \\ 1 & 5 & / & / & / \\ 1 & 3 & 5 & / & / \\ / & / & / & / & / \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

with max vectors:

$$\texttt{zero\_max} = \begin{bmatrix} 2 & 3 & 2 & 5 & 0 \end{bmatrix} \quad \texttt{nonzero\_max} = \begin{bmatrix} 3 & 2 & 3 & 0 & 5 \end{bmatrix}$$

In Stan, zero-counts in year t can be be easily indexed from Y as `Y[t, zero_idx[t, 1:zero_max[t]]]` and nonzeros as `Y[t, nonzero_idx[t, 1:nonzero_max[t]]]`. Essentially, these sparse matrices are an efficient way to store ragged arrays, which are not supported natively in a C-based language like Stan. The R code for generating these matrices and vectors from matrix Y in:

```
n_T = nrow(Y)
N = ncol(Y)
zero_max = array(rep(0,n_T))
nonzero_max = array(rep(0,n_T))
zero_idx = matrix(0, nrow = n_T, ncol = N)
nonzero_idx = matrix(0, nrow = n_T, ncol = N)

for (t in 1:n_T) {
  for (n in 1:N){
    if (Y[t,n] == 0) {
      zero_max[t] = zero_max[t] +  1
      zero_idx[t, zero_max[t]] = n
    }
    else {
      nonzero_max[t] = nonzero_max[t] + 1
      nonzero_idx[t, nonzero_max[t]] = n
    }
  }
}
```

Then, we can turn to writing a log probability mass function describing equations 3.7 and 3.8 that takes advantage of this vectorization. First, recall that π is on the logit scale and λ is on the log scale, and we wish to evaluate the probability on the log scale. Assuming that π and λ have been transformed using their corresponding inverse-link functions, this yields likelihood:

$$\log(P(y_i = 0)) = \log\left(\pi_i + (1 - \pi_i)e^{\lambda_i}\right) \tag{C.4}$$

$$\log(P(y_i = k)) = \log\left((1 - \pi_i)e^{\lambda_i}\lambda_i^k / k!\right) \tag{C.5}$$

Equation C.4 is problematic in that it does not simplify to built-in Stan probability statements, but can be written in a way that is efficiently vectorized[4]. Luckily, equation C.5

---

[4] In theory, an additional optimization of the Zero-likelihood involves use of the `log-sum-exp` trick, which

simply evaluates to two independent expressions of Bernoulli probability and Poisson probability on the log scale, respectively. This means they can be evaluated in Stan as:

```
vector[N] pi_inv_logit;
vector[N] lambda_exp;

for (t in 1:T) {
    pi_inv_logit = inv_logit(pi[t]);
    lambda_exp = exp(lambda[t]);

    // Zeros
    target += sum(log(
           pi_inv_logit[zero_idx[t, 1:zero_max[t]]] +
        (1-pi_inv_logit[zero_idx[t, 1:zero_max[t]]]) .*
        exp(-lambda_exp[zero_idx[t, 1:zero_max[t]]])
      ));

    // Nonzeros
    target += bernoulli_lpmf(
                rep_array(0, nonzero_max[t]) |
                pi_inv_logit[nonzero_idx[t,1:nonzero_max[t]]]
          ) + poisson_lpmf(
                y[t, nonzero_idx[t, 1:nonzero_max[t]]] |
                lambda_exp[nonzero_idx[t, 1:nonzero_max[t]]]
          );
}
```

In a test case, this vectorized model evaluated 100 warmup iterations and 100 sampling iterations in 1099 seconds, more than 10 times faster than the non-vectorized example.

C.4 Knorr-Held Spatio-Temporal Models

The other priors outlined in Knorr-Held (2000) are, respectively:

- Type I interaction, where all interaction terms are *a priori* independent:

  $$P(\delta|\sigma_\delta) \propto exp\left(-\frac{\sigma_\delta}{2}\sum_{i\in I,t\in T}(\delta_{it})^2\right) \tag{C.6}$$

  which is suitable if the space-time interaction does not have any structure.

