The burden of Yellow Fever in Brazil: Quantifying disease mortality and producing short-term forecasts

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Joseph L Servadio

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Matteo Convertino, PhD

July 2020

Acknowledgements

This work was funded through letters of agreement from the Pan American Health Organization as well as the University of Minnesota Doctoral Dissertation Fellowship. I am grateful to have been able to work on this research and that it was made possible through these funding sources.

I also am very grateful to the Pan American Health Organization and the Brazilian Ministries of Health for allowing me to use the Yellow Fever case data for my dissertation, as well as to Fernando Nardi and Antonio Annis at Universita per Stranieri di Perugia for providing the drainage density data used in this dissertation. Also, the use of resources from the Minnesota Supercomputing Institute was invaluable to this work.

I would like to thank my committee members, Matteo Convertion, Claudia Muñoz-Zanzi, Mark Fiecas, and Craig Hedberg, for always making me think and helping me make this work something I am truly proud of. Outside of my dissertation, I appreciate their advice, opportunities to build connections and collaborations, and letters of recommendation.

The faculty and staff through our Division have been incredibly helpful to me throughout my time at the University of Minnesota. I am grateful for all the help, advice, and collaborations from Khosi Nkosi, Frank Strahan, Shelly Ring, Debb Grove, Bridget Brennan, Bruce Alexander, Susan Arnold, and Jesse Berman. Outside of our division, I am grateful for other faculty connections that have enriched my experience as a doctoral student, including Dominic Travis, Tiffany Wolf, and the faculty from the Sustainable Healthy Cities research group, both within and outside the University of Minnesota.

Also deserving thanks are the current and former students in the Environmental Health Division. From collaborating to commiserating on topics within my dissertation, other projects, or simply navigating through grad school, their friendships have all made this a better experience. I feel lucky to have been in a cohort with Cate Babcanec, Melanie Firestone, Ashley Hernandez, Christy Rosebush, and Scott Vial, as well as to have spent time in the Division with Natalie Bontrager, Shannon Cigan, Ryan Gavin, Ben Khoo, Yang Liu, Michelle Meyer, Austin Rau, Maria Sundaram, and Madhura Vachon. Outside of our division, I am grateful for my friendships through the School of Public Health and College of Veterinary Medicine, including Lauren Bernstein, Jessica Deere, Emily Groene, Jim Kincheloe, and Sara Lammert.

Throughout my time here, I have been able to collaborate with numerous researchers, both within and outside of the University of Minnesota. These include groups from the UMN Humphrey School, UMN College of Veterinary Medicine, Georgia Institute of Technology, North Carolina State University, and Universidad Complutense. I am grateful to have been able to expand my collaborative network and work with such a diverse group of researchers.

The final group I want to express my gratitude for consists of all of my family and friends outside the University of Minnesota for reminding me that there is a world outside of my academics and for consistently supporting my endeavors over the past several years.

Abstract

Yellow Fever is a mosquito-borne viral disease impacting much of South America and sub-Saharan Africa. It is endemic in several nations, causing hundreds of thousands of annual infections and tens of thousands of deaths. In Brazil, cases in recent decades have been seen in western areas of the nation, typically in regions adjacent to the Amazon Rainforest. A major outbreak beginning in December 2016, however, saw a major increase in cases, particularly in southeastern states. As a result, there is interest in finding ways to predict when and where future Yellow Fever cases are expected. Also of interest is the ability to anticipate future fatalities by finding the proportion of cases that are fatal. Several mechanisms influence risk of Yellow Fever cases, including human activities and environmental conditions. The latter is comprised of many characteristics outside of human control, representing a component of risk that cannot be targeted for direct intervention, but only used for preparations. Using environmental conditions to predict future Yellow Fever burden typically employs the use of either mathematical or statistical models to quantify relationships between environmental predictors and disease burden. In developing such models, several assumptions are made out of necessity. One particular assumption, relating to the time units used in developing a model, is not commonly investigated for its potential impact on describing disease dynamics. In order to investigate these various topics, four studies were conducted in order to predict Yellow Fever burden using various environmental conditions, estimate fatality risk among severe Yellow Fever cases, and examine assumptions of time unit sizes when describing disease incidence probabilistically. The findings of the various studies show sensitivity when changing time unit sizes, offer an update of the estimated fatality risk among severe Yellow Fever cases, and estimate potential for Yellow Fever case burden throughout Brazil both using annual environmental trends and weekly weather patterns. Methodological contributions are discussed.

Table of Contents

List of Tables

List of Figures

Preface

The discipline of environmental health encompasses numerous topics with underlying commonalities. While the field can be defined in numerous ways, one applicable definition is the relations between people's health and their surroundings. This can take several forms, as one's surroundings can encompass their natural surroundings as well as their built surroundings.

Gaining knowledge in the ways in which a person's environment can affect their health requires rigorous research. From diligent data collection to properly designed analyses to designing interventions, multiple forms of research exist aiming to determine how the population's health can relate to various aspects of their surroundings. One notable part of the research process is the use of quantitative analyses in order to measure or estimate the mechanisms or dynamics behind various health effects.

Along this topic, this dissertation seeks to emphasize the use of rigorous quantitative analyses in order to better understand the dynamics and determinants of a particular health effect: Yellow Fever. A common disease in the tropics [1], the subsequent sections of this dissertation focus on the dynamics of Yellow Fever as well as the use of the natural environment to predict its occurrence. The studies presented here show the use of quantitative analyses in order to design better studies, understand the disease's progression, and inform future preparedness. These topics, while focused on Yellow Fever in these studies, are relevant within other infectious disease topics as well as other topics throughout environmental epidemiology. The methodological contributions of these studies are of interest beyond the study of this one disease.

While the primary focus within this dissertation is Yellow Fever dynamics, research in other areas within the field of environmental health have been undertaken in

tandem with the studies that will be presented in the subsequent sections. While the topics vary in their contextual subject matters, similarities and connections among these topics exist, and the research practices and methods that contributed to many of these studies were considered while designing and executing the four studies presented in this dissertation. An overview of these projects is presented within this preface along with their contributions to environmental epidemiology as well as their commonalities.

In contrast to the studies comprising this dissertation, which focus on the natural environment for human health outcomes, many other studies place emphasis on the built environment, as industrialization and urbanization are believed to impact various aspects of human health [2]. Many studies have examined connections between various aspects of the built environment and human health [3, 4]. Among the many mechanisms that may exist, access to greenspace in an urban setting, defined as deliberately undeveloped land, as well as air pollution exposure, are components of a person's surroundings that can affect health [5, 6]. The existence of numerous characteristics of a person's built environment that might impact health outcomes motivates the need for data reduction methods in order to reduce a high number of variables into a smaller set that adequately represent the whole. This was done by developing a network model based on Transfer Entropy between pairs of variables [7]. Such a network, suitable for time series data, indicates which variables can explain others and can inform a reduction in the number of variables, leaving a subset that are adequate representations of the collection. When applied to a city health metric, the method showed that some cities' aggregate health perceptions may differ from reality [7].

Another important note in examining effects of the built environment on human health are issues of equity [8]. Many works have existed that seek to identify how certain demographic groups are disproportionately harmed by various exposures [9]. A study aiming to quantify the severity of these inequities in Atlanta was

conducted, with the additional aim of identifying potential benefits of some aspects of the built environment [10]. A spatial statistical model was used to consider interrelatedness among nearby locations. When comparing exposure to two pollutants, two respiratory health outcomes, and two cardiovascular health outcomes, among demographic groups in the Atlanta metropolitan area, areas with a majority African American population were found to have greater exposures to both pollutants and higher prevalence of all four health outcomes. Characteristics of the built environment, such as parks and tree canopy, were not shown to significantly associate with lower pollution exposure or better health [10].

Other hazardous exposures may not exist within the community, but within an individual's home. Chemical inhalation in the home can pose as a notable health hazard [11]. Ventilation and dispersion of an airborne contaminant is important for protecting those that may be exposed, whether at home or in a work environment. A study examined the concentrations of a household cleaner in an enclosed environment. Using a mechanistic model rather than a statistical model relied on measured data regarding evaporation and dispersion of chemicals [12].

Chemical hazards can also be found in the natural environment. Many harmful chemicals have been found in bodies of water, including those that were previously believed to be unaffected by human activity [13]. Numerous hypothesized sources exist, including wastewater treatment, human contamination, and through evaporation and precipitation [13, 14]. A study was conducted aiming to identify chemical contamination among many bodies of water in Minnesota that are used by indigenous populations for subsistence fishing. Sampling Lake Superior at multiple points as well as several inland lakes found that, among over 150 chemicals tested, every site found at least one in fish, water, or sediment samples [15]. Ongoing continuations from this work aiming to identify anthropogenic activities, such as impervious surface, healthcare facilities, or presence of increased human

populations, can explain the presence of these chemicals. Penalized regression methods are used to adjust for correlated predictors.

Hazards used in locations for subsistence fishing are a major concern, but other food sources exist, including restaurants and food retailers. However, access to healthy food is not uniform throughout the United States, presenting further issues of equity in health. A study was conducted to examine food deserts, defined broadly as locations with inadequate access to affordable and healthy food [16]. Connections between food access and related health outcomes such as diabetes and obesity were not clear in the Minneapolis-St. Paul metropolitan area, and further investigation of definitions of food deserts proved problematic, highlighting the need for locally specific research in defining food access and equity [16].

Outside of considerations of food scarcity, the safety of purchased food products is of concern. Listeriosis is a disease caused by a bacterial infection. While rare, this disease has high severity [17]. An ongoing study to identify foods most commonly associated with sporadic listeriosis cases was conducted to update the list of foods considered most risky for this bacterial infection. By combining multiple data sets, food consumption patterns were estimated comparing sporadic listeriosis cases in Minnesota to a demographically matched control population. Among food items known to be related to listeriosis, melons and certain soft cheeses were found to be the most associated with infection.

Other infectious disease studies have occurred throughout the course of this dissertation as well. The yearly influenza forecasting challenge led by the United States Centers for Disease Control and Prevention invites infectious disease modelers to attempt to forecast the ongoing influenza season each year using influenza-like illness data from healthcare facilities. Forecasting targets include the beginning of the influenza season, the week with the highest incidence of influenza-like illness, and the incidence of influenza-like illness one through four weeks in the

future [18]. Incidence is estimated by the percentage of outpatient hospitalizations with patients presenting with symptoms defined by influenza-like illness, without requiring a test. Using a Bayesian hierarchical model and B-splines to use the trajectories of previous flu seasons to forecast the current season proved effective, performing generally on par with models using far more resources external data sources.

Another infectious disease study highlighting the use of disease data without other sources focused on Visceral Leishmaniasis in Brazil. In order to determine if surveillance of Visceral Leishmaniasis over a subset of the nation could adequately estimate national burden as well as determine if surveillance per municipality is advantageous over surveillance by state, probability distributions were fit to municipality-resolution incidence nationwide and by state as well as state-resolution incidence nationwide. Without imposing model assumptions or other data sources, it was proven that the dynamics of this disease differ when examining incidence by state or by municipality. Furthermore, few states have a municipality-resolution incidence that adequately describes the nation [19]. The results of this study emphasize the need for high quality disease surveillance by municipality and nationwide.

These various studies presented not only substantive questions of interest, but also methodological questions. The study focusing on Visceral Leishmaniasis presented a new method for providing quantitative evidence for an intuitive concept. The study of listeriosis required novel uses of multiple data sets to inform dietary habits. Developing a city health index motivated the development of Optimal Information Networks in order to reduce large quantities of data. The other studies also involved advanced analytic methods such as spatial statistical models, penalized regression, and high performance computing.

Involvement in numerous works relating to a variety of components of environmental epidemiology show how similarities can exist across research contexts. It also highlights the importance of multidisciplinary research groups, as very few of these studies could exist without them. An underlying theme among all of these studies is the use of translatable methods for researching a variety of topics relating the surrounding environment and health outcomes.

Chapter 1 Introduction

Vector-borne diseases account for high morbidity and mortality worldwide [20]. Numerous diseases, caused by various viruses, parasites, and bacteria and spread by arthropod vectors, are among the top health concerns globally. World regions facing the greatest impact from vector-borne diseases include Latin America, sub-Saharan Africa, South Asia, and Southeast Asia, though some also pose a threat in temperate regions [21]. Many vector-borne diseases, due to their reliance on arthropod vectors, are sensitive to short-term and long-term environmental changes [22, 23]. Furthermore, the existence of nonhuman reservoirs of many diseases leads to further influence by factors of the environment [24]. Research relating various vector-borne diseases to global climate change has largely comprised of long-term projections regarding changes in burden and changes in locations at greatest risk [25–27].

One particular vector-borne disease of international concern is Yellow Fever (YF), which is a disease caused by a flavivirus spread by mosquitoes [1, 28]. Each year, an estimated 200,000 people are infected with YF globally, and 30,000 of them die [28]. The virus is believed to have originated in Africa and later migrated to South America [29, 30]. Currently, several nations in Latin America and, primarily, sub-Saharan Africa are endemic for YF [1, 30, 31]. It is one of many arthropod-borne viruses (arboviruses) endemic to the tropics.

Upon infections with Yellow Fever Virus (YFV), many cases are asymptomatic, while others present illness with flu-like symptoms including fever, muscle aches, and nausea within three to six days [1]. These symptoms, indicating a mild case, persist typically for three to four days before subsiding [32]. A subset of these cases, after mild symptoms subside, develop severe disease, where fever returns with jaundice and hemorrhaging [33]. This can lead to organ failure [33, 34] and is fatal in a high

percentage of cases [35]. While no treatments exist for YF [36], a safe and effective vaccine exists [37, 38], which is believed to confer lifelong immunity. Six worldwide facilities create three versions of the vaccine [37]. International Health Regulations allow some nations to require YF vaccination upon entry [32, 39].

In Africa, the primary vector is the *Aedes* mosquito, which also spreads Chikungunya and dengue viruses [40–42]. *Aedes* mosquitoes also spread YF in South America, with the addition of *Haemagogus* and *Sabethes* mosquitoes [1, 31, 40]. In South America, nonhuman primates of the *Alouetta* species, known as howler monkeys, are a well-known reservoir of YF [24, 43, 44]. Mosquitoes infected with YFV when laying eggs are capable of transmitting the virus to their offspring [31]. Due to this fact and the persistent nonhuman primate reservoir, total eradication of the disease is believed to be impossible [45, 46].

In South America, two types of transmission exist for YF: sylvatic and urban. Sylvatic transmission occurs when a person is infected by a mosquito that acquired the virus from a nonhuman primate in the wild, typically through *Haemagogus* or *Sabathes* mosquitoes [47–49]. These transmissions tend to occur in or near the Amazon Rainforest. Urban transmission occurs from mosquito bites where the mosquito acquired the virus from another human, which typically occurs in or near developed areas, typically through *Aedes* mosquitoes [50, 51]. During the last decades, sylvatic transmission accounted for most of the YF cases seen in South America, including those from outbreaks [48].

Globally, 90% of YF cases occur in African nations [52, 53]. Among Latin American nations, Brazil, Peru, and Colombia have seen the highest numbers of cases in recent decades [54]. Unlike other viruses spread by mosquitoes, YFV has not become endemic in Asia [30], and imported cases in Europe and Asia have remained isolated [37, 55]. Cases in the United States have remained rare and have predominantly occurred prior to the 20th Century [56, 57].

The most recent outbreak of YF in Brazil began in December 2016 [58–60]. More cases were seen in Brazil in the one year following the start of the outbreak than had been seen in more than a decade prior. The outbreak was primarily focused in the southeastern part of the country, including the states of Espírito Santo, Minas Gerais, Rio de Janeiro, and São Paulo [60]. A concurrent outbreak among nonhuman primates in Espírito Santo was also observed [61]. These states include coastal locations where, prior to the outbreak, vaccination was not recommended for visitors prior to travel [60, 62]. Prior to this outbreak, the last major outbreak of human YF in Brazil was seen between 1935 and 1940, after which endemic cases were seen regularly [40]. An outbreak of cases in nonhuman primates was also observed in Brazil in 2008 and 2009 [30], and other minor human outbreaks have been observed in South America [63, 64] and Africa [65, 66] over the last several years. Prior to the outbreak in Brazil, a large outbreak was observed in Angola and the Democratic Republic of Congo [50, 51, 67].

Efforts to control the epidemic in Brazil have included increased nonhuman primate surveillance and vaccination, achieving high rates of vaccination in targeted municipalities and rates near 50% in other states [68]. Current regular measures to prevent or reduce YF incidence primarily include vaccination and mosquito control, whether by reducing mosquito populations [69–71] or preventing access to breeding grounds [72]. Previous measures to control YF have also included vaccination and mosquito control. High vaccine adherence was seen in central Brazil during the 1970s and 1980s, but decreased in the 1990s and 2000s [73]. *Aedes* mosquitoes, the predominant vector for urban YF transmission in Brazil, were believed to have been eradicated from Brazil by the late 1950s due to eradication campaigns, but they have reappeared in the country during the 1970s [74].

A substantial volume of previous research exists to understand YF through laboratory methods. A common theme of research involves development or

improvement of testing methods, aiming to improve accuracy or efficiency of detection [75–80]. While many have been proposed, eleven PCR tests with a high sensitivity and specificity have been made commercially available as of 2018 [81], though these have been recently made commercial [82]. Other tests used include types of enzyme-linked immunosorbent assay (ELISA) tests [83], including the most commonly used test by the United States Centers for Disease Control and Prevention [81]. Other studies have aimed to develop treatments and therapies for YFV-infected patients [84, 85].

Other research pertaining to YF, and other related diseases, involves various forms of epidemiological and statistical modeling. Primarily, models used for YF and other infectious disease research include mathematical models [86–90], using known or estimated parameters regarding disease transmission in order to estimate potential outcomes under current conditions or other scenarios, or statistical models [50, 54, 91, 92], where outcome data are used along with exposure or other predictor data in order to estimate associations or determinants of the outcome. Examples of mathematical and statistical models used in YF research have shown that both can prove useful for understanding the disease's dynamics.

Previous modeling studies concerning YF have aimed to estimate disease burden [35, 45], identify areas at greatest risk of seeing YF [93, 94], or determine effectiveness of vaccination strategies [95]. Statistical models in particular have been used to estimate risk of disease presence [50] or disease incidence [93]. A common theme in many of these studies is the use of environmental factors when developing models used. Previous connections have been examined between several environmental variables and risk of YF [50] or other arboviruses [96, 97], including temperature [91], precipitation [88], and elevation [54].

Modeling studies rely on several assumptions as part of their design [98]. Some assumptions made are a direct consequence of data availability, while others are

based on subjective decisions. Assumptions in mathematical models can include the model structure itself, such as the compartments to include if using a compartmental model or the parameters included and estimated, though these can also be informed by previous works. Assumptions unique to statistical models can include the combination of predictor variables and the way in which they may interact. Some of these assumptions can be tested with various metrics.

Other assumptions include operational characteristics of the data used and the model format. For studies using time series analyses, such assumptions can include the time between repeated measures or whether the spatial area represented by the data is representative of the population upon which inference is being drawn. Another key assumption in modeling includes the distributional fit of particular variables. This can be seen in mathematical models, for example, if the probability of transitioning between two compartments in a compartmental model is assumed to be random. The transition probability would be assumed to follow a particular distribution and then be drawn at iterations to determine how individuals transition. This can also be seen in statistical models in choosing the link function of the outcome variable in a generalized linear model. The outcome variable is assumed to follow a particular distribution, and the statistical model is designed around that assumption; assumption of a Normal or Beta distributed outcome, for example, would lead to differences in model design by using an identity or logit link respectively.

In recent years, advancements in technology have allowed researchers to access large quantities of data. The use of large data sets has been increasing in appeal in infectious disease contexts [99]. This is also seen with Electronic Health Records, where large quantities of information can be stored [100]. Environmental variables can be measured and estimated using sophisticated sensor technology, using small areal units over long periods of time [101–104]. While storage and computational

capacities have improved as well over time, the existence of such large datasets can lead to challenges in data management and computation [105].

Connections have been established between arboviruses, including YF, and environmental factors. Both long-term climate patterns and short-term weather events have been used to estimate locations at greatest risk of seeing YF, predicting future disease occurrence, or targeting interventions. Typical environmental factors used to predict YF occurrence have included temperature [86], precipitation [92], mosquito habitats [93], nonhuman primate habitats [54], and vegetation [106].

The connection between the environment and disease dynamics is known to be complex, without a distinct consensus regarding the mechanisms [107, 108]. Many studies projecting long-term changes in disease risk make note of changing mosquito habitats in relation to global climate change [109], suggesting either an expansion or shift in habitats [110, 111]. Specifically pertaining to mosquito-borne diseases, optimal conditions such as temperature have been found for life cycles and disease incubation [112, 113]. Precipitation has also been shown to have complex relationships with disease transmission; increased rainfall may increase disease burden in some contexts [114], but some outbreaks have also been associated with drought conditions by creating standing water in river beds [112].

Underreporting of infectious disease data is a well-established issue and has presented a challenge in research [93, 115–117]. Many studies aiming to elucidate disease dynamics acknowledge that collected data likely only represent a subset of the true number of infections. This can exist due to undetected asymptomatic or symptomatic infections as well as underreporting, leading to only the most severe cases being noticed [116, 117]. Additional sources of underreporting involve patient motivations to seek healthcare [118, 119] or improper reporting from healthcare facilities [120]. Some studies exist aiming to estimate the true burden of disease,

and therefore address underreporting, for YF [35, 45], as well as other diseases [121, 122].

Considering the challenges in disease modeling and need for informed preparedness to reduce YF burden, this dissertation aims to address assumptions made prior to modeling disease dynamics, provide updated information regarding YF fatality, and develop a useable warning system to identify riskiest locations for short and long term YF burden in Brazil. These different aspects of YF dynamics represent a variety of challenges and knowledge gaps in best practices for preparing for disease incidence. The work is divided into four specific aims.

Chapter 2 addresses decisions regarding time units in the denominators for calculating disease incidence. Modeling studies commonly define the units of time being considered, such as days, weeks, or months, without considering whether there is sensitivity to this choice. Rather than imposing a specific model form, daily, weekly, and monthly incidence of YF in Brazil were fit to probability distributions and then compared. The results of this study show that the distributional fits for the three time units, which are representative of the unknown underlying processes defining disease incidence, are not equivalent. It follows that different processes likely exist that govern incidence over a day, a week, and a month.

Chapter 3 addresses the case fatality risk among severe YF cases. The currently cited estimates of the case fatality risk are based on a small number of previous works as well as reported YF case data, which may suffer from underreporting issues. A systematized literature review [123] was conducted with a meta-analysis to provide an updated estimate of the fatality risk. Based on the included studies, the fatality risk among severe YF cases is approximately 40%, lower than the 50% estimate previously stated [1]. Notable differences were found between studies in Africa and South America.

Chapter 4 uses long-term environmental patterns to identify locations within Brazil that are most vulnerable to YF burden. Here, vulnerability is defined by the ability to see high incidence in a year. Using locations where YF has been seen between 2000 and 2017, several environmental predictors were found to accurately predict vulnerability, defined in three categories. The results identify specific locations that are most likely to experience high YF burden over a long period of time.

Chapter 5 uses short-term environmental patterns to produce short-term forecasts of YF incidence in Brazil. Using short-term weather patterns as well as the vulnerability risks developed in Chapter 4, predictions of YF presence and incidence were developed for forecasts two to seven weeks into the future. The results in this study show the ability of using only environmental factors to produce accurate forecasts.

The combined results from all four studies show both methodological and substantive insights related to disease modeling. In addition, an updated fatality risk is important for anticipating disease severity among future outbreaks. Creating a usable forecasting system for Brazil can allow preventative measures to focus on locations that would benefit the most.

Chapter 2 Disease incidence sensitivity to temporal resolution: The case of Yellow Fever in Brazil

Summary: Mathematical and statistical models for infectious disease research commonly rely on specified time units, which constitute their temporal resolution. Decisions on temporal resolution are commonly constrained by data availability or arbitrarily chosen, though they may potentially impact model results. This study aims to determine if the choice of temporal resolution for infectious disease incidence can affect its probabilistic dynamics, which can represent underlying mechanisms governing incidence that models can be designed to seek. Yellow Fever in Brazil poses an interesting case study, as the disease is endemic in much of the tropics, and Brazil experienced a major epidemic starting in December 2016, showing a major increase in cases. To avoid imposing specific model assumptions, pattern sensitivity to temporal resolution was determined through distribution fitting of daily, weekly, and monthly Yellow Fever incidence. Probability distributions were fit to incidence data at all three resolutions by finding the distribution and parameter set minimizing the Kolmogorov-Smirnov statistic. Aggregated empirical distributions were generated from fitted distributions to allow comparisons of distributional fits across resolutions. This process was completed separately using data prior to and following the start of the 2016 – 2017 outbreak. Among all three resolutions, incidence was best fit by the Gamma distribution with different parameters. Prior to the outbreak, there was little evidence suggesting that the distributional fit at one resolution would not adequately describe incidence at another. After the start of the outbreak, differences in distributions were greater, indicating that the distributional fits at different resolutions, which represent underlying data generating mechanisms, are not equivalent. These results suggest that the temporal resolution used for incidence can impact the way the data are described probabilistically. Implications for modeling are discussed.

2.1. Introduction

Many topics in infectious disease research involve the use of mathematical and statistical models, which are commonly used to estimate the true burden of disease [45], evaluate predictors of disease risk, evaluate an intervention [124], or forecast future disease incidence [18, 125]. Mathematical models largely include various compartmental models, where individuals transition between infectious disease states such as susceptible, latent, and infected, through specified probabilities and contact rates. These models rely on known or assumed parameters that can determine transitioning through the compartments and can serve to predict outcomes under various scenarios. Statistical models commonly use case or incidence data to determine factors that may be associated with increased risk of a disease or higher incidence. These models intake predictor and outcome data in order to estimate determinants or associations.

When using either mathematical or statistical models to study infectious diseases, units of time, reflected in time steps or amount of time between repeated observations, are an important consideration for model structure. Time units are important when determining the denominator to calculate person-time for incidence, the amount of time between iterations of a mathematical model, or frequency of collection for predictor data. Time units relate to the concept of temporal resolution, which refers to the size of time units considered, such as hours, days, or months.

Temporal resolution has been considered in various research contexts. Ecologists have examined the choice of temporal resolution when using predictive models in the contexts of predicting determinants of species life cycles [126] or in predicting vegetation development [127]. Both studies examined predictive models at different resolutions to determine that models using higher resolution data produced more accurate predictions. In the context of human health, temporal resolution has been

addressed in neuroimaging [128] and other scanning methods [129–131] for the purpose of increasing sensitivity of diagnoses using repeated scans. These studies found that higher temporal resolution, assessed by having shorter time spans between scans, allows diagnoses to be more accurate.

Other studies rely on the notion that higher resolution data are more useful for accurate analyses and aim to improve methods for acquiring data at a high temporal resolution through mathematical estimation or technological improvements [132–136]. Focusing on the ability to acquire high-resolution data emphasizes its desirability in human health contexts.

When working with mathematical and statistical infectious disease models, the choice of temporal resolution is typically an assumed part of the model design rather than considered as a choice that can impact model outputs. Often, the choice of temporal resolution in modeling studies is limited by data availability rather than made as an explicit decision, but the question of whether model results are sensitive to the resolution used is not commonly examined. Instead, temporal resolution is assumed or decided arbitrarily; this is represented through the length of time between iterations of a mathematical model, which can also be determined by the disease of interest [137, 138], or through the amount of time represented by observations (such as hours, days, or months) in the data set for a discrete-time statistical model [139]. In the context of infectious disease research, some attention has been given to spatial resolution, or the size of spatial units considered, where sensitivity to the choice of spatial resolution was addressed by repeating analyses at two resolutions and comparing results [140, 141]. This method, however, relies on the specific assumptions of the model used.

There are advantages and disadvantages to using infectious disease data at a higher or lower temporal resolution in models. Classifying data as high-resolution is determined based on subject matter and is relative to typical data patterns. For

example, in infectious disease epidemiology, daily resolution data would be considered high resolution [137], while in neuroimaging, scan data per second could be considered high resolution [130]. It is often assumed that having relatively high-resolution data will lead to greater accuracy in results, whether the objective is to estimate disease burden, to identify predictors of disease incidence, or other topical concerns. However, the methods for collecting accurate high-resolution data for infectious disease research require high surveillance capacity and dedicated resources for surveillance and data collection. Also, using high-resolution data is computationally expensive and requires extensive computing resources. Using lower resolution data is more efficient for computation and data management purposes, but potentially masks heterogeneous effects and patterns within larger time units. This conceptually is analogous to ecological fallacy [142] in that aggregation over time can lead to averaging of exposures and outcomes, which can adversely affect model results.

In the context of a mathematical model, the decision to change the temporal resolution can lead to changes in model form, such as the parameters describing transitions between compartments. If, for example, using weekly iterations instead of daily iterations in a model, a probability distribution used to represent transition probabilities would be expected to change, but the way in which they change is not commonly investigated. Currently, the only way this has been examined is by repeating specific analyses [126, 127]; there does not exist a method for doing this that does not impose the assumptions of a particular model. In the context of a statistical model, for example, a generalized linear model for the disease incidence, the link function for the outcome variable depends on its assumed distribution. If the temporal resolution associated with person-time for calculating incidence changes, (e.g. from person-week to person-month incidence), the most appropriate distributional family, and therefore link function, could differ. Furthermore, the most relevant predictors could differ when analyzing data at different temporal resolutions.

To address this knowledge gap, this study aims to determine if disease incidence data at different temporal resolutions are equivalently represented through probability distributions, which can represent underlying mechanisms of disease occurrence. The study uses Yellow Fever incidence in Brazilian municipalities as a case study. Yellow Fever (YF) is a flavivirus endemic to sub-Saharan Africa and South America [1]. The virus is spread in South America by *Aedes* and *Haemagogus* genera of mosquito, with monkey species as reservoirs [1, 143]. Brazil has comprehensive YF data and provides an interesting case study having experienced a major outbreak of YF in 2016-2017 [58, 60].

Probability distributions were fit to daily, weekly, and monthly YF incidence. Empirical distributions were generated using these fitted distributions to ensure appropriate comparisons across resolutions. Comparisons were made to determine whether the distributional fits across different temporal resolutions were interchangeable and can represent data at other resolutions. Potential implications in modeling are discussed.

2.2. Methods

2.2.1. Data sources

The setting for this study is Brazil, the largest nation in South America by both land area and population. The country is subdivided into 26 states plus a federal district or into 5,570 municipalities including the federal district [144]. Brazil undergoes passive surveillance for YF as a notifiable disease [145].

Yellow Fever case data were provided by the Pan American Health Organization and the Brazilian Ministries of Health, which include reported cases to the Ministries of

Health between January 2000 and March 2018. The reported cases list the date of report as well as the state and municipality where observed. Annual population data are publicly available for each municipality from the Brazilian Institute for Geography and Statistics for all years between 2000 and 2018 with the exceptions of 2007 and 2010 [144]. The populations for those years were singly imputed as the arithmetic mean of the two adjacent years' populations. Case counts and populations were matched to municipalities to allow daily, weekly, and monthly incidence per 100,000 population to be calculated for each municipality. Because population data are available yearly, it is assumed that population is held constant for the entire calendar year, giving each day, week, and month in a year the same population for person-time for incidence calculation.

An outbreak that occurred in 2016 and 2017 [58] posed challenges in fitting distributions over the entire time period of observation. As seen in Figure 2.1, case counts nationwide after the outbreak began were much higher than for previous years. It is likely that case detection and notification differed during the outbreak, and this likely applies to the months after the outbreak as well. As a result, data were split into two time periods: January 2000 to November 2016 and December 2016 to March 2018. The same process for fitting distributions was applied to both time periods separately.

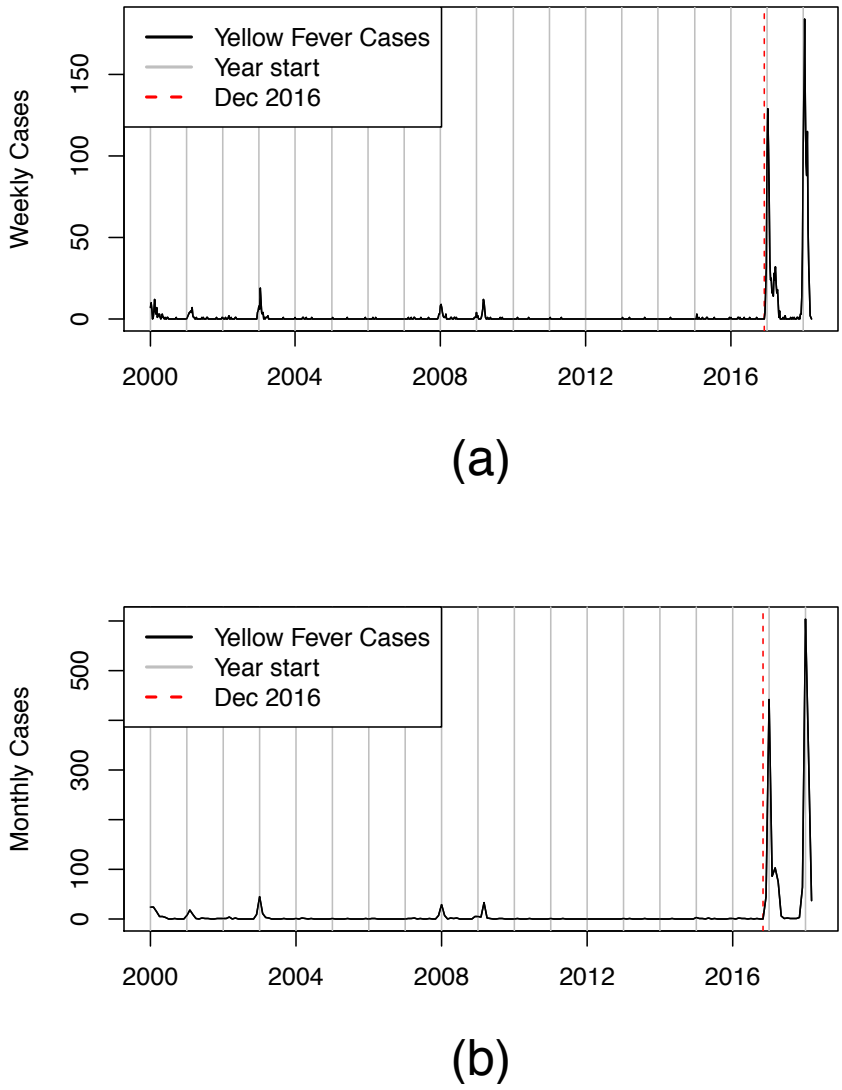


Figure 2.1. Total Yellow Fever cases in Brazil by (a) week and (b) month, Jan 2000 – Mar 2018. Vertical lines represent the first observation each calendar year. The beginning of the outbreak in December 2016 is noted.

2.2.2. Analytic approach

In order to determine if incidence data at different temporal resolutions can be interchanged to represent disease incidence without detriment, best fitting

probability distributions were fit to daily, weekly, and monthly incidence by municipality. This is relevant in the context of modeling because the distributional fit of data can inform either the link function of a generalized linear statistical model [146] or assumed behaviors represented in a mathematical model [147, 148]. The distributional fit of YF incidence represents the unknown underlying mechanism that a model would be designed to elucidate. If distributional fits at different temporal resolutions are not equivalent, then this may imply that different underlying data generating mechanisms exist, and models aiming to learn about the mechanism at one resolution would not be appropriate for use at another.

Among the municipalities in Brazil, many days, weeks, and months were observed without any cases; more than 99% of daily and weekly incidence observations were zero. As a result of the high number of observations with zero incidence, YF incidence was assumed to follow a mixture distribution consisting of a binomial process for seeing non-zero incidence and another process following a continuous probability distribution for positive incidence. Only municipalities where at least one YF case was seen between January 2000 and March 2018 included in distribution fitting to account for the possibility that municipalities that did not see any cases may have experienced undetected cases rather than truly having zero incidence. A total of 466 municipalities were included in analyses (Figure 2.2).

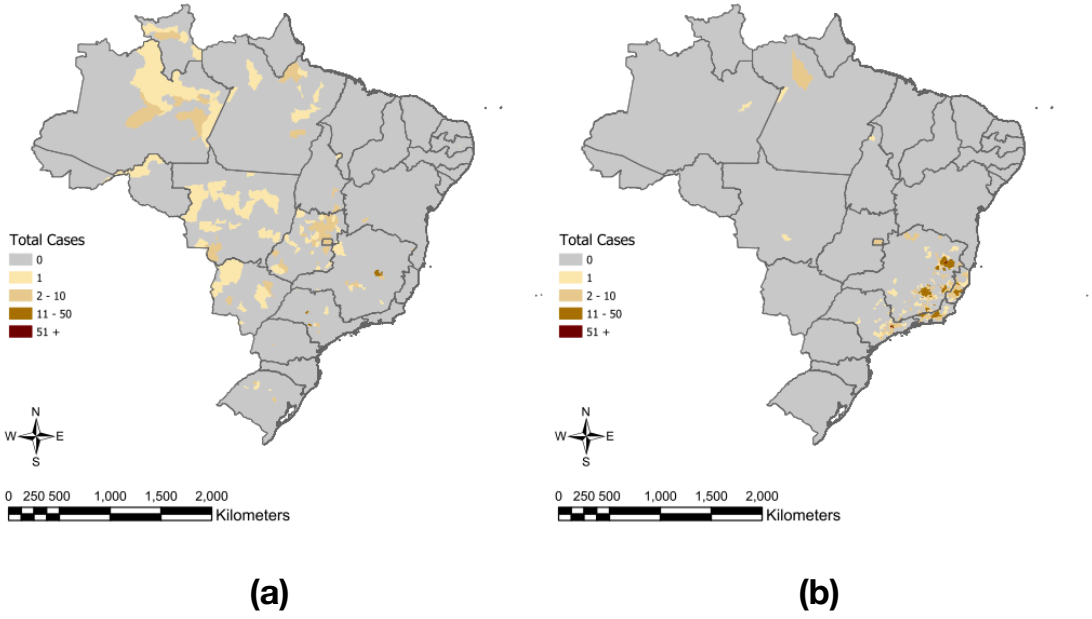


Figure 2.2. Total cases of Yellow Fever by municipality from (a) Jan 2000 – Nov 2016 and (b) Dec 2018 – Mar 2018.

To incorporate the binomial process in distribution fitting, the estimated probability of seeing zero cases was assigned the proportion of incidence values equaling zero, and a fitted continuous distribution was weighted by the complement of that value. For example, if incidence data at a particular temporal resolution is found to have a proportion *p* of incidence values equaling zero, and positive incidence follows a probability density function (pdf) denoted with *g(x)*, then the mixture distribution takes the form

$$\boldsymbol{f(x)} = \begin{cases} \boldsymbol{p}, & \boldsymbol{x = 0} \\ \boldsymbol{(1-p)g(x)}, & \boldsymbol{x > 0} \end{cases} \qquad \text{(Equation 2.1)}$$

The methods used in this study, primarily distribution fitting of incidence data at each temporal resolution, aimed to determine the form of *g(x)* that would provide the best fit to the positive incidence data.

Candidate distributions for positive incidence, represented by *g(x)* in Equation 2.1, included the following continuous distributions: Lognormal, Truncated Normal, Exponential, Gamma, Power Law, Uniform, and Triangle [149]. Among all distributions, the distributional family and parameter set that minimized the Kolmogorov-Smirnov (KS) statistic was selected as the best fitting distribution using the 'ks.test()' function in base R [150].

For each candidate distribution, optimal parameters to minimize the KS statistic were evaluated through a simulated annealing algorithm [151] using 100,000 iterations. To assure that the simulated annealing algorithm converged to a true global minimum, three chains of the algorithm were run. Convergence was determined by the following three criteria, where the algorithm was considered to converge once one was met, assessed in order:

1. Parameter values among the optimal parameter sets differ by less than 0.01
2. Parameter values among the optimal parameter sets differ by less than 0.05 and minimum KS statistics among parameter sets differ by less than 0.01
3. Minimum KS statistics differ by less than 0.0001

If the algorithm did not converge, it was repeated using the parameter values that minimized the KS statistic during the previous algorithm run as initial values for 200,000 iterations. This process was repeated until convergence was reached, yielding fitted distributions for positive YF incidence.

2.2.3. Empirical distributions

Once best-fitting distributions were determined for daily, weekly, and monthly incidence, they were compared to determine if they could be used interchangeably

in a modeling context. Directly comparing the distributional fits themselves would not be appropriate due to being fit using data with different denominators. In order to make appropriate comparisons, empirical distributions were generated to use for comparison, with the intent of generating an empirical distribution for a lower temporal resolution that would be expected based on the fitted distribution at a higher temporal resolution. From the fitted daily distribution, empirical weekly and monthly distributions were generated which could be compared to the fitted weekly and monthly distribution. Additionally, from the fitted weekly distribution, another empirical monthly distribution was generated to allow this comparison.

Empirical distributions were generated from the fitted distributions by randomly drawing and then summing samples from the fitted distributions. The empirical weekly distribution was generated by repeatedly summing seven samples from the fitted daily distribution as described in Equation 2.1. Summing drawn incidence values is possible because populations are assumed constant over the year.

Two empirical monthly distributions were created, using both the fitted daily and the fitted weekly distributions. Creating an empirical monthly distribution from the fitted daily distribution involved, for each sample, drawing a number of days in the month based on the numbers of days in each month of a non-leap year, and then drawing and summing that many samples from the fitted daily distribution. To resolve temporal misalignment, with nondiscrete numbers of weeks in a month, months were randomly assigned either four or five weeks. To create an empirical monthly distribution from the fitted weekly distribution, either four or five draws from the weekly distribution were summed.

The number of samples drawn for each empirical distribution was equal to the number of observations in the data being examined. For example, when using the Jan 2000 – Nov 2016 fitted daily distribution to generate an empirical weekly distribution, the number of summed draws of seven values was equal to the number

of weeks during that time period multiplied by the number of municipalities that observed at least one YF case. Using these in places of the fitted distributions used to generate them allows the different resolutions to be compared while using comparable units attached to the distributions.

2.2.4. Comparing distributions

The three empirical distributions for weekly and monthly incidence were compared to their respective fitted distributions by isolating positive values and comparing KS statistics of the sample with the analogous fitted distributions. The empirical weekly distribution was compared to the fitted weekly distribution, and the two empirical monthly distributions were compared to the fitted monthly distribution. The proportions of zero values were not compared because they were taken directly from the data and therefore, by design, directly related. The process for fitting empirical distributions and calculating the KS statistic was repeated 100 times to provide multiple KS values for each comparison. The median KS statistic was used for interpretation along with the middle 95% of samples.

A threshold for the KS statistic value was found to use the KS statistic to conclude if distributions are interchangeable instead of the KS hypothesis test due to the artificially large sample size. The number of samples in the empirical distributions is chosen to be large, leading to small KS values providing significant results, even if the value of the statistic may not be meaningful. In order to determine a threshold, 1,000 random values were generated for the shape and rate parameters for the Gamma distribution, both independently drawn uniformly between zero and five. For each of the 1,000 parameter sets, 1,000 draws from the Gamma distribution with those parameters were generated and then compared to the generating distribution with a one-sample KS statistic. This provides values of the KS statistic that can result from random noise. The maximum value of the KS statistic was 0.103,

with the 95th percentile of samples being 0.060. Therefore, a KS statistic above 0.100 was selected as a threshold to determine if compared distributions are notably different. This value is somewhat conservative as it is based on a simulation count less than the number of observations in the data and is approximately equal to the maximum value of these simulated KS statistics.

All analyses were performed using R version 3.6.0 [150]. The 'poweRlaw' package [152] and 'triangle' package [149] were used to fit the Power Law and Triangle distributions, respectively.

2.3. Results

More nonzero observations of YF incidence were seen between Dec 2016 and Mar 2018 compared to the time period between Jan 2000 and Nov 2016. Among positive values, higher daily and weekly incidence values were seen between Jan 2000 and Nov 2016, and higher monthly incidence values were seen between Dec 2016 and Mar 2018 (Table 2.1).

Table 2.1. Descriptive statistics for positive daily, weekly, and monthly Yellow Fever incidence before and after the beginning of the outbreak.

	Jan 2000 – Nov 2016			Dec 2016 – Mar 2018		
	Daily incidence	Weekly incidence	Monthly incidence	Daily incidence	Weekly incidence	Monthly incidence
n	312	261	206	1433	893	524
Mean	10.229	12.277	15.527	6.154	9.827	16.863
sd	12.543	17.409	26.572	7.469	13.914	27.058
Min	0.039	0.039	0.040	0.008	0.008	0.008
10%	1.173	1.130	1.130	0.532	0.600	0.610
25%	2.511	2.470	2.620	1.131	1.663	2.377
50%	5.500	5.511	6.382	3.267	5.112	7.342
75%	13.717	14.656	15.848	7.950	12.377	18.937
90%	28.868	31.716	33.495	15.517	24.176	40.472
Max	86.843	117.165	190.739	65.147	154.253	236.888

2.3.1. Fitted distributions

For both time periods, before and after the start of the YF outbreak, positive incidence among all three temporal resolutions followed the Gamma distribution with differing parameters. This distributional fit is a contrast from the more commonly seen distributions applied to incidence data such as lognormal or a binomial distribution for occurrence. The Gamma distribution has a support between zero and infinity, noninclusive, and uses two parameters: a shape parameter α and a rate parameter β. The mean of a Gamma distribution is $\frac{\alpha}{\beta}$, and its variance is $\frac{\alpha}{\beta^2}$. Gamma distributional fits for each data set represent the parameter sets that most closely match the observed data. Through the KS statistic, the maximum difference of cumulative density of the observed data and of the distribution is minimized under the parameterizations shown in Table 2.2.

The fitted distributions are shown in Table 2.2 and Figure 2.3 (a and b), and the full results of all distributions are shown in Appendix A. The pdf of the fitted distributions for positive incidence specified in Table 2.2, which are conditional distributions for having nonzero incidence, are presented in Figure 2.3. The algorithm did not converge for fitting the power law distribution to daily incidence for Dec 2016 – Mar 2018, and the algorithm did not converge for fitting the triangle distribution to weekly and monthly incidence for Dec 2016 – Mar 2018.

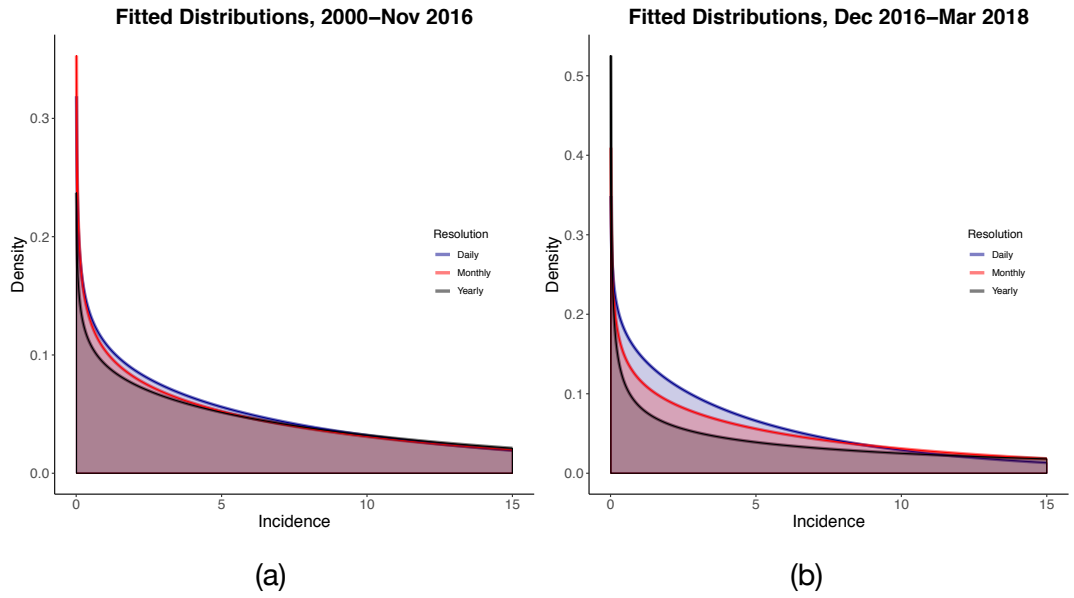


Figure 2.3. Fitted distributions for daily, weekly, and monthly positive incidence of Yellow Fever in Brazil for (a) Jan 2000 – Nov 2016 and (b) Dec 2016 – Mar 2018

Table 2.2. Fitted distributions to incidence of Yellow Fever. Distributions are represented as a mixture distribution, with probability mass assigned to zero and distributional family from minimized Kolmogorov-Smirnov statistic

	Resolution	Distribution	Mean	Variance
Jan 2000 – Nov 2016	Daily	$p(x) = \begin{cases} 0.9999, & x = 0 \\ Gamma(0.787, 0.083), & x > 0 \end{cases}$	6.282	47.958
	Weekly	$p(x) = \begin{cases} 0.9994, & x = 0 \\ Gamma(0.749, 0.070), & x > 0 \end{cases}$	9.278	117.449
	Monthly	$p(x) = \begin{cases} 0.9979, & x = 0 \\ Gamma(0.809, 0.068), & x > 0 \end{cases}$	14.372	334.235
Dec 2016 – Mar 2018	Daily	$p(x) = \begin{cases} 0.9938, & x = 0 \\ Gamma(0.847, 0.143), & x > 0 \end{cases}$	5.923	41.420
	Weekly	$p(x) = \begin{cases} 0.9724, & x = 0 \\ Gamma(0.747, 0.084), & x > 0 \end{cases}$	8.893	105.867
	Monthly	$p(x) = \begin{cases} 0.9301, & x = 0 \\ Gamma(0.609, 0.034), & x > 0 \end{cases}$	17.912	526.817

These results show that the Gamma distribution, a less commonly used distribution for representing infectious disease incidence, is a practical alternative for accurately

describing incidence. Though it does only apply to nonzero incidence values, its use in accurately describing positive incidence values is useful.

2.3.2. Comparisons across temporal resolutions

2.3.2.1 January 2000 – November 2016

The median KS statistic comparing the distributions fitting daily and weekly incidence, which compares the fitted weekly distribution to the empirical weekly distribution generated from the fitted daily distribution, was 0.062, with the middle 95% of samples lying between 0.037 and 0.107 (Table 2.3). This does not offer strong evidence that the two distributions are notably different. Similarly, the median KS statistic that compares the fitted monthly distribution to the empirical monthly distribution generated from the fitted daily distribution was 0.096, with the middle 95% of samples lying between 0.058 and 00.148. This provides weak evidence that the two distributions may not be appropriately used interchangeably. The median KS statistic comparing the fitted monthly distribution to the empirical monthly distribution generated with the fitted weekly distribution was 0.069, with the middle 95% of samples lying between 0.032 and 0.118, which does not offer strong evidence of differences between these distributions. These results are shown in Table 2.3.

2.3.2.2 December 2016 – March 2018

The KS statistic comparing the distributional fits of daily and weekly YF incidence compare the fitted weekly distribution to the empirical weekly distribution generated with the fitted daily distribution. The median KS statistic was 0.124, with the middle 95% of samples lying between 0.105 and 0.147. This indicates that these

two temporal resolutions do not follow interchangeable distributional fits; changing the temporal resolution of a model from one to the other would likely impact the model results. The median KS statistic comparing the fitted monthly distribution to the empirical monthly distribution that was generated with the fitted daily distribution was 0.283, with the middle 95% of samples lying between 0.264 and 0.295, offering strong evidence that the two resolutions do not provide interchangeable distributional fits. Similarly, the median KS statistic comparing the fitted monthly distribution to the empirical monthly distribution that was generated with the fitted weekly distribution was 0.280, with the middle 95% of samples lying between 0.262 and 0.302, offering strong evidence that the two resolutions do not provide interchangeable distributional fits.

Table 2.3. Results from KS statistics comparing distributional fits at different temporal resolutions. Empirical distributions were used to represent aggregation of the fitted distributions for incidence at higher resolutions.

	Fitted distribution	Empirical distribution	Median KS value	Middle 95%
Jan 2000 – Nov 2016	Daily incidence	Weekly incidence based on daily incidence	0.062	(0.037, 0.107)
	Daily incidence	Monthly incidence based on daily incidence	0.096	(0.058, 0.148)
	Weekly incidence	Monthly incidence based on weekly incidence	0.069	(0.032, 0.118)
Dec 2016 – Mar 2018	Daily incidence	Weekly incidence based on daily incidence	0.124	(0.105, 0.147)
	Daily incidence	Monthly incidence based on daily incidence	0.283	(0.264, 0.295)
	Weekly incidence	Monthly incidence based on weekly incidence	0.280	(0.262, 0.302)

2.3.3. Example of modeling implications

In order to show how the results presented here can be reflected in modeling results, a simple data example was run. Weekly case counts of YF, aggregated for all

of Brazil, were fit to Poisson models by week and by month. Two predictors were generated from the weekly case data and then aggregated to match monthly data, where monthly predictors are the means of groups of weekly predictors. The model parameters, shown in Table 2.4, show the same directional effect, but not at an expected magnitude. The dissimilarity in magnitudes of parameter estimates indicates that the predictors do not predict weekly and monthly cases in the same way. The parameter estimates also did not scale identically or by the same magnitude of time aggregation. One parameter has a larger magnitude in the monthly model compared to the weekly, model. The other has a smaller magnitude in the monthly model compared to the weekly model. Similar results are seen when using the Dec 2016 – Mar 2018 data, producing predictors in the same way (Table 2.5).

Table 2.4. Output from Poisson regression models using observed nationwide weekly and monthly Yellow Fever case counts in Brazil between Jan 2000 and Nov 2016.

Model parameter	Estimate (SE) using weekly cases	Estimate (SE) using monthly cases
Intercept	-13.209 (1.014)	-10.557 (2.008)
X_1	4.150 (0.249)	5.691 (0.493)
X_2	2.082 (0.575)	0.545 (1.188)

Table 2.5. Output from Poisson regression models using observed nationwide weekly and monthly Yellow Fever case counts in Brazil between Dec 2016 and Mar 2018.

Model parameter	Estimate (SE) using weekly cases	Estimate (SE) using monthly cases
Intercept	1.066 (0.297)	2.357 (0.354)
X_1	3.150 (0.205)	2.725 (0.206)
X_2	-2.548 (0.301)	-1.963 (1.188)

2.4. Discussion

This study aimed to determine if distributional fits of disease incidence data at different temporal resolutions are interchangeable, which can represent differences in underlying mechanisms guiding incidence. Using distributional fits rather than a model that incorporates other data sources minimizes the assumptions made when comparing resolutions since only incidence data were used. Comparisons of daily, weekly, and monthly YF incidence were made both before and after the start of the 2016-2017 outbreak. From the fitted distributions, empirical distributions were generated so that proper comparisons could be made across temporal resolutions, which use distributions sharing the same person-time denominator. Using data collected prior to the start of the outbreak, there was weak evidence supporting the claim that the fitted distributions are not interchangeable. After the start of the outbreak, when greater incidence was seen, there was strong evidence to conclude that distributional fits were not interchangeable.

The results of this study have potential modeling implications, as described previously. When developing a mathematical or statistical model for infectious disease research, the probability distribution representing an outcome, such as disease incidence, is assumed through a link function of a generalized linear statistical model or a compartment transition probability in a mathematical model. If the distributional fits, assumed to represent the true mechanisms guiding incidence, differ across resolutions, then the models, which are designed to elucidate these mechanisms, may not be universally appropriate across multiple temporal resolutions. The simple examples presented earlier show how the parameter estimates from a statistical model change when aggregating data. The change seen was not uniform across parameters or intuitive based on the aggregation of data. Though this example uses the same data used in distribution fitting, though in a different form, it serves as a demonstrative example of how the

concepts presented in the methods used in this study can translate into modeling applications.

The finding of how distributional fits change when aggregating incidence data over time units in an incomparable manner suggests that, with respect to temporal resolution, aggregated data may not follow a distribution equivalent to the sum of the distribution being aggregated. This can be described in the context of daily and weekly incidence: if daily incidence is found to follow a distribution denoted $f(x)$, data aggregated to represent weekly incidence do not necessarily follow $f(7x)$. In the context of disease modeling, this may be seen in a mathematical model where a weekly iteration cannot be accurately simulated by combining seven daily iterations. In a statistical model, while the distributional family remains the same, the change in parameterization suggests that the predictors for incidence may change in either magnitude or relevance. Under such circumstances, it would be shown that predictors relevant to disease incidence at one temporal resolution may not be as influential or have the same effect at a different resolution.

Higher incidence values were seen among daily and weekly incidence before the outbreak, though fewer nonzero values were seen. Higher values of monthly incidence were seen after the outbreak began (Table 2.1). This fact, paired with the notable difference in the proportion of observations with incidence of zero before and after the outbreak began, suggests that surveillance quality may differ between the two time periods with regards to accuracy of matching dates to cases. It is possible that YF surveillance was more thorough after the outbreak began and that cases were reported with greater regularity. In the earlier years of data collection, multiple cases may have been reported together on the same date, despite having actually occurred on different dates. This further motivates the examination of the two time periods separately, as a change in surveillance quality can impact analytic results.

For many municipalities, few YF cases were seen prior to December 2016, leading to the number of cases in a municipality in a week often being the same the as the number of cases in a day since there was only one day in a week with observed YF cases. After the outbreak began in December 2016, there were more municipalities with observed cases more than one day per week, leading to the distributions for incidence being less similar when comparing fitted distributions for daily and weekly incidence. Each week where a municipality sees only one day with cases increases the similarity of the daily and weekly data. It follows that having a greater number of data points in common would lead to a more similar distributional fit. This could explain why greater similarity between distributions, as indicated by the KS values in Table 2.3, was seen in this earlier time period. This may also explain why, even though evidence of differences in distributional fits across resolutions in the earlier time period was weaker, the example model, which aggregated cases nationwide, saw a greater difference in model parameters than expected. Aggregating the data nationwide likely led to fewer similar values. Also, the example models did not only regress on positive values, as zeroes were included in the Poisson regressions.

The results from comparing incidences before the beginning of the outbreak indicate that distributions fit using temporal units of closer sizes (day/week and week/month) showed greater similarity than those fit using temporal units of more different sizes (day/month) (Table 2.3). Greater amounts of aggregation can lead to stronger differences in distribution, as seen in a larger KS statistic between the daily and monthly distributions using the Dec 2016 – Mar 2018 data compared to the KS statistics between the daily and weekly distributions during the same time period (Table 2.3), making the potential effects of using a different resolution more severe.

2.4.1. Significance and context

While no other studies, to the authors' knowledge, have placed the topic of temporal resolution sensitivity as the main focus, other works have addressed this issue and similar issues. When examining environmentally sensitive infectious diseases, including arthropod-borne viruses such as YF virus, desire for high temporal resolution has been expressed due to the connections between disease risk and both weather patterns as well as arthropod population dynamics [103, 153, 154]. While the results of this study do not directly support desire for high-resolution data specifically, they do suggest that the temporal resolution of a model should be, when possible, selected with attention to the desired context, as arbitrarily changing a temporal resolution is likely to impact the results of a model and the overall conclusions that may result from it, as indicated by the simple example shown. Similar results have been seen in ecology studies comparing models at different temporal resolutions, where exact quantitative results may change across temporal resolutions, even if qualitative findings remain similar [155, 156]. Studies of temporal resolution in ecology, similar to epidemiology, remain rare [156]. Similar methods to those of this study have been used to assess sensitivity to spatial resolution using Visceral Leishmaniasis incidence in Brazil, where aggregating municipality-resolution incidence data to the state resolution showed significant changes in distributional fit [19].

Surveillance groups, including government organizations, should consider the findings of this study to motivate record-keeping and surveillance strategies. Because changing the temporal resolution of the data prior to distribution fitting can lead to distributional fits that are not interchangeable, it is imperative to maintain accurate and precise records. Previous studies highlight feasibility and data availability when describing selection of temporal resolution [157, 158], so assuring that high-resolution data are available is crucial for accuracy in analyses. If case data are available at a daily resolution but without accuracy regarding the date

of infection or diagnosis, then the utility of analyses with daily incidence is diminished.

2.4.2. Contributions

This study offers a thorough examination of sensitivity to temporal resolution as the primary focus of analysis rather than as a sensitivity analysis for a larger group of analyses. Placing focus on this topic is important for examining assumptions often made without thorough investigation. These results provide evidence for the importance of being attentive to decisions typically made prior to conducting statistical analyses.

This study shows the use of distribution fitting for disease incidence to vector-borne infectious disease research. This process is not uncommon in animal epidemiology [159], but is uncommon in epidemiologic research for human diseases. Also of note is the consideration of multiple distributional families to provide the best fit to the observed data. Common practices in infectious disease epidemiologic research using statistical models involve selecting a distributional family for the outcome a priori, based on exploratory analyses, or selecting from a narrower range of distributional families. This study considered a broad range of distributional families, including some that are less conventional, and one of the less commonly used distributional families was found to provide the best fit for positive incidence.

An additional point of novelty from the results of this study is the use of the Gamma distribution to characterize disease incidence [19]. Mathematical models of infectious diseases have previously used the Gamma distribution to describe disease growth and progression rather than incidence [160–162]. Common statistical models for infectious disease cases or incidence typically use binary [91, 92] or Poisson [93, 163] processes. Though uncommon, the Gamma distribution has

advantages for future research relating to infectious diseases, as has been shown in other contexts such as oncology [164] or rainfall hydrology [165].

2.4.3. Limitations

This study and its results rely heavily on the quality of the YF data available. The methods used to fit distributions assume that accurate and precise dates were associated to cases. Reported YF case data represent some, not all, of the actual cases [166, 167]. Since the data used for this study were collected via passive surveillance, it is also likely that reliability of surveillance differed among locations within Brazil.

Another limitation of this study is the ability to generalize the results to the municipalities where YF was not observed. It is assumed that, if these municipalities were to observe any YF cases, the incidence of YF would follow the same distribution as the incidence observed in the municipalities used for distribution fitting. On a related note, the method of incorporating the zero values in distribution fitting assumes that the probability of seeing an incidence of zero is the same for all municipalities. However, the primary aim of this study was to examine sensitivity to the choice of temporal resolution rather than to infer disease incidence in new locations.

Also of note is the assumption of a constant distribution over the entire period of observation as well as over all municipalities observed. The impact of this assumption was mitigated by separating the total time of observation into two time periods, which allowed the possibilities of both differences in processes determining YF incidence in the presence and absence of an outbreak as well as differences in surveillance diligence, relating to differences in motivations to report cases or accurately diagnose cases.

When generating the empirical distributions for comparisons, the process of drawing from the higher resolution fitted distributions assumes that, for all municipalities and time periods, draws of incidence values were independent. It is likely that nearby municipalities and adjacent time periods will not be entirely independent in their YF incidence, which can be seen visually in Figure 2.1. However, these factors should be considered in model fitting. The purpose of distribution fitting in this study was to determine if examining the aggregation of incidences at different resolutions would necessitate notably different model/analytic considerations. The next steps after considering the results of this study, model construction, should consider the possibilities of spatial autocorrelation and temporal autocorrelation.

2.4.4. Recommendations

Based on the findings of this study, it is recommended that, when collecting data regarding disease incidence, accurate records be maintained with regards to the timing of cases. While this does pose challenges, as dates of infection, diagnosis, and reporting are likely different, accounting for this uncertainty would be beneficial. Providing artificial precision in the timing of cases is harmful for research by imposing improper confidence in less accurate data.

Another recommendation is to carefully consider the temporal resolution when conducting analyses. If confident in the quality of data at multiple resolutions, the motivations of the research question and stakeholders should be considered. Because the results of this study showed the same distributional family to most closely fit YF incidence across three different temporal resolutions, repeating analyses may not be problematic, but differences in parameterizations suggest that

it is unlikely that model results will remain the same when repeating with multiple resolutions, potentially changing the conclusions drawn.

Despite the observed sensitivity to the choice of temporal resolution during the later time period, it does not necessarily follow that one temporal resolution should be considered more "correct" than others. While small units have been viewed as desirable [132–136], the potential for imprecision in recorded case dates, particularly prior to the outbreak, could lead to inaccuracy in exchange for the increased precision. The decision to use a particular temporal resolution should be guided by research questions of interest and known quality of data. However, the results from this study should be considered, particularly when a temporal resolution is being decided, because it should be noted that the choice of temporal resolution has potential to impact results and conclusions.

These findings are of interest to researchers using analytic models for infectious diseases. Statistical models rely on assumptions made for disease incidence for selection of model type and appropriate link function. Mathematical models with stochastic components may benefit from the use of the Gamma distribution for accurate probabilistic representation of disease incidence. In either context, it is notable to mention that the temporal resolution used is important and likely to impact model results.

A final recommendation is to account for epidemics and outbreaks when examining data over long periods of time. The assessment of sensitivity to temporal resolution in this study showed differing results when stratifying time periods based on the 2016-2017 YF outbreak in Brazil (Table 2.3). Analyses in other contexts are likely to find similar differences in both incidence characterization and factors that determine both presence and magnitude of incidence.

2.5. Conclusions

The findings of this study show that probabilistic representation of disease incidence changes with different temporal resolutions when examining YF incidence through distribution fitting, particularly after the onset of major outbreaks. Methods and results in this study are translatable to other infectious diseases, and in particular resolution-dependence sensitivity is a common problem that likely affect epidemic dynamics with major outbreaks. Similar findings have been shown in biodiversity research where some ecological patterns vary with resolution but resolution-aware function can rescale data with loss of accuracy [168]. The findings of this study have implications for model design when conducting modeling studies.

Care should be taken when making the decision for temporal resolution in analyses of infectious disease data, as using different resolutions likely leads to inference made for different processes that act at different scales. Because vector-borne infectious diseases such as YF result from a combination of anthropogenic [169] and environmental [91, 125, 154] factors, it is reasonable to consider that the importance of factors affecting disease incidence over a day may differ from those affecting incidence over a week or a month. This seems more likely when considering epidemic periods with major outbreaks. Furthermore, due the processes driving dynamics of epidemics likely differing from those of endemic, separating data of these two dynamical regimes is recommended if causality investigation is the goal of the study. Other approaches may look into resolution-rescaling functions to guarantee the same prediction accuracy or models with varying resolutions over time, such as with higher and lower resolution over large outbreaks and during baseline disease periods, respectively. While accounting for these considerations, always of note is the quality of surveillance data, which affects the ability to use particular temporal resolutions or analytic methods.

Chapter 3 Estimating case fatality risk of severe Yellow Fever cases: A systematized literature review and meta-analysis

Summary: Case fatality risk (CFR), commonly referred to as a case fatality ratio or case fatality rate, represents the probability of a disease case being fatal. It is often estimated for various diseases through analysis of surveillance data, case reports, or health record examinations. The WHO reports that severe cases of Yellow Fever, a mosquito-borne viral disease, have a CFR of approximately 50%. This fatality risk estimate has not been confirmed or updated via a comprehensive review of literature. This study aims to estimate the case fatality risk of severe Yellow Fever cases through a systematized literature review and meta-analysis. A search strategy was implemented in PubMed and Ovid Medline in June 2019, seeking reported severe case counts, defined by fever and either jaundice or hemorrhaging, and the number of those that were fatal. The search yielded 842 studies, and title/abstract review followed by full text review produced 11 articles reporting 24 proportions of fatal cases suitable for meta-analysis. Data were analyzed through an intercept-only logistic meta-regression with random effects for study. The estimated CFR was 0.39 (95% CI: 0.30, 0.48). Stratifying by continent, study type, and symptom definitions showed that South America observed a higher CFR than Africa, studies reporting surveillance case data observed higher CFR than studies investigating outbreaks, and no differences in CFR were seen when under two symptom definitions of severe cases. This study provides the first comprehensive literature review to estimate the CFR of severe Yellow Fever cases, which can provide insight into outbreak preparedness and quantification of underreporting.

3.1. Introduction

Evaluations of infectious disease severity often account for both morbidity and mortality. The latter can be represented by the case fatality risk (CFR), defined as the probability of a case of a disease being fatal [170, 171]. It is, in its simplest form, estimated by a quotient of the number of fatal cases and the total number of cases, as have been reported for disease outbreaks [172–174]. The CFR can be used in these contexts to understand the severity of a disease and implement appropriate policy in the event of an outbreak [175].

Case fatality risks have also been described as case fatality ratios or case fatality rates without difference in definition [176, 177]. The case fatality risk can differ from a case fatality rate in that the case fatality risk does not explicitly specify a time period, whereas a case fatality rate implies a period of time [171, 178]. It is common, however, for the three terms to be used interchangeably, regardless of whether time periods are taken into consideration [178].

An estimate for the case fatality risk of a disease based on observed fatal cases is typically included when reporting results from an outbreak investigation [173, 179]. Outbreak investigations for various diseases tend to report the CFR during the time period of observation, though this is not often their primary study aim. Other works aiming to estimate case fatality risk have done so by observing hospital records for the proportion of fatal cases [180] or by tracking outcomes for confirmed cases of diseases [176, 181]. Many of these studies aimed not only to estimate CFR for various diseases, but also to examine risk factors for fatality in order to identify individuals most at risk [177, 180] or examine changes in CFR over time [182]. While valuable for the certainty of information among those recruited, such studies can experience limitations such as ascertaining only the most severe cases or not observing fatal cases who died after data collection [117]. In ascertaining only the most severe cases, the denominator used to calculate the CFR is not representative

of all cases. If more severe cases are more likely to be seen in the denominator of the CFR calculation, then the CFR will likely be overestimated due to the denominator representing a subset of cases. Another limitation seen in studies reporting CFRs is the use of unconfirmed cases. Reported CFRs of suspected cases without confirmation may lead to a CFR that also included non-cases in its calculation. Depending on the proportion of such cases that are fatal, this could lead to an overestimate or an underestimate of the CFR.

Other studies have aimed to estimate CFR by collecting data through literature review [35, 170] and, in many cases, meta-analysis. Some articles have estimated an overall case fatality risk by pooling numerators and denominators across studies [183], pooling study results with a random effects meta analytic method [184–186] , or generalized linear models [187], though not all reviews combined results [188].

One disease with a reported case fatality risk is Yellow Fever (YF), a disease caused by a *flavivirus* spread by multiple genera of mosquito [143] and endemic in sub-Saharan Africa and South America [1]. The disease is asymptomatic in a large number of cases, and symptomatic cases present with flu-like symptoms such as fever and body aches [1]. In a smaller number of cases, severe disease develops within a few days, with more severe symptoms including jaundice and hemorrhaging. The World Health Organization (WHO) reports that approximately half of severe cases are fatal [1].

The WHO figure of half of severe YF cases being fatal is often cited, but little work exists aiming to evaluate or update this estimate. A 2014 study by Johanssen et al aimed to estimate the proportions of cases that are asymptomatic, mild, and severe, and also estimated a CFR among severe cases which aligned with the WHO estimate of 50% [15]. No other works exist that offer an update of this estimate using a systematic method.

This study aims to estimate the case fatality risk among severe Yellow Fever cases through a systematized literature review [123] and meta-analysis. Through literature review, articles were found that contained denominators of severe YF cases and numerators of those cases that were fatal. The CFR was estimated among all studies and then stratified by continent, symptom definitions, and by study type, which is based on whether an outbreak investigation was described in the study or whether cases were reported without known data collection methods. The results of this study offer an updated estimate of the CFR for severe YF using a comprehensive search of relevant literature.

In conducting the literature review, parts of the search strategy aimed to not only estimate the proportion of fatal severe cases, but also the proportions of severe and mild cases to estimate burden of disease through proportions of asymptomatic, mild, and severe cases. There were insufficient studies for reliable estimates for disease burden; only methods and results for estimating the CFR among severe cases are presented here

3.2. Methods

A systematized literature review [123] was conducted to collect relevant data to estimate the case fatality risk of severe YF cases. A systematized literature review follows many of the same procedures as a systematic literature review, allowing the relaxing of some practices such as multiple independent reviewers. The aim of the literature review was to find articles containing proportions of observed severe YF cases that are fatal, including a numerator and denominator. Severe YF cases in this study were defined by the WHO definition of a febrile patient presenting jaundice and/or hemorrhaging [1].

3.2.1. Literature review

This study adhered to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines [189]. The PRISMA checklist [190] for this review can be found in Appendix B. A search strategy was developed for PubMed and Ovid Medline aiming to identify published articles in English that report a denominator of total severe YF cases and a numerator of fatal cases. There were no restrictions regarding the year and location of the study, though studies were expected a priori to only be set in nations where YF is endemic.

The search strategy, as run in PubMed, was as follows:

("Yellow Fever" in the title, abstract, or a MeSH term)
AND
(fatal*, severe, severit*, mortality, asymptomatic, symptomatic, diagnosis, misdiagnosis, outbreak, or cases in multiple places)
AND NOT
("Vaccine" in title or abstract)

with asterisks denoting wildcards. "Vaccine" was excluded because of the high number of studies that aimed to study vaccine efficacy or describe vaccination campaigns, which are outside the aims of this study. Some studies excluded due to the "vaccine" criterion were examined and found not to lie within the scope of the analyses for the study aims. The search strategy was run in June 2019, and yielded a total of 485 articles from Ovid Medline and 605 articles from PubMed. After removing duplicates, 842 unique articles underwent title and abstract screening.

Articles found via the database search were screened by title and abstract to remove those without relevant information. Articles were excluded if the topic was a disease other than YF or if the title and abstract did not mention investigating or reporting

cases of YF. Excluded articles primarily focused on other diseases, laboratory diagnosis methods, or vaccine efficacy research. Following title and abstract screening, 164 articles remained for full text review.

In full text review, remaining articles were screened for relevant information for the purpose of this study: identifying the proportion of severe YF cases that are fatal. Articles were included if full text was available in English and contained both a denominator of total severe YF cases and a numerator of fatal cases. The symptoms outlined by the WHO of fever with at least one of jaundice and hemorrhaging were needed. Articles where some information was not stated, but could be calculated, such as an article stating a CFR and a denominator, were included. Articles that were not included most commonly included editorials, single case studies, reports of capacity building for YF prevention, or laboratory detection methods. Of the articles undergoing full text review, 14 were not available in full text, seven were not available in English, and 117 did not contain relevant information. During the full text review, four articles were added from the references of the articles read, yielding 30 articles in total. Following data collection, the 30 included studies prior to isolating severe cases were re-read to confirm the information extracted.

The final set of inclusion criteria only included studies that described severe YF cases. Severe YF cases were defined through the WHO descriptions of symptoms, and included either jaundice or hemorrhaging along with fever [1]. The requirement of explicitly mentioning fever was relaxed if YF was confirmed and either jaundice or hemorrhaging was stated. Only 12 studies contained explicit symptomatic descriptions of cases (Figure 3.1, Appendix B).

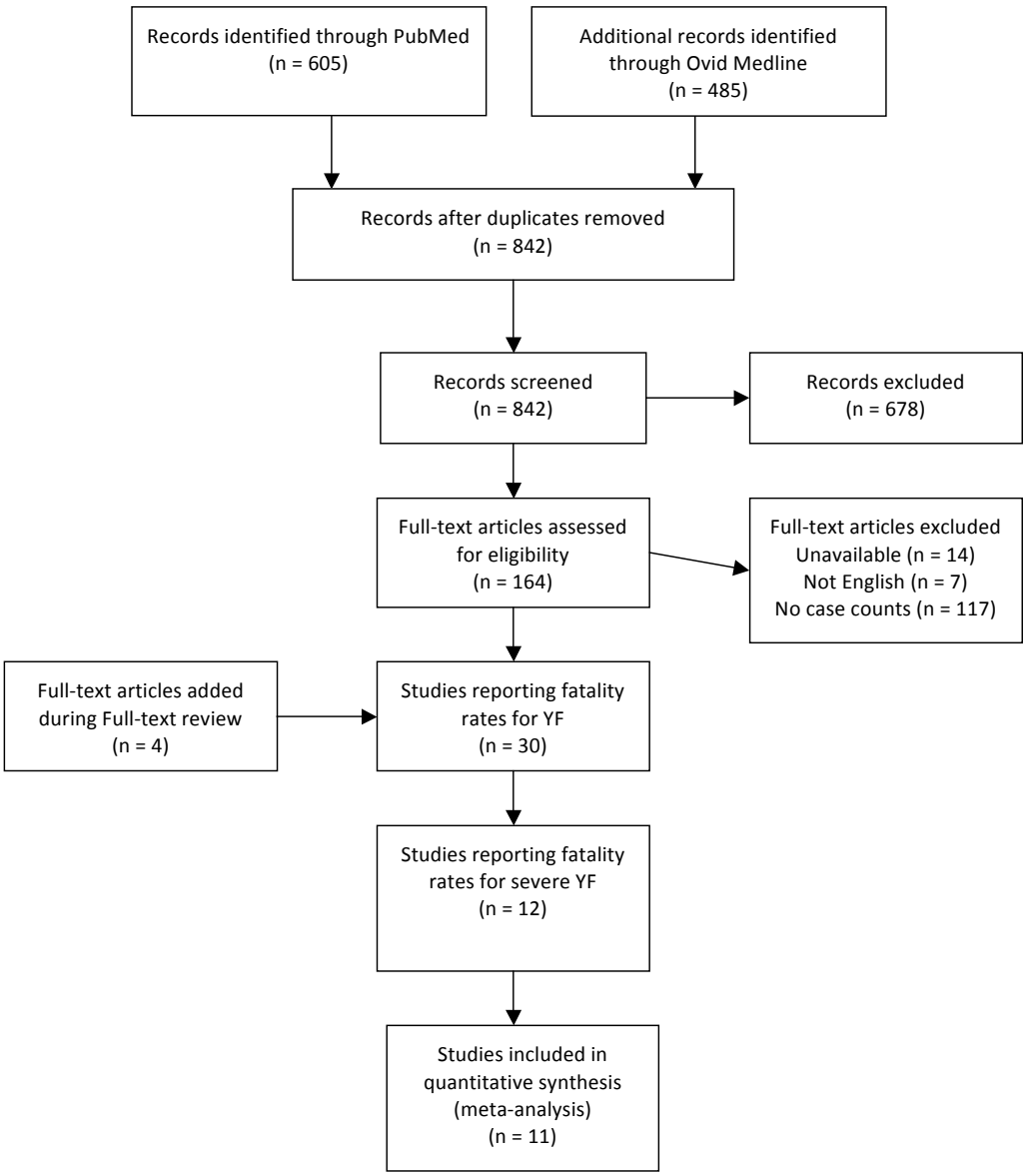


Figure 3.1. Flow chart of search strategy and studies screened and evaluated for inclusion in meta-analysis.