- Type II interaction, where each spatial unit follows a 1st order random walk independent of its neighbors:

  $$P(\delta|\sigma_\delta) \propto exp\left(-\frac{\sigma_\delta}{2}\sum_{i\in I,t\in[2:T]}(\delta_{it}-\delta_{i,t-1})^2\right) \tag{C.7}$$

  which is suitable if temporal trends differ between spatial units and the temporal trends do not have any structure in space.

---

provides computationally efficient evaluation of `log(a+b) = log(exp(a) + exp(b))`, but this remains unexplored at present.

- Type III interaction, where interaction effects follow and intrinsic autoregression such as the type laid out in equation 3.1, but are indepdent at each time:

$$P(\delta|\sigma_\delta) \propto exp\left(-\frac{\sigma_\delta}{2}\sum_{j\sim i,t\in T}(\delta_{it}-\delta_{jt})^2\right) \tag{C.8}$$

  which is suitable if the spatial trends differ between time points, but the temporal trends do not have any structure in space.

- Type IV interaction, perhaps the most methodologically and conceptually interesting, where effects are totally dependent over space and time:

$$P(\delta|\sigma_\delta) \propto exp\left(-\frac{\sigma_\delta}{2}\sum_{j\sim i,t\in[2:T]}(\delta_{it}-\delta_{jt}-\delta_{i,t-1}+\delta_{j,t-1})^2\right) \tag{C.9}$$

  Which defines a space-time Markov random field and is suitable if temporal trends are different from location to location but are more likely to be similar in adjacent locations. This prior can be written in Stan as:

```
real knorr_held_type4_lpdf(vector delta_t, vector delta_tm1, int N, int[] node1, int[] node2) {
    return -0.5 * dot_self(delta_t[node1] - delta_t[node2] - delta_tm1[node1] + delta_tm1[node2]) +
              normal_lpdf(sum(delta_t) | 0, 0.001*N) ;
}
```

  where `delta_t` is the value of δ at time t, `delta_tm1` is the value of δ at time $t-1$, `node1` and `node2` indicate adjacent pairs of nodes, and the `normal_lpdf` statement indicates a soft sum-to-0 constraints for δ_t, as done for the ICAR prior above.

At face value, interaction Type IV would be the most useful for our purposes, however in model development we found that this model both was under-identified and over-smoothed random variation in the data. Instead, we opt for Type I priors, which are both simpler to estimate and more easily identified. Theoretically, type IV interactions are comparing not only the first degree neighbors—each observation with its spatial neighbors and previous observation—but also the 2nd order neighbors—the spatial neighbors of temporal neighbors, or equivalently, the temporal neighbors of spatial neighbors (Knorr-Held 2000). Essentially, this prior is an extension of the pairwise-differences CAR prior (eq C.2) to the temporal dimension. Where the pairwise CAR prior focuses on the differences between adjacent units, the Knorr-Held Type IV prior includes the differences between adjacent units in the current time period and the prior time period. Knorr-Held remark that such a model may be useful for modeling the spatio-temporal spread of both infectious diseases and non-infectious diseases where the underlying risk has a spatio-temporal pattern, as is appropriate for our application to Chagas disease. For the first time point $t = 1$, the ‘previous’ time period $t = 0$ is unavailable, so for this case only the ‘previous’ time period is instead taken to be a 0 vector, at which point the model simplifies to the ICAR prior. If the full model is specified as in equation 3.5, it is likely that identifiability will be poor without highly informative priors, as was the case for the BYM model above. For the present application to Chagas disease, it may be possible to disregard the time trends α and γ, since most locations begin and remain absent Chagas disease throughout the duration of study.