In addition to the numbers of total and fatal cases, descriptive information was collected for each study. The country and year of data collection were reported for each study. Case definition was collected based on symptoms in order to confirm

that cases listed in each study were severe cases. The symptoms used to identify cases were listed for each study. It also was noted whether the authors of each study were active in recruiting YF patients of stating reported case data. If the authors reported determining a case definition, being present in data collection, or stated details of an outbreak investigation, then the study was classified as an "investigative" study. Others, where case counts are reported without specifically describing active measures to identify cases, were classified as "reporting" studies. Also collected was whether the cases in each study were confirmed or suspected. Cases were considered confirmed if the articles stated laboratory confirmation, whether through stating specific laboratory methods or by stating that cases were confirmed. Cases were assumed suspected if the article explicitly stated cases were suspected, laboratory confirmation was not mentioned, or no other information was given regarding whether cases were confirmed or suspected. No studies reported a combination of suspected and confirmed cases.

Because all included studies list total and fatal counts of severe YF cases without the CFR being the primary focus of the study, the sources of bias in the individual studies remained consistent. Two major sources of bias are underreporting of cases within studies and deaths occurring after followup [117]. Underreported cases, if nonfatal, would lead to a smaller denominator and an overestimated CFR. If the underreported cases were as fatal as reported cases, no bias would be observed. However, because the primary focus of the studies was not to report or estimate the CFR of YF, issues of publication bias, where tests of statistical significance affect reporting of results [191], may not be present, since no included articles used statistical significance tests for CFR estimates.

Collected data were inputted into a Microsoft Excel (2013) spreadsheet during full text review, with information to be collected determined prior to data collection.

3.2.2. Data analysis

Case fatality risk was estimated using a meta-analysis for proportions. The metaprop() function from the 'meta' R package [192] was used to run an intercept-only logistic meta-regression with random effects for study. Model inputs included the observed proportion of fatal cases for each study, $\hat{p}$, as well as its standard error, estimated by $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, where n is the denominator of the study. Only proportions not equal to zero or one were included, as these would produce standard errors of zero. Estimates for CFR were found for confirmed, suspected, and all severe YF cases. Stratified case fatality risks were estimated for differences in continent, through collection of the country of each study. Stratified CFRs were also estimated by study type, separating investigative and reporting studies defined previously. A third stratified analysis separated case definitions, where studies only using fever and jaundice for severe YF case definitions were compared to studies using other symptoms, whether in addition to or instead of jaundice. Values of I^2 were calculated to describe heterogeneity [193]. Analyses were conducted using R version 3.6.0 [150].

3.3. Results

3.3.1. Papers found

A total of 12 studies were found through the literature review, two of which were added from the references of the 164 articles that underwent full-text screening. The 12 papers contained a total of 30 proportions of fatal severe cases; 24 of which, present in 11 of the studies, were not equal to zero or one and therefore included in the meta-analysis (Figure 3.1). A total of eight articles provided ten proportions of CFR among confirmed severe cases of YF, and another three articles provided 14 proportions among severe suspected cases.

Articles investigating YF reported cases in both Africa and South America. The countries with more than one included fatality proportion were Nigeria (3 papers, 9 proportions), Ghana (1 paper, 6 proportions), and Brazil (2 papers, 3 proportions).

Articles reporting multiple proportions in the same country during the same year(s) reported different proportions for different locations within countries.

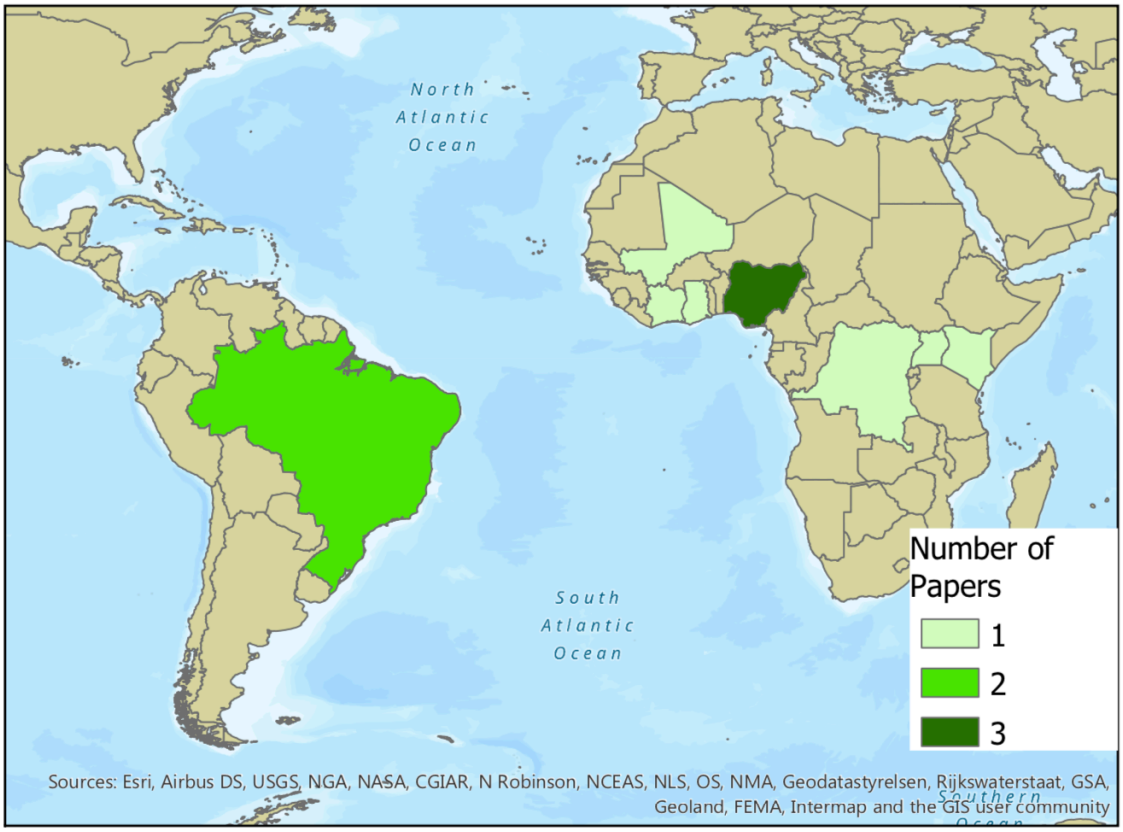
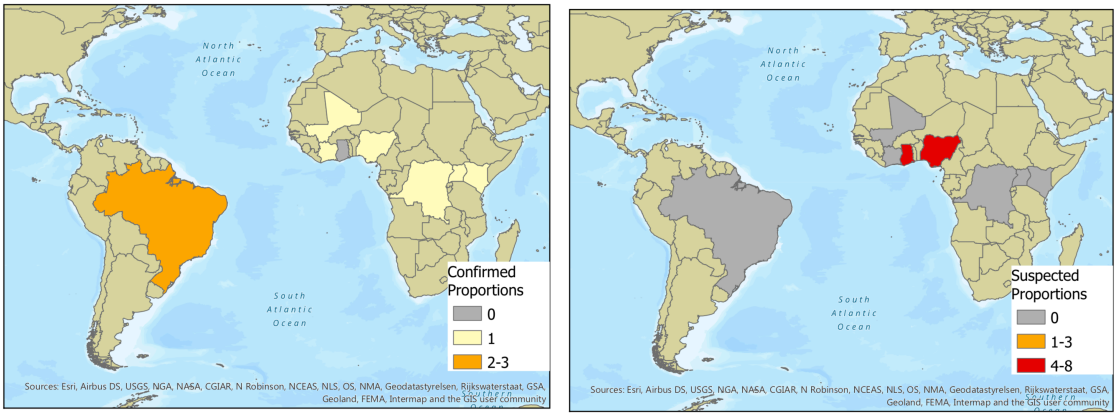


Figure 3.2. Numbers of articles from literature review containing proportions of fatal severe Yellow Fever cases by country.



(a) (b)

Figure 3.3. Numbers of proportions of fatal severe Yellow Fever cases found through literature review by nation among (a) confirmed and (b) suspected severe Yellow Fever cases

3.3.2. Estimates of case fatality risk

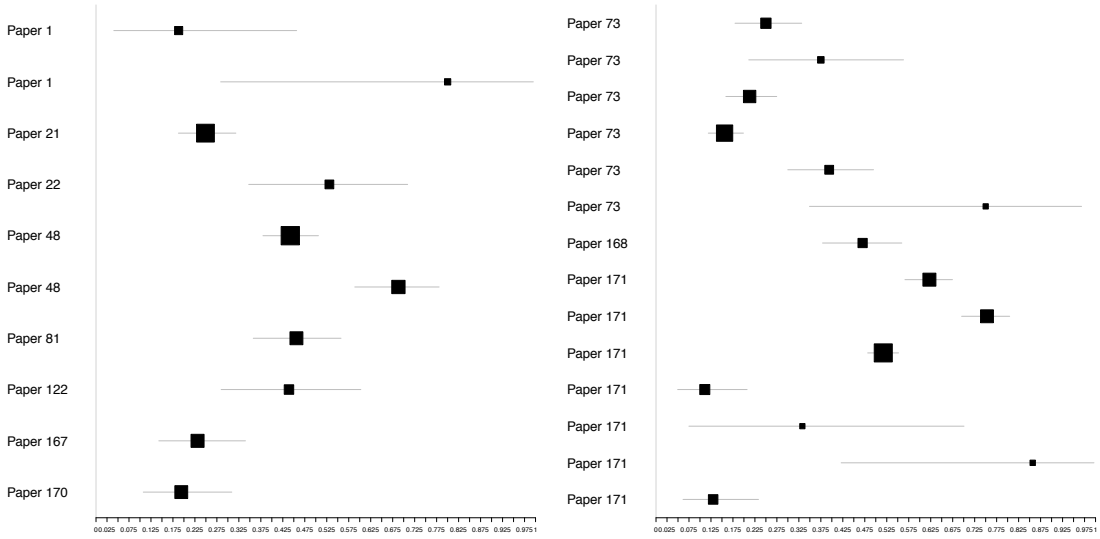
Forest plots are shown in Figure 3.4, separated by confirmed and suspected YF cases. The individual numerators and denominators in each study are shown in Table 3.1. Variation in the reported proportion of fatal severe YF cases is seen among both confirmed and suspected YF.

Table 3.1. Included studies from literature review that include numbers of fatal and nonfatal severe Yellow Fever cases with descriptive data.

Paper number	Authors	Country	Year	Lab confirmed	Study type	Numerator	Denominator
1	WHO 2008 [194]	Cote d'Ivoire	2006	Yes	Report	3	16
1	WHO 2008 [194]	Mali	2006	Yes	Report	4	5
21	Wamala 2012 [195]	Uganda	2010-2011	Yes	Investigation	45	181
22	de Filippis 2002 [196]	Brazil	2002	Yes	Investigation	17	32
48	Tuboi 2007 [197]	Brazil	1998-2007	Yes	Report	67	97
48	Tuboi 2007 [197]	Brazil	1998-2003	Yes	Report	111	251
73	Agadzi 1984 [198]	Ghana	1977-1978	Not stated	Investigation	34	136
73	Agadzi 1984 [198]	Ghana	1978-1979	Not stated	Investigation	44	207
73	Agadzi 1984 [198]	Ghana	1978	Not stated	Investigation	12	32
73	Agadzi 1984 [198]	Ghana	1978	Not stated	Investigation	53	340
73	Agadzi 1984 [198]	Ghana	1979	Not stated	Investigation	41	104
73	Agadzi 1984 [198]	Ghana	1980	Not stated	Investigation	6	8
81	Jones 1972 [199]	Nigeria	1969	Yes	Investigation	47	103
122	WHO 1993 [200]	Kenya	1993	Yes	Report	18	41
167	Ingelbeen 2018 [201]	Democratic Republic of Congo	2017	Yes	Report	18	78
168	deCock 1988 [202]	Nigeria	1986	Not stated	Investigation	59	126
170	Monath 1980 [203]	Gambia	1978-1979	Yes	Investigation	13	67
171	Nasidi 1989 [204]	Nigeria	1987	No	Investigation	202	325
171	Nasidi 1989 [204]	Nigeria	1987	No	Investigation	192	255
171	Nasidi 1989 [204]	Nigeria	1987	No	Investigation	416	805

Table 3.1 cont.

171	Nasidi 1989 [204]	Nigeria	1987	No	Investigation	8	72
171	Nasidi 1989 [204]	Nigeria	1987	No	Investigation	3	9
171	Nasidi 1989 [204]	Nigeria	1987	No	Investigation	6	7
171	Nasidi 1989 [204]	Nigeria	1987	No	Investigation	9	69



(a) (b)

Figure 3.4. Forest plots of case fatality risk estimates among (a) confirmed and (b) suspected severe Yellow Fever cases.

The estimated CFR among all severe cases was 0.39, with a 95% confidence interval of [0.30, 0.48]. Separating CFR by case confirmation yielded no differences in CFR (Table 3.2). The CFR for confirmed cases was 0.39 (95% CI: [0.28, 0.51]), and the CFR for suspected cases was also 0.39 (95% CI: [0.26, 0.53]). Notable heterogeneity was seen among studies, as indicated by I^2 values.

Table 3.2. Estimated case fatality risk for severe Yellow Fever infection among confirmed, suspected, and combined severe cases.

Confirmed	Suspected	Combined
0.39	0.39	0.39
[0.28, 0.51]	[0.26, 0.53]	[0.30, 0.48]
$I^2 = 0.89$	$I^2 = 0.97$	$I^2 = 0.96$

3.3.3. Stratified case fatality risks

The CFRs for severe cases were stratified by characteristics of the articles to account for potential heterogeneities in either YF dynamics or data collection and reporting. First, CFRs were stratified by continent. Combining proportions of both confirmed and suspected severe cases, the estimated CFR among severe cases in African countries was 0.36 (95% CI: [0.27, 0.47], n = 21), and the estimated CFR among South American countries was 0.55 (95% CI: [0.42, 0.68], n = 3).

An additional stratification considered the study types of each article. Investigative articles and reporting articles, as described previously, were defined by whether data collection methods were stated as part of the article and whether the researchers were involved in data collection. Among investigative studies, the estimated CFR was 0.37 (95% CI: [0.27, 0.49], n = 18), and among reporting articles, the estimated CFR was 0.44 (95% CI: [0.28, 0.61], n = 6).

Estimates of CFR were also stratified by symptom definitions. While symptom definitions varied across studies (Appendix B), many studies based YF diagnosis on fever and jaundice only. Others included other symptoms such as hemorrhaging, organ failure, and abdominal pain. Proportions were stratified by whether any other symptoms besides fever and jaundice were considered in case definitions, since

many of the studies reported only including fever and jaundice to diagnose severe YF cases. Among studies that only used fever and jaundice to diagnose YF, the estimated CFR was 0.39 (95% CI: [0.25, 0.56], n = 12). Among studies that did consider other symptoms, the estimated CFR was 0.38 (95% CI: [0.29, 0.49], n = 12).

3.4. Discussion

This study aimed to systematically evaluate the case fatality risk of severe YF cases. A systematized literature review was conducted in order to find reported proportions of fatal cases of severe YF. Using 24 proportions recorded from 11 articles, the estimated CFR was 39%, which was consistent among both confirmed and suspected severe cases. Separating the CFR by continent showed a notably higher CFR in South America compared to Africa, separating the CFR among severe cases by article type showed a lower CFR among investigative studies when compared to passively reported cases, and separating the CFR by inclusion of symptoms beyond fever and jaundice did not lead to different estimates. The estimated CFR of 39% is lower than the common estimate of half of severe cases being fatal [1, 35]. While there is uncertainty around both estimates, a difference of approximately ten percentage points may be notable for clinical decision-making and perceived severity of YF.

The drastic difference in estimated CFR between South America and Africa may potentially result from differences in data collection as well as differences in YF dynamics. Other systematic reviews estimating CFR have seen geographic differences, including separating Hong Kong and other regions from mainland China in estimating CFR for hand, foot, and mouth disease [205] and separating WHO world regions to stratify CFR for Salmonella infections [206]. The differences seen across geographic locations in these studies were less pronounced than the difference in CFR between continents found in this study. Different strains of Yellow

Fever are found between South America and Africa [46, 207] as well as different primary mosquito vectors and nonhuman primate reservoirs [208]. Clinical care also differs across countries. Though International Health Regulations require reporting of YF cases [44], implementation and surveillance quality may differ between the two continents, as well as differences in healthcare seeking behaviors, which can lead to differences in severe cases represented across each continent. The small number of proportions representing CFR in South America may also account for the difference in estimated CFR across continents, as having three South American proportions makes the CFR estimate more sensitive to any single study providing a non-representative sample of severe YF cases. Additionally, the proportions found in African nations primarily represented a geographic subset within the continent, primarily representing western African nations. Further heterogeneity may also be seen within the continent.

Another notable difference in CFR was observed when separating investigative studies, where researchers were involved in case recruitment, and reporting studies, which typically report passive surveillance cases. Studies in the former group were more frequently investigating ongoing outbreaks. Testing strategies may differ during outbreaks as well as increase awareness of YF. Because YF shares symptoms with many other tropical diseases [209, 210], there is a possibility for YF cases to be diagnosed as another disease without testing. Additionally, an outbreak may impact healthcare seeking behavior [211], leading to differences in observed numerators or denominators.

Inclusion of symptoms beyond fever and jaundice to classify severe YF cases did not lead to a difference in estimated CFR. This indicates that specific symptoms cannot be used to predict whether a severe YF case is more or less likely to be fatal. If particular symptoms of severe YF are not predictive of fatality, this finding indicates that the different symptom definitions may lead to similar case ascertainment since CFRs were the same. Due to low sample sizes of particular symptoms, comparisons

of other specific symptoms were not feasible. Doing so may provide further insight of whether any specific symptoms can be used to predict fatality.

Across studies, there was significant heterogeneity among the CFRs reported (Table 3.2). Among other reviews estimating CFR for other infectious diseases, high heterogeneities have also been seen [206, 212, 213]. Potential sources of heterogeneity across the different studies include differences in surveillance resources as well as differences in healthcare infrastructure across the various settings of the studies.

The results of this study show a reported CFR that is notably lower than the estimate of 50% reported by the WHO [1] and by Johansson et al [35]. If the WHO estimate is based on reported case counts by passive surveillance, then this may be an overestimate, as this study shows that articles passively reporting cases produce higher CFR estimates compared to articles described as investigative studies. This does not, however, suggest that the estimated CFR from this study will apply to every outbreak situation. The CFRs of other diseases have been seen to change over time [182] and may differ in relation to industrialization [214]. The estimate yielded in this study should be used as an average CFR and broad recommendation.

This study's literature review yielded a total of 12 relevant studies, 11 of which provided 24 proportions of fatal severe Yellow Fever cases that were included in meta-analysis. Other studies using systematic review methods to estimate the CFR of other diseases commonly have more available papers relevant to the study aims [183, 206, 212], though others have had similarly lower article counts [205, 213, 215]. Many of these studies also used the I^2 statistic to consider heterogeneity, with many of them similarly showing high heterogeneity [20–22]. Stratification by geography was also seen in other studies [205, 206].

3.4.1. Use for estimating burden

Having a reliable estimate for CFR can prove useful for attempting to quantify underreporting of YF cases. Case reports with higher proportions of fatal cases may suffer from underreporting under the assumption that the CFR found in this study is broadly applicable to other incidence of YF. For example, using data provided from the Pan American Health Organization and the Brazilian Ministries of Health, 157 fatal among 327 confirmed cases of Yellow fever were observed in Brazil between 2000 and 2014, which may include non-severe cases. Under the assumptions that the 39% CFR found in this study is applicable to these data, no fatal cases were unreported, and only severe cases become fatal, estimates of actual case counts can be produced. By multiplying the 157 fatal cases by the inverse of the estimated CFR, an estimate of 403 severe YF cases is obtained. If all 327 reported cases were severe cases, then approximately 19% of severe cases were undetected. This serves as a minimum proportion of underreported cases rather than an estimate [35] since this assumes all fatal cases were observed and all reported cases were severe.

3.4.2. Strengths and novelty

This study benefits from the use of a comprehensive strategy for literature review, which maximizes the completeness of data available for analysis. Conducting a literature review rather than estimating CFR solely from surveillance data allowed multiple outbreak investigations to contribute to the data analysis. As a result, studies with researchers taking a more active role in surveillance of YF cases, which may have greater accuracy, were included.

This study stratified CFR by the methods of the individual studies into investigative and reporting studies. Separating the studies by the researchers' involvement in patient recruitment and assessing symptoms of cases showed that studies with

researchers involved in an investigation reported a lower proportion of fatal cases compared to articles reporting surveillance statistics. Both types of studies could experience different limitations to accuracy. While investigative studies may be less likely to have an inaccurate CFR due to underreporting, they may be more likely to have an inaccurate CFR due to deaths occurring after study followup [117]. These two matters are likely to lead to an overestimate and underestimate, respectively, of CFR.

3.4.3. Limitations and future directions

The results of this study rely on the reported data from the 11 studies used in analysis. Because the purpose of these articles was not necessarily to offer estimates of the CFR for Yellow Fever, it is possible that maximizing the accuracy of fatal and nonfatal case counts was not the highest priority. The studies detailing outbreak investigations do report numbers of severe and fatal cases, but the purpose was not to assure generalizable accuracy of the CFR. This possibility is even stronger among reporting studies. Since underreporting of infectious disease cases is a well established issue [93], it is likely that the proportions used in this analysis are also subject to issues of data quality. Furthermore, because accurate estimation of CFR was not the primary focus of these studies, case definitions across some studies lacked precision. Some studies did mention symptoms as inclusion criteria, while others listed cases that were seen among patients. Such reporting can lead to ambiguity in interpretation of severe YF case definitions.

Within this review, the results are limited by the heterogeneity in studies and the assumption that different world regions are expected to have similar CFRs. In combining the studies from South America and Africa, where stratified CFRs differed notably, it is assumed that the differences observed are artifacts of the individual studies rather than indicative of actual differences in CFR across the two

continents. This heterogeneity may result from actual differences in probability of fatality; risk of fatality may differ by population demographics [170, 216] or national industrialization [214], as seen in other diseases.

3.5. Conclusions

Among severe cases of YF, the CFR is estimated to be approximately 39% based on the results of a systematized literature review. This is lower than the frequently cited CFR for severe cases, indicating that the previous estimate is either a cautious estimate or based on underreported data. Further research is needed to distinguish among asymptomatic, mild, and severe YF infections in order to most accurately estimate the total burden of disease. Stratified CFR estimates for Africa and South America showed notably different estimates. Different study types, defined by the involvement of the researchers in patient diagnosis, were shown to produce different stratified estimates of CFR, but differences in symptoms used to define severe YF cases did not. This estimate of CFR can be used to estimate the severity of future outbreaks of YF or to estimate minimum quantities of underreported cases.

Chapter 4 Environmental determinants predicting long-term vulnerability to Yellow Fever burden

Summary: Yellow Fever is an endemic mosquito-borne disease in Brazil. Many locations within Brazil have not observed any cases within the most recent decades despite having characteristics suitable for disease transmission; these locations may be at risk of observing high future incidence. Environmental factors, including ecological habitat type or annual precipitation patterns, are known to impact human populations, mosquito vectors, and nonhuman primate reservoirs, indicating potential for accurate development of risk profiles for Yellow Fever. This study aims to use long-term environmental characteristics to estimate vulnerability to Yellow Fever, defined as capability of observing high yearly incidence, among all Brazilian municipalities. Long-term is defined as a year or longer in duration. Yearly incidence between 2000 and 2017 were available from 466 municipalities that observed any Yellow Fever cases between January 2000 and March 2018. Vulnerability was defined by three categories: zero, up to 10, and greater than 10 cases per 100,000 people. A cumulative logit model was fit using various environmental predictors that are not expected to fluctuate over short periods of time such as days or weeks. Separate models were run for an endemic time period (2000 – 2016) and epidemic time period (2017) to account for a major outbreak. Results from the model that most accurately fit the data were then applied to all municipalities to identify most vulnerable locations to high Yellow Fever burden. Percent of days with precipitation yearly, ecological habitat type, population density, drainage density, elevation, and yearly presence of Yellow Fever in nearby municipalities were included in the best fitting model. Precipitation and nearby Yellow Fever cases were found to be most influential to the model through global sensitivity analysis. Results of this study identify municipalities in Brazil most likely to have a high vulnerability to Yellow Fever, offering prioritized locations for prevention.

4.1. Introduction

Yellow Fever (YF), a disease caused by Yellow Fever Virus, is a prevalent public health concern in much of the tropics. It is endemic in sub-Saharan Africa and South America and spread by multiple genera of mosquitoes [1, 31]. Infections of the virus can present a febrile illness, with a fraction of cases becoming severe and developing symptoms including jaundice and hemorrhaging, many of which are fatal [1, 217]. The existence of nonhuman primates as a reservoir and the ability for mosquitoes to transmit the virus transovarially [218, 219] lead to the belief that the virus will never be globally eradicated [45]. Infections can, however, be reduced in humans due to the existence of a safe and effective vaccine [218, 220].

Due to the anticipated presence of YF in parts of the tropics in perpetuity, there is interest in determining risk factors for YF in order to prioritize preparedness. Such preparations can include vaccine prioritization, vector control, surveillance improvements, or prioritized messaging. One set of determinants that can inform risk of YF burden includes characteristics of the natural environment [54, 111, 221, 222]. Mosquito populations have been shown to be sensitive to climatic factors such as temperature and rainfall [112], and different ecological habitat types are more hospitable for mosquitoes [54]. Climatic and environmental factors can also impact activities of humans [223] and nonhuman primates [224], furthering their effects on YF dynamics.

In recent decades, YF has been seen in much of Brazil in areas adjacent to the Amazon Rainforest [54]. An outbreak, which began in December 2016 [58], saw high case counts in southeastern areas of the country, including areas where the vaccine was not previously recommended for travelers [62]. However, even following this outbreak, many locations in Brazil have not seen any YF cases in the most recent decades. This could be because of epidemiologic silence, in which infections may have occurred without being detected by health or surveillance

systems [115, 116, 225], or because infections did not occur during the time period. Without knowing whether these municipalities do not observe YF cases due to epidemiologic silence or lack of risk, it is important to examine YF risk across all municipalities, including the epidemiologically silent ones, to determine those that are more likely to see a high disease burden. This can be used as a way to define vulnerability to YF infections and motivate preparations, whether through improved surveillance and testing, vaccination, or improved healthcare infrastructure.

Vulnerability to infectious diseases has been previously defined in multiple ways. Previous studies defining vulnerability describe susceptibility to outbreak cases, likelihood of seeing high case burden, or inability to adequately care for observed cases [226]. Definitions of vulnerability to infectious diseases commonly rely on socioeconomic factors [226–229], and many studies measuring vulnerability emphasize the vulnerability of a nation or other large region [227, 228]. For diseases such as YF, these socioeconomic factors are relevant to disease risk. However, due to the nature of the mosquito vector and nonhuman primate reservoirs, YF is an environmentally sensitive disease, meaning that factors of the natural environment are also of great concern for assessing vulnerability [54, 91, 226]. Furthermore, assessing vulnerability for more localized areas is of interest due to differences in local environmental and climatic factors throughout large nations as well as the small flight range of the mosquitos that transmit YF Virus [230]. Using features of the environment, such as presence of water, elevation, and ecological habitat types, can prove useful in estimating vulnerability.

Numerous factors can contribute to vulnerability to high YF burden, many of which can be described in broad groups. One group consists of characteristics of the surrounding environment. This can include characteristics that cannot be directly modified by human activities, such as climate, ecological characteristics, or geomorphology [54], as well as those that can be modified, such as urban form and land use [231]. Another group consists of human activities that can affect YF

vulnerability. These can include vaccination or mosquito control. A third group consists of human socioeconomics. Factors within a community such as poverty or healthcare infrastructure can affect susceptibility to YF [232].

This study aims to assess vulnerability to YF burden throughout all municipalities of Brazil using long-term environmental features. This primarily focuses on the first group of factors affecting YF vulnerability described above, representing environmental characteristics that cannot be targeted for immediate intervention. In this study, "long-term" is defined as comprising at least one year to contrast with environmental characteristics that can fluctuate over shorter periods of time such as over a week or month. Vulnerability in this study is defined by an ability to observe a high burden of YF over a year and predicted based on long-term environmental characteristics. Long-term environmental characteristics are not expected to fluctuate over time, as this study aims to isolate a component of total vulnerability that is less likely to fluctuate.

Vulnerability was defined categorically, and then predictors were fit to these categories among municipalities where YF was observed. These results were then extended to municipalities where YF had not been observed in order to produce vulnerability estimates by municipality nationwide. The results of this study provide estimates of long-term risk for YF burden based on environmental characteristics, which allow locations within Brazil that most closely resemble locations with greatest disease burden to be identified and targeted for public health interventions.

4.2. Methods and materials

4.2.1. Study setting

The study region is Brazil, the largest nation in South America both by land area and by population. The nation is subdivided into 5,570 municipalities [144], which provide the spatial units of analysis. Data were divided into two time periods due to an outbreak that began in December 2016 [58]; this accounts for potential changes in surveillance or disease determinants during an outbreak. The first time period, representing vulnerability to YF outside of an outbreak, spanned between 2000 and 2016. The second time period, representing vulnerability to YF during and immediately following an outbreak, spanned 2017. Figure 4.1 shows the trend of yearly YF cases during the entire time period. Monthly YF case counts are also shown in the two separate time periods to show seasonal trends of YF burden. Throughout the remainder of this study, the earlier time period, 2000 – 2017, will be referred to as the "endemic" period, and the later time period, 2017, will be referred to as the "epidemic" period. These labels indicate the motivation behind the separating the time periods and describe the difference in YF dynamics between the two.

In this study, vulnerability to YF was defined using yearly incidence, defined as the number of cases per 100,000 population in a year for each municipality. Data were analyzed by municipality-year. Yearly incidence can be influenced both by seeing cases frequently and by seeing high numbers of concurrent cases over the course of a year because both situations can lead to high case counts in a year, but seeing yearly incidence on its own does not distinguish between the two. Vulnerability was defined in three categories: minimal vulnerability (denoted category 0), defined by observing zero cases per year; moderate vulnerability (denoted category 1), defined by observing up to 10 cases per 100,000 population annually; and high vulnerability(denoted category 2), defined by observing more than 10 cases per

100,000 population annually (Table 4.1). Using 10 cases per 100,000 population provided an intuitive, interpretable cutoff and approximately led to similar sample sizes for municipalities with moderate and higher vulnerability.

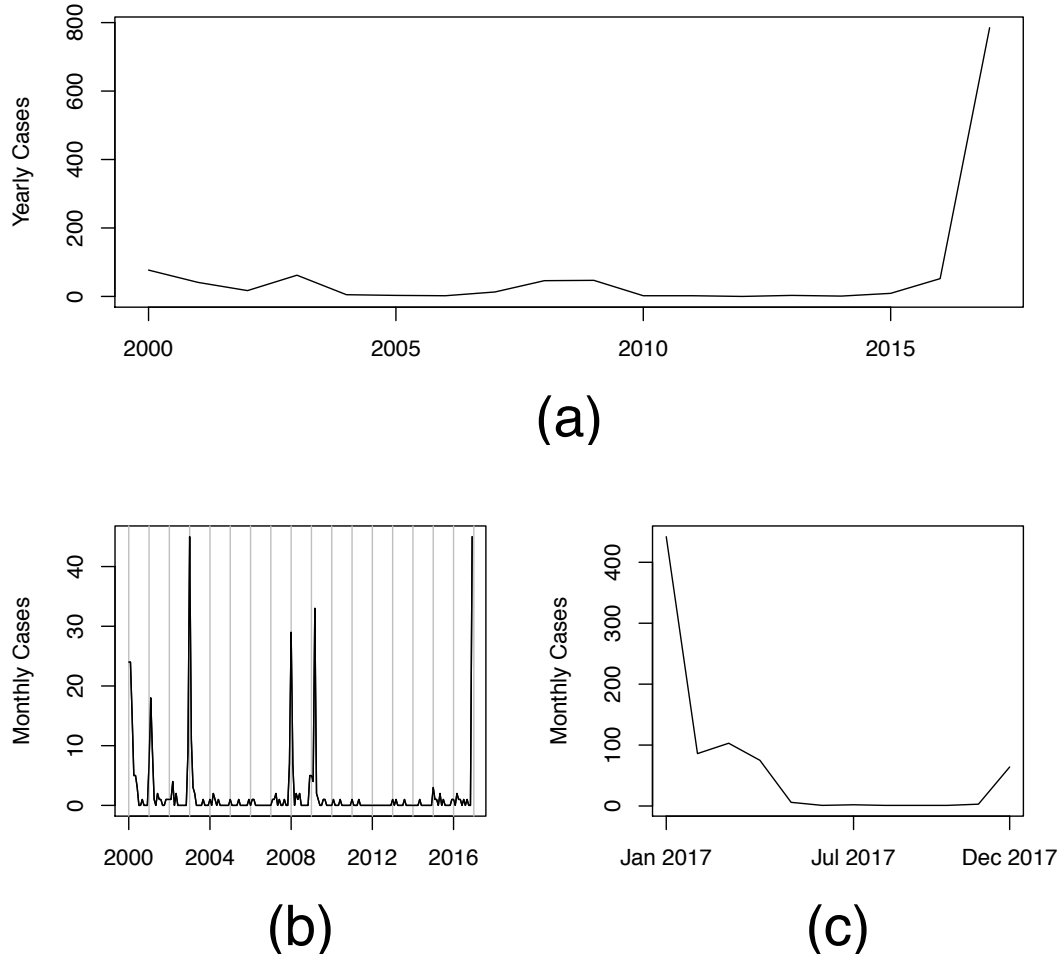


Figure 4.1. Time series of total Yellow Fever cases in Brazil. (a) Yearly cases, 2000 – 2017. (b) Monthly cases, 2000 – 2016 (endemic period). (c) Monthly cases, 2017 (epidemic period). Monthly cases are shown to show heterogeneity in cases throughout a year.

Table 4.1. Vulnerability categories among municipalities where Yellow Fever cases were observed in Brazil, 2000 – 2017.

Category number	Category name	Incidence	Number of municipality-years, 2000-2016	Number of municipalities, 2017
0	Minimal Vulnerability	0 cases per 100,000 population	7727	304
1	Moderate Vulnerability	>0-10 cases per 100,000 population	109	75
2	Higher Vulnerability	> 10 cases per 100,000 population	86	87

4.2.2. Data sources

4.2.2.1. Yellow Fever incidence data

Yellow Fever case data were provided by the Pan American Health Organization and the Brazilian Ministries of Health. Case data were collected through passive surveillance activities, with healthcare providers reporting confirmed cases to the Ministries of Health. For each case, the date of report and the municipality associated with the case are included. Case data are available between January 2000 and March 2018. Case data through 2017 were used in analysis to include only complete years.

Population data for Brazilian municipalities are publically available through the Brazilian Institute for Geography and Statistics [144]. Annual population estimates are available for all years between 2000 and 2017, with the exceptions of 2007 and 2010. For these two years, the arithmetic mean of the two adjacent years was used

to singly impute missing values. Using population data and case data, annual incidence values per 100,000 population were calculated for each municipality-year. Incidence values were converted into vulnerability categories defined in Table 4.1.

4.2.2.2. Environmental predictors

Environmental variables were included with the intention of predicting vulnerability. Selection was based on representativeness of YF determinants that both are unable to be targeted by immediate intervention and represent characteristics of the surrounding environment. The aim of this study was to provide vulnerability estimates that are based on characteristics that are less likely to change, leading to more permanent estimates of this particular set of contributing factors. In isolating the group of predictors of YF vulnerability describing the surrounding environment that cannot be targeted for intervention, the vulnerability estimates produced in this study represent a part of total YF vulnerability, one that is unable to immediately controlled by human activities for disease prevention.

Shapefiles for Brazilian municipalities were available from the Database of Global Administrative Areas version 3.6 [233]. The shapefiles, which are current as of 2019, contain the land area in square kilometers for each municipality. Using this and the population data described previously, population density per square kilometer was calculated for each municipality. Population density can represent urban form, as developed urban locations are more likely to be population dense.

Precipitation has been previously linked to risk of YF [86, 92, 114, 234]. Precipitation data are publically available through the Modern-Era Retrospective analysis for Research and Applications, version 2, through the United States National Aeronautics and Space Administration [101]. Rainfall data are available in 0.5-degree by 0.625-degree grids, represented as millimeters of precipitation per hour [101], and are available for the entire time period between January 1, 2000

and December 31, 2017. These data were aggregated to represent the percentage of days for each municipality each year that experienced any rainfall. Many mosquito species can lay eggs in small amounts of water [235], so using the relative frequency of opportunities for standing water to exist is of interest for YF vulnerability. This variable can represent the frequency of opportunities for mosquito breeding grounds to exist. Grids were matched to municipalities using ArcGIS Pro version 2.2.0 [236] using a spatial join, allowing the percentage of days with rainfall per year to be measured per municipality.

Average elevation in meters, which has also been shown to relate to YF risk [54], is publically available through the Advanced Spaceborne Thermal Emission and Reflection Radiometer Digital Elevation Model from the United States National Aeronautics and Space Administration and Japan's Ministry of Economy, Trade, and Industry [104]. These data are available in 30-meter by 30-meter grids. Elevation grids were matched to Brazilian municipalities using a spatial join in ArcGIS Pro, allowing the average elevation for each municipality to be measured.

Standing water provides opportunities for mosquito breeding [235], and has been shown to relate to YF or other mosquito-borne viruses through river drainage [237], storm drains [238], and irrigation systems [239]. Drainage density, defined as the ratio of the length of rivers and streams to land area, was modeled for Brazil accounting for elevation and known hydrological features in the area. This represents relative amounts of water movement through municipalities, which can also create areas of standing water to serve as mosquito breeding grounds. Drainage density was estimated for water basins, where basins are defined to have a minimum area of 1000 km^2. The average drainage density, weighted by proportion of land overlap between basins and municipalities, was computed for municipalities using a spatial join in ArcGIS Pro. Drainage density is also assumed to remain constant over the entire study period.

Ecological habitat types, which have been related to YF risk [54], are publically available from the World Wildlife Foundation [240]. These habitat types are referred to herein as ecoregions, and can represent access to forested areas. A total of seven ecoregions, reflecting ecohydrological and geomorphological factors, are present in Brazil: Deserts and xeric shrublands; Flooded grasslands and savannas; Mangroves; Montane grasslands and shrublands; Tropical and subtropical dry broadleaf forests; Tropical and subtropical grasslands, savannas, and shrublands; and Tropical and subtropical moist broadleaf forests [240]. These habitats are assumed to remain constant throughout the entire study period, and were matched to municipalities to assign to each municipality the habitat with the highest percentage of land cover. Due to positivity concerns, and for parsimony and interpretability of results, ecoregion was treated as a binary indicator variable for a municipality primarily belonging to tropical and subtropical dry broadleaf forests or tropical and subtropical grasslands, savannas, and shrublands.

Another factor considered in assessing a location's vulnerability to YF is occurrence of the disease in nearby locations, which can account for movements and disease transmission among humans, non-human primates, and mosquitoes. This was included in analyses by assigning each municipality a list of neighboring municipalities. Two definitions of neighbors were considered for each municipality: neighbor by border and neighbor by radius. For any given municipality, neighbor by border was the count of all other municipalities sharing its land border. Similarly, neighbor by radius counted the number of other municipalities with a centroid located within a 50-kilometer radius from a given municipality's centroid. This radius is wider than typical flight radii of mosquitoes found in Brazil [230], but also considers expanded human movement. For each definition, a binary variable was created to indicate whether at least one of the neighboring municipalities saw at least one YF case. Due to differences in sizes of municipalities and differing numbers of neighbors, combining all neighbors into one unit and using a binary indicator removes the issue of different denominators for calculating a percentage or count.