C.5 Stan model code, edited slightly for clarity

```
functions {
  real icar_normal_lpdf(vector phi, int N, array[] int node1, array[] int node2) {
    // Soft sum-to-zero constraint
    return -0.5 * dot_self(phi[node1] - phi[node2]) + normal_lpdf(sum(phi) | 0, 0.001*N);
 }
}
data {
  // Number of municipalities
  int<lower=0> N;
  // Number of years
  int<lower=0> T;
  // Number of adjacent edges
  int<lower=0> N_edges;
  // node1[i] adjacent to node2[i]
  array[N_edges] int<lower=1, upper=N> node1;
  // and node1[i] < node2[i]
  array[N_edges] int<lower=1, upper=N> node2;
  // count outcomes
  array[T,N] int y;
  // Population exposure
  array[T,N] int E;
  // Scaling factor-- scales variance of spatial effects
  real<lower=0> scaling_factor;
  // indices of zero counts
  array[T,N] int zero_idx;
  // Max number of zero counts
  array[T] int zero_max;
  // indices of nonzero counts
  array[T,N] int nonzero_idx;
  // max number of nonzero counts
  array[T] int nonzero_max;
}
transformed data {
  // Logged population
  array[T] vector[N] log_E;

  for (t in 1:T) {
    log_E[t] = to_vector(log(E[t,1:N]));
  }
}
parameters {
  // Bernoulli part: Knorr-Held model
  // Intercept
  real mu_pi;
  real mu_lambda;

  // Structured temporal trend
  vector[T] alpha_pi;
  vector[T] alpha_lambda;
  real<lower=1e-10, upper=10> sigma_alpha_pi;
  real<lower=1e-10, upper=10> sigma_alpha_lambda;

  // Structured spatial pattern
  vector[N] phi_pi;
```

```
  // vector[N] phi_lambda;

  // Unstructured spatial pattern
  vector[N] theta_pi;
  vector[N] theta_lambda;

  // Proportion of spatial/aspatial error
  real<lower=0, upper=1> rho_pi;
  // real<lower=0, upper=1> rho_lambda;
  real<lower=1e-10, upper=10> sigma_convolved_pi;
  real<lower=1e-10, upper=10> sigma_convolved_lambda;

  // Knorr-Held Type I spatio-temporal interaction
  array[T] vector[N] delta_pi;
  array[T] vector[N] delta_lambda;
  real<lower=1e-10, upper=10> sigma_delta_lambda;
  real<lower=1e-10, upper=10> sigma_delta_pi;


}
transformed parameters{
  array[T] vector[N] pi;   // Bernoulli GLM term
  array[T] vector[N] lambda;  // Poisson GLM term

  for (t in 1:T) {
    pi[t] = inv_logit(mu_pi +
            alpha_pi[t] +
            sigma_convolved_pi * (
              sqrt(rho_pi/scaling_factor) * phi_pi + sqrt(1-rho_pi)*theta_pi
            ) +
            sigma_delta_pi * delta_pi[t]);
    lambda[t] = exp(log_E[t] + mu_lambda +
              alpha_lambda[t] +
              sigma_convolved_lambda * (
                // sqrt(rho_lambda/scaling_factor) *
                // phi_lambda
                theta_lambda
                // sqrt(1-rho_lambda)*theta_lambda
              ) +
              sigma_delta_lambda * delta_lambda[t]);
  }
}
model {
  // Intercepts
  mu_pi ~ normal(-10, 10);
  mu_lambda ~ normal(-5, 10);

  // Structured temporal trend
  alpha_pi[1] ~ normal(0, sigma_alpha_pi);
  alpha_pi[2:T] ~ normal(alpha_pi[1:(T-1)], sigma_alpha_pi);
  sigma_alpha_pi ~ gamma(2, 1);

  alpha_lambda[1] ~ normal(0, sigma_alpha_lambda);
  alpha_lambda[2:T] ~ normal(alpha_lambda[1:(T-1)], sigma_alpha_lambda);
  sigma_alpha_lambda ~ gamma(2, 1);

  // Structured spatial patten
  phi_pi ~ icar_normal(N, node1, node2);
  // phi_lambda ~ icar_normal(N, node1, node2);

  // Unstructured spatial error
  theta_pi ~ std_normal();
  theta_lambda ~ std_normal();

  // Prior on Rho
```

```
  rho_pi ~ beta(.5, .5);
  // rho_lambda ~ beta(.5, .5);

  // Convolved variance
  sigma_convolved_pi ~ gamma(2,1);
  sigma_convolved_lambda ~ gamma(2,1);

  for (t in 1:T){
    // Interaction
    delta_pi[t] ~ std_normal();
    delta_lambda[t] ~ std_normal();

  }
  sigma_delta_pi ~ gamma(2, 1);
  sigma_delta_lambda ~ gamma(2, 1);



  // Likelihood
  for (t in 1:T) {

    // Vectorized ZIP
    // Zeros
    if (zero_max[t] > 0) {
      target += log(
             pi[t, zero_idx[t, 1:zero_max[t]]] +
          (1 - pi[t, zero_idx[t, 1:zero_max[t]]]) .*
          exp(-lambda[t, zero_idx[t, 1:zero_max[t]]])
        );
    }

    // Nonzeros
    if (nonzero_max[t] > 0) {
      target += bernoulli_lpmf(
                  rep_array(0, nonzero_max[t]) |
                  pi[t, nonzero_idx[t,1:nonzero_max[t]]]
             );
      target += poisson_lpmf(
                  y[t, nonzero_idx[t, 1:nonzero_max[t]]] |
                  lambda[t, nonzero_idx[t, 1:nonzero_max[t]]]
             );
    }
  }
}
```