4.2.3. Data analysis

In order to reduce uncertainty of the epidemiologically silent municipalities regarding a true absence of cases or undetected cases, only the 466 municipalities where any YF cases were reported in the available data were included in analysis. These represent the municipalities where at least one case was seen in the YF data set between January 2000 and March 2018 (Figure 4.2). While January – March 2018 data were present, this time period was not included in analysis to only use data from full calendar years, but municipalities observing YF cases during this time period were included in analysis under the assumption that these municipalities would have detected cases that would have occurred prior.

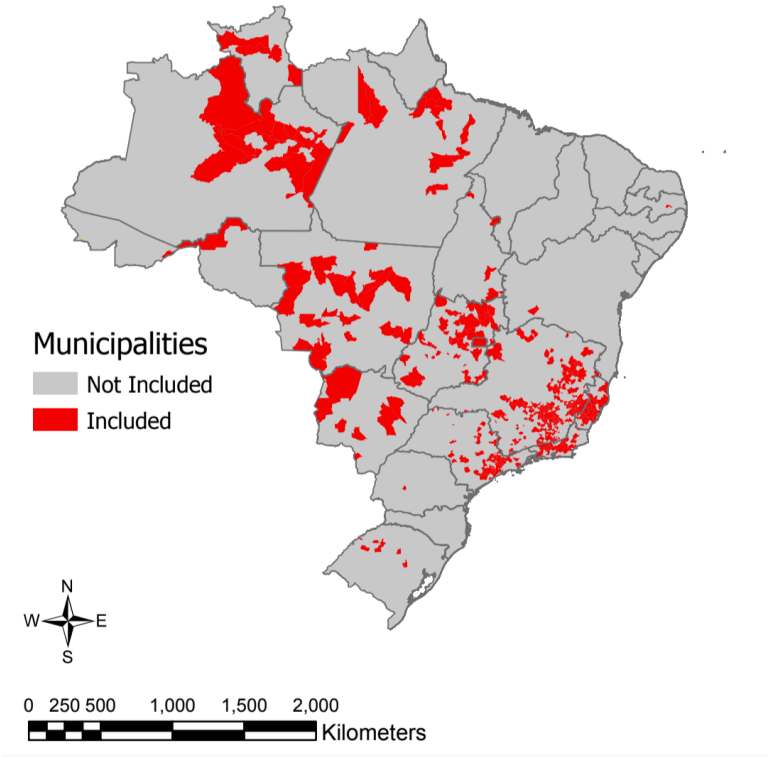


Figure 4.2. Municipalities of Brazil where any Yellow Fever cases were seen between Jan 2000 and Mar 2018. The highlighted municipalities' vulnerability data were used in model fitting for both the endemic and epidemic time periods.

In the endemic time period described previously (2000 – 2016), the 466 municipalities contribute 17 years of observation. Because the data are analyzed by municipality-year, the same municipality can be assigned different vulnerability categories in different years, which leads to the inclusion of all three categories in the data set. Only including municipality-years that observed any YF would, by design, omit any observations in the minimal vulnerability category because it is defined by having zero cases.

During the epidemic time period (2017), the 466 municipalities contribute one year of observation. Municipalities seeing YF cases prior to 2017 and in the first three months of 2018 are included to represent observations in the minimal vulnerability category. YF incidence or vulnerability measured in 2018 were not included because the entire year is not observed in the data. While the data from these months contribute to the analyses in determining which municipalities are included, case and vulnerability data from these months are not included for model fitting.

For both time periods, including only the municipalities where YF was seen at least once increases confidence that absence of cases represents a true absence of cases rather than an artifact of epidemiologic silence. This assumes a constant ability to detect cases in these municipalities during the time period of observation.

The outcome representing vulnerability to YF (minimal, moderate, and higher vulnerability) by municipality-year was fit to a cumulative logit regression model to produce estimated probabilities associated with the three ordinal categories. Two separate models were developed for the two time periods described previously: the endemic period (2000 - 2016) and the epidemic period (2017). Both analyses were conducted separately over the 466 municipalities described previously.

Potential predictors for the model include the previously described environmental, landscape, and ecological data. The predictors considered include percentage of days with rain (continuous, linear and quadratic), ecoregion (binary indicator for tropical and subtropical moist broadleaf forests or tropical and subtropical grasslands, savannas, and shrublands), population density (continuous), elevation (continuous), and YF occurrence in neighboring municipalities (binary, based on both definitions separately). Because of the study design and aims of analyses, the model parameters are selected into a parameter set providing most accurate predictions, so the terms of the cumulative logit models are not designed for explicit causal interpretation. They instead represent predictive effects to be considered with the other predictors rather than isolated, unconfounded, causal effects. The aim was not to elucidate causal mechanisms, but to accurately predict YF vulnerability based on unchanging environmental features.

The aim of model fitting was to find the set of predictors that produce the most accurate fits of vulnerability. The previously mentioned predictors are candidate predictors for the model, with inclusion reliant on improvement of predictive ability. The best predictor set for the cumulative logit model was selected based on the most accurate predictions of vulnerability category, representing the closest fit to the observed data. A total of 143 parameter combinations were considered, which includes all combinations of inclusion and exclusion of the predictors described previously, omitting combinations that include the quadratic term for percentage of rainy days without including the linear term as well as combinations that include both indicator variables for YF presence in neighboring municipalities.

Closest fit was determined based on a model fit score. The parameter set providing the optimal score was considered to produce the most accurate predictions of vulnerability. Fitted values from the cumulative logit model produce estimated probabilities for each municipality to belong to each of the three vulnerability categories. For each municipality-year, the fitted probability associated with the

true, observed vulnerability category was identified, and these values were averaged among all three vulnerability groups. For example, among all municipality-years assigned to the minimal vulnerability category, the average of those municipalities' fitted probabilities for minimal vulnerability was calculated to determine the model's ability to predict minimal vulnerability. Repeating this process for the other two categories provides similar evaluations of fit for moderate and higher vulnerability.

The sum of the three averages for each vulnerability category produced a model score. This score places equal weight on ability to predict from each of the three vulnerability categories, which is designed to improve the fit for the moderate and higher vulnerability categories. Because most of the observations observe an incidence of zero in some years, models placing high probability on the category of zero vulnerability would be favored, regardless of ability to predict higher incidence. Models were run using the 'MASS' package [241] in R version 3.6.0 [150]. Multicollinearity was considered through the Variance Inflation Factor (VIF), which uses the 'car' package [242] in R, to determine if the model parameter estimates are affected by comparing the estimated variance of the full model to the variance of a model using each variable individually.

4.2.4. Estimating vulnerability

The final models were used to predict vulnerability across all municipalities. The predictors were measured for all municipalities, so the model could be applied to generate predicted values in the form of probabilities associated with belonging to each of the three vulnerability categories. This was done using the models from both time periods, showing how vulnerability estimates may differ during an outbreak period.

4.3. Results

Tables 4.2 and 4.3 show descriptive statistics of all predictor variables considered in analysis. All variables in these tables were eligible for inclusion into the model for either time period, with the caveat that only one of the two definitions for neighboring municipality, if any, can be included.

Table 4.2. Descriptive statistics of predictor variables for Yellow Fever vulnerability models during the endemic time period. Mean and standard deviation values are shown for continuous predictors, and counts of positive values with percentages are shown for binary variables.

Predictor	mean n	sd %	min	25%	50%	75%	max
Ecoregion	7280	98.7	0	1	1	1	1
Cases 50 km	599	7.6	0	0	0	0	1
Cases border	269	3.4	0	0	0	0	1
% Rain days	49.3	12.7	22.4	42.1	47.3	52.6	100
Drainage	0.177	0.034	0.117	0.159	0.174	0.185	0.663
Elevation	401.223	179.067	7.949	284.361	433.972	526.545	906.159
Population density	9.764	36.434	0.007	0.53	1.269	3.286	431.637

Table 4.3. Descriptive statistics of predictor variables for Yellow Fever vulnerability models for the epidemic time period. Mean and standard deviation values are shown for continuous predictors, and counts of positive values with percentages are shown for binary variables.

Predictor	mean n	sd %	min	25%	50%	75%	max
Ecoregion	460	98.7	0	1	1	1	1
Cases 50 km	342	73.4	0	0	0	1	1
Cases border	216	46.4	0	0	1	1	1
% Rain days	44.3	12.7	22.7	36.4	43.4	46.8	98.6
Drainage	0.177	0.034	0.117	0.159	0.174	0.185	0.663
Elevation	401.223	179.067	7.949	284.361	433.972	526.545	906.159
Population density	10.748	39.546	0.009	0.551	1.378	3.516	433.765

4.3.1. Model fit and validation

The model representing the endemic time period (2000 - 2016) found that the best fitting predictor set consisted of percentage of days with rain as a linear and quadratic term, drainage density, population density, elevation, and case presence in municipalities within 50 kilometers. The presence of these variables underline the importance of ecohydrological, geomorphological and population factors in YF dynamics. The model fit score was 1.35, with 48 of 142 other models having a fit score of at least 90% of this score. The minimum score was 1.00. When excluding the quadratic term for precipitation, the highest VIF value among predictors was 1.202, suggesting that the parameter estimates are unlikely to suffer from multicollinearity. When including the quadratic term, only the two precipitation parameters showed a high VIF, as they are correlated by design.

The model representing the epidemic time period (2017) found that that best fitting predictor set consisted of percentage of days with rain as a linear and quadratic term, ecological habitat, drainage density, population density, elevation, and case presence in bordering municipalities. The model fit score for this model was 1.37, with 80 of 142 other models having a fit score of at least 90% of this score. The minimum score was 1.00. Excluding the quadratic term for precipitation, the maximum VIF value was 1.394, indicating little impact from multicollinearity. When including the quadratic term, only the two precipitation parameters showed a high VIF, as they are correlated by design.

Table 4.4. Parameter estimates for vulnerability for Yellow Fever in Brazil in the endemic and epidemic time periods. Empty cells indicate that a predictor was not included in the best-fitting model.

Predictor	Estimate (SE), endemic period	Estimate (SE), epidemic period
Ecoregion	-0.117 (0.619)	1.132 (1.312)
Pct days with rain	-4.263 (3.608)	-19.71 (5.864)
Pct days with rain^2	0.873 (2.840)	11.830 (5.583)
Population density	-10.451 (9.681)	-6.773 (3.608)
Drainage density	-9.296 (4.193)	-12.715 (4.858)
Elevation	-2.429 (0.747)	-1.720 (0.699)
YF in neighboring mun (border)	--	2.217 (0.250)
YF in neighboring mun (50 km)	22.297 (0.812)	--
Int 0|1	18.404 (0.812)	-6.161 (2.106)
Int 1|2	19.499 (0.811)	-4.931 (2.093)

Parameter estimates in Table 4.4 show the log odds ratio of a municipality being in a higher fitted vulnerability category. The fitted model for the endemic time period shows a parabolic trend between the percentage of days with precipitation and the log odds ratio associated with a higher category with a minimum value outside the range of the percentage of days with precipitation, producing a monotonic relationship. Municipalities with lower population density, lower drainage density, lower elevation, and not located in tropical and subtropical moist broadleaf forests or tropical and subtropical grasslands, savannas, and shrublands ecoregions are also associated with higher probabilities of higher vulnerability categories. Municipalities with any YF occurrence in municipalities within 50 kilometers also have a higher log odds ratio associated with a higher vulnerability category.

The model for the epidemic time period shows a similarly shaped parabolic trend between the percentage of days with precipitation and log odds ratio of observing a higher vulnerability category. The minimum log odds ratio is seen in this model when 83.3% of days in a year see precipitation. Municipalities with lower population density, lower drainage density, and lower elevation are also associated

with higher probabilities of higher vulnerability categories. Municipalities located in tropical and subtropical moist broadleaf forests or tropical and subtropical grasslands, savannas, and shrublands ecoregions and with any YF occurrence in bordering municipalities also have higher log odds ratio associated with a higher vulnerability category.

4.3.2. Estimated vulnerability

Fitted probabilities of vulnerability categories among municipalities are shown in Figures 4.3, 4.4, and 4.5. During the endemic period, locations with the greatest probability of having highest vulnerability were located primarily in the western part of Brazil, including areas near and within the Amazon rainforest. Municipalities with the greatest probability of moderate vulnerability are located in the same areas, largely including the same or nearby municipalities to those with the greatest probability for higher vulnerability. The municipalities with the greatest predicted probability of minimal vulnerability, conversely, are primarily located in southern and northeastern regions of the country.

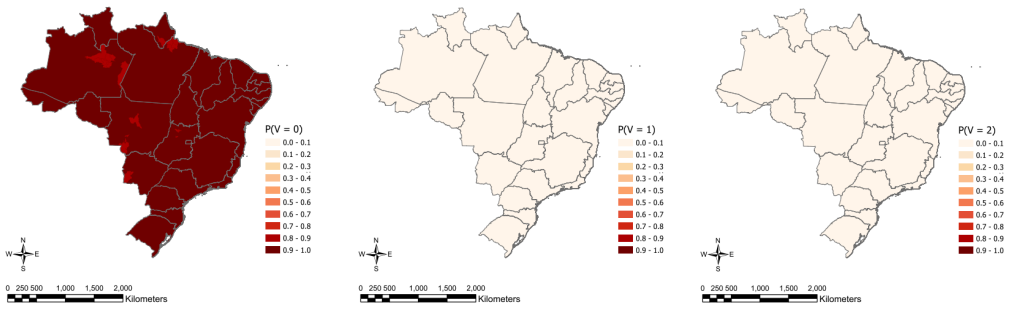


Figure 4.3. Estimated probabilities associated with (a) minimal, (b) moderate, and (c) higher Yellow Fever vulnerability among municipalities in Brazil for the endemic time period.

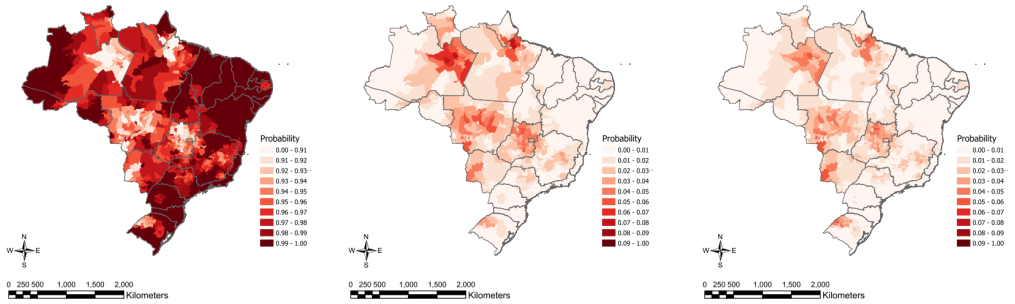


Figure 4.4. Estimated probabilities associated with (a) minimal, (b) moderate, and (c) higher Yellow Fever vulnerability among municipalities in Brazil for the endemic time period. Scales are adjusted in this figure to show heterogeneities across municipalities.

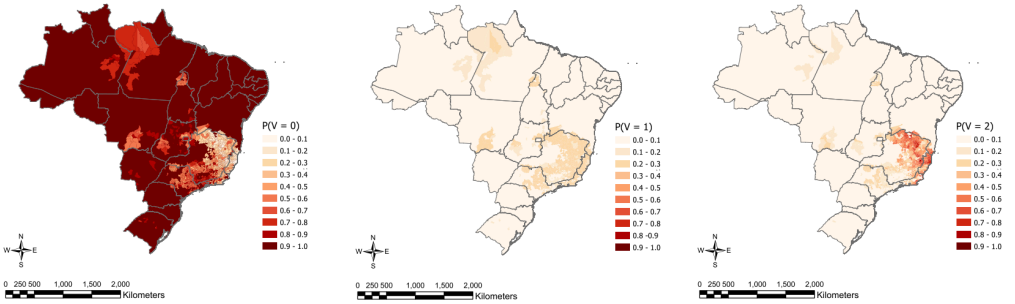


Figure 4.5. Estimated probabilities associated with (a) minimal, (b) moderate, and (c) higher Yellow Fever vulnerability among municipalities in Brazil for the epidemic time period.

In the epidemic period, the locations of high vulnerability were largely seen in the southeastern states where the outbreak was primarily concentrated. Particularly in this time period, municipalities surrounding those with observed cases, were most likely to be in the highest vulnerability category. Municipalities with greatest

probability for moderate vulnerability are also concentrated in the southeastern part of Brazil as well as isolated locations in Mato Grosso, Goiás, Tocantins, and Pará states. Municipalities with greatest probability associated with minimal vulnerability are primarily found in southern, central, northeastern, and western areas of Brazil.

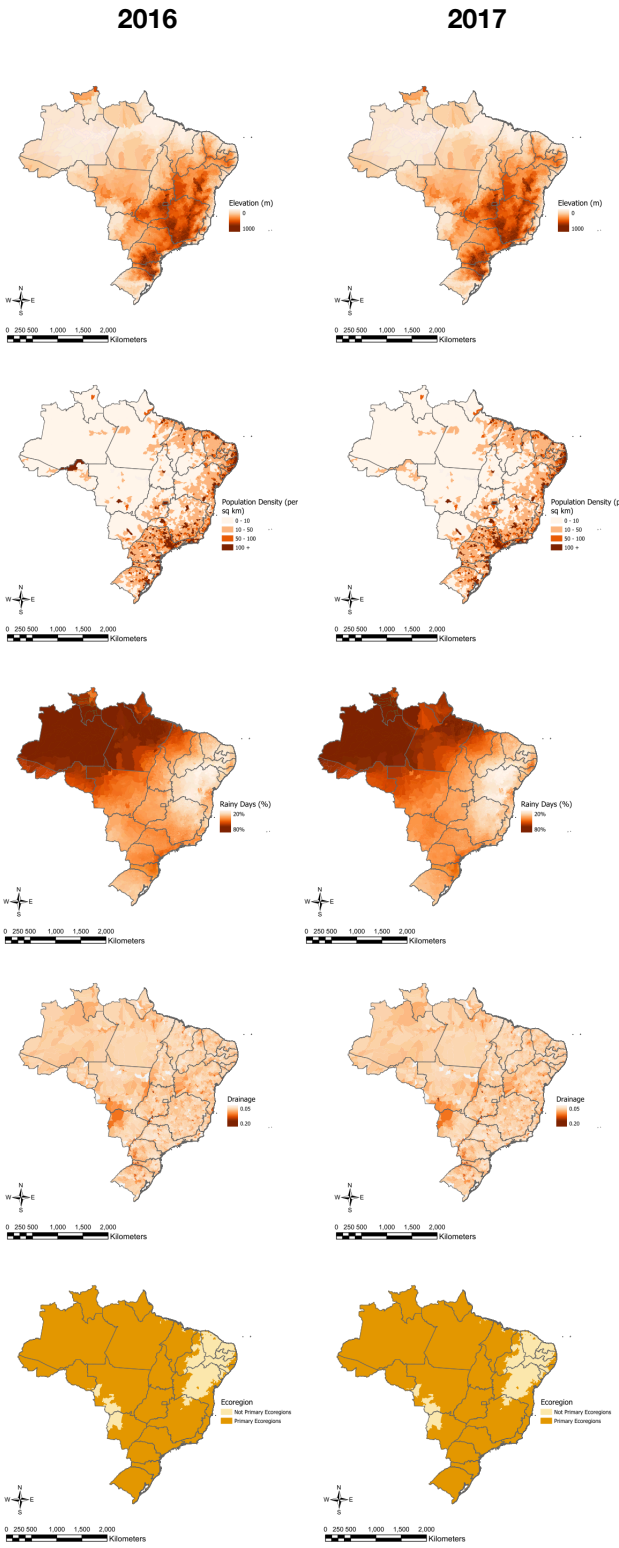


Figure 4.6. Environmental predictors contributing to Yellow Fever vulnerability models, in 2016 and 2017.

4.3.3. Sensitivity analyses for critical threshold and predictor influence

A sensitivity analysis was conducted to determine if model results are sensitive to the value of incidence used as a cut point between the moderate and higher vulnerability categories. Using the same predictor sets, the two cumulative logit models were re-run using each integer between one and 35 as the cut point between the two categories. The model fit score used for determining the best parameter set was used for comparing model fit using the different cut points, as shown in Figure 4.7. Results for the endemic time period show that cut points between nine and 12 all show nearly equal maximum model fit scores. Results from the epidemic tie period show that cutoffs up to 19 all lie within a wide range of models with better fit. The results of this sensitivity analysis indicate that keeping a cutoff of 10, which is an intuitive and easily interpretable value, is an appropriate choice for these analyses. This also reveals that values within these ranges can serve as a critical threshold for discriminating patterns whose dynamics (also in terms of environmental drivers) is very different.

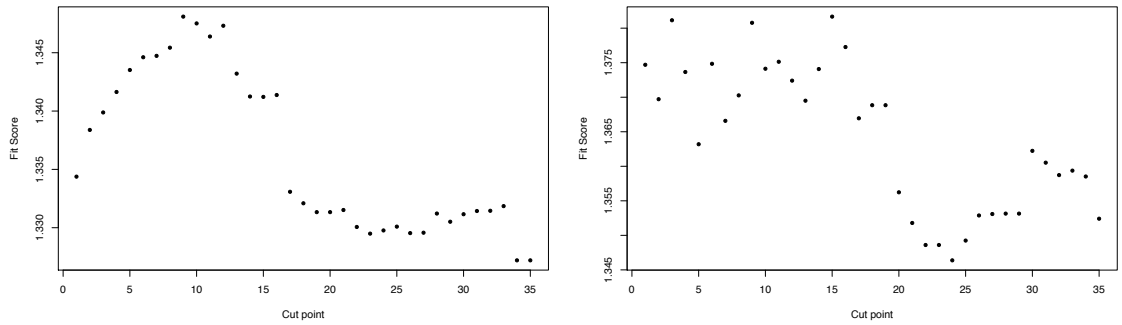


Figure 4.7. Model fit score of best-fitting model under each of the proposed cut points for (a) the endemic time period and (b) the epidemic time period

A global sensitivity analysis [243, 244] was conducted to determine if the parameter set was primarily driven by any single input variables and for assessing variable interdependence. Sobol indices [245, 246]were adapted to the cumulative logit model to determine the most influential variables contributing to the probabilities associated with being in each of the three vulnerability categories. The results of these analyses indicate if any predictors have greatest relative influence on vulnerability to YF compared to the others.

The adapted first order Sobol indices were generated for each variable through 1,000 repetitions of the following steps:

1. Assign the variable a constant value drawn uniformly between the minimum and maximum observed value.
2. With the variable held constant, calculate fitted probabilities for each vulnerability category from the fitted model
3. Calculate the mean of all fitted probabilities for each category.

The estimated Sobol index was calculated from these means for each vulnerability category by calculating their variance and then standardizing by dividing by their sum. Second-order indices were calculated using the same method, but holding each variable constant along with each of the other variables, one at a time. Second-order indices for each predictor are then summed together, which include first-order indices.

The first-order indices show the most influential variables by themselves, and the second-order indices do this considering potential interactions between pairs. For the endemic time period, the indicator variable for cases within 50 km was the most influential parameter through both first and second order indices for all three vulnerability categories (Table 4.5, Figure 4.8) For the epidemic time period, percentage of days with rain was most influential (Table 4.6, Figure 4.8). This indicates that precipitation variability is indeed a critical variable also affecting all

other landscape features such as drainage density that is indeed physically varying when precipitation changes with implications for the ecological habitat.

Table 4.5. Adapted Sobol indices for each vulnerability category in the endemic time period. Normalized first-order and second-order indices are presented. Numeric categories of vulnerability are defined in Table 4.1.

	P(V = 0)		P(V = 1)		P(V = 2)	
Variable	First-order	First-order + Second-order	First-order	First-order + Second-order	First-order	First-order + Second-order
Ecoregion	< 0.01	0.34	< 0.01	0.16	< 0.01	0.19
% Days with rain	0.11	0.39	0.02	0.24	0.11	0.32
% Days with rain^2	0.01	0.27	< 0.01	0.35	0.01	0.37
Population density	0.01	0.01	0.01	0.02	0.01	0.02
Drainage density	0.04	0.04	0.03	0.04	0.04	0.07
Elevation	0.04	0.06	0.01	0.04	0.04	0.06
Cases within 50 km	0.78	0.66	0.92	0.92	0.78	0.77

Table 4.6. Adapted Sobol indices for each vulnerability category in the epidemic time period. Normalized first-order and second-order indices are presented. Numeric categories of vulnerability are defined in Table 4.1.

	P(V = 0)		P(V = 1)		P(V = 2)	
Variable	First-order	First-order + Second-order	First-order	First-order + Second-order	First-order	First-order + Second-order
Ecoregion	< 0.01	0.24	0.01	0.20	0.01	0.24
% Days with rain	0.19	0.36	0.35	0.39	0.33	0.42
% Days with rain^2	0.79	0.79	0.18	0.29	0.55	0.67
Population density	< 0.01	0.09	0.09	0.17	0.02	0.11
Drainage density	< 0.01	0.10	0.22	0.28	0.04	0.14
Elevation	0.01	0.13	0.01	0.21	0.02	0.14
Cases in bordering mun	0.01	0.16	0.13	0.28	0.04	0.16

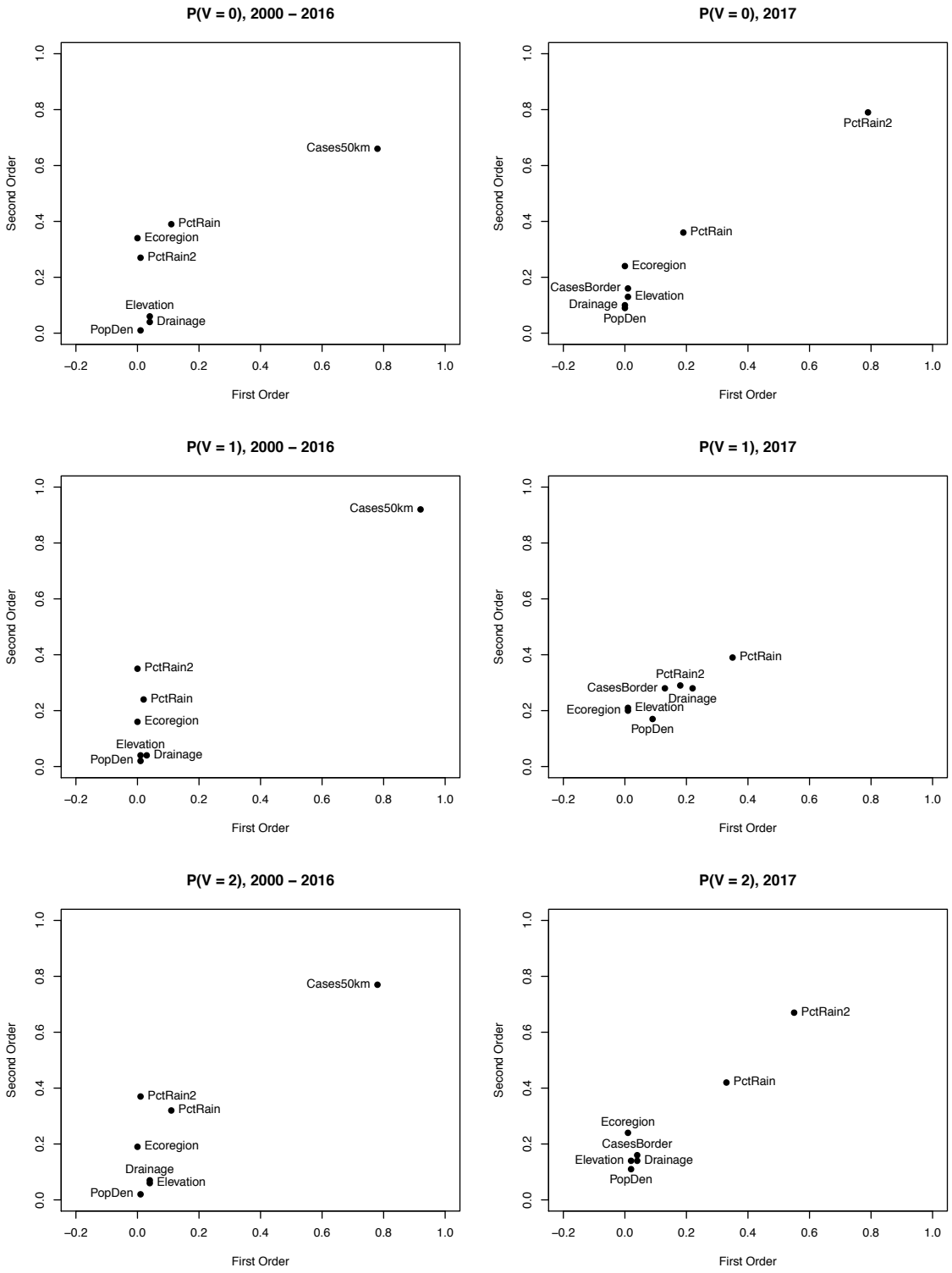


Figure 4.8. Sobol indices for all three estimated probabilities from the cumulative logit model for Yellow Fever vulnerability under both the endemic and epidemic time periods.

4.4. Discussion

This study aimed to identify municipalities of Brazil with the highest vulnerability to long-term burden of YF, defined by categories of yearly incidence. Best-fitting models were determined using long-term environmental factors, which represent characteristics of the surrounding environment that cannot be targets for immediate intervention and are not expected to fluctuate over the course of a year. These predictors represent a substantial component of total vulnerability. The models predict vulnerability based on information from municipalities where YF had been seen, and model results were then also applied to municipalities where YF had not previously been seen to produce estimates of vulnerability, by municipality, across all of Brazil. Several environmental factors, largely influenced by yearly precipitation patterns or observing YF cases in nearby municipalities, provided accurate fits. Predictions from the model show, using the model for the endemic time period (2000 – 2016), that western municipalities comprise most of those with highest risk for high vulnerability. The model for the epidemic time period (2017) shows that southeastern municipalities comprise most of those with highest risk for high vulnerability. The differences in results produced by the two models shows that vulnerability estimates can be notably affected by an ongoing outbreak.

The predictor set that best fit vulnerability in the endemic time period was nearly identical to that for the epidemic time period, suggesting that the set of parameters that best fit vulnerability outside of a large outbreak is very similar to that for fitting vulnerability during a large outbreak. The fitted model estimates and predicted locations that were considered most vulnerable, however, were different between the two periods (Table 4.4, Figures 4.3-4.5). This suggests that the presence of an outbreak may lead to different effects of risk factors for YF. While many of the same predictors contribute to accurate predictions, some human behavioral mechanisms

may change during an outbreak [247] which may in turn indirectly affect parameter estimates.

Occurrence of YF in nearby locations was an influential parameter in the predictor set for vulnerability during the endemic time period. This finding suggests that there is a spatial relationship in YF transmission. This is supported in previous work [248] and by the data; visual examination of the locations where YF was seen (Figure 4.2) shows that municipalities where cases were seen tend to be near each other. This may emphasize the importance of local human activities on YF dynamics in Brazil or result from nearby locations having similar climate conditions (Figure 4.6).

Previous works addressing vulnerability to infectious diseases assess vulnerability in terms of socioeconomic factors or across larger spatial units such as states or nations. In these situations, vulnerability has been defined in terms of history of infection [226] or through the connection between underserved populations and disease risk [249, 250]. This study differs by focusing on environmental factors in assessing vulnerability rather than human characteristics. These represent different components of total vulnerability, as all of these factors contribute. In this study, the focus remained on environmental characteristics that cannot be targeted for intervention; other relevant determinants such as healthcare infrastructure, vaccination, or mosquito control [40] can undergo interventions to control YF vulnerability over the course of a year.

The resulting models in this study were applied to the epidemiologically silent municipalities in order to assess their own vulnerabilities through these predictors used. This provided a nationwide examination of YF vulnerability to allow most vulnerable municipalities to be identified. Similar extensions of results have been done by other studies by applying model results to locations outside of the data set used [93], while others avoided extrapolating and applied results only to locations providing data [94]. Having nationwide vulnerability estimates are beneficial

because knowledge of the risks of high YF burden among currently epidemiologically silent municipalities may lead to preparedness measures that apply to future YF occurrence.

For YF and other mosquito-borne diseases, previous studies have considered various environmental determinants and predictors of incidence. Many of the environmental factors associated with risk of mosquito-borne diseases included temperature [251, 252], precipitation [253, 254], elevation [255], ecoregions [54], and vegetation [256]. Many of these or similar variables were included in this study. The results of this study are consistent with many of these findings in that long-term environmental factors found to relate to YF burden were similar to those identified from the previous studies.

This study explicitly uses environmental predictors for the purpose of long-term vulnerability to YF, defined in this study as ability to see high incidence over a year. While this is not the only way to determine vulnerability, as evidenced by the other studies using other determinants [226, 249, 250], this vulnerability measurement is beneficial for predicting locations that may experience greater disease burden based on factors which cannot be changed by immediate human activities. Since the environmental factors considered in this study cannot be targets for intervention, these estimates can serve to indicate which locations are most vulnerable to high YF burden based on unchangeable characteristics, indicating which locations most need other preparedness measures to intervene on vulnerability determinants that can be targeted for intervention.

The included predictors represent a subset of all predictors of yearly YF burden, which is used to represent vulnerability in this study. Other predictors not included in this study, including socioeconomics, urban form, or forest vegetation, can frequently be changed through human activities. Some were estimated through other variables used, such as using population density as a proxy for urban form and

ecoregions for forest exposure. By excluding those from analysis, vulnerability in this study represents the component of total vulnerability that is less likely to change over time.

4.4.1. Implications for preparedness

The results of this study identify municipalities that are expected to be more likely to experience high burden of YF, making them the most beneficial targets for prevention strategies for YF. Municipalities with the highest predicted probabilities for the highest vulnerability category are those that are most likely to experience high burden, whether by high case counts at once or by frequent occurrences of cases. In either situation, preventative measures would be most important in these municipalities to prevent YF infection. Such potential strategies can include vaccination campaigns [73], early warning systems [257, 258], or vector control strategies [259, 260], all of which have been employed in the past. One of the most notable warning signs for YF would be that it has been observed nearby, indicating that having seen YF infections in nearby municipalities should be considered an indication of a need for preventative measures.

Applying the model from each time period to inform preparedness currently would yield different priorities for prevention measures. Early reports from the 2018-2019 season of YF (defined as December through May) indicate that this season is likely to see fewer cases compared to the 2016-2017 and 2017-2018 seasons [261]. While YF dynamics may currently more closely resemble those from 2000 – 2016, lingering effects of the outbreak may also influence current vulnerability. Considering estimates from both models in tandem or obtaining current data to update the model would be recommended for applying the results of this study to practice.

Estimating vulnerability in the epidemiologically silent municipalities shows where high burden of YF can be seen, even if it has not been seen previously. Many possibilities exist for why these municipalities had not seen YF. Differences in reporting rates of YF have been noted previously [93], posing the possibility of these locations having undetected cases [116, 120]. Additionally, YF has remained rare in Brazil prior to the recent major outbreak, so epidemiologically silent municipalities may be so by chance. Fitted probabilities of having higher vulnerability remain low among most municipalities, suggesting that during the years of observation, YF may not have been present in the epidemiologically silent municipalities. There may also be additional factors not included in this study's model, such as human intervention [73, 106], that actually would prevent YF from being seen. In such a situation, these municipalities would see zero cases as an actual instance of zero cases and not be vulnerable.

4.4.2. Strengths and novelty

This study offers insight into YF vulnerability at a high spatial resolution across the entire nation of Brazil. While many other studies compare nations for their disease risks [228, 262], this study identifies the municipalities within one country with characteristics that make them appear at greater vulnerability to YF burden. Other studies using small areal units for similar purposes commonly focus on a more localized region [263–265] rather than a large nation.

By splitting the analyses into endemic and epidemic time periods based on the start of the 2016-2017 outbreak, this study shows how relative disease vulnerability can differ between endemic and epidemic time periods. Outside of the outbreak, municipalities found to be most likely to be at the highest vulnerability category were primarily located in the western areas of Brazil, which is consistent with previous vaccination recommendations [62]. Following the start of the outbreak, vulnerability was seen most in southeastern states, similarly to the outbreak

locations. This is likely due to the high ability for observation of YF in neighboring municipalities to predict vulnerability categories.

The use of a global sensitivity analysis for identifying influential model predictors is beneficial for contextualizing results and determining if any particular predictors among the predictor set are of particular importance in relating to YF vulnerability. Showing the strongest predictors can offer insight into most effective preventative measures. For example, the model in the earlier time period showed that observed cases in nearby municipalities was the most influential predictor. Preventative measures based on this particular predictor, such as early warning systems indicating nearby cases, may prove most effective.

4.4.3. Limitations and future directions

This study relies on the accuracy of the collected data when producing the vulnerability estimates. The data for YF cases were collected through passive surveillance activities, and inaccurate data collection could impact results. Only including municipalities where YF cases were seen aimed to reduce inaccuracies due to underreporting of cases. This assumes that, during the entire time period, all included municipalities were equally capable of detecting cases across all years.

This study is also limited by interrelationships among predictors. The purpose of the modeling was to provide accurate predictions rather than isolate individual effects, which leads to model parameters that, while in concert can accurately predict vulnerability to YF, are less likely to have reliable individual interpretations. Though there was little evidence of multicollinearity, selection of predictors did not consider how they might influence each other through interactive or confounding effects. However, the purpose of this study was to develop accurate predictions, not to estimate individual, unconfounded causal effects. The study was not designed for such interpretations of coefficients.

In focusing on the component of total vulnerability based on environmental predictors, other relevant factors contributing to YF burden that are not environmentally related were not considered. These can include many important human-related characteristics such as vaccination [93] or socioeconomic status [91, 106]. Instead, this study focused on the components of total vulnerability that are outside of human control and cannot be directly intervened.

In estimating vulnerability among the epidemiologically silent municipalities, the assumption is made that these municipalities are capable of observing YF cases. It is possible that, due to factors not accounted for in this study, some of these municipalities are not at all vulnerable to YF. The predictions made in this study assume this is not the case. Additionally, by focusing on predictors less likely to change, the vulnerability estimates under these circumstances likely represent vulnerability in the absence of any characteristics not accounted for here that make them not susceptible to YF.

In addition to the preventative measures that can be informed by the results of this study, future research topics can benefit from these results. This study incorporates long-term factors of the natural environment to predict long-term vulnerability to YF burden. These predicted vulnerability categories can prove useful when predicting short-term risk of YF burden. In determining short-term risk of YF burden, informing the vulnerability in the long term to YF burden may be helpful in providing a baseline for short-term risk. Other studies can investigate the other facets of total vulnerability to YF not pursued in this study. Doing so would allow various aspects of vulnerability to be examined and potentially combined.

4.5. Conclusions

Based on long-term environmental predictors, several municipalities within Brazil were identified as those most likely to experience higher burden of YF. These municipalities were identified through having characteristics similar to those where greatest YF burden was seen. Rainfall and presence of YF cases in nearby municipalities were found to be influential variables in the generation of these predictions and physically responsible for the modification of ecoregions via ecohydrological landscape changes. These results can be used to inform preparedness strategies in Brazil for future YF burden, including healthcare capacity, mosquito control, or risk communication, particularly when climate change scenarios are considered about the future. Knowledge of environmental factors predicting disease burden is crucial for preparedness because human activities cannot immediately control them, but can adapt to them in a proactive ecosystem health perspective where alterations of the habitat can be engineered in a way that do not harm nature and people.