C.5.1 Main Model Additional Figures

C.5.2 Climate Model Additional Figures

Table C.1: MCMC convergence diagnostics for main smoothing model selected parameters

Parameter	$\hat{R}$ 5th Quantile	$\hat{R}$ 95th Quantile	ESS 5th Quantile	ESS 95th Quantile
π	1.00	1.01	941.00	3,708.96
λ	1.00	1.00	6,116.68	10,733.15
ϕ_π	1.00	1.00	993.54	3,845.31
θ_π	1.00	1.00	8,329.99	12,173.52
θ_λ	1.00	1.00	4,293.10	11,373.93
α_π	1.00	1.01	430.58	635.58
α_λ	1.00	1.01	364.32	537.66
δ_π	1.03	1.03	316.20	316.20
δ_λ	1.00	1.00	734.21	734.21

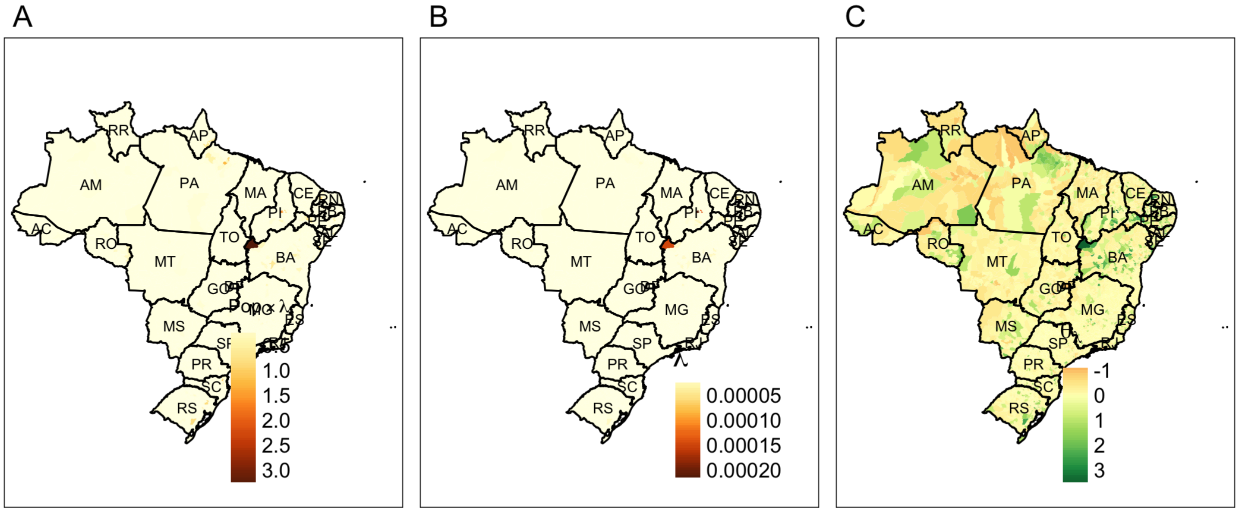


Figure C.1: Spatial process in the Poisson term, without temporal effects. A: overall rate of Chagas, calculated as $Pop_i \times \lambda_i$, where $\lambda_i = exp(\mu_\lambda + \theta_i * \sigma_i)$; B: per-capita rate of Chagas λ, net of population; C: spatial heterogeneity term θ_λ, with $N(0,1)$ prior.

Variable name	Description
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (×100)
BIO4	Temperature Seasonality (standard deviation ×100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

Table C.2: WorldClim suite of Bioclimatic variables downloaded from the Copernicus Climate Change Service's Global Bioclimatic Indicators from 1950-2100 Derived from Climate Projections dataset.

Variable	PC1	PC2	PC3	PC4	PC5	PC6
Annual Mean Temperature	0.316	-0.083	0.019	-0.066	0.065	-0.143
Mean Diurnal Range	-0.010	0.286	-0.494	-0.137	0.409	0.018
Isothermality	0.171	0.082	-0.133	0.408	0.662	0.139
Temperature Seasonality	-0.271	0.104	-0.074	-0.294	-0.159	-0.127
Max Temperature of Warmest Month	0.261	0	-0.186	-0.370	0.010	-0.216
Min Temperature of Coldest Month	0.302	-0.155	0.130	-0.032	-0.016	-0.147
Temperature Annual Range	-0.114	0.224	-0.417	-0.412	0.036	-0.054
Mean Temperature of Wettest Quarter	0.295	-0.038	0.003	0.061	0.056	-0.457
Mean Temperature of Driest Quarter	0.296	-0.113	0.073	-0.172	0.082	0.067
Mean Temperature of Warmest Quarter	0.292	-0.075	-0.005	-0.260	-0.030	-0.217
Mean Temperature of Coldest Quarter	0.315	-0.099	0.030	0.012	0.077	-0.061
Annual Precipitation	-0.175	-0.424	-0.170	-0.015	0.122	-0.128
Precipitation of Wettest Month	-0.051	-0.411	-0.370	0.127	-0.172	0.027
Precipitation of Driest Month	-0.223	-0.173	0.257	-0.220	0.316	-0.062
Precipitation Seasonality	0.222	0.099	-0.310	0.125	-0.268	0.222
Precipitation of Wettest Quarter	-0.052	-0.428	-0.359	0.149	-0.092	-0.054
Precipitation of Driest Quarter	-0.257	-0.156	0.177	-0.214	0.308	-0.143
Precipitation of Warmest Quarter	-0.251	-0.050	-0.109	0.297	0.049	-0.540
Precipitation of Coldest Quarter	0.041	-0.438	-0.013	-0.282	0.159	0.472
Standard Deviation	3.084	1.809	1.515	1.323	0.882	0.716

Table C.3: Principal Components 1-6 of the 19 WorldClim Bioclimatic Variables for median municpality-years in Brazil, 2000-2019.