Chapter 5 Short-term forecasts for Yellow Fever burden based on meteorological and ecological conditions

Summary: Yellow Fever is a mosquito-borne disease endemic to South America and sub-Saharan Africa. Following several years with low incidence, a large outbreak was observed in Brazil, which began in December 2016. This outbreak emphasized the need for knowledge of future disease occurrence and incidence to improve preparedness. This study aims to use meteorological and ecological conditions to produce spatially explicit, short-term forecasts, defined as predictions up to eight weeks into the future, of weekly Yellow Fever occurrence and incidence in Brazil. Gamma hurdle models were fit using several short-term and long-term environmental factors, including temperature, precipitation, humidity, drainage density, and ecoregions. The minimum, maximum, mean, and range of daily temperature, precipitation, and humidity in each week were considered, as well as temporal lag periods between two and eight weeks. Separate models were created using municipality-level data for two time periods: an endemic period, representing Jan 2000 – Nov 2016, and an epidemic period, representing Dec 2016 – Mar 2018. The two time periods were split by the start of the 2016-2017 outbreak. Different sets of predictors were included in the best fitting binomial and Gamma steps of each of the two models. For both time periods, accurate forecasts were produced for probability of Yellow Fever occurrence as well as estimated incidence, conditioned on seeing any cases. Temperature, humidity, and previous Yellow Fever occurrence were the most influential variables in prediction, evaluated through global sensitivity analysis. Minimum, mean, and ranges of weekly temperature, precipitation, and humidity values provided most accurate forecasts. Temporal lag periods providing the best fits ranged between two and seven weeks. Results from this study show the use of implementing short-term forecasting systems for Yellow Fever using meteorological and ecological conditions, relying on the relationships between such conditions and disease risk.

5.1. Introduction

Yellow Fever (YF) is a mosquito-borne viral disease endemic to Latin America and sub-Saharan Africa. It is spread by multiple genera of mosquito and infects both humans and non-human primates [1, 266]. Symptomatic cases of YF typically present flu-like symptoms such as fever, nausea, and body aches, which typically persist for less than a week [266]. A subset of these cases then become severe, with symptoms including jaundice and hemorrhaging, and many severe cases are fatal due to organ damage [1, 31, 35]. Though there are no treatments for YF [38], human cases can be prevented through administration of a safe and effective vaccine [44].

Despite the existing vaccine and efforts to reduce virus transmission, cases have persisted in endemic regions [40], motivating further research in understanding the dynamics of YF. It is believed that the virus will never be globally eradicated due to the presence of nonhuman primate reservoirs and the ability for infected mosquitoes to transmit the virus to their offspring [45]. Preparations, including warning systems to the public, vaccination strategies, and mosquito control, are necessary to prevent or reduce disease burden.

The need for preparation is highlighted by recent outbreaks of YF. Small outbreaks have been observed in many nations in South America [63, 64] and Africa [65, 66]. A large outbreak was observed in Brazil beginning in December 2016 [58, 59], following the Zika outbreak that previously impacted Eastern Brazil [267]. In December 2016 and the six months that followed, Brazil observed more YF cases than had been observed over the previous decade. Prior to this outbreak, most YF cases were observed in Western Brazil, near the Amazon rainforest [54]. During the outbreak, cases were instead seen in southeastern parts of the country, including areas surrounding the cities of Rio de Janeiro and São Paulo [60]. The YF vaccine was not previously advised for many parts of Brazil that were most impacted [62], potentially increasing the outbreak's severity. Knowledge of when and where YF

cases are expected is crucial for preparation efforts and other public health measures in endemic nations.

One group of mechanisms affecting risk of YF and other mosquito-borne diseases includes environmental patterns [268]. Multiple environmental factors, including meteorological conditions such as temperature and precipitation as well as ecological conditions such as habitat types and elevation, have been shown to relate to risk of different mosquito-borne diseases [111, 269, 270], including YF [54, 91]. Environmental factors can affect YF risk, as they relate to mosquito populations through life cycles [113] and pathogen incubation [112], nonhuman primates through habitat suitability [224], and humans through behavioral adjustments [223] and mobility patterns. Some factors, such as elevation and ecological habitat type, are not expected to fluctuate over short periods of time such as days or weeks, while others, such as temperature and precipitation, do experience such fluctuations.

In forecasting pertaining to arthropod-borne viruses (arboviruses), long-term projections of disease burden over ten or more years have been produced [109, 271]. Because of the environmentally sensitive nature of arboviruses, climate events such as increases in temperature or ocean warming are often used to produce such long-term projections, particularly with the focus on global climate change [88, 109]. These long-term projections are typically used to motivate policy changes or changes to surveillance activities [272]. In contrast, short-term forecasting refers to producing predictions of disease events in advance for up to several weeks into the future. This can involve producing predictions of occurrence or incidence of a disease for a small period of time such as a week occurring in the future [18], typically by identifying a particular previous time, or lag period, that provides the best-informed predictions. Infectious disease forecasting research has less commonly aimed to produce short-term forecasts for arboviruses, though some examples exist [268, 273]. It is more common to see short-term forecasting for human-to-human communicable diseases [18, 274], commonly in the midst of an

epidemic or pandemic [275, 276], and some of these studies incorporate environmental factors [277]. Providing short-term disease forecasts motivates immediate decisions for prevention of new cases by targeting locations in greatest need [273, 278]. In the event of limited resources, which can include vaccines, medications, sanitizing products, or mosquito repellent, information regarding the most likely specific locations for new cases or highest case counts can aid in prioritizing resource allocation and prevent the greatest number of cases and deaths.

Short-term meteorological conditions with daily or weekly variation, including temperature, precipitation, and humidity, are important predictors of YF incidence by impacting mosquito breeding sites [279], life cycles [280], and disease incubation [112]; therefore, they are important for guiding immediate public health priorities regarding specific localities needing to implement preparedness measures such as assessing healthcare infrastructure or initiating vector control. This knowledge for decision-making can be gained through disease forecasting efforts using short-term weather events to produce short-term forecasts for YF burden.

This study aims to evaluate the ability of meteorological and ecological conditions to produce weekly forecasts of YF presence and incidence, using municipalities of Brazil as a case study. Statistical models were developed aiming to maximize the accuracy of forecasts using various meteorological and ecological conditions as predictors. The results of this study show the ability of such patterns to reliably forecast disease occurrence and disease incidence with a usable set of models for producing current, updatable forecasts. The resulting models in this study also produce forecasts each week for the probability of YF occurrence and estimated incidence, if seen, among municipalities in Brazil. Because the forecasts are expected to show different probabilities of disease occurrence and different estimated incidences across municipalities, those with highest probability of presence or estimated burden can be prioritized for preventative measures.

5.2. Methods and materials

5.2.1. Data sources

5.2.1.1. Yellow Fever incidence data

The setting of this study is Brazil, the largest nation in South America by both land area and population. The nation is subdivided into 26 states plus a federal district, containing 5,570 municipalities [281]. Yellow Fever case data were provided by the Pan American Health Organization and the Brazilian Ministries of Health. Data were available between January 2000 and March 2018. All confirmed cases were associated with a municipality and a date of report. The data for this study were analyzed by municipality-week.

Annual population data per municipality are publically available through the Brazilian Institute for Geography and Statistics [281]. Estimates are available for each year between 2000 and 2018, with the exceptions of 2007 and 2010. For these years, the arithmetic mean of the two adjacent years' populations was used to singly impute population estimates for these years. Using these population estimates and the YF case data, weekly incidence of YF per 100,000 population was estimated for each municipality.

Weekly burden of YF, operationalized through both occurrence, referring to seeing at least one case, as well as incidence, represented by cases per 100,000 population, was the outcome for forecast modeling. Previous YF burden was considered as a predictor to account for potential autoregressive tendencies [282]. For a target municipality-week for forecasting, previous burden was considered by incorporating a predictor indicating the observed incidence in a previous week as well as a predictor indicating whether there was at least one case in a previous

week. The optimal lag of these variables was investigated through the model fitting methods. Previous YF occurrence one-week prior is not always feasible or complete, as reporting delays and back-filling known to occur with infectious disease surveillance [283]; therefore, YF occurrence and incidence up to several weeks prior were considered as potential model predictors for future YF burden.

5.2.1.2. Environmental predictors

Environmental predictors were selected based on known relationships with YF risk, whether through direct studies of YF or similar diseases; connections to mosquito population dynamics; or connections to disease incubation. Meteorological predictors considered for modeling include precipitation [86, 92, 114, 234], temperature [86, 112], and humidity [284, 285]. Ecological predictors include water drainage [237], and ecological habitat type [54].

Temperature, precipitation, and humidity data are publically available from the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) from the United States National Aeronautics and Space Administration [101, 102]. This data set contains hourly temperature in Kelvin (degrees Celsius + 273.15), hourly precipitation in millimeters, and hourly humidity in kilograms of water per kilogram of air worldwide in 0.5-degree by 0.625-degree grids [101, 102]. The data for each grid were aggregated into daily averages for temperature and humidity and daily cumulative totals for precipitation, and then the grids were matched to Brazilian municipalities through a spatial join in ArcGIS Pro version 2.2.0 [236] using a shapefile for Brazilian municipalities available from the Database of Global Administrative Areas, version 3.6 [233], providing a daily value for each municipality between January 1, 2000 and March 31, 2018.

Previous studies have shown that many meteorological conditions can affect YF incidence in a variety of ways. Various studies have found associations between

weather maxima [111, 114], minima [286], means [114], and ranges [287] of meteorological conditions and mosquito-borne diseases. This study considered all four statistics for each of temperature, precipitation, and humidity. The weekly maximum and minimum values for temperature and humidity represent the highest and lowest average daily values in each week. The weekly range represents the difference between these two extreme values. The weekly mean for temperature and for humidity represents the average of the seven average daily values. Similarly, the weekly maximum and minimum values for rainfall represent the highest and lowest cumulative daily rainfall values in each week. The weekly range represents the difference between these two extreme values. The weekly rainfall mean represents the average of the cumulative rainfall daily values. Among the three meteorological variables, the same form was used within a model for parsimony. Using these four statistics of the meteorological variables is beneficial for practicality, as they represent values that are commonly collected and easy to interpret within the context of a statistical model.

Drainage density was modeled for Brazil using elevation and hydrological characteristics of the area. It is defined as the ratio of total length of rivers and streams to the total land area. This was calculated for known water basins with a minimum area of 1000 km^2. The water basins were matched to municipalities in ArcGIS Pro in order to compute a spatially-weighted average drainage density for each municipality.

Terrestrial ecological habitat types, referred to herein as ecoregions, are publically available from the World Wildlife Foundation [240]. A total of seven ecoregions are found in Brazil: Deserts and xeric shrublands; Flooded grasslands and savannas; Mangroves; Montane grasslands and shrublands; Tropical and subtropical dry broadleaf forests; Tropical and subtropical grasslands, savannas, and shrublands; and Tropical and subtropical moist broadleaf forests [240]. These were matched to municipalities using a spatial join to represent the most prevalent ecoregion within

each municipality. Because the majority of YF cases were seen in only two of the ecoregions (Tropical and subtropical moist broadleaf forests; and Tropical and subtropical grasslands, savannas, and shrublands), a new binary variable was created to indicate whether a municipality belonged to one of these two most common ecoregions

An additional predictor corresponded to previous work done to produce estimates of municipality-level YF vulnerability probabilities based on precipitation, population density, elevation, water drainage, and ecoregion. Each municipality was assigned predicted probabilities of being in each of three pre-defined yearly vulnerability categories: minimal vulnerability, labeled as category 0 (0 cases per 100,000 population); moderate vulnerability, labeled as category 1 (0-10 cases per 100,000 population); and higher vulnerability, labeled as category 2 (more than 10 cases per 100,000 population) (Table 4.1). These fitted probabilities were estimated using a cumulative logit model that used percentage of days with rain, drainage density, population density, elevation, ecoregion, and YF occurrence in nearby municipalities as predictors. These predictors were chosen because they provided the closest fit to vulnerability category among municipalities that observed YF cases between Jan 2000 and Mar 2018. The estimates were then applied to all municipalities in Brazil. Using these three fitted probabilities as model predictors accounts for multiple potentially influential factors parsimoniously, reducing potential predictors. The three probabilities for each municipality sum to one by design.

5.2.1.3. Temporal lags

Temporal lags for the three meteorological variables as well as the two variables for previous weeks' YF incidence and occurrence within each municipality were used to allow time between produced forecasts and implementation of potential prevention strategies as well as to consider delayed effects of meteorological changes. As

expected for vector-borne diseases, the effects of meteorological events on mosquito populations, and therefore disease incidence, are assumed not to occur instantaneously [288, 289]. Temporal lags are defined as the number of weeks between the week associated with the observed data and the week being targeted for forecasting. For example, a lag period of *l* weeks means that, in order to produce a forecast for week *t,* data from week *t-l* were used as predictors.

Lag periods between two and eight weeks were considered, using the same lag for all lagged variables. The same lag period was used for the three weather variables and the two variables for previous YF for model parsimony. For practical applications, requiring data from one week to predict YF occurrence or incidence is beneficial for data management. Though biological processes governing YF risk may be best explained through different lags for the different predictors, such as differences previously found with temperature and precipitation [91], the assumption that a single lag period for all lagged variables can provide accurate forecasts focuses solely on predictive ability rather than isolating specific effects of weather patterns on YF risk. Using the range between two and eight weeks conforms to the definition of "short-term" mentioned previously, as disease forecasting weeks into the future is a much smaller timespan compared to disease forecasting years into the future, as described previously.

5.2.2. Model fitting

The aim of model fitting was to develop a set of models that produce accurate forecasts of YF occurrence and incidence. In order to validate the predictive ability of the model, the data were split into a training data set and a testing data set. Due to an outbreak that began in Dec 2016, the data and analyses were split into two time periods: Jan 2000 – Nov 2016 and Dec 2016 – Mar 2018. For the remainder of this study, the earlier time period, Jan 2000 – Nov 2016, will be referred to as the "endemic" time period, and the later time period, Dec 2016 – Mar 2018, will be

referred to as the "epidemic" time period. These definitions are intended to align with those from the previous study (Chapter 4); the differences in this study are that Dec 2016 is now considered part of the epidemic time period, and the months of Jan – Mar 2018 are added as part of the epidemic time period.

Model fitting in the two time periods was done separately and independently of each other. In the endemic time period, the training set consisted of data from Jan 2000 to Dec 2013, and the testing set consisted of data from Jan 2014 to Nov 2016. In the epidemic time period, the training set consisted of data from Dec 2016 to Dec 2017, and the testing set consisted of data from Jan 2018 to Mar 2018. The separation of training and testing data sets in both time periods assured that at least one calendar year was present in training data, allowing potential seasonal cycles to be accounted for.

Due to the fact that many municipalities in Brazil did not see any YF cases during the entire time range in the data (Jan 2000 – Mar 2018), possibly as a result of a true absence of cases or a reduced ability to detect cases, only the 466 municipalities where at least one YF case was seen between January 2000 and March 2018 were included in model fitting and validation. In order to keep the same data contributing to models using different lag periods, only weeks with available lagged data for two-week through eight-week lags were included in model fitting. These 466 municipalities were used for model fitting in both the endemic and epidemic time periods.

The following variables were used as potential model predictors: temperature (continuous, linear and quadratic), precipitation (continuous, linear and quadratic), humidity (continuous, linear and quadratic), month (indicators for each month with January as the reference), previous occurrence of YF (binary), previous incidence of YF (continuous), drainage density (continuous), ecological habitat (binary indicator for Tropical and subtropical moist broadleaf forests or Tropical and subtropical

grasslands, savannas, and shrublands), and fitted probabilities for each of the three vulnerability categories (continuous). For temperature, humidity, and precipitation, each of the maximum, minimum, mean, and range daily values for each week was considered, with all three variables using the same form. For temperature, humidity, precipitation, previous incidence, and previous occurrence, lags between two and eight weeks were considered, with all variables using the same lag period.

The model form was a Gamma hurdle model [290, 291], which incorporates a binomial step predicting probability of observing YF occurrence and a Gamma step predicting incidence under the condition that any YF occurrence is seen. This model form was selected because there was a high percentage of municipality-weeks with zero cases of YF, and incidence per 100,000 population can take nondiscrete values. The Gamma distribution is defined to be positive and continuous, and it has been used previously to describe disease incidence in Chapter 2 and other previous work [19]; the Gamma distribution has also has been used to describe disease growth [160, 161]. Gamma hurdle models are not commonly used for disease incidence, but have been used in ecology applications [292]. The hurdle model in general has been used for infectious disease research previously for relating meteorological conditions to disease risk [293], and the motivations for using a hurdle model are conceptually similar to those for using a zero-inflated Poisson model [294, 295].

The same predictors were considered for both the binomial and Gamma steps of the model for each time period, allowing the predictor sets to differ between steps within the same model. The form of the meteorological conditions (minimum, maximum, mean, range) was allowed to differ between the two steps of each hurdle model as well as the optimal lag period.

All combinations of the previously mentioned parameters were included, omitting combinations that are not practical. Such combinations would include those that use all three vulnerability categories as predictors, as they sum to one by design; those

that include any vulnerability categories as well as either ecoregion or drainage density, as the vulnerability probabilities are a function of ecological habitat and drainage density, among other predictors; or those that include quadratic terms without appropriate linear terms. A total of 1,295 combinations of parameters were included for each combination of meteorological variable form (maximum, minimum, mean, range) and lag period (two-week through eight-week) for both time periods' models.

5.2.3. Model evaluation

The best-fitting model was selected for each time period by assessing the predictive ability of both the binomial step and the Gamma step individually. The binomial step was evaluated through Receiver Operator Characteristic (ROC) curves and associated areas under the curve (AUC). The ROC curves were created by comparing predicted probabilities of YF occurrence to observed YF occurrence in the testing data. The model parameter set that produced the highest AUC was selected as the predictors for binomial step of the model. The parameter set for the Gamma step producing the minimum mean absolute error (MAE) was selected, comparing predicted nonzero incidence values to observed nonzero incidence values in the testing data. The MAE is defined as

$$MAE = \frac{1}{n}\sum|\widehat{y_i} - y_i| \text{ (Equation 5.1)}$$

where y_i values represent the n observed nonzero incidence values and $\widehat{y_i}$ represent the corresponding fitted values from the model. All analyses were run using R version 3.6.0 [150]. Visual maps were made using ArcGIS Pro version 2.2.0 [236].

5.2.4. Producing forecasts

From the best fitting model for the epidemic and for the endemic time period, forecasts for both YF occurrence and incidence can be produced weekly. The

forecasts for occurrence and incidence should be used jointly, since the Gamma step predicting incidence is conditioned on seeing any YF cases. The results of the binomial step of the model are interpreted as the predicted probability of seeing at least one YF case in the week forecasted. The Gamma step predictions are interpreted as the expected incidence that would be observed in the forecasted week conditioned on observing at least one case. The two steps are considered as separate components of a single model.

Example forecasts were generated for two selected weeks nationwide for each time period. For each, the first week in the testing data was used, representing the first full weeks of January 2014 and January 2018. These weeks are also when YF cases typically begin to rise and observe a peak, providing interesting demonstrative examples. Also forecasted were the first weeks beyond the range of the testing data sets: the first weeks of December 2016 and April 2018. Forecasts were generated from the fitted models by applying data from the appropriately lagged weeks. For example, if the optimal lag period for a model is two weeks, than data from the week of March 19-25, 2018 would be used to produce the forecast for the week of April 2-8, 2018. Forecasts for all other weeks during the testing periods for each model were also generated for selected municipalities to show longitudinal forecasts.

5.3. Results

The time series of weekly YF cases throughout the entire study period shows the severity of the 2016-2017 outbreak, having notably higher case counts compared to previous years (Figure 5.1). This justifies separating the two time periods, as incidence values between the time periods differ greatly and may be explained by different determinants. The seasonal patterns of YF are also seen throughout all years of observation, with cases predominantly occurring during the early months of each year. Tables 5.1 and 5.2 show descriptive statistics of the YF data and the predictors considered in the models for the endemic and epidemic time periods.

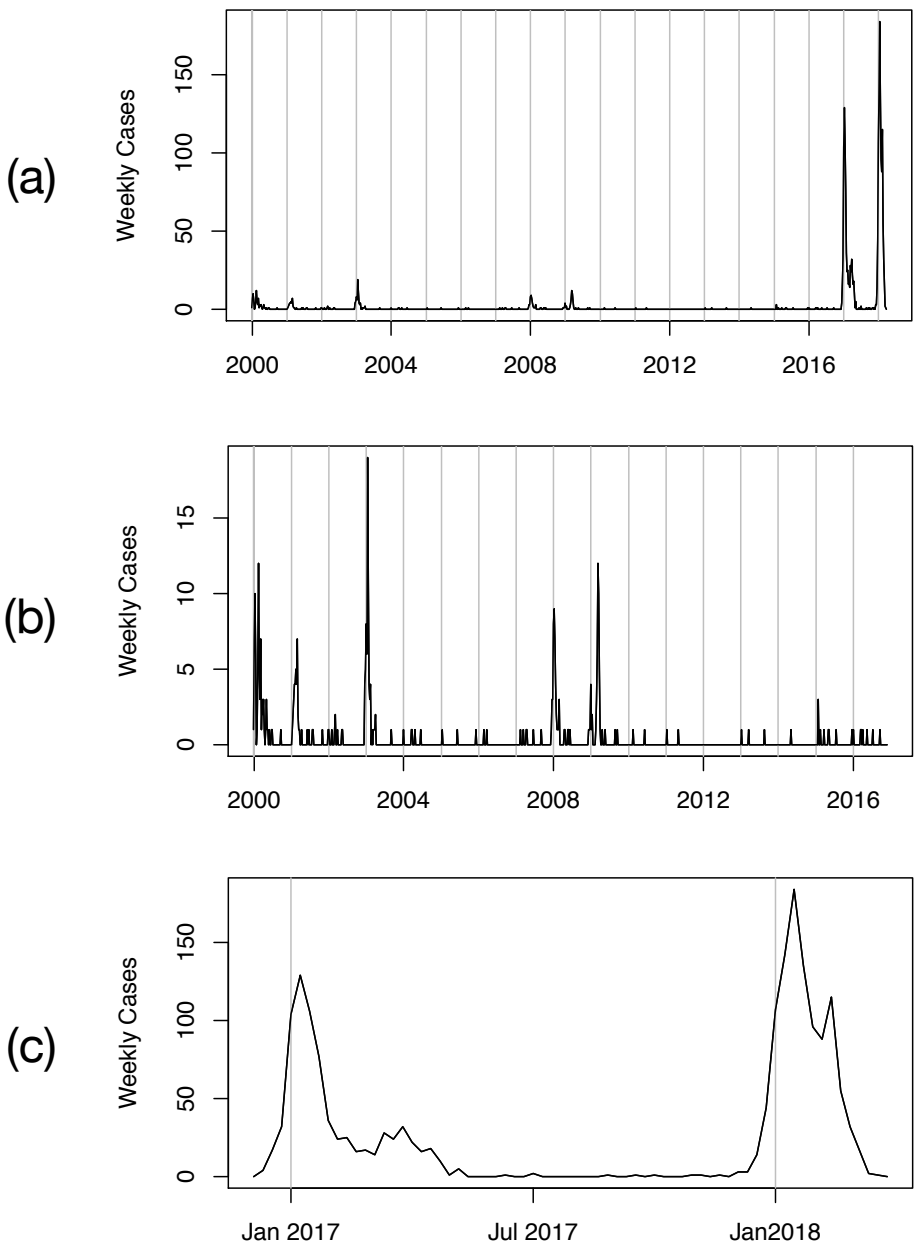


Figure 5.1. Total weekly Yellow Fever cases in Brazil in (a) the entire time period of observation (Jan 2000 – Mar 2018), (b) the endemic time period (Jan 2000 – Nov 2016), and (c) the epidemic time period (Dec 2016 – Mar 2018). Vertical lines indicate the first week of each calendar year.

Table 5.1. Summary statistics of Yellow Fever outcome variables and predictors considered in forecast models. Observed data are per municipality-week in the endemic time period (Jan 2000 – Nov 2016) (n = 411,994). Binary variables (YF occurrence and ecoregion) are represented with counts and percentages of positive values, and continuous variables are represented with means and standard deviations.

	mean n	(sd) (%)	min	25%	50%	75%	max
YF occurrence	258	0.063%	0	0	0	0	1
Nonzero YF incidence (n = 258)	12.174	17.415	0.039	2.48	5.511	14.54	117.165
Min Tmp	20.238	3.655	1.457	17.883	20.399	22.752	34.133
Max Tmp	23.599	2.832	7.994	21.745	23.591	25.444	36.913
Mean Tmp	21.94	3.108	5.066	19.908	21.97	24.012	35.783
Range Tmp	3.361	2.131	0.004	1.775	2.848	4.453	18.935
Min Rain	0.001	0.001	0.000	0.000	0.000	0.001	0.026
Max Rain	0.006	0.007	0.000	0.001	0.004	0.008	0.110
Mean Rain	0.002	0.002	0.000	0.000	0.002	0.003	0.028
Range Rain	0.005	0.006	0.000	0.001	0.003	0.007	0.108
Min Hum	0.013	0.004	0.002	0.010	0.013	0.015	0.025
Max Hum	0.015	0.003	0.005	0.013	0.016	0.018	0.026
Mean Hum	0.014	0.003	0.004	0.011	0.014	0.017	0.025
Range Hum	0.003	0.001	0.000	0.002	0.002	0.003	0.014
Drainage	0.092	0.017	0.061	0.082	0.090	0.096	0.343
P (Vuln = 0)	0.975	0.028	0.867	0.959	0.982	1.000	1.000
P (Vuln = 1)	0.014	0.016	0.000	0.000	0.010	0.023	0.076
P (Vuln = 2)	0.011	0.012	0.000	0.000	0.008	0.018	0.061
Ecoregion	406640	98.712%	0.000	1	1	1	1.000

Table 5.2. Summary statistics of Yellow Fever outcome variables and predictors considered in forecast models. Observed data are per municipality-week in the epidemic time period (Dec 2016 – Mar 2018) (n = 32,154). Binary variables (YF occurrence and ecoregion) are represented with counts and percentages of positive values, and continuous variables are represented with means and standard deviations.

	mean n	(sd) (%)	min	25%	50%	75%	max
YF occurrence	893	2.777%	0	0	0	0	1
Nonzero YF incidence (n = 893)	9.828	13.914	0.008	1.663	5.112	12.377	154.253
Min Tmp	20.886	3.411	3.303	18.705	21.1	23.2	32.763
Max Tmp	23.944	2.788	13.891	22.33	24.082	25.693	35.475
Mean Tmp	22.431	2.997	9.25	20.587	22.568	24.383	33.979
Range Tmp	3.058	1.733	0.167	1.77	2.73	3.945	14.494
Min Rain	0.001	0.001	0	0	0	0.001	0.008
Max Rain	0.006	0.007	0	0.001	0.004	0.009	0.061
Mean Rain	0.002	0.002	0	0.001	0.002	0.004	0.021
Range Rain	0.006	0.007	0	0.001	0.004	0.008	0.061
Min Hum	0.013	0.004	0.003	0.01	0.013	0.016	0.024
Max Hum	0.015	0.003	0.006	0.013	0.016	0.018	0.026
Mean Hum	0.014	0.003	0.005	0.012	0.015	0.017	0.024
Range Hum	0.002	0.001	0	0.001	0.002	0.003	0.011
Drainage	0.092	0.017	0.061	0.082	0.09	0.096	0.343
P (Vuln = 0)	0.651	0.279	0.062	0.388	0.762	0.903	1
P (Vuln = 1)	0.164	0.101	0	0.067	0.153	0.27	0.298
P (Vuln = 2)	0.185	0.195	0	0.03	0.084	0.315	0.816
Ecoregion	31740	98.712%	0	1	1	1	1

5.3.1. Model results

In the best-fitting hurdle model for the endemic time period, the binomial step included as predictors minimum temperature as a linear and quadratic term, minimum rainfall, minimum humidity, previous occurrence of YF, and estimated probability of moderate YF vulnerability. A 5-week lag for temperature, rainfall, humidity, and previous occurrence of YF provided the best fit. The AUC was 0.926.

The Gamma step during this time period included range of precipitation, range of humidity as a linear and quadratic term, previous occurrence of YF, and ecoregion. A seven-week lag period was used for precipitation, humidity, and previous occurrence of YF. The MAE was 5.075 (Table 5.3). The model for this time period takes the form

$$
\begin{aligned}
&logit\big(P(YF_t > 0)\big) = \\
&\qquad -6205.134 + 41.923\ MinTemp_{t-5} - 0.071\ MinTemp_{t-5}{}^2 \\
&\qquad + 13.998\ MinRain_{t-5} + 153.312\ MinHum_{t-5} \\
&\qquad + 3.483\ I(YF_{t-5} > 0) + 30.198\ P(Vuln = 1) \\
&\log(E(YF_t\ |YF_t > 0)) = \\
&\qquad 0.883 + 10.837\ RangeRain_{t-7} + 584.418\ RangeHum_{t-7} \\
&\qquad - 82599.353\ RangeHum_{t-7}^2 - 0.100\ I(YF_{t-7} > 0) \\
&\qquad +0.691\ Ecoregion
\end{aligned}
$$

(Equation 5.2)

where YF_t refers to YF incidence at week t in a particular municipality and predictor variables refer to the same municipality. The function $I(\cdot)$ represents an indicator function returning a value of one if the statement is true and zero otherwise.

The best-fitting hurdle model for the epidemic time period included, in the binomial step, mean temperature, month, previous occurrence of YF, and ecoregion. A two-week lag period was used for temperature and previous occurrence of YF. The AUC was 0.842. The Gamma step of the model during this time period included minimum precipitation, minimum humidity as a linear and quadratic term, previous occurrence of YF, and drainage density. A seven-week lag period was used for precipitation, humidity, and previous occurrence of YF. The SAE was 6.772 (Table 5.3). The model for this time period takes the form

$$
\begin{aligned}
&logit\big(P(YF_t > 0)\big) = \\
&\qquad -7.560 - 0.038\, MeanTemp - 0.876\, I(Month_t = Feb) \\
&\qquad - 0.897\, I(Month_t = Mar) - 1.278\, I(Month_t = Apr) \\
&\qquad - 3.227\, I(Month_t = May) - 4.679\, I(Month_t = Jun) \\
&\qquad - 4.309\, I(Month_t = Jul) - 4.707\, I(Month_t = Aug) \\
&\qquad - 4.670\, I(Month_t = Sep) - 4.121\, I(Month_t = Oct) \\
&\qquad - 3.925\, I(Month_t = Nov) - 1.187\, I(Month_t = Dec) \\
&\qquad + 3.006\, I(YF_{t-2} > 0) + 0.718\, Ecoregion \\
&\log(E(YF_t \,|YF_t > 0)) = \\
&\qquad -5.787 + 241.296\, MinRain_{t-7} + 1119.615\, MinHum_{t-7} \\
&\qquad - 42145.567\, RangeHum_{t-7}^2 - 0.507\, I(YF_{t-7} > 0) \\
&\qquad + 9.530\, DrainageDens
\end{aligned}
$$

(Equation 5.3)

where, as in Equation 5.2, YF_t refers to YF incidence at week t in a particular municipality, $I(\cdot)$ is used as an indicator function, and predictor variables refer to the same municipality.

Table 5.3. Estimated model parameters for best-fitting forecast models for Yellow Fever in Brazil in the endemic and epidemic time periods. Cells are left empty when a particular variable was not included in the model. The form of temperature, precipitation, and humidity (minimum, maximum, mean, range) and model fit statistics are indicated along the bottom rows.

	Endemic time period				Epidemic time period			
	Binomial		Gamma		Binomial		Gamma	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	-6205.134	1099.265	0.883	0.614	7.560	8.787	-5.787	2.106
Temp	41.923[a]	7.436			-0.038[f]	0.029		
Temp^2	-0.071[a]	0.013						
Rain	13.998[b]	77.125	10.837[d]	13.885			241.946[b]	112.496
Rain^2								
Hum	153.312[c]	36.189	584.418[e]	304.979			1119.650[c]	290.476
Hum^2			-82599.353[b]	49789.737			-42145.567[c]	10425.169
Month-Feb					-0.876	0.158		
Month-Mar					-0.897	0.167		
Month-Apr					-1.278	0.195		
Month-May					-3.257	0.427		
Month-Jun					-4.679	1.009		
Month-Jul					-4.309	0.732		
Month-Aug					-4.707	1.012		
Month-Sep					-4.670	1.008		
Month-Oct					-4.121	0.714		
Month-Nov					-3.925	0.714		
Month-Dec					-1.187	0.162		
Previous occurrence	3.483	0.323	-0.100	0.865	3.006	0.131	-0.507	0.292
Previous incidence								
P(Vuln = 0)								
P(Vuln = 1)	30.198	4.033						
P(Vuln = 2)								
Ecoregion			0.691	0.541	0.718	0.727		
Drainage density							9.530	6.493
Weather form	Minimum		Range		Mean		Minimum	
Lag period	5-week		7-week		2-week		7-week	
Model fit	AUC = 0.926		MAE = 5.075		AUC = 0.842		MAE = 6.772	

Table 5.3 cont.

[a]Minimum weekly temperature is represented by the minimum of the seven daily average temperatures of the week

[b]Minimum weekly precipitation is represented by the minimum of the seven daily cumulative precipitation totals of the week

[c]Minimum weekly humidity is represented by the minimum of the seven daily average humidity values of the week

[d]Range of precipitation is represented by the difference between the maximum and minimum of the seven daily cumulative precipitation totals of the week

[e]Range of weekly humidity is represented by the difference between the maximum and minimum of the seven daily average humidity values of the week

[f]Mean of weekly temperature is represented by the average of the seven average daily temperatures of the week

The parameter estimates shown in Equations 5.2 and 5.3 as well as Table 5.3 show the predictive effect of each variable on estimated probability of YF occurrence through the binomial step of the hurdle models and, if YF is present, the estimated incidence of YF through the Gamma step of the hurdle models. During the endemic time period, the binomial step shows that predicted probability of seeing any YF cases in a given week is higher when the minimum daily rainfall total of the week five weeks prior is higher. The probability is also higher if any YF cases were seen five weeks prior and among municipalities classified as of "moderate YF vulnerability" (having higher fitted probability of seeing up to 10 cases per 100,0000 population annually). The weekly minimum daily temperature five weeks prior showed a parabolic relationship with risk of observing any YF cases. The negative sign of the quadratic parameter indicates that there is a peak temperature at which risk is maximized. Based on the linear and quadratic parameter estimates in the model, maximum probability of YF presence was seen when the weekly minimum daily temperature five weeks prior was 22.082 degrees Celsius (295.232 Kelvin)(Figure 5.2).

From the endemic time period, the Gamma step of the hurdle model shows that, if YF is seen, incidence is expected to be higher if the difference between the maximum and minimum daily rainfall totals for the week is higher seven weeks

prior. The model also predicts a lower incidence if any YF was seen in the same municipality seven weeks prior, and that higher incidence is predicted if the municipality lies in either tropical and subtropical moist broadleaf forests or tropical and subtropical grasslands, savannas, and shrubland ecoregions. The parabolic relationship with humidity shows that there is a humidity range where maximum incidence is predicted. This would occur when the difference between the maximum and minimum days' humidity values of the week seven weeks prior is 0.003 kilograms of water per kilogram of air (Figure 5.3).

During the epidemic time period, from Dec 2016 to Mar 2018, the binomial step of the hurdle model shows that the predicted probabilities of seeing any YF cases is lower when the mean of the daily average temperatures two weeks prior is higher. If the municipality also saw any YF cases two weeks prior, the probability of seeing any cases is increased. If the municipality is located within either tropical and subtropical moist broadleaf forests or tropical and subtropical grasslands, savannas, and shrubland ecoregions, or if the week forecasted is in January, then predicted probability of YF cases is also higher (Figure 5.4).

In the epidemic time period, the Gamma step shows that, if any YF cases are seen, predicted incidence is expected to be higher if the weekly minimum daily rainfall total seven weeks prior is higher and if the municipality sees greater drainage density. Predicted incidence is lower if any YF cases were seen in the same municipality seven weeks prior. The minimum daily average humidity of the week shows a parabolic trend with incidence peaking if the minimum daily humidity seven weeks prior is 0.013 kilograms of water per kilogram of air (Figure 5.5). This is approximately the median value of this variable in the data (Table 5.2).

The two steps of the hurdle model can be implemented independently, meaning that the Gamma step estimating incidence does not require a particular output from the binomial step to estimate conditional incidence. Even when the predicted

probability of YF occurrence is very low or equal to zero, the Gamma step can be used to predict incidence conditioned on observing any YF cases. As a result, even municipality-weeks with very low estimated probabilities of YF occurrence have associated estimates of incidence, even if the probability of nonzero incidence is very low.

The use of the lag variables in the models, indicated for each step of the hurdle model in Table 5.3, shows how previous meteorological conditions and observed YF occurrence impact future YF risk, whether through probability of disease occurrence or through the incidence value if it is observed. Within each model step, the same lag period was used for previous YF and for the weather variables for parsimony, allowing different lag periods to be used for each of the two steps of the same model. For example, during the endemic time period, probability of disease occurrence is best predicted using data from five weeks prior, but the expected incidence if the disease does occur is best predicted using data from seven weeks prior. In order to fully predict occurrence and incidence using this model for week *t,* data from week *t-7* would be used to predict locations with highest incidence if YF is seen. Then, two weeks later, in week *t-5*, most accurate predictions can be made regarding the locations most likely to see any YF cases in week *t*. Similarly, during the Dec 2016 – Mar 2018 time period, the probability of disease occurrence is best predicted using data from two weeks prior, but the expected incidence if the disease does occur is best predicted using data from seven weeks prior.

While the lags presented in Table 5.3, showing different lag periods for predicting probability of YF occurrence and predicting YF incidence in each model, provide the most accurate predictions, it is noteworthy to consider a desire for stakeholders to produce predictions for probability of YF occurrence and incidence of YF using the same lag period. This will lead to loss of accuracy, as one or both of the two steps of the hurdle model will need to use a non-optimal lag period for this to occur. A

sensitivity analysis shown in section 5.3.4 shows the potential loss of accuracy if such adjustments were to be made.

The plots in Figures 5.2-5.5 show the individual relationships between each predictor and the outcomes of the two model steps (predicted probability of YF occurrence and estimated incidence if YF does occur). For each model step, each variable was plotted independently against the outcome. In doing so, every other continuous variable was assigned its mean, and every other binary variable was assigned zero. The continuous variables' entire ranges from the data were used in the figures. The parabolic trends of temperature and humidity, when appropriate, are seen, noting optimal conditions from previous weeks that maximize either the probability of YF occurrence or the expected incidence if YF is seen.