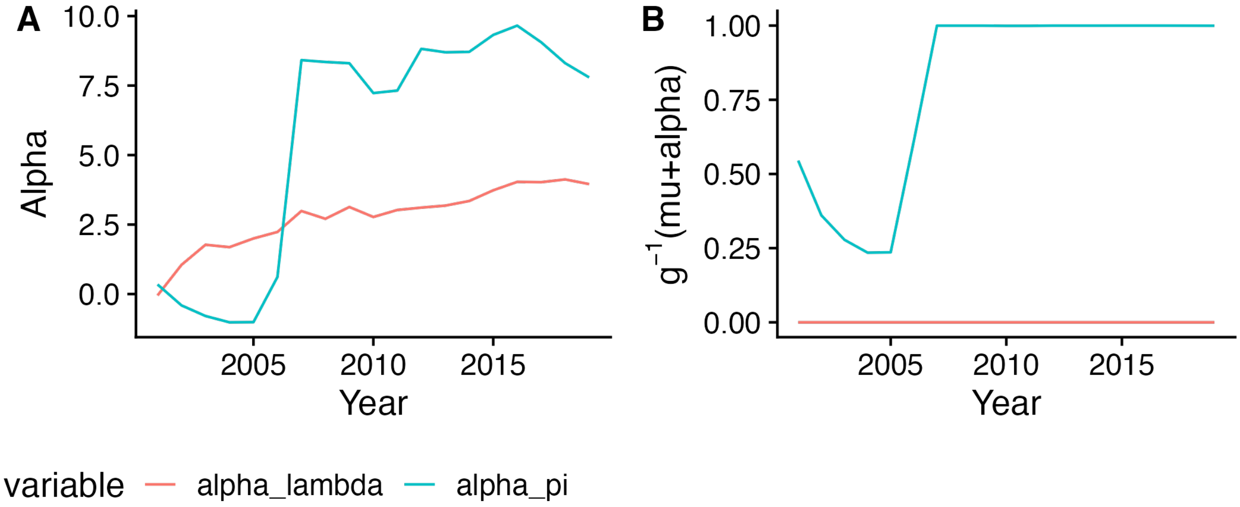


Figure C.2: Global AR(1) time trend for Bernoulli and Poisson processes, on the (A) crude scale and (B) transformed scale, where the transformed scale is $logit^{-1}(\mu_\lambda + \alpha_\pi)$ for the Bernoulli probability and $exp(\mu_\lambda + \alpha_\lambda)$ for the Poisson process. While the Poisson process always stays near 0, indicating that the rate of Chagas conditional on its presence in an area is stable over time, the global temporal trend of the Bernoulli parameter indicating probability of non-exposure drops initially, recovering to 100% by 2008. This implies that over the period of study, Chagas disease became much less global and more local in presentation.

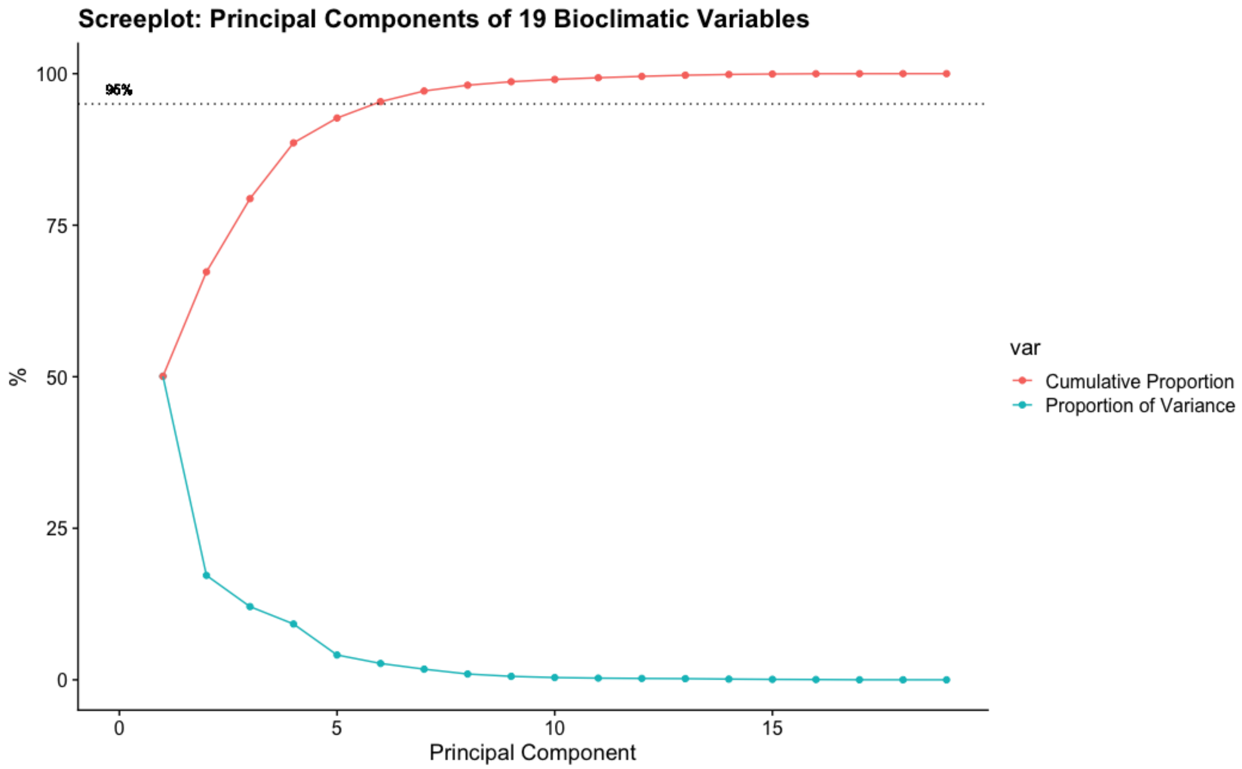


Figure C.3: Screeplot of variance and cumulative variance explained by the first n principal components, with 95% of cumulative variance indicated by the dotted line.