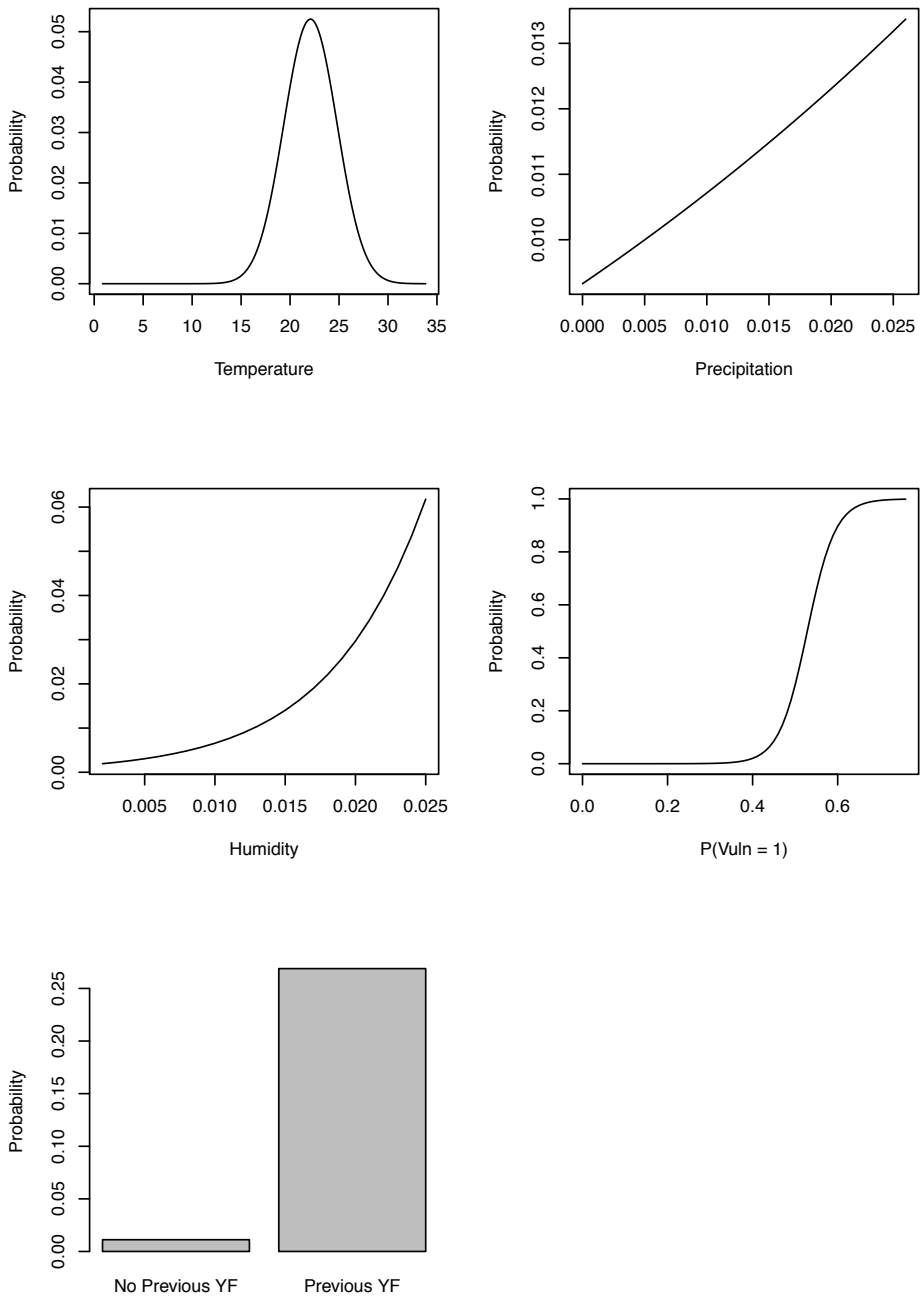


Figure 5.2. Individual effects of each predictor on predicted probability of Yellow Fever occurrence in Brazil using the binomial step of the hurdle model for the endemic time period

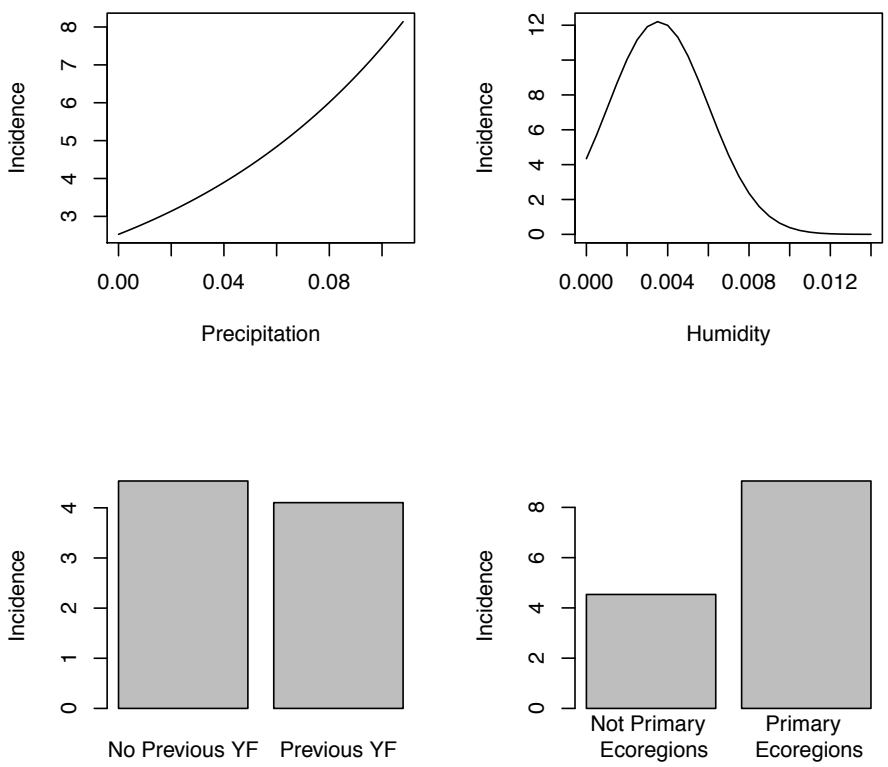


Figure 5.3. Individual effects of each predictor on predicted incidence of Yellow Fever in Brazil if it is present using the Gamma step of the hurdle model for the endemic time period.

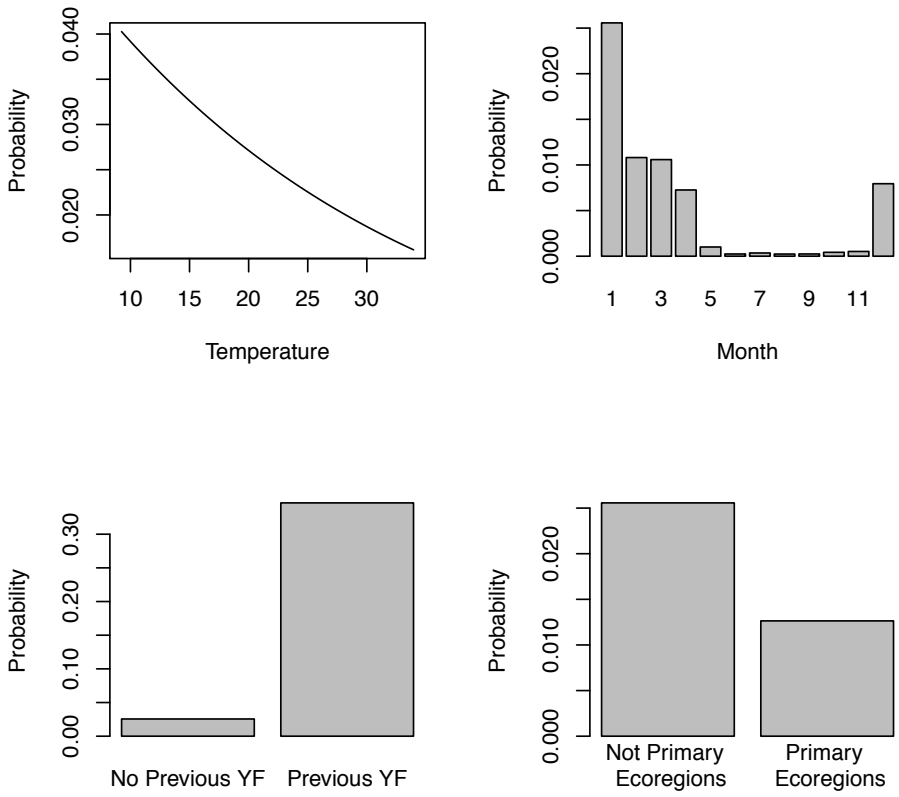


Figure 5.4. Individual effects of each predictor on predicted probability of Yellow Fever occurrence in Brazil using the binomial step of the hurdle model for the epidemic time period.

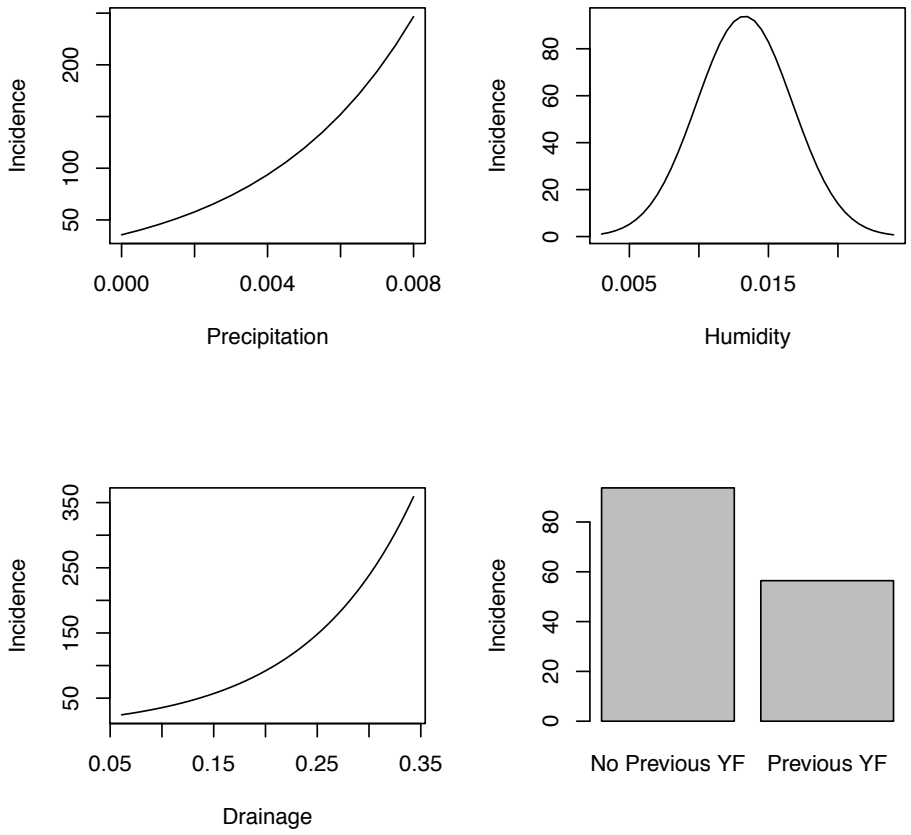


Figure 5.5. Individual effects of each predictor on predicted incidence of Yellow Fever in Brazil if it is present using the Gamma step of the hurdle model for the epidemic time period.

Many of the plots in Figures 5.2-5.5 show predicted probabilities of YF occurrence and predicted incidence values that are notably higher than actual forecasted estimates shown in Figures 5.8-5.14. The range of each predictor shown in Figures 5.2-5.5 match the ranges of the data, indicated in Tables 5.1 and 5.2. This suggests that it is unlikely to see the mean values of each all predictors in a given municipality-week. However, the interpretation of the shapes of each plot shown in Figures 5.2-5.5 are indicative of the trend seen in the models between each predictor and the outcomes.

5.3.2. Model diagnostics

Model diagnostics were run in order to assess the models' forecasting abilities. This was done by comparing model fit criteria, AUC and MAE, under shorter training periods and under specific testing periods. To examine differences in length of training period on forecast accuracy, the training data sets were truncated by removing weeks at the beginning of each time period. Every week was used as a starting period for the training data set, provided that a minimum of 20 weeks was present. Figure 5.6 shows the model fit statistics by the starting points of the truncated training periods for both steps of both models. The binomial step of the model for the endemic time period shows a decrease in model fit as the training period becomes shorter, seeing a large drop in AUC when the training period begins after April 2013. The Gamma step of this model shows a notable loss of model fit when the training period begins after 2009. The binomial step of the model for the epidemic time period shows consistent AUC values, with a decrease in model fit when the training data begins after February 2017. The Gamma step of this model sees a decrease in model fit when the training data begin after April 2017.

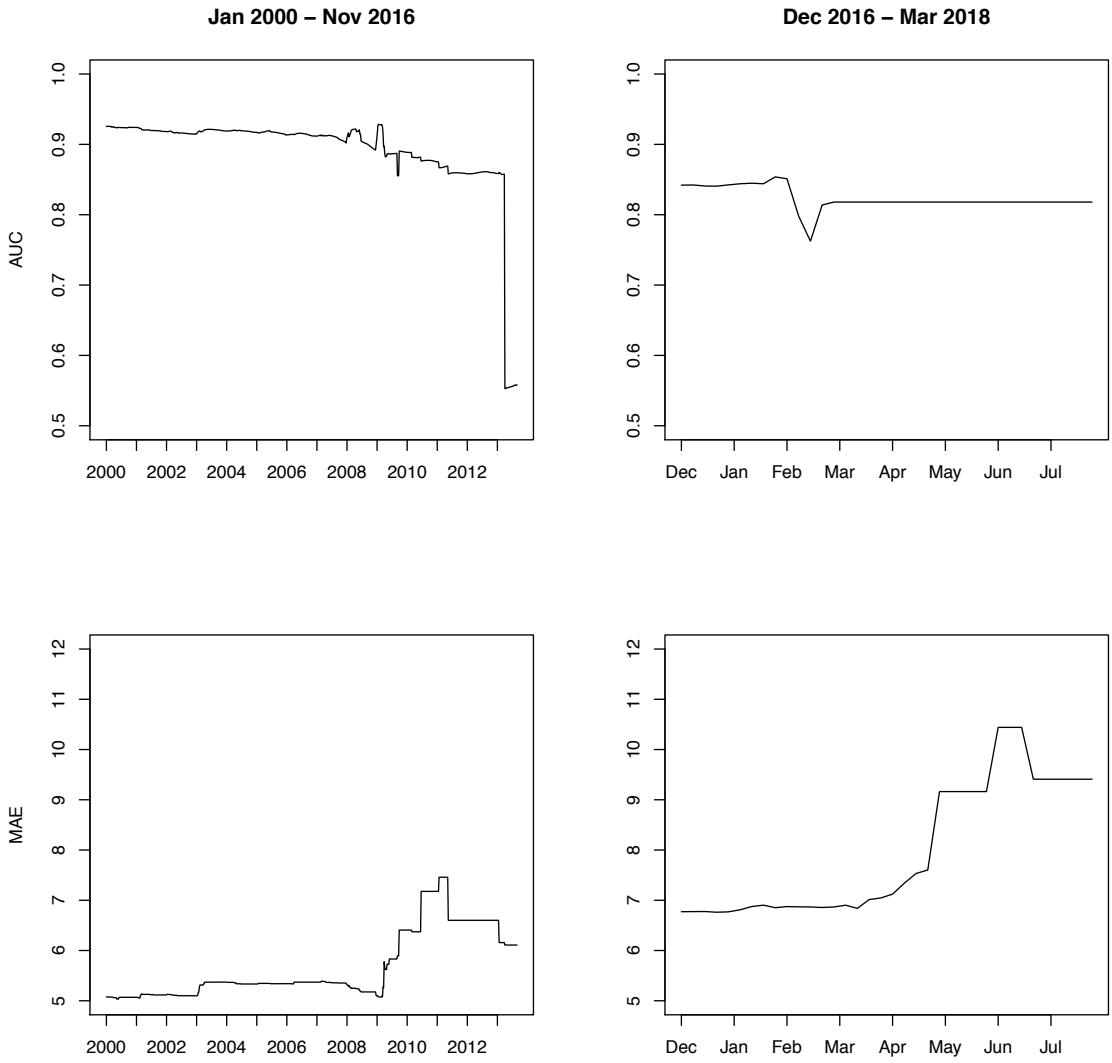


Figure 5.6. Model fit by beginning week of truncated training periods. The points on the horizontal axis represent a training period starting at that time, and changes in model fit criteria are seen when the training periods of each model are shortened.

Other diagnostics were conducted to determine if the models have better predictive accuracy during certain parts of the year. Keeping the original training and testing data sets and using the models described in Equations 5.2 and 5.3 and Table 5.3, the model fit statistics, AUC and MAE, were fit to the testing data isolated in specific calendar months across all years. This accounts for seasonality of YF while allowing flexibility of peak weeks and starts of seasonal increases. The month-specific model

fit criteria, shown in Figure 5.7, show relative forecasting ability. The binomial step of the model for the endemic time period shows consistent AUC across months. No data are provided for June, August, October, and November because no cases were seen in these months across all years of the testing data. The Gamma step shows that forecasts in February through May are most accurate, with forecasts in December having the least accuracy. The binomial step of the model for the epidemic time period shows similar consistency in AUC across months. The Gamma step of this model shows that forecasts in March are more accurate than those in January and February.

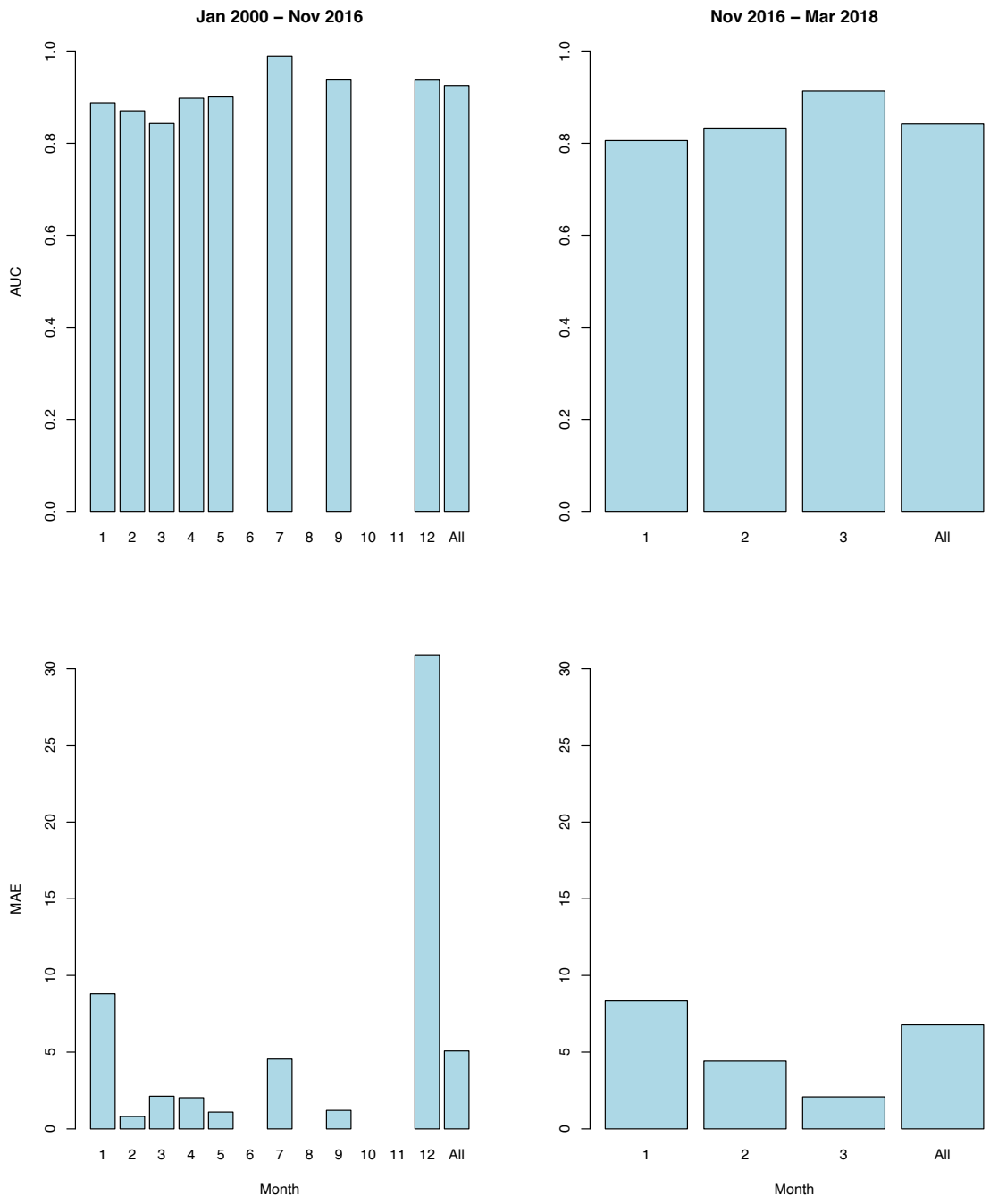


Figure 5.7. Month specific model fit statistics for Yellow Fever forecasting models.

5.3.3. Example forecasts

Figures 5.8-5.11 show examples of nationwide forecasts that can be produced with the models reported along with the spatially varying predictors that contributed to the forecasts. The models from both time periods were used to forecast the probability of YF occurrence and incidence, if seen, for the first week of each respective testing period as well as the first week beyond the respective testing data sets. From the model for the endemic time period, example forecasts were provided for the weeks of Jan 6-12, 2014 and for Dec 5-11, 2016. The forecasts for the week of Jan 6-12, 2014 used data from the week of Dec 2-8, 2013 (five weeks prior) to produce the predicted probability of YF occurrence with the binomial step of the model as well as data from Nov 18-24, 2013 (seven weeks prior) to produce predicted incidence of YF with the Gamma step of the model. From the model for the epidemic time period, example forecasts were created for the weeks of Jan 1-7, 2018 and Apr 2-8, 2018. The forecasts for the week of Jan 1-7, 2018 used data from the week of Dec 18-24, 2017 (two weeks prior) to produce the predicted probability of YF occurrence with the binomial step of the model as well as data from Nov 13-19, 2017 (seven weeks prior) to produce predicted incidence of YF with the Gamma step of the model. These weeks also serve as interesting demonstrative examples, as YF incidence typically is highest or increasing around these times of year (Figure 5.1).

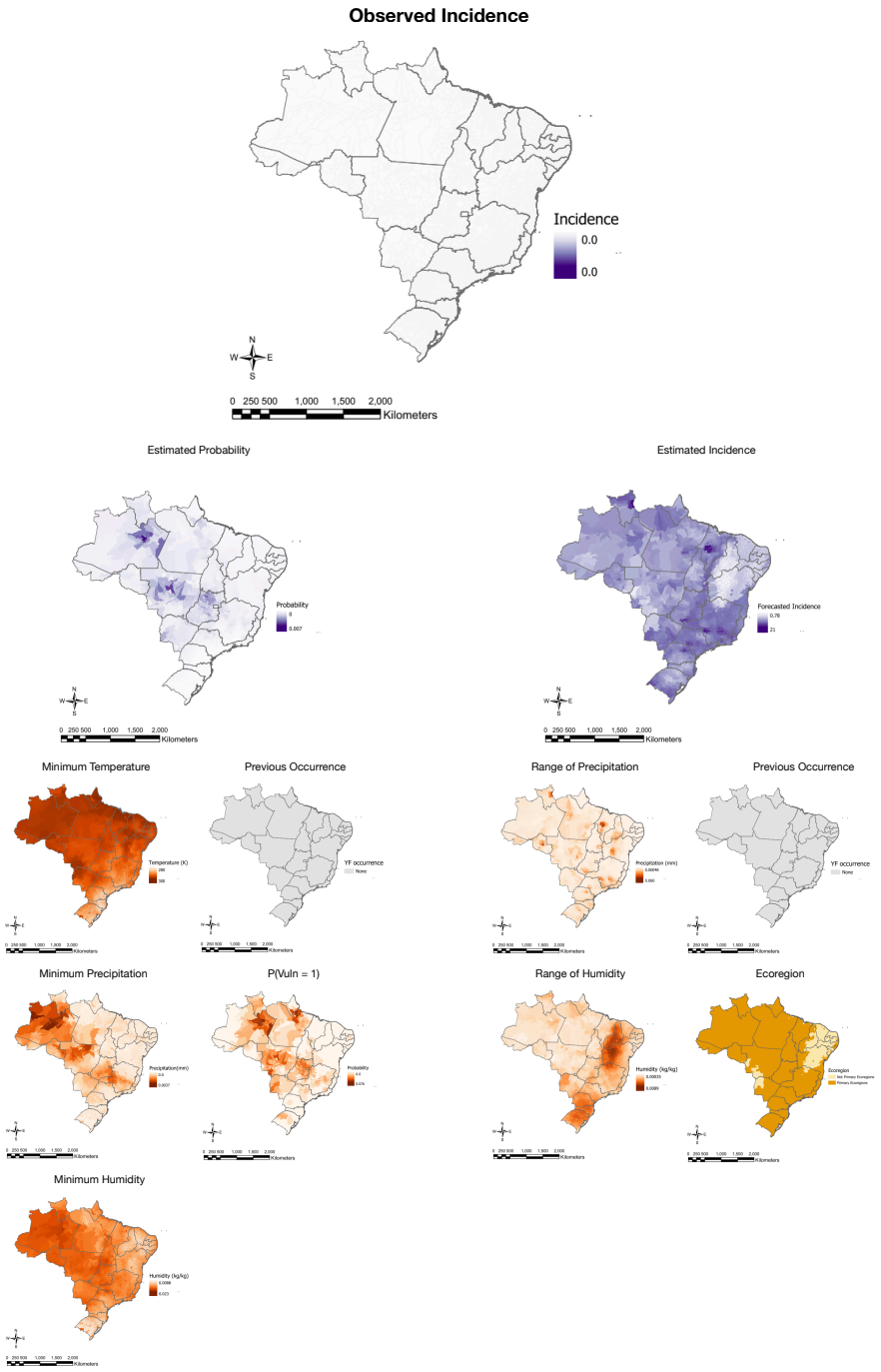


Figure 5.8. Forecasts for the week of Jan 6-12, 2014 generated by the model for the endemic time period. Observed incidence, forecasted probability of occurrence, forecasted incidence, and spatially varying predictors are shown. Predictors contributing to forecasts are shown in maps directly underneath the corresponding forecast maps. Lagged predictors used in the binomial step of the model represent Dec 2-8, 2013 (five weeks prior), and lagged predictors used in the Gamma step represent Nov 18-24, 2013 (seven weeks prior).

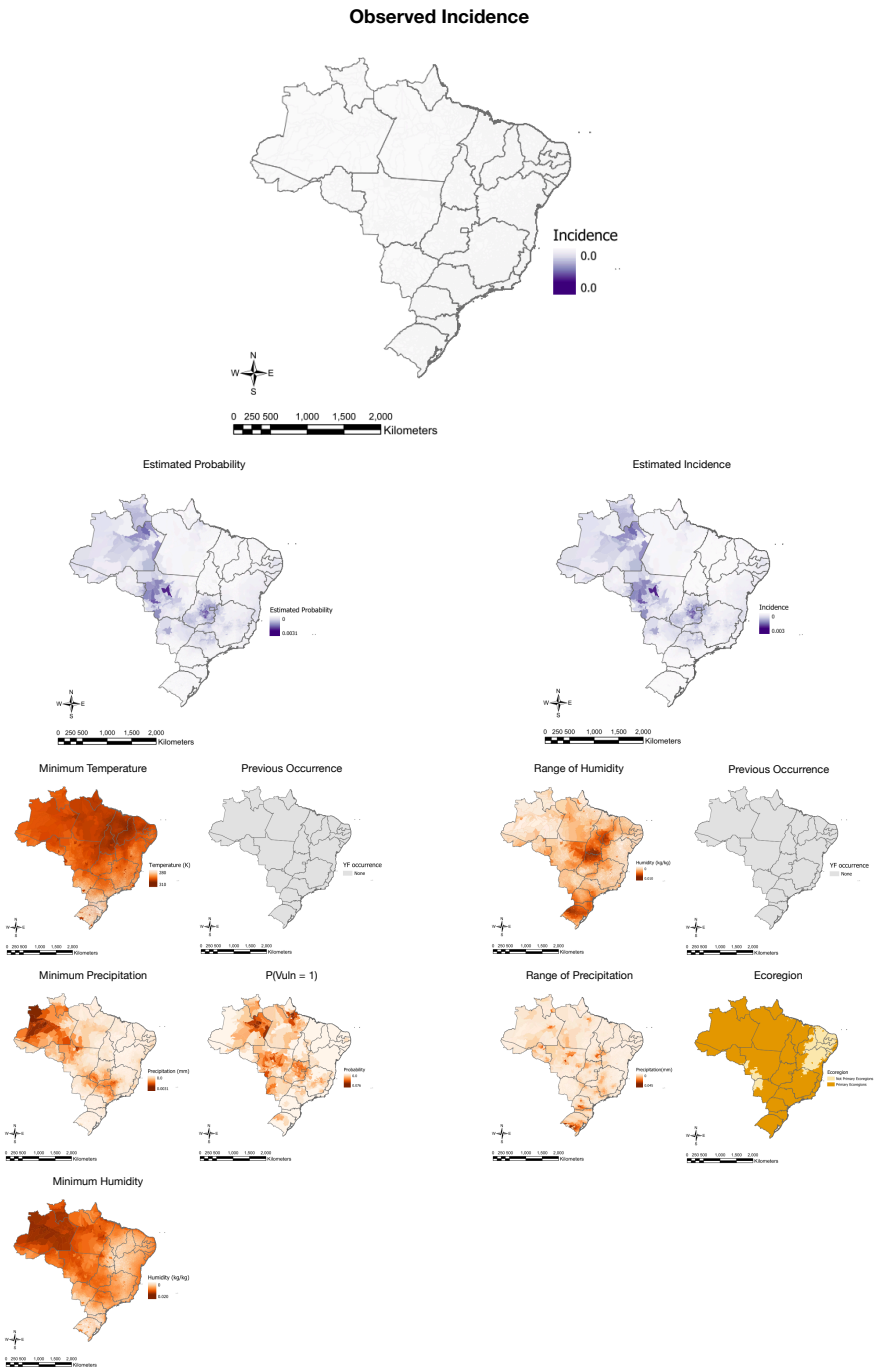


Figure 5.9. Forecasts for during the week of Dec 5-11, 2016 generated by the model for the endemic time period. Observed incidence, forecasted probability of occurrence, forecasted incidence, and spatially varying predictors are shown. Predictors contributing to forecasts are shown in maps directly underneath the corresponding forecast maps. Lagged predictors used in the binomial step of the model represent Oct 31-Nov 6, 2013 (five weeks prior), and lagged predictors used in the Gamma step represent Oct 17-23, 2013 (seven weeks prior).

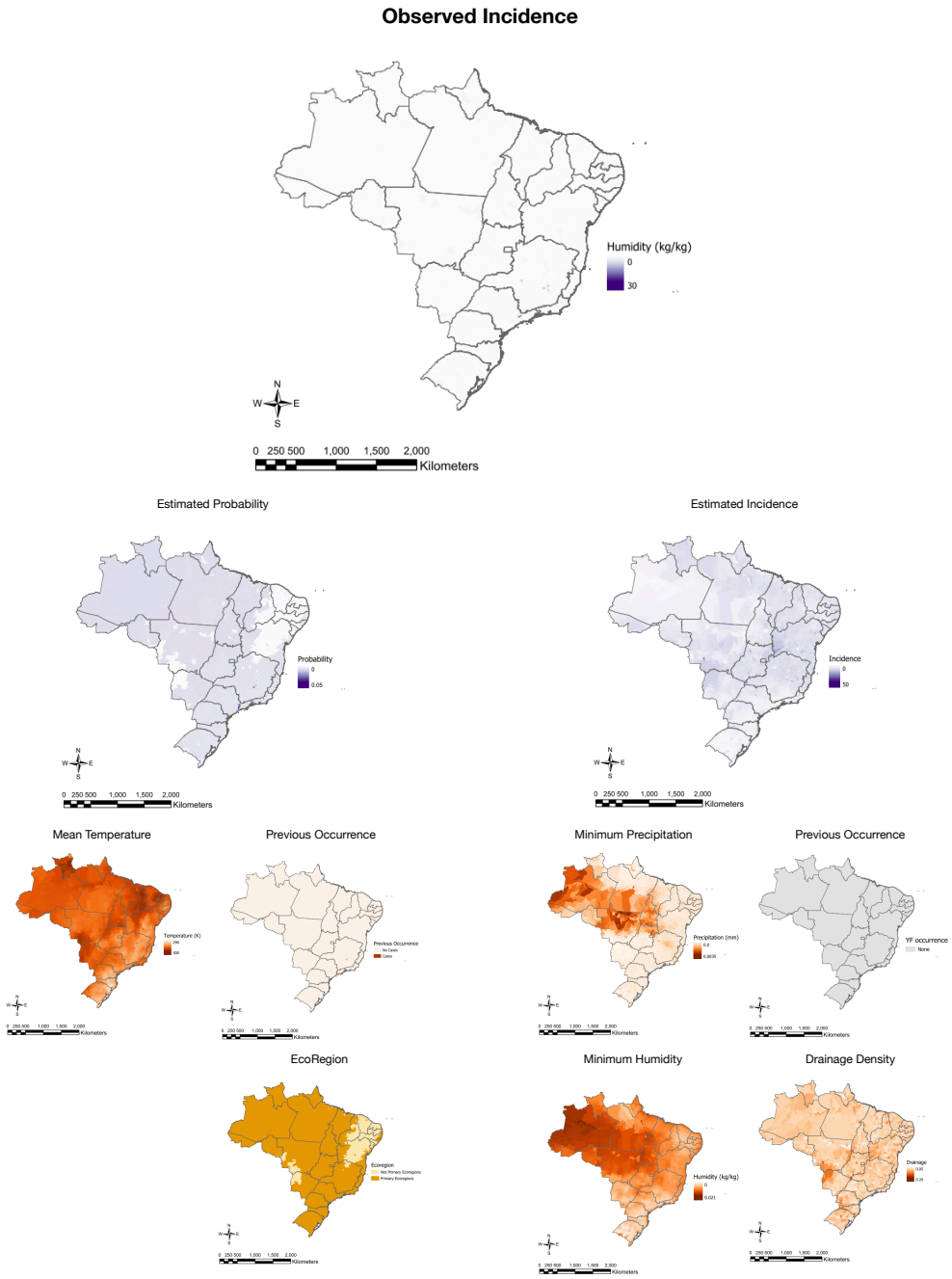


Figure 5.10. Forecasts for during the week of Jan 1-7, 2018 generated using the model for the epidemic time period Observed incidence, forecasted probability of occurrence, forecasted incidence, and spatially varying predictors are shown. Predictors contributing to forecasts are shown in maps directly underneath the corresponding forecast maps. Lagged predictors used in the binomial step of the model represent Dec 18-24, 2017 (two weeks prior), and lagged predictors used in the Gamma step represent Nov 13-19, 2013 (seven weeks prior).

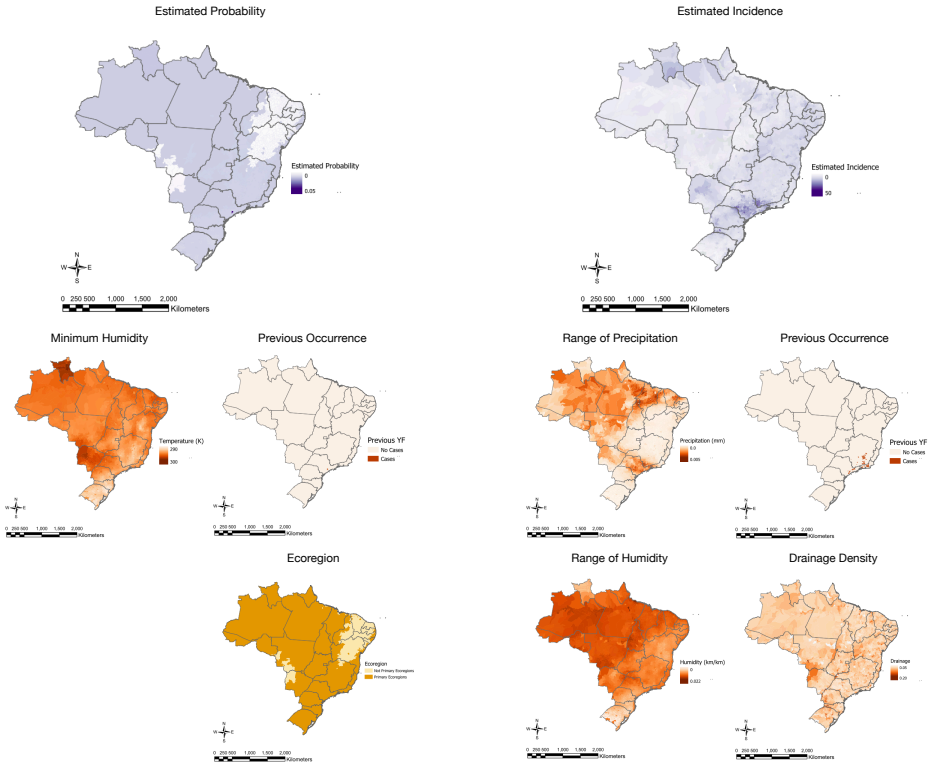


Figure 5.11. Forecasts for during the week of Apr 2-8, 2018 generated using the model for the epidemic time period. Forecasted probability of occurrence, forecasted incidence, and spatially varying predictors are shown. Predictors contributing to forecasts are shown in maps directly underneath the corresponding forecast maps. Observed incidence is not present in the data set. Lagged predictors used in the binomial step of the model represent Mar 19-25, 2018 (two weeks prior), and lagged predictors used in the Gamma step represent Feb 12-18, 2018 (seven weeks prior).

The forecasts using the model from the endemic time period both show greater probability of YF occurrence in western municipalities, consistent with the locations of most cases during those years (Figure 5.8, Figure 5.10). By contrast, the forecasts using the model from the epidemic time predict greater probability in eastern municipalities as well. The forecast plots for incidence show the expected magnitude of YF cases if it is seen in the forecasted weeks, allowing comparisons regarding the municipalities that are most likely to observe YF cases or higher case counts. The endemic time period shows more variability across locations compared

to the epidemic time period, while the forecasts during the epidemic time period tend to appear more consistent across locations.

To examine the precision of the forecasts shown in Figures 5.8-5.11, 95% confidence intervals were produced for forecasted values. Forecast precision for ten selected municipalities in different states are shown through error bars in Figure 5.12. Using forecasts from each of the weeks described previously, the estimated probability of YF occurrence and the estimated YF incidence, if seen, was plotted for each of the selected municipalities with error bars. The error bars show high precision among the estimates, as the error bars are typically short.