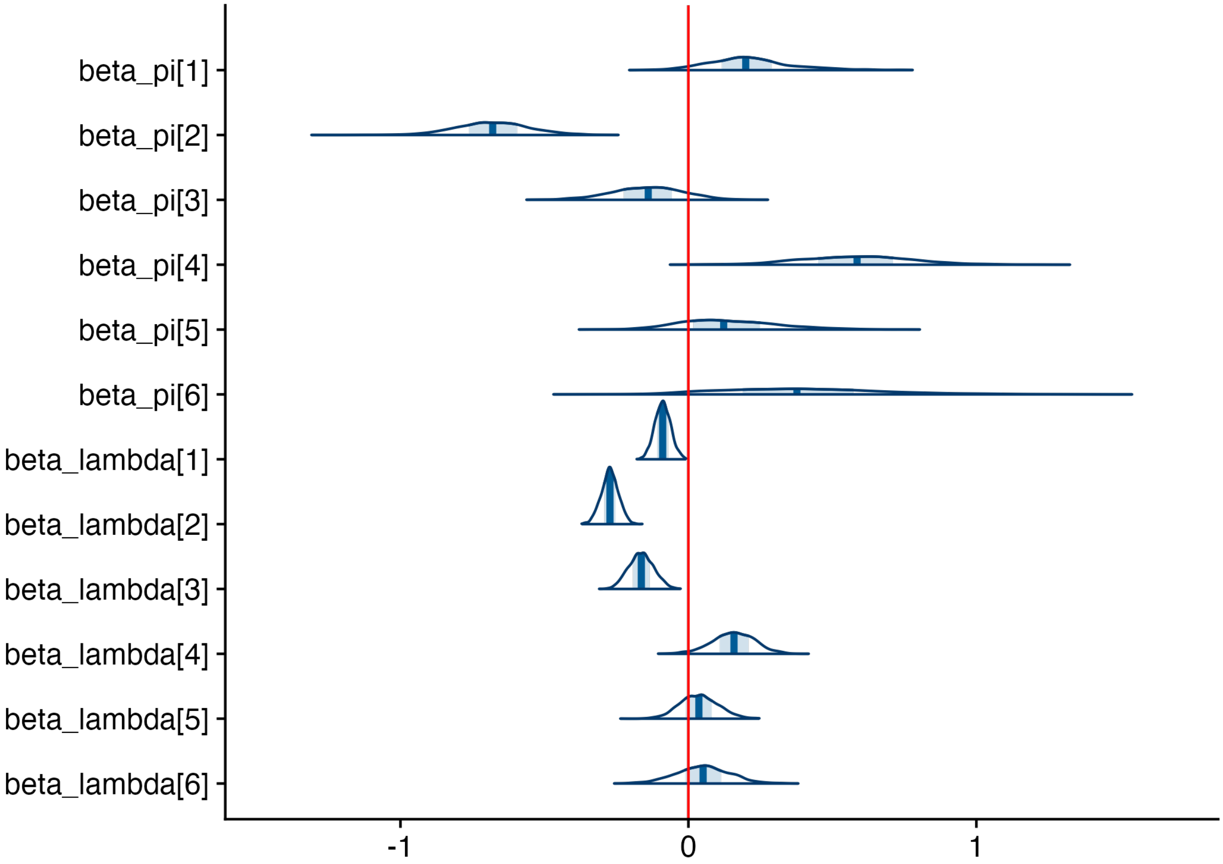


Figure C.4: Plot of Climate Covariates, specified for both the Bernoulli process (π) and Poisson process (λ) as each municipality-year's location in principal component