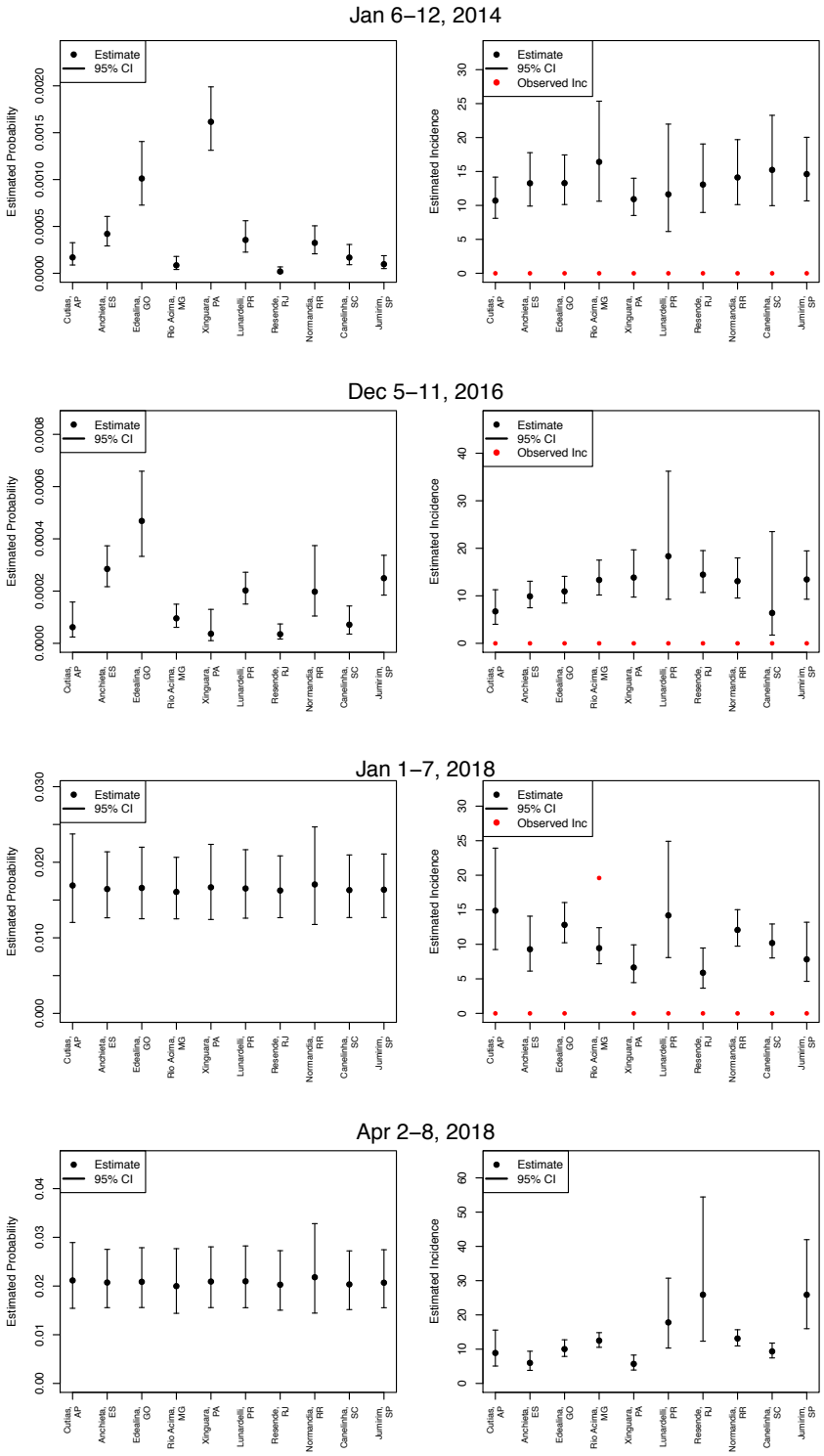


Figure 5.12. Forecasted probabilities of Yellow Fever occurrence and incidence for ten municipalities in Brazil with error bars representing 95% confidence intervals. Weeks shown are Jan 6-12, 2014; Dec 5-11, 2016; Jan 1-7, 2018; and Apr 2-8, 2018. Observed incidences are shown, when available, with incidence estimate plots.

Forecasts were also produced over the entire testing periods for each model to show predicted probabilities of YF incidence and predicted incidence over more than the two selected weeks shown previously. The testing period for the endemic time period ranges from Jan 2014 to Nov 2016, and the testing period for the epidemic time period ranges from Jan 2018 to Mar 2018. The municipalities shown represent a mixture of municipalities that saw YF cases during the testing period and those that did not.

During the endemic time period, observed YF cases were very rare during the testing period. When cases were seen, the predicted probability of YF occurrence five weeks after was notably higher, though no municipalities saw YF cases in more than one week during this time period. Throughout this entire time period, the municipalities that saw any YF cases tended to have higher predicted probabilities of YF occurrence compared to those that did not see any cases. Also, predicted incidence among municipalities that saw any cases were typically higher compared to those that did not see any during the testing period (Figure 5.13).

During the epidemic time period, more municipality-weeks with cases were observed. Similarly to the endemic time period, notable increases in predicted probability of YF occurrence were seen in the weeks following observed cases by the two-week lag period. Also similar to the previous time period was the trend of predicting higher incidence when probability of YF occurrence was higher. When YF cases did occur, some municipality-weeks' incidence values were overestimated while others were underestimated. More overestimations of incidence were seen (Figure 5.14).

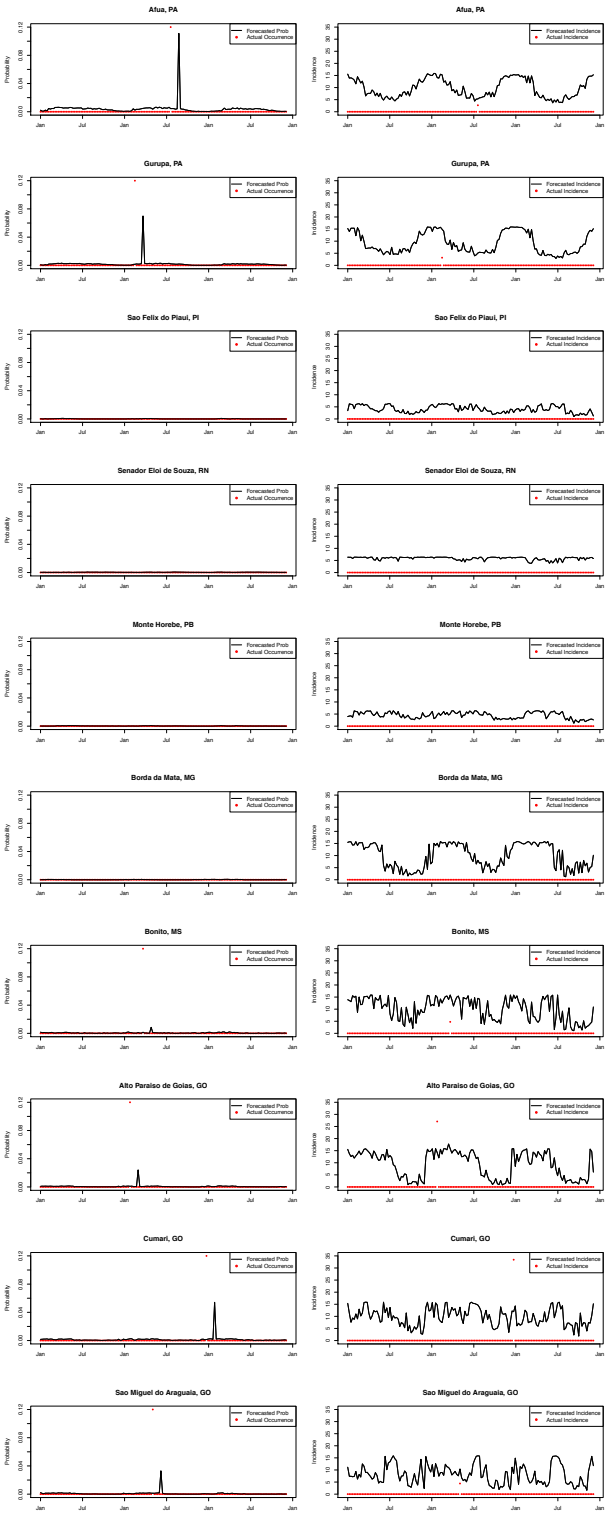


Figure 5.13. Predicted probabilities of Yellow Fever occurrence and incidence between Jan 2014 and Nov 2016 in ten selected Brazilian municipalities using the model for the endemic time period.

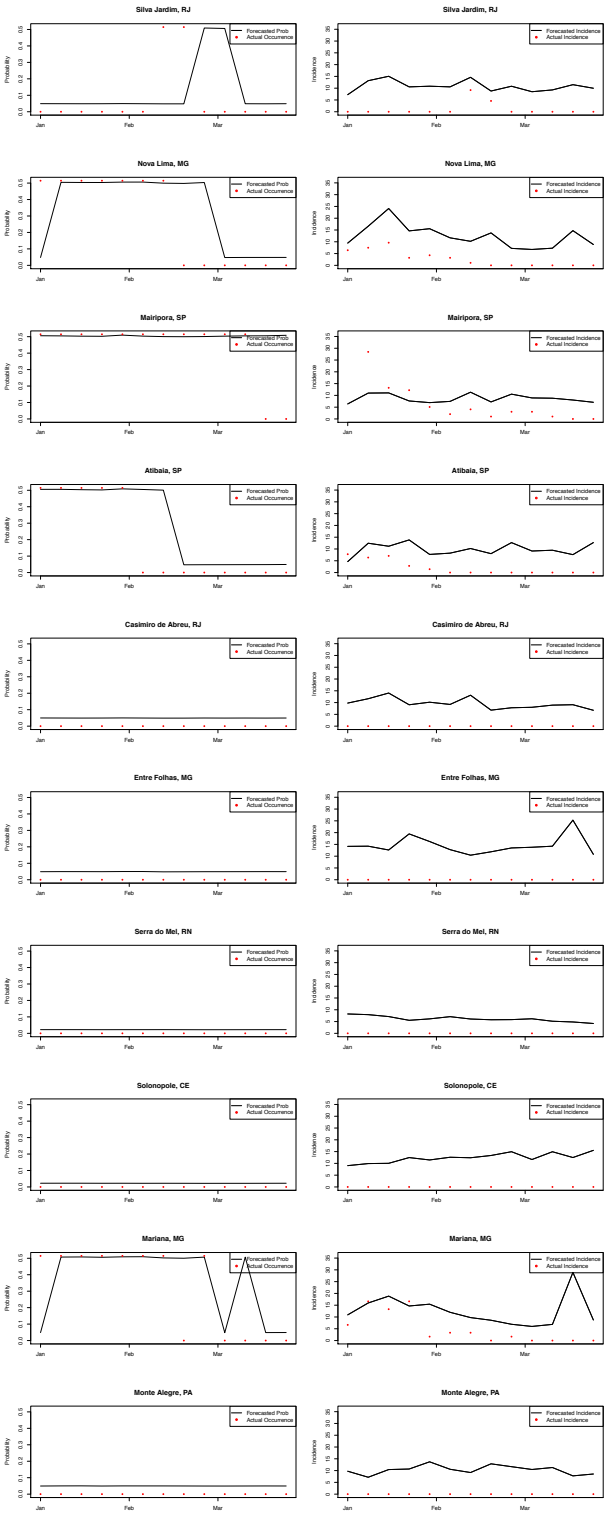


Figure 5.14. Predicted probabilities of Yellow Fever occurrence and incidence between Jan 2018 and Mar 2018 in ten selected Brazilian municipalities using the model for the epidemic time period.

5.3.4. Sensitivity analyses

A sensitivity analysis was conducted to determine the importance of the optimal lag periods. This scenario considers the possibility of stakeholders requiring a different period of time between forecast production and time of target forecast than is provided by the best-fitting models. For example, using the forecasting model from the epidemic time period (Nov 2016 – Mar 2018), a two-week lag period provides the best forecasts for the occurrence of YF, but best predictions of incidence for the same target week can be made seven weeks before the target, leaving five weeks between. This sensitivity analysis aids stakeholders who may want to make forecasts of YF occurrence and YF incidence simultaneously. While the two steps of the model can be run individually to predict probability of YF occurrence and YF incidence asynchronously, running the two steps simultaneously may be more desirable in certain contexts.

Using the parameter sets described previously and in Table 5.3, the models were refit using two-week through eight-week lags, and the model fit criteria, AUC and MAE, were evaluated for each lag period. Results of these analyses show that using the same predictor set in different lag periods can provide substantial decreases in model fit quality, though there are some alternate lag periods that can provide a similar fit (Figure 5.15). The differences in model fit were more pronounced for the binomial step of each model compared to the Gamma step.

As an example, changing the lag period for the binomial step for the model for the endemic time period (Jan 2000 – Nov 2016) from five weeks to four weeks would cause a small decrease in AUC, but changing the lag period to eight weeks would lead to a more substantial decrease. If aligning the lag periods of the two model steps, changing the lag period of the binomial step to seven weeks to match the Gamma step would reduce the AUC from 0.926 to 0.889 (Table 5.4). Conversely,

changing the lag period of the Gamma step to five weeks to match the binomial step would increase the MAE from 5.074 to 5.196. While the changes in the two metrics are not able to be directly compared (since AUC is bounded by 0 and 1 while MAE can take any real nonnegative value), a qualitative assessment would suggest that, if desired, using a five-week lag for both would lead to better fits than a seven-week lag for both steps if the asynchronous lag periods were to be avoided.

Table 5.4. Model fit criteria (AUC and MAE) for Yellow Fever forecasting models using optimal and non-optimal lag periods for the binomial and Gamma steps of the models for each time period.

Lag period	Endemic time period		Epidemic time period	
	Binomial step	Gamma step	Binomial step	Gamma step
2-week	0.868	5.167	0.842	6.924
3-week	0.892	5.261	0.788	6.873
4-week	0.902	5.156	0.779	6.805
5-week	0.926	5.196	0.778	6.816
6-week	0.875	5.193	0.762	6.795
7-week	0.889	5.075	0.773	6.772
8-week	0.860	5.183	0.758	6.828

When using the models developed here to produce forecasts for YF, these sensitivity analyses show potential loss in accuracy when using different lag periods than the ones that produced best-fitting models. If stakeholders want to use the two steps of the hurdle models to produce forecasts for a particular target week at the same time instead of during the optimal weeks, the results of this sensitivity analysis shows the potential loss in accuracy in exchange. The decision of which is more important lies with the stakeholders as any model is related to targeted objectives.

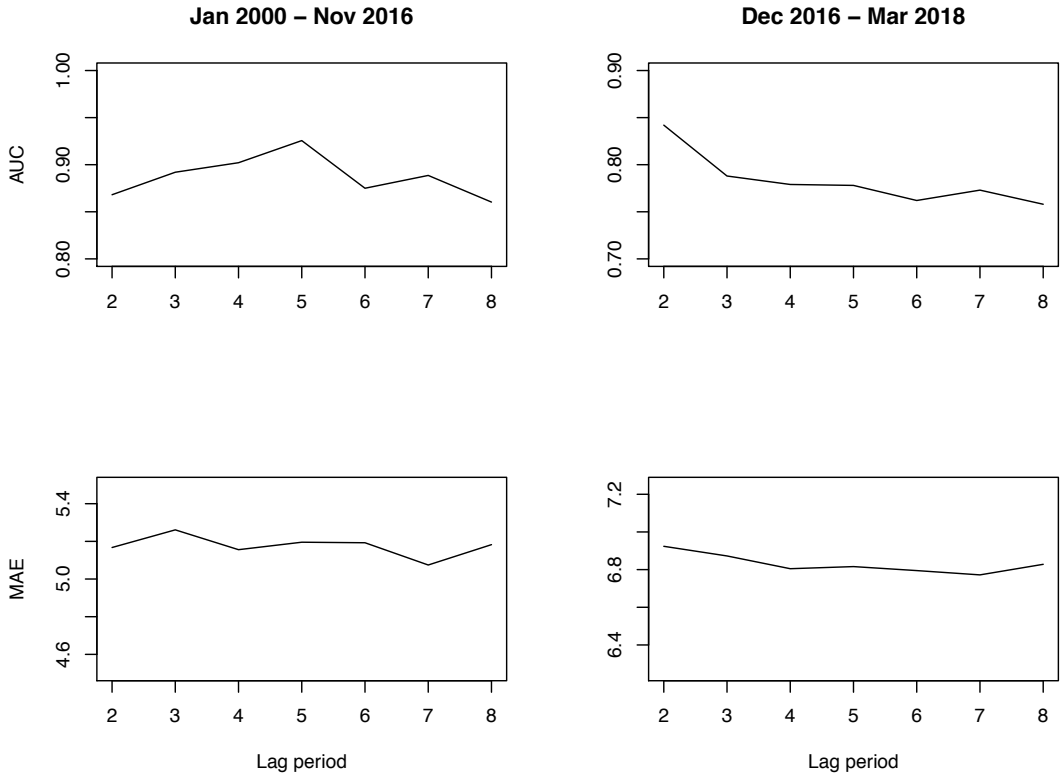


Figure 5.15. Differences in model fit resulting from using the same predictor set with different lag periods.

A global sensitivity analysis was conducted to determine which, if any, variables are most influential in each model step. First and second-order Sobol indices [244, 245] were calculated for each model step to measure relative influence. The Sobol indices were calculated for each variable by repeatedly holding the variable to a randomly drawn constant, fitting the model step to the testing data set, and then recording the mean outcome (fitted probability of case occurrence or fitted incidence). The variance of 1,000 iterations provided the first-order Sobol indices for each variable after being standardized by dividing by their sum. This was also repeated among pairs of variables, and the second-order indices for each variable were calculated by standardizing the sum of all values using that variable. Second-order Sobol indices also contain the first-order indices [245].

Across models, temperature, humidity, and previous YF occurrence were typically the most influential variables in the model (Table 5.5, Figure 5.16). Second order indices show how each variable interacts with others. All forecasting models highlight linear dynamics of YF considering the higher magnitude of first order versus second order Sobol indices. Humidity has the highest influence, both alone and in interactions with other variables, when predicting YF incidence in either time period. Temperature and previous YF occurrence are the most influential predictors, both independently and interactively with other predictors, when predicting probabilities of future YF occurrence.

Table 5.5. Sobol indices for fitted forecast models.

	Endemic time period				Epidemic time period			
	Binomial		Gamma		Binomial		Gamma	
	First-order	Second-order	First-order	Second-order	First-order	Second-order	First-order	Second-order
Temp	0.491	0.305			0.003	0.098		
$Temp^2$	0.509	0.311						
Rain	0.000	0.095	0.008	0.159			0.000	0.023
$Rain^2$								
Hum	0.000	0.095	0.890	0.508			0.185	0.447
Hum^2			0.063	0.104			0.815	0.492
Month					0.056	0.315		
Previous occurrence	0.000	0.097	0.000	0.121	0.939	0.527	0.000	0.012
Previous incidence								
P(Vuln = 0)								
P(Vuln = 1)	0.000	0.097						
P(Vuln = 2)								
Ecoregion			0.039	0.109	0.111	0.060		
Drainage density							0.000	0.025

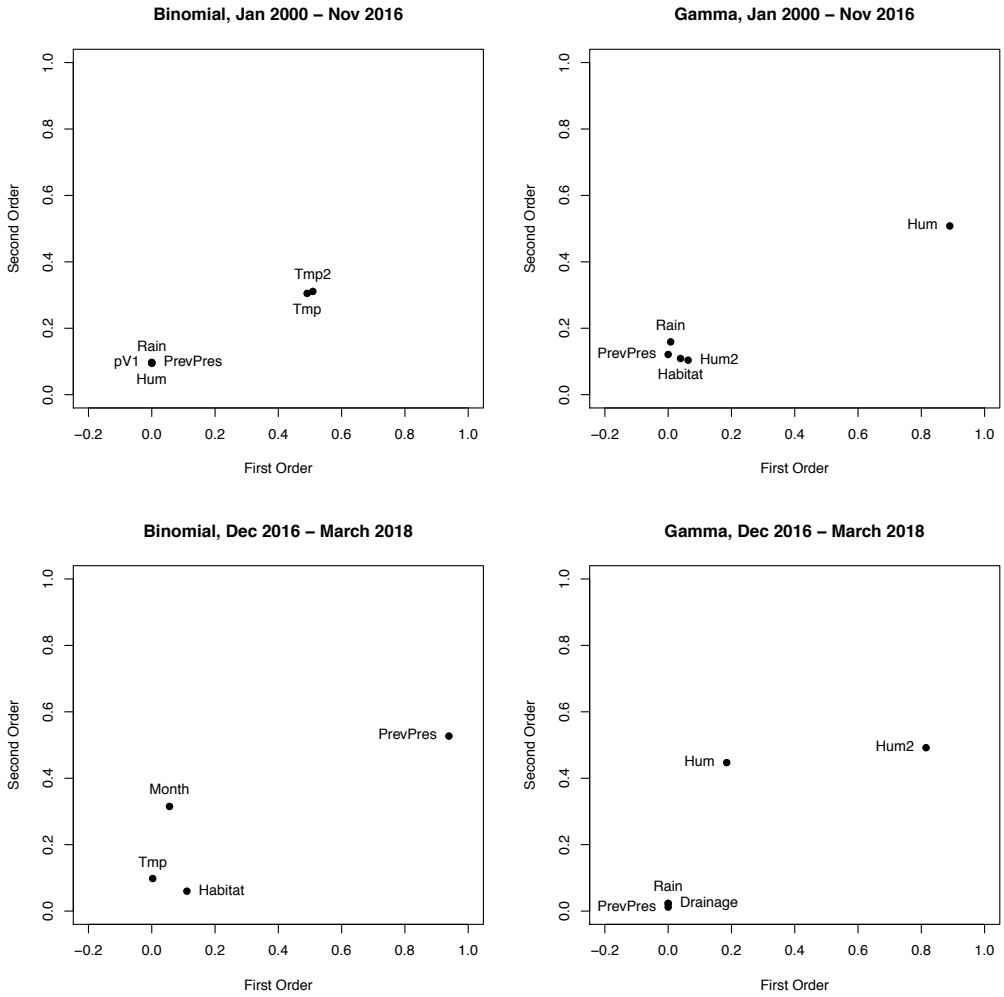


Figure 5.16. First and second order Sobol Indices for Yellow Fever forecast models. Higher values of Sobol indices indicate greater relative influence for a predictor over model predictions.

5.4. Discussion

This study aimed to develop short-term forecasting models to predict YF occurrence and incidence in Brazil using weekly environmental factors. These models were intended to forecast several weeks into the future to inform more immediate YF risk. Dynamic variables characterizing weather, particularly temperature and humidity, were shown to be influential predictors of both occurrence and incidence of YF among Brazilian municipalities; previous occurrence of YF was also an

influential predictor. Different sets of predictors provided the most accurate forecasts in the endemic (Jan 2000 – Nov 2016) and epidemic (Dec 2016 – Mar 2018) time periods, which represent time before and after the start of the 2016-2017 outbreak. The models demonstrated high predictive accuracy as defined through MAE and AUC values with the testing data, suggesting that environmental factors can be reliably used for forecasting YF burden in Brazil. Seasonality is a known characteristic of YF incidence, with cases primarily occurring between January and April [296]. The strong predictive ability of these environmental factors for future YF burden may in part result from a shared seasonality.

Differences seen in the optimal parameter sets and lag periods across the endemic and epidemic time periods suggest that some environmental factors may contribute differently to YF risk during an outbreak compared to outside of an outbreak. This may, however, also be impacted by differences in the relative lengths of the training and testing periods. In the endemic time period, the training data consists of 168 months, followed by a 35-month testing period. In the epidemic time period, the training data consists of 13 months, followed by a three-month testing period. The different lengths of both times, and small difference in the ratio between the two, may contribute to the model fitting.

Additionally, the differences seen between the two steps (binomial and Gamma) within the same hurdle model within the same time period suggests that there may be different contributing factors to seeing any YF cases and to seeing a greater incidence when observing cases. The same candidate predictors were considered for both steps, and the predictor sets that produced the most accurate forecasts are shown in Table 5.3. This was also seen with the temporal lag periods; three different lag periods were observed among the four model components (Table 5.3). Previous research has noted the complex dynamics between environmental factors and vector-borne disease risk [108, 297]. Results from this study support this notion that the relationships between environmental characteristics and disease risk are

not straightforward and may change in different circumstances, which has been observed in Zika research [293].

Based on the Sobol indices for model parameters, temperature, humidity, and previous YF occurrence were shown to be most influential predictors in the forecasting models (Table 5.5, Figure 5.16). Connections between these variables have been made previously to YF [114, 298], other mosquito-borne diseases [299], and mosquito vectors in general [285]. Temperature and humidity predictors included both linear and quadratic terms in three of the four combined instances of their inclusions, supporting hypotheses of nonlinear trends connecting disease risk and environmental factors [107, 293].

By producing forecasts with the models developed in this study, specific municipalities can be identified that are most likely to see any YF cases or are expected to have higher incidence if YF cases are to be seen, as seen in Figures 5.8-5.11. This allows for specific, targeted interventions or preventative measures. Using weather and other environmental predictors, which are not able to be targets of interventions, provides insight onto the ability to forecast YF occurrence and incidence based on uncontrollable phenomena, magnifying the importance of other preventative measures. Because the predictors used in this study have been seen to relate to mosquito life cycles and activities [300, 301], the forecasts produced in this study likely primarily represent risk from mosquito populations. Understanding this risk is important for early warning systems and other precautions.

Implementing the models for contemporary forecasting would require consideration of how the current dynamics of YF in Brazil compare to the two time periods used in modeling. That is, determining which model to use today for forecasting YF would require assumptions to be made. Reported case counts of YF early in the 2018 – 2019 YF season were much lower than during the beginning of 2017 and 2018 [261], suggesting that current YF dynamics may more closely

resemble those during the endemic time period. However, it is also possible that effects of this outbreak are relevant to current YF dynamics, indicating that the model from the endemic time period by itself may not be entirely appropriate.

The model for the endemic time period showed greater accuracy compared to the model for the epidemic time period based on both the AUC and MAE (Table 5.3). The data contributing to those models have the advantage of having multiple full years of data for both the training and testing data sets. The epidemic time period, due to having a narrower time window, used thirteen months for training data and three months for testing data. Having longer time periods for both may yield different, more accurate forecasts. However, it is also possible that, during an epidemic, disease occurrence and incidence become more unpredictable, and producing accurate forecasts becomes more challenging.

The example forecasts for incidence of YF, given occurrence, yielded overall similar trends in relative severity across Brazil, as seen in Figures 5.8-5.11. Using previous occurrence of YF as a predictor is likely a contributor to this finding since the training data sets for the two models showed YF incidence in different locations of Brazil. This is consistent with other disease prediction efforts, which often assume that recent disease incidence will inform current or future incidence [282].

Previous works have, similarly to this study, established connections between YF risk and environmental factors. Some have used similar environmental predictors, including temperature, precipitation [86, 91, 92, 114], and ecoregion [54]. Humidity, a notable predictor in many of the models in this study, is less commonly seen in YF studies, despite relating to mosquito life and virus incubation [301]. Other studies on this topic also include temporal lags to represent delayed effects in the associations between environmental events and YF risk [86]. Other studies, unlike the present study, also include predictors such as mosquito habitats [50, 86, 92, 93], nonhuman primate habitats [54, 86, 92], vegetation [91–93, 106], and human

demographics [91, 106]. Some predictors included in other studies, such as elevation [54, 93] and population density [86, 92], contribute to the vulnerability probabilities computed previously and used as predictors.

The model forms used in other works predicting YF risk bear similarities and differences to the Gamma hurdle model used in this study. Many other studies use binary regression in order to compare occurrence and absence of YF [50, 54, 91, 92]. Similar methods such as regression trees [93] are used in other works. Others, rather than using statistical models, use mathematical or mechanistic models [86, 87, 90], including in conjunction with a statistical model [106].

The objective of the modeling in this study was to produce accurate short-term forecasts for YF in order to predict locations where it is likely to be seen in the near future, along with an estimate of incidence if seen. Similarly motivated work has been seen with dengue, where precipitation and temperature were used to distinguish between epidemic and non-epidemic years in seven cities in Brazil [302]. A machine learning algorithm was employed to identify a time window for temperature and precipitation that produces the best prediction, allowing variation in both length of window and time between the time window and the prediction time period. Results identified the time of year to best predict whether the city would experience a dengue epidemic [302]. While able to estimate municipality-specific effects, using a limited set of municipalities makes results less generalizable. Other similar studies focused on YF. One used mechanistic and statistical models together, incorporating weather patterns, primate populations, mosquito populations, and human populations in order to predict YF spillover from nonhuman primates to humans by municipality-month in Brazil [86]. The study also predicted YF over the entire nation of Brazil using weather data, but additionally incorporated mosquito, nonhuman primate, and human vaccination data. The study only predicted spillover from other populations without estimating incidence [86]. Another used generalized linear models incorporating environmental and

geographic data, such as vegetation and land use, to predict YF occurrence in several nations in Africa. The fitted model was then used, with data including demographics and vaccination coverage, to predict disease burden in the forms of death counts, severe case counts, and disability adjust life years lost [106]. Estimates were made at a fine spatial resolution across several countries, providing one-time estimates [106]. Other works relating environmental factors to YF risk have aimed to predict seasonal trends [91] or produce one-time risk associations [54, 93, 114].

5.4.1. Implications for use

These analyses were developed to consider use for preparedness planning, whether through government or research-based groups. The models developed in this study primarily rely on publically available data sets and are adaptable for practical use. A minimum of two weeks was considered for lag times to represent adequate time to prepare and implement any responses to the provided forecasts. As seen through the sensitivity analysis shown in Figure 5.15, if other lag times are required or desired outside the optimal lag time found during model fitting, there may be substantial loss of model fit for the binomial step of the model, but less loss of fit for the Gamma step of each model.

One challenge of implementing this model for practical use is the required computational infrastructure and data management, as weekly data updates are required for predictors. While the longitudinal data used in model fitting would not be necessary for regular updated forecasts, the raw files for environmental variables can be large and require cleaning in order to be suitable for use. This would provide challenges in resource-constrained environments.

5.4.2. Strengths and novelty

The analyses conducted in this study were intended to be able to provide frequent updates for YF risk, with the ability to regularly update the current locations that are expected to need targeted interventions. Other studies examining environmental risk factors for YF infection provide a snapshot [54, 114] or overall comparisons of risk by location [73], but this study instead aimed to provide recommendations that can be more easily updated regularly.

The use of a two-step Gamma hurdle model is a point of novelty in this study. While hurdle models have been used before in vector-borne disease research [293], they are not consistently used, despite the advantage of being able to separate predictors of occurrence from predictors of incidence, as highlighted in this study. They are also useful when encountering zero-inflated data, as was seen in this study. Gamma regression models are also uncommon, despite the Gamma distribution being observed to closely fit disease incidence data.

This study also considered four different forms to operationalize weather variables as well as seven different weekly lag lengths. While consideration of extreme values and ranges is discussed previously, few works seek to fit models using multiple forms to determine which provides the best fit.

5.4.3. Limitations and future directions

While this study aimed to focus on environmental predictors and made use of publically available data, other factors not accounted for in these analyses could yield more accurate predictions and explain more of the dynamics of YF spread [293], though potentially with a loss of usability. Other works have used human mobility [50], mosquito movements, and nonhuman primate movement [86, 92]. However, the models presented in this study were shown to produce accurate forecasts as measured by AUC and MAE. Adding other predictors on its own is not a

guarantee of improved accuracy, as evidenced by the fact that saturated models in this study did not provide the most accurate forecasts. Furthermore, there have been established connections between the weather patterns examined in this study and movements and behaviors among humans, mosquitoes, and nonhuman primates, suggesting that the environmental factors used in this study may, in part, account for some of these, assuming that these effects on YF risk might be downstream of environmental determinants.

In aiming to determine the utility of primarily environmental predictors and focusing on YF determinants that are not potential targets of intervention, certain determinants of YF risk related to human activities are not accounted for in the models. These can include vaccination, socioeconomic status [106], and specific human mobility patterns [50]. However, while these factors can be highly influential in YF risk prediction, the results of this study, focusing solely on the environmental predictors, showed that predictive models without human activities are still useful for provide accurate forecasts.

The design of this study focused on forecasting results and predictive ability. As a result, the predictor sets were not chosen with the intent of producing individually interpretable effect estimates, but rather to show how the variables predict YF burden in concert. Even in the absence of strong multicollinearity, the predictors were not selected in order to produce an unconfounded effect estimate for any particular predictor due to the study design and objective of producing accurate forecasts. While the parameter estimates are not without meaning, the specific values are meant to be interpreted for prediction along with the other predictors rather than as a single, isolated causal mechanism.

An extension of this work could seek the specific effect of a particular weather event including a storm or a heat wave. In such a situation, determining other predictors to assure an accurate effect estimate is prioritized above predictive ability. This

study has the opposite aim, prioritizing predictive ability above specific causal effect estimates. Results form an extension of this nature could yield specific targeting warning systems relating increased YF risk to atypical weather events.

5.5. Conclusions

Weekly occurrence and incidence of YF can be forecasted in the short term with accuracy using only meteorological and ecological conditions along with previous cases information. Required quantities of training data and forecasting ability throughout different parts of the year were examined. Specifically, forecasts based on precipitation, humidity, and previous occurrence of cases in particular were accurate both before and after the start of a major epidemic. These forecasts can be produced between two and seven weeks prior to the target week. The results from these models are usable for informing preventative measures to prevent YF cases by accounting for risk factors outside of human control. The methods used in this study are translatable to similar diseases and demonstrate the use of forecasting models for informing public health practice.

Chapter 6 Conclusions

The aim of this dissertation was to provide insight into the use of quantitative analyses to improve public health outcomes through improving preparations for infectious disease prevention. Focusing specifically on YF in Brazil as an applicaiton, additional perspective was produced regarding the likelihood of fatality as well as the ability for environmental predictors to estimate both long-term and short-term burden. Additionally, assumptions used in modeling were highlighted in this work, both as a focal point of one study as well as through sensitivity analyses of others. The findings within these studies are of interest to substantive infectious disease epidemiologists and methodological researchers alike, as new substantive information was generated through these studies while also exploring relevant concerns in the methods used to produce such information.

Chapter 2 placed the choice of temporal resolution, through daily, weekly, and monthly YF incidence, as the focus of the study. By fitting probability distributions to incidence rather than imposing a model using other data sources, the characteristics of incidence data themselves served as the focal point. The probability distributions are assumed to represent the underlying data generating processes of incidence, and comparing the distributional fits across different resolutions shows whether they are interchangeable. Particularly during an epidemic, the distributional fits representing disease incidence are not interchangeable across temporal resolutions. This may have implications for future modeling studies in that, if different data generating processes exist at different resolutions, models attempting to study them may not be appropriate at multiple resolutions.

Chapter 3 provided an update to the estimate of the case fatality risk among severe YF cases through a systematized literature review. Compiling data from multiple

sources in the literature showed that the estimated CFR is lower than what is typically stated by the WHO and in other literature. Geographic heterogeneity is possible, as reported fatality rates in South America were typically higher than those in Africa.

Chapter 4 defined vulnerability to YF as the ability of seeing high incidence over a year, representing long-term burden. It was predicted using static features of the natural environment and estimated over the nation of Brazil by municipality. The results highlight municipalities that, despite not having seen YF cases in recent years, are more likely to experience high YF burden due to characteristics of the environment that are not immediately changeable.

Chapter 5 produced statistical models using variable features of the environment to produce short-term forecasts of YF presence and incidence in Brazil. Using only environmental factors successfully produced forecasts that were accurate in comparison to observed data. With optimal forecast times ranging from two to seven weeks ahead, the lagged effects of weather events on YF risk were highlighted as well as the summary statistics of weather patterns that best predict YF burden.

While many of the studies described in this dissertation are distinct from each other, many of the concepts introduced are applicable to each other. For example, the studies in chapters 4 and 5 use different temporal resolutions. While the model forms are also notably different, it can be seen that the models differ regarding the inclusions of some parameters that were considered in both models. This serves as anecdotal support of the notion from Chapter 2 that incidence at different temporal resolutions likely are guided by different mechanisms, as implied by differences in distributional fit.

6.1. Substantive contributions

The study in Chapter 2 provides quantitative evidence to a concept that is overlooked, but also intuitive in nature. While the notion that altering the temporal resolution of data can affect probabilistic representations of it may seem intuitive, this study quantitatively shows the extent to which this exists. It also shows that this is not always the case in situations where disease occurrence may be rare. Viewing this study as an examination of what may be a temporal analogue to ecological fallacy [142] or the modifiable areal unit problem [303] serves the purpose of drawing attention to this topic. While the recommendations from this study are not that one temporal resolution is necessarily advantageous over others, it does suggest that this is a topic that should be considered when designing studies or quantitative analytic plans.

Updating the estimate for CFR in Chapter 3 is important for understanding the seriousness of YF as a disease. Having a more current estimate can prove useful in the event of a future YF outbreak. Knowledge of the proportion of severe cases that are fatal can better aid preparedness in regards to anticipating the number of deaths that may occur. The results from this study also note that there are few studies in the existing literature that provide usable data for this purpose. Furthermore, there is insufficient literature for a thorough estimation of the proportions of YF cases that are asymptomatic, mild, and severe. Highlighting these points in this study can serve as motivations for future works.

Estimating YF vulnerability across Brazil in Chapter 4 provides information regarding likely locations for high future incidence. The results from this study provide long-term targets for YF prevention measures, such as preventative vaccination [73], mosquito control [259, 260], or communication [257, 258]. The predictors used for this model are not modifiable by human means, emphasizing the importance of other preventative measures. Rather than identifying specific control

strategies, this study indicates which locations, through their environments, are predisposed to higher YF burden. The model-fitting results show that nearby presence of YF is an important predictor of categorical vulnerability, emphasizing the importance of observing YF not only within a municipality, but also across nearby municipalities.

The YF forecast models produced in Chapter 5 contribute to immediate preparedness efforts by identifying the most likely locations to see YF presence in the near future as well as those most likely to see high disease burden if present. Similarly to Chapter 4, natural phenomena were used as predictors of YF presence and incidence, showing locations that are likely to see YF burden due to factors that cannot be prevented. Through the model fitting process, previous presence of YF was shown to be a strong predictor of future burden, emphasizing the autoregressive nature of this disease.

The connections made between YF risk and the environment are consistent with previous works. Other studies have found the environmental factors used in this study to be related to risk of YF [252, 304] and other similar diseases [253, 255]. Specific findings of each study differ, highlighting the complex nature of the relationships between the environment and disease risk [108, 297]. The estimated CFR among severe YF cases of approximately 40% is lower than previously reported estimates [1, 35], but the estimates are not majorly dissimilar.

6.2. Methodological contributions

The studies in Chapters 2, 4, and 5 make note of the YF epidemic that began in Brail in Dec 2016 [58, 59]. This epidemic, and the potential differences in mechanisms determining the dynamics of YF during and after it, was considered in these studies by separating the data into two time periods, representing time before and after its

start. Based on the results of these studies and the differences when using the two time periods separately, it is recommended in other study designs to stratify infectious disease data analyses to separate epidemics from endemic periods.

Chapters 2 and 5 make use of the Gamma distribution for describing YF incidence. This distribution has advantages over others for describing disease incidence such as placing high probability on small values while having a positive support. Despite its advantages over other distributions, it is infrequently used in infectious disease contexts. It has been used in ecology [165] and economics [305] for reasons similar to those used in Chapters 2 and 5.

Similarly, the forecast models in Chapter 5 consist of Gamma hurdle models to predict both probability of YF presence as well as estimated YF incidence if present. The use of this model is uncommon across disciplines, including infectious disease research. Hurdle models themselves are atypical in infectious disease research, though they have the advantage of considering different mechanisms governing presence and incidence of a disease. The results in Chapter 5 show that, in both time periods considered, the parameter sets providing the most accurate forecasts differed between the binomial process and the Gamma process. Future works related to the topics presented in these studies may benefit from considering this model design.