space of the 19 WorldClim Bioclimatic Variables.

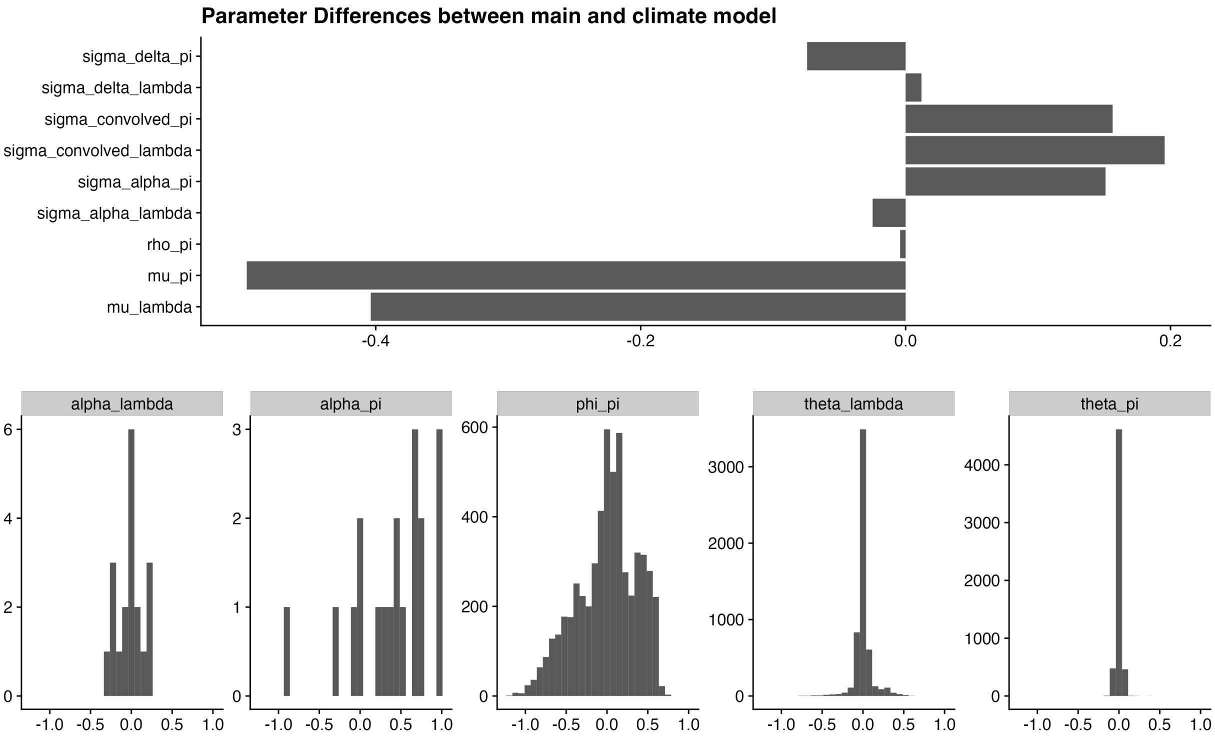


Figure C.5: Differences in parameter estimates between the main smoothing model and the climate covariate model.