Together, the studies in Chapters 4 and 5 emphasize the differentiation between long-term baseline risk and short-term future risk. The two both rely on the notion of short-term and long-term data and predictions. Considering a span of a year or longer was considered long-term in these studies, and within several weeks was considered short-term. Chapter 4 uses environmental predictors that are unlikely to fluctuate over short periods of time. Conversely, Chapter 5 specifically uses predictors that change regularly. While both contribute to the understanding of YF dynamics and the elucidation of the complex connections between disease burden

and the environment, the approaches for the two differ. From a methodological standpoint, this is important in developing studies and designing analyses, as the different perspectives taken in these two studies led to differences in analytic plans.

6.3. Future directions within arboviral research

The findings of the studies presented in this dissertation can be used in the development of future works on similar topics. Whether YF specific research or based in other arboviruses, these findings can lead to extensions for future research. Both the methodological and substantive contributions of these works can inform the directions of future works in this field.

In Chapter 5, multiple weather predictors were considered for forecasting YF burden. The aim of the study focused on accurate prediction rather than estimation of specific associations between particular weather events and YF risk. Further work could aim to identify any specific events that would be most associated with YF presence or severity. For example, some works have identified a relationship between rainfall and disease risk [86, 92, 114]. A future study could aim to identify if any anomalies, such as major storms [306], droughts [307], or high frequency of rain [284], may be a specific predictor of YF risk.

Other related directions can aim to determine if specific conditions or events can alter the overall severity of a season of YF. In Brazil, YF is mostly seen between the months of January and April [296], with the severity of the season varying across years. There may exist factors that influence whether a season is more or less severe than others. Conditions over the previous months or years may contribute notably to the following season's risk. Alternately, there may be a particular time of year when climate and weather patterns are most influential for the following season.

The study presented in Chapter 3 focused on the CFR among severe YF cases. An initial aim also existed to estimate the proportions of asymptomatic, mild, and severe cases in order to estimate the burden of disease [35], but the literature search produced too few studies for adequate analysis. Future studies can be designed to collect this data. For example, in the outbreak investigations used in this chapter, fatal and total case counts were reported. If more studies separated these case counts by mild and severe cases, the proportion of symptomatic cases that develop severe disease could be estimated. Furthermore, widespread testing among those without symptoms can offer an estimate of the number of asymptomatic YF cases.

Incorporating human activities into YF risk estimation can provide insight into the ways in which various human behaviors can affect risk of YF or other related diseases. Some studies include demographics [91, 106], but few examine individual practices such as personal use of mosquito repellent or active removal of breeding grounds. Since mosquitoes rely on standing water for breeding grounds [308, 309], comparing behaviors related to management of water can show if and how individual behaviors can affect disease risk. A challenge in such a study would be to determine whether certain activities are done in order to diligently prevent infection or in response to increased infection.

Acquiring more recent YF case data would likely improve the predictive modeling in Chapters 4 and 5. The results of both of those studies differ when using the data from Jan 2000 – Nov 2016 compared to Dec 2016 – Mar 2018. Having data from more recent months would be of use for having more data in the later time period. Furthermore, future years' data would be of interest to determine which set of models between the two time periods would be more accurate and applicable to today's YF risk. It is possible that, following the outbreak, YF prediction is better when using the model from the earlier data, which represent YF dynamics outside

an epidemic. Alternately, lasting effects of the epidemic may make the models from the later time period more applicable to today's YF risk.

6.4. Future directions across infectious disease topics

Outside of YF and other arboviruses, the methods and results of the studies presented here motivate further directions of infectious disease research. Other vector-borne diseases, such as malaria, leishmaniasis, and Lyme disease, are parasitic or bacterial and spread by mosquitos or other arthropod vectors. Connections drawn between environmental conditions and disease risk may motivate similar investigations for these diseases. The relationships seen between environmental factors and YF forecasting may be similar or different among these other vector-borne diseases.

The methods used in this study may also be of interest. Hurdle models, considering disease presence and incidence separately, can show whether the mechanisms predicting any cases or number of cases differ by including different predictors or estimating different effect sizes. Outside of vector-borne diseases, hurdle models can provide strong contributions to examine this finding among a variety of infectious diseases, including human communicable diseases.

6.5. Final conclusions

The studies described in this dissertation highlight the important connections between infectious disease dynamics and the natural environment, as well as address methodological concerns when studying the topic. These studies provide a comprehensive investigation into YF dynamics, including an evaluation of some modeling assumptions, an updated estimate of fatality, and accurate predictions of

risk across Brazil based on environmental predictors. The methods are translatable to other disease contexts and show the use of less commonly used statistical methods in this field. Knowledge of the public health impacts of the environment are important for controlling disease burden globally and reducing preventable diseases.

Bibliography

1. World Health Organization. Yellow fever [Internet]. Fact Sheets. 2018 [cited 2018 Jul 21]. Available from: http://www.who.int/en/news-room/fact-sheets/detail/yellow-fever
2. Singh KP, Gupta S, Rai P. Identifying pollution sources and predicting urban air quality using ensemble learning methods. Atmos Environ. 2013 Dec 1;80:426–37.
3. Nowak DJ, Hirabayashi S, Bodine A, Hoehn R. Modeled PM2.5 removal by trees in ten U.S. cities and associated health effects. Environ Pollut. 2013 Jul 1;178:395–402.
4. Kirwa K, Eliot MN, Wang Y, Adams MA, Morgan CG, Kerr J, et al. Residential proximity to major roadways and prevalent hypertension among postmenopausal women: Results from the women's health initiative San Diego cohort. J Am Heart Assoc [Internet]. 2014 [cited 2020 Jun 23];3(5). Available from: https://pubmed.ncbi.nlm.nih.gov/25274494/
5. Landrigan PJ, Fuller R, Acosta NJR, Adeyi O, Arnold R, Basu N (Nil), et al. The Lancet Commission on pollution and health [Internet]. Vol. 391, The Lancet. Lancet Publishing Group; 2018 [cited 2020 Jun 23]. p. 462–512. Available from: https://pubmed.ncbi.nlm.nih.gov/29056410/
6. Taylor M. Health and the built environment. Perspect Public Health [Internet]. 2016 [cited 2020 Jun 23];137(1):2. Available from: https://pubmed.ncbi.nlm.nih.gov/28074688/
7. Servadio JL, Convertino M. Optimal information networks: Application for data-driven integrated health in populations [Internet]. Science Advances American Association for the Advancement of Science; Feb 2, 2018 p. e1701088. Available from: http://advances.sciencemag.org/lookup/doi/10.1126/sciadv.1701088
8. Forno E, Celedón JC. Health Disparities in Asthma. Am J Respir Crit Care Med [Internet]. 2012 May 15 [cited 2018 Dec 16];185(10):1033–5. Available from: http://www.atsjournals.org/doi/abs/10.1164/rccm.201202-0350ED
9. Hajat A, Hsia C, O'Neill MS. Socioeconomic Disparities and Air Pollution Exposure: a Global Review. Curr Environ Heal Reports [Internet]. 2015 Dec 18 [cited 2018 Dec 16];2(4):440–50. Available from: http://link.springer.com/10.1007/s40572-015-0069-5
10. Servadio JL, Lawal AS, Davis T, Bates J, Russell AG, Ramaswami A, et al. Demographic Inequities in Health Outcomes and Air Pollution Exposure in the Atlanta Area and its Relationship to Urban Infrastructure. J Urban Heal. 2018 Nov 26;1–16.
11. Nielsen GD, Larsen ST, Olsen O, Løvik M, Poulsen LK, Glue C, et al. Do indoor chemicals promote development of airway allergy? Indoor Air [Internet]. 2007 Jun [cited 2020 Jun 23];17(3):236–55. Available from: https://pubmed.ncbi.nlm.nih.gov/17542836/

12. Arnold S, Ramachandran G, Kaup H, Servadio J. Estimating the time-varying generation rate of acetic acid from an all-purpose floor cleaner. J Expo Sci Environ Epidemiol [Internet]. 2020 Mar 1 [cited 2020 Jun 23];30(2):374–82. Available from: https://pubmed.ncbi.nlm.nih.gov/31089245/
13. Ferrey ML, Heiskary S, Grace R, Hamilton MC, Lueck A. Pharmaceuticals and other anthropogenic tracers in surface water: A randomized survey of 50 Minnesota lakes. Environ Toxicol Chem [Internet]. 2015 Nov 1 [cited 2020 Jun 23];34(11):2475–88. Available from: https://pubmed.ncbi.nlm.nih.gov/26468892/
14. Newton S, Bidleman T, Bergknut M, Racine J, Laudon H, Giesler R, et al. Atmospheric deposition of persistent organic pollutants and chemicals of emerging concern at two sites in northern Sweden. Environ Sci Process Impacts [Internet]. 2014 Feb [cited 2020 Jun 23];16(2):298–305. Available from: https://pubmed.ncbi.nlm.nih.gov/24385192/
15. Deere JR, Moore S, Ferrey M, Jankowski MD, Primus A, Convertino M, et al. Occurrence of contaminants of emerging concern in aquatic ecosystems utilized by Minnesota tribal communities. Sci Total Environ. 2020 Jul 1;724:138057.
16. Groene EA, Hernandez AM, Servadio JL, Jang S. Food access in context: Testing spatial alternatives to food desert measures in the Minneapolis-St. Paul metro area (poster). In: Society of Epidemiologic Research Annual Meeting. Baltimore, MD; 2019.
17. Centers for Disease Control and Prevention. Listeria [Internet]. 2020 [cited 2020 Jun 23]. Available from: https://www.cdc.gov/listeria/index.html
18. Lutz CS, Huynh MP, Schroeder M, Anyatonwu S, Dahlgren FS, Danyluk G, et al. Applying infectious disease forecasting to public health: A path forward using influenza forecasting examples. BMC Public Health. 2019 Dec 10;19(1).
19. Servadio JL, Machado G, Alvarez J, de Ferreira Lima Júnior FE, Vieira Alves R, Convertino M, et al. Information differences across spatial resolutions and scales for disease surveillance and analysis: The case of Visceral Leishmaniasis in Brazil. Carvalho LH, editor. PLoS One [Internet]. 2020 Jul 17 [cited 2020 Jul 17];15(7):e0235920. Available from: https://dx.plos.org/10.1371/journal.pone.0235920
20. Danasekaran R, Mani G, Annadurai K, Ramasamy J. Small Bite, Big Threat: The Burden of Vector-borne Diseases. Iran J Public Health [Internet]. 2014 Jul [cited 2020 Jun 10];43(7):1014–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25909072
21. World Health Organization. Vector-borne Diseases. Health and Environment Linkages Initiative. World Health Organization; 2005.
22. NH O. Climate Change and Vector-Borne Diseases of Public Health Significance. FEMS Microbiol Lett. 2017;364(19).
23. Jones RT, Tusting LS, Smith HMP, Segbaya S, Macdonald MB, Bangs MJ, et al. The impact of industrial activities on vector-borne disease transmission. Vol. 188, Acta Tropica. Elsevier B.V.; 2018. p. 142–51.

24. Valentine MJ, Murdock CC, Kelly PJ. Sylvatic cycles of arboviruses in non-human primates. Parasites and Vectors. 2019 Oct 2;12(1).
25. Caminade C, McIntyre KM, Jones AE. Impact of recent and future climate change on vector-borne diseases. Vol. 1436, Annals of the New York Academy of Sciences. Blackwell Publishing Inc.; 2019. p. 157–73.
26. Baylis M. Potential impact of climate change on emerging vector-borne and other infections in the UK. Vol. 16, Environmental Health: A Global Access Science Source. BioMed Central Ltd.; 2017.
27. Wilcox BA, Echaubard P, De Garine-Wichatitsky M, Ramirez B. Vector-borne disease and climate change adaptation in African dryland social-ecological systems. Vol. 8, Infectious Diseases of Poverty. BioMed Central Ltd.; 2019.
28. Centers for Disease Control and Prevention. Yellow Fever [Internet]. [cited 2019 Feb 25]. Available from: https://www.cdc.gov/yellowfever/symptoms/index.html
29. Bryant JE, Holmes EC, Barrett ADT. Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. PLoS Pathog. 2007 May;3(5):e75.
30. Monath TP. Yellow fever: an update. LANCET Infect Dis. 2001;1:11–20.
31. Monath TP, Vasconcelos PFCC. Yellow fever. J Clin Virol [Internet]. 2015 Mar 1 [cited 2018 Jul 21];64:160–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25453327
32. Rollins D, Ramsey R, Parsh B. Yellow fever. Nursing (Lond). 2017 Sep;47(9):69–70.
33. Litvoc MN, Novaes CTG, Lopes MIBF, World Health Organization. Yellow fever [Internet]. Feb, 2018 p. 106–13. Available from: http://www.who.int/en/news-room/fact-sheets/detail/yellow-fever
34. Lopes RL, Pinto JR, Junior GB da S, Santos AKT, Souza MTO, Daher EDF. Kidney involvement in yellow fever: A review. Vol. 61, Revista do Instituto de Medicina Tropical de Sao Paulo. Instituto de Medicina Tropical de Sao Paulo; 2019.
35. Johansson MA, Vasconcelos PFC, Staples JE. The whole iceberg: estimating the incidence of yellow fever virus infection from the number of severe cases. Trans R Soc Trop Med Hyg [Internet]. 2014 Aug [cited 2018 Jul 22];108(8):482–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24980556
36. Ho Y-L, Joelsons D, Leite GFC, Malbouisson LMS, Song ATW, Perondi B, et al. Severe Yellow Fever in Brazil: Clinical Characteristics and Management. J Travel Med. 2019 Jun;25(5).
37. Barrett ADT. Yellow fever live attenuated vaccine: A very successful live attenuated vaccine but still we have problems controlling the disease. Vol. 35, Vaccine. Elsevier Ltd; 2017. p. 5951–5.
38. Collins ND, Barrett ADT. Live Attenuated Yellow Fever 17D Vaccine: A Legacy Vaccine Still Controlling Outbreaks In Modern Day. Vol. 19, Current Infectious Disease Reports. Current Medicine Group LLC 1; 2017. p. 14.

39. World Health Organization. International Health Regulations, Second Edition. 2005.
40. Chippaux JP, Chippaux A. Yellow fever in Africa and the Americas: A historical and epidemiological perspective. Vol. 24, Journal of Venomous Animals and Toxins Including Tropical Diseases. BioMed Central Ltd.; 2018.
41. Burt FJ, Chen W, Miner JJ, Lenschow DJ, Merits A, Schnettler E, et al. Chikungunya virus: an update on the biology and pathogenesis of this emerging pathogen. Vol. 17, The Lancet Infectious Diseases. Lancet Publishing Group; 2017. p. e107–17.
42. Guzman MG, Gubler DJ, Izquierdo A, Martinez E, Halstead SB. Dengue infection. Nat Rev Dis Prim. 2016 Aug 18;2.
43. de Almeida MAB, Dos Santos E, da Cruz Cardoso J, da Fonseca DF, Noll CA, Silveira VR, et al. Yellow fever outbreak affecting Alouatta populations in southern Brazil (Rio Grande do Sul State), 2008-2009. Am J Primatol. 2012 Jan;74(1):68–76.
44. Gardner CL, Ryman KD. Yellow fever: A reemerging threat. Vol. 30, Clinics in Laboratory Medicine. NIH Public Access; 2010. p. 237–60.
45. Garske T, Van Kerkhove MD, Yactayo S, Ronveaux O, Lewis RF, Staples JE, et al. Yellow Fever in Africa: Estimating the Burden of Disease and Impact of Mass Vaccination from Outbreak and Serological Data. PLoS Med. 2014 May;11(5):e1001638.
46. Barrett ADT, Higgs S. Yellow Fever: A Disease that Has Yet to be Conquered. Annu Rev Entomol. 2007 Jan;52(1):209–29.
47. Dexheimer Paploski IA, Souza RL, Tauro LB, Cardoso CW, Mugabe VA, Pereira Simoes Alves AB, et al. Epizootic Outbreak of Yellow Fever Virus and Risk for Human Disease in Salvador, Brazil. Vol. 168, Annals of internal medicine. United States; 2018. p. 301–2.
48. Faria NR, Kraemer MUG, Hill S, Jesus JG de, Aguiar RS de, Iani FCM, et al. Genomic and epidemiological monitoring of yellow fever virus transmission potential. bioRxiv [Internet]. 2018 Apr 16 [cited 2018 Jul 21];299842. Available from: https://www.biorxiv.org/content/early/2018/04/16/299842
49. Cardoso J da C, de Almeida MAB, dos Santos E, da Fonseca DF, Sallum MAM, Noll CA, et al. Yellow fever virus in Haemagogus leucocelaenus and Aedes serratus mosquitoes, Southern Brazil, 2008. Emerg Infect Dis. 2010 Dec;16(12):1918–24.
50. Kraemer MUG, Faria NR, Reiner RC, Golding N, Nikolay B, Stasse S, et al. Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: a modelling study. Lancet Infect Dis. 2017 Mar 1;17(3):330–8.
51. Grobbelaar AA, Weyer J, Moolla N, Jansen van Vuren P, Moises F, Paweska JT. Resurgence of Yellow Fever in Angola, 2015-2016. Vol. 22, Emerging infectious diseases. United States; 2016. p. 1854–5.
52. Tolle MA. Mosquito-borne diseases. Curr Probl Pediatr Adolesc Health Care.

2009 Apr;39(4):97–140.
53. Tomori O. Yellow Fever in Africa: Public Health Impact and Prospects for Control in the 21st Century. Biomedica. 2002 Jun;22(2):178–210.
54. Hamrick PN, Aldighieri S, Machado G, Leonel DG, Vilca LM, Uriona S, et al. Geographic patterns and environmental factors associated with human yellow fever presence in the Americas. Barker CM, editor. PLoS Negl Trop Dis [Internet]. 2017 Sep 8 [cited 2018 Jul 21];11(9):e0005897. Available from: http://dx.plos.org/10.1371/journal.pntd.0005897
55. Dou XF, Zheng Y, Lyu YN, Li J, Li XY, Chen LJ, et al. [The first confirmed imported case of yellow fever in China]. Zhonghua Liu Xing Bing Xue Za Zhi [Internet]. 2016;37(6):788–90. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med12&NEWS=N&AN=27346103
56. Eckert J. Every prospect of a healthy summer: the 1839 outbreak of yellow fever in Charleston, South Carolina. Trans Stud Coll Physicians Phila [Internet]. 1992;14(2):167–75. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=1621279
57. Eckert J. In the days of the epidemic: the 1793 yellow fever outbreak in Philadelphia as seen by physicians. Trans Stud Coll Physicians Phila [Internet]. 1993;15(5):31–8. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=8128536
58. Goldani LZ. Yellow fever outbreak in Brazil, 2017. Brazilian J Infect Dis. 2017 Mar 1;21(2):123–4.
59. Moreira-Soto A, Torres MC, Lima de Mendonça MC, Mares-Guia MA, dos Santos Rodrigues CD, Fabri AA, et al. Evidence for multiple sylvatic transmission cycles during the 2016-2017 yellow fever virus outbreak, Brazil. Clinical Microbiology and Infection. 2018;
60. Possas C, Lourenço-de-Oliveira R, Tauil PL, Pinheiro F de P, Pissinatti A, Cunha RV da, et al. Yellow fever outbreak in Brazil: the puzzle of rapid viral spread and challenges for immunisation. Vol. 113, Memorias do Instituto Oswaldo Cruz. NLM (Medline); 2018. p. e180278.
61. Fernandes NCC de A, Cunha MS, Guerra JM, Ressio RA, Cirqueira CDS, Iglezias SD, et al. Outbreak of Yellow Fever among Nonhuman Primates, Espirito Santo, Brazil, 2017. Emerg Infect Dis. 2017 Dec;23(12):2038–41.
62. Centers for Disease Control and Prevention. Areas with Risk of Yellow Fever Virus Transmission in South America [Internet]. 2018 [cited 2020 Apr 24]. Available from: https://www.cdc.gov/yellowfever/maps/south_america.html
63. de Goes P, Guimaraes JC, Machado RD, Lobo GG, Andrade CM, Bastos RA, et al. On an outbreak of sylvan yellow fever verified in the State of Goias (Brazil) in the period from 1972--1973. An Microbiol (Rio J). 22:9–34.
64. Barros ML, Boecken G. Jungle yellow fever in the central Amazon. Vol. 348, Lancet (London, England). England; 1996. p. 969–70.

65. Outbreak news. Yellow fever, Uganda. Relev Epidemiol Hebd. 2011 Jan;86(5):37–8.
66. Addy PA, Minami K, Agadzi VK. Recent yellow fever epidemics in Ghana (1969-1983). East Afr Med J. 1986;63(6):422–34.
67. Green A. Yellow fever continues to spread in Angola. Vol. 387, Lancet (London, England). England; 2016. p. 2493.
68. World Health Organization. Yellow fever – Brazil [Internet]. Disease Outbreak News. 2019 [cited 2020 Jun 5]. Available from: https://www.who.int/csr/don/11-february-2019-yellow-fever-brazil/en/
69. Govindarajan M, Rajeswary M. Ovicidal and adulticidal potential of leaf and seed extract of Albizia lebbeck (L.) Benth. (Family: Fabaceae) against Culex quinquefasciatus, Aedes aegypti, and Anopheles stephensi (Diptera: Culicidae). Parasitol Res. 2015 May;114(5):1949–61.
70. Gratz NG, Jany WC. What role for insecticides in vector control programs? Am J Trop Med Hyg. 1994;50(6 Suppl):11–20.
71. Preet S, Sneha A. Biochemical evidence of efficacy of potash alum for the control of dengue vector Aedes aegypti (Linnaeus). Parasitol Res. 2011 Jun;108(6):1533–9.
72. Paploski IAD, Rodrigues MS, Mugabe VA, Kikuti M, Tavares AS, Reis MG, et al. Storm drains as larval development and adult resting sites for Aedes aegypti and Aedes albopictus in Salvador, Brazil. Parasit Vectors. 2016 Jul;9(1):419.
73. Shearer FM, Moyes CL, Pigott DM, Brady OJ, Marinho F, Deshpande A, et al. Global yellow fever vaccination coverage from 1970 to 2016: an adjusted retrospective analysis. Lancet Infect Dis. 2017 Nov 1;17(11):1209–17.
74. Kotsakiozi P, Gloria-Soria A, Caccone A, Evans B, Schama R, Martins AJ, et al. Tracking the return of Aedes aegypti to Brazil, the major vector of the dengue, chikungunya and Zika viruses. PLoS Negl Trop Dis. 2017 Jul 25;11(7).
75. Bae H-G, Nitsche A, Teichmann A, Biel SS, Niedrig M. Detection of yellow fever virus: a comparison of quantitative real-time PCR and plaque assay. J Virol Methods. 2003 Jun;110(2):185–91.
76. Drosten C, Gottig S, Schilling S, Asper M, Panning M, Schmitz H, et al. Rapid detection and quantification of RNA of Ebola and Marburg viruses, Lassa virus, Crimean-Congo hemorrhagic fever virus, Rift Valley fever virus, dengue virus, and yellow fever virus by real-time reverse transcription-PCR. J Clin Microbiol. 2002 Jul;40(7):2323–30.
77. Liu J, Ochieng C, Wiersma S, Stroher U, Towner JS, Whitmer S, et al. Development of a TaqMan Array Card for Acute-Febrile-Illness Outbreak Investigation and Surveillance of Emerging Pathogens, Including Ebola Virus. J Clin Microbiol. 2016 Jan;54(1):49–58.
78. Nordstrom H, Falk KI, Lindegren G, Mouzavi-Jazi M, Walden A, Elgh F, et al. DNA microarray technique for detection and identification of seven flaviviruses pathogenic for man. J Med Virol. 2005 Dec;77(4):528–40.
79. Reusken CBEM, Knoester M, GeurtsvanKessel C, Koopmans M, Knapen DG, Bierman WFW, et al. Urine as Sample Type for Molecular Diagnosis of Natural

Yellow Fever Virus Infections. Vol. 55, Journal of clinical microbiology. United States; 2017. p. 3294–6.
80. Sanchez-Seco MP, Rosario D, Hernandez L, Domingo C, Valdes K, Guzman MG, et al. Detection and subtyping of dengue 1-4 and yellow fever viruses by means of a multiplex RT-nested-PCR using degenerated primers. Trop Med Int Health. 2006 Sep;11(9):1432–41.
81. Domingo C, Charrel RN, Schmidt-Chanasit J, Zeller H, Reusken C. Yellow fever in the diagnostics laboratory review-article. Vol. 7, Emerging Microbes and Infections. Nature Publishing Group; 2018.
82. Mendez MC, Domingo C, Tenorio A, Pardo LC, Rey GJ, Mendez JA. Development of a reverse transcription polymerase chain reaction method for yellow fever virus detection. Biomedica. 2013 Sep;33 Suppl 1:190–6.
83. Domingo C, Escadafal C, Rumer L, Méndez JA, García P, Sall AA, et al. First international external quality assessment study on molecular and serological methods for yellow fever diagnosis. PLoS One. 2012 May 3;7(5).
84. Assuncao-Miranda I, Cruz-Oliveira C, Neris RLS, Figueiredo CM, Pereira LPS, Rodrigues D, et al. Inactivation of Dengue and Yellow Fever viruses by heme, cobalt-protoporphyrin IX and tin-protoporphyrin IX. J Appl Microbiol. 2016 Mar;120(3):790–804.
85. Julander JG, Morrey JD, Blatt LM, Shafer K, Sidwell RW. Comparison of the inhibitory effects of interferon alfacon-1 and ribavirin on yellow fever virus infection in a hamster model. Antiviral Res [Internet]. 2007;73(2):140–6. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=17049380
86. Childs ML, Nova N, Colvin J, Mordecai EA. Mosquito and primate ecology predict human risk of yellow fever virus spillover in Brazil. Philos Trans R Soc B Biol Sci. 2019 Sep 30;374(1782).
87. Bonin CRB, Fernandes GC, dos Santos RW, Lobosco M. A qualitatively validated mathematical-computational model of the immune response to the yellow fever vaccine. BMC Immunol. 2018 May 25;19(1).
88. Zhao S, Stone L, Gao D, He D. Modelling the large-scale yellow fever outbreak in Luanda, Angola, and the impact of vaccination. PLoS Negl Trop Dis. 2018 Jan 1;12(1).
89. Ribeiro AF, Tengan C, Sato HK, Spinola R, Mascheretti M, França ACC, et al. A public health risk assessment for yellow fever vaccination: A model exemplified by an outbreak in the state of São Paulo, Brazil. Mem Inst Oswaldo Cruz. 2015 Apr 1;110(2):230–4.
90. Glover A, White A. A vector-host model to assess the impact of superinfection exclusion on vaccination strategies using dengue and yellow fever as case studies. J Theor Biol. 2020 Jan 7;484.
91. Hamlet A, Jean KK, Perea W, Yactayo S, Biey J, Van Kerkhove M, et al. The seasonal influence of climate and environment on yellow fever transmission across Africa. PLoS Negl Trop Dis. 2018 Mar 15;12(3):e0006284.

92. Kaul RB, Evans M V., Murdock CC, Drake JM. Spatio-temporal spillover risk of yellow fever in Brazil. Parasit Vectors. 2018 Aug 29;11(1):488.
93. Shearer FM, Longbottom J, Browne AJ, Pigott DM, Brady OJ, Kraemer MUGG, et al. Existing and potential infection risk zones of yellow fever worldwide: a modelling analysis. Lancet Glob Heal. 2018 Mar 1;6(3):e270–8.
94. Briand S, Beresniak A, Nguyen T, Yonli T, Duru G, Kambire C, et al. Assessment of yellow fever epidemic risk: An original multi-criteria modeling approach. PLoS Negl Trop Dis. 2009 Jul;3(7).
95. Ndeffo-Mbah ML, Pandey A. Global Risk and Elimination of Yellow Fever Epidemics. J Infect Dis. 2019 Sep 23;
96. Jayaraj VJ, Avoi R, Gopalakrishnan N, Raja DB, Umasa Y. Developing a dengue prediction model based on climate in Tawau, Malaysia. Acta Trop. 2019 Sep 1;197.
97. Kakarla SG, Mopuri R, Mutheneni SR, Bhimala KR, Kumaraswamy S, Kadiri MR, et al. Temperature dependent transmission potential model for chikungunya in India. Sci Total Environ. 2019 Jan 10;647:66–74.
98. Holmes DT, Buhr K. Mathematical Modeling: Assumptions Affect Results. Clin Chem [Internet]. 2006 Aug 1 [cited 2020 Jun 11];52(8):1606–8. Available from: https://academic.oup.com/clinchem/article/52/8/1606/5626980
99. Bansal S, Chowell G, Dimonsen L, Vispignani A, Viboud C. Big Data for Infectious Disease Surveillance and Modeling. J Infect Dis. 2016 Dec;214(S4):S375–9.
100. Evans RS. Electronic Health Records: Then, Now, and in the Future. Yearb Med Inform. 2016 May 20;Suppl 1(Suppl 1):S48–61.
101. Gelaro R, McCarty W, Suárez MJ, Todling R, Molod A, Takacs L, et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J Clim [Internet]. 2017 Jul 20 [cited 2018 Jul 31];30(14):5419–54. Available from: http://journals.ametsoc.org/doi/10.1175/JCLI-D-16-0758.1
102. NASA. Global Modeling and Assimilation Office MERRA-2: File Specification [Internet]. 2016 [cited 2018 Jul 31]. Available from: http://gmao.gsfc.nasa.gov/pubs/office_notes.
103. Liu H, Weng Q. Enhancing temporal resolution of satellite imagery for public health studies: A case study of West Nile Virus outbreak in Los Angeles in 2007. Remote Sens Environ. 2012 Feb 15;117:57–71.
104. NASA, Ministry of Economics, Trade, and Industry. ASTER Global Digital Elevation Model Version 3 [Internet]. 2019. Available from: https://cmr.earthdata.nasa.gov/search/concepts/C1575731655-LPDAAC_ECS.html
105. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and Opportunities of Big Data in Health Care: A Systematic Review. JMIR Med Informatics. 2016 Nov 21;4(4):e38.
106. Jean K, Hamlet A, Benzler J, Cibrelus L, Gaythorpe KAM, Sall A, et al. Eliminating yellow fever epidemics in Africa: Vaccine demand forecast and impact modelling. Althouse B, editor. PLoS Negl Trop Dis [Internet]. 2020 May

7 [cited 2020 Jun 1];14(5):e0008304. Available from: https://dx.plos.org/10.1371/journal.pntd.0008304
107. Githeko AK, Lindsay SW, Confalonieri UE, Patz JA. Climate change and vector-borne diseases: A regional analysis. Bull World Health Organ. 2000;78(9):1136–47.
108. Tabachnick WJ. Challenges in predicting climate and environmental effects on vector-borne disease episystems in a changing world. J Exp Biol. 2010 Mar 15;213(6):946–54.
109. Semenza JC, Suk JE. Vector-borne diseases and climate change: a European perspective. FEMS Microbiol Lett. 2018;365(2).
110. Lafferty KD. The ecology of climate change and infectious diseases. Ecology [Internet]. 2009 Apr [cited 2018 Jul 21];90(4):888–900. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19449681
111. Servadio JL, Rosenthal SR, Carlson L, Bauer C. Climate patterns and mosquito-borne disease outbreaks in South and Southeast Asia. J Infect Public Health. 2018;11(4):566–71.
112. Gage KL, Burkot TR, Eisen RJ, Hayes EB. Climate and Vectorborne Diseases. Am J Prev Med. 2008 Nov;35(5):436–50.
113. Bayoh MN, Lindsay SW. Effect of temperature on the development of the aquatic stages of Anopheles gambiae sensu stricto (Diptera: Culicidae) . Bull Entomol Res. 2003 Sep;93(5):375–81.
114. de Almeida MAB, dos Santos E, Cardoso J da C, da Silva LG, Rabelo RM, Bicca-Marques JC. Predicting Yellow Fever Through Species Distribution Modeling of Virus, Vector, and Monkeys. Ecohealth. 2019 Mar 15;16(1):95–108.
115. Doyle TJ, Glynn MK, Groseclose SL. Completeness of notifiable infectious disease reporting in the United States: An analytical literature review. Am J Epidemiol. 2002 May 1;155(9):866–74.
116. Gibbons CL, Mangen M-JJ, Plass D, Havelaar AH, Brooke RJ, Kramarz P, et al. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. BMC Public Health [Internet]. 2014 Dec 11 [cited 2018 Jul 21];14(1):147. Available from: http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-14-147
117. Lipsitch M, Donnelly CA, Fraser C, Blake IM, Cori A, Dorigatti I, et al. Potential Biases in Estimating Absolute and Relative Case-Fatality Risks during Outbreaks. PLoS Negl Trop Dis [Internet]. 2015 [cited 2019 Jun 17];9(7):e0003846. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26181387
118. MacDougall L, Majowicz S, Doré K, Flint J, Thomas K, Kovacs S, et al. Under-reporting of infectious gastrointestinal illness in British Columbia, Canada: Who is counted in provincial communicable disease statistics? Epidemiol Infect. 2008 Feb;136(2):248–56.
119. Reed C, Angulo FJ, Swerdlow DL, Lipsitch M, Meltzer MI, Jernigan D, et al. Estimates of the prevalence of pandemic (H1N1) 2009, United States, april-

july 2009. Emerg Infect Dis. 2009 Dec;15(12):2004–7.
120. Abdool Karim SS, Dilraj A. Reasons for Under-Reporting of Notifiable Conditions. South African Med J . 1996 Jul;86(7):834–6.
121. Van Lier A, McDonald SA, Bouwknegt M, Van Der Sande M, Bijkerk P, Van Benthem B, et al. Disease burden of 32 infectious diseases in the Netherlands, 2007-2011. PLoS One. 2016 Apr 1;11(4).
122. Iuliano AD, Roguski KM, Chang HH, Muscatello DJ, Palekar R, Tempia S, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet. 2018 Mar 31;391(10127):1285–300.
123. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Heal Inf Libr J [Internet]. 2009 Jun 1 [cited 2020 Apr 18];26(2):91–108. Available from: http://doi.wiley.com/10.1111/j.1471-1842.2009.00848.x
124. Domingues CMAS, Verani JR, Montenegro Renoiner EI, de Cunto Brandileone MC, Flannery B, de Oliveira LH, et al. Effectiveness of ten-valent pneumococcal conjugate vaccine against invasive pneumococcal disease in Brazil: A matched case-control study. Lancet Respir Med. 2014;2(6):464–71.
125. Ford TE, Colwell RR, Rose JB, Morse SS, Rogers DJ, Yates TL. Using satellite images of environmental changes to predict infectious disease outbreaks. Vol. 15, Emerging Infectious Diseases. Centers for Disease Control and Prevention (CDC); 2009. p. 1341–6.
126. Jan PL, Farcy O, Boireau J, Le Texier E, Baudoin A, Le Gouar P, et al. Which temporal resolution to consider when investigating the impact of climatic data on population dynamics? The case of the lesser horseshoe bat (Rhinolophus hipposideros). Oecologia. 2017 Aug 1;184(4):749–61.
127. Waldner F, Horan H, Chen Y, Hochman Z. High temporal resolution of leaf area data improves empirical estimation of grain yield. Sci Rep. 2019 Dec 31;9(1):15714.
128. Yoo PE, John SE, Farquharson S, Cleary JO, Wong YT, Ng A, et al. 7T-fMRI: Faster temporal resolution yields optimal BOLD sensitivity for functional network imaging specifically at high spatial resolution. Neuroimage. 2018 Jan 1;164:214–29.
129. Othman AE, Falkner F, Weiss J, Kruck S, Grimm R, Martirosian P, et al. Effect of temporal resolution on diagnostic performance of dynamic contrast-enhanced magnetic resonance imaging of the prostate. Invest Radiol. 2016;51(5):290–6.
130. Wang G. High temporal-resolution dynamic PET image reconstruction using a new spatiotemporal kernel method. IEEE Trans Med Imaging. 2019 Mar 1;38(3):664–74.
131. Dorniak K, Heiberg E, Hellmann M, Rawicz-Zegrzda D, Wesierska M, Galaska R, et al. Required temporal resolution for accurate thoracic aortic pulse wave velocity measurements by phase-contrast magnetic resonance imaging and comparison with clinical standard applanation tonometry. BMC Cardiovasc Disord. 2016;16(1).
132. Mortensen KH, Tann O. Computed tomography in paediatric heart disease.

Vol. 91, British Journal of Radiology. British Institute of Radiology; 2018.
133. Mathew RC, Löffler AI, Salerno M. Role of Cardiac Magnetic Resonance Imaging in Valvular Heart Disease: Diagnosis, Assessment, and Management. Vol. 20, Current Cardiology Reports. Current Medicine Group LLC 1; 2018.
134. Downs NJ, Parisi A V., Butler H, Rawlings A, Elrahoumi RS. An Inexpensive High-Temporal Resolution Electronic Sun Journal for Monitoring Personal Day to Day Sun Exposure Patterns. Front Public Heal. 2017 Nov 16;5.
135. Muzic RF*, Saidel GM. Distributed Versus Compartment Models for PET Receptor Studies. IEEE Trans Med Imaging. 2003;22(1).
136. Bernini A, Bolzoni L, Casagrandi R. When resolution does matter: Modelling indirect contacts in dairy farms at different levels of detail. PLoS One. 2019;14(10).
137. Kretzschmar M, Wallinga J. Mathematical Models in Infectious Disease Epidemiology. In: Modern Infectious Disease Epidemiology [Internet]. Nature Publishing Group; 2009 [cited 2020 Jun 22]. p. 209–21. Available from: /pmc/articles/PMC7178885/?report=abstract
138. Huppert A, Katriel G. Mathematical modelling and prediction in infectious disease epidemiology. Vol. 19, Clinical Microbiology and Infection. Blackwell Publishing Ltd; 2013. p. 999–1005.
139. Wakefield J, Dong TQ, Minin VN. Spatio-Temporal analysis of surveillance data. In: Handbook of infectious disease data analysis. 1st ed. Boca Raton, FL: Chapman & Hall; 2019. p. 455–76.
140. Jones SG, Kulldorff M. Influence of Spatial Resolution on Space-Time Disease Cluster Detection. PLoS One. 2012;7(10):e48036.
141. Mills HL, Riley S. The Spatial Resolution of Epidemic Peaks. PLoS Comput Biol. 2014 Apr 10;10(4):e1003561.
142. Piantadosi S, Byar DP, Green SB. The ecological fallacy. Am J Epidemiol. 1988;127(5):893–904.
143. Couto-Lima D, Madec Y, Bersot MI, Campos SS, Motta M de A, Santos FBDFB dos, et al. Potential risk of re-emergence of urban transmission of Yellow Fever virus in Brazil facilitated by competent Aedes populations. Sci Rep. 2017 Jul 7;7(1):4848.
144. Instituto Brasileiro de Grografia e Estatistica, Instituto Brasileiro de Geografia Estastistica. Estimativas da População. 2019.
145. Romano APM, Costa ZGA, Ramos DG, Andrade MA, de Jayme VS, de Almeida MAB, et al. Yellow Fever Outbreaks in Unvaccinated Populations, Brazil, 2008-2009. PLoS Negl Trop Dis. 2014;8(3).
146. Breslow NE. Generalized linear models: Checking assumptions and strenghthening conclusions. Stat Appl. 1996;8(1):23–41.
147. Ball F, O'Neill P. The distribution of general final state random variables for stochastic epidemic models. J Appl Probab [Internet]. 1999 [cited 2020 Jun 22];36(2):473–91. Available from: /core/journals/journal-of-applied-probability/article/distribution-of-general-final-state-random-variables-for-stochastic-epidemic-models/D685556D125E03466B82E73382380C51

148. Allen LJS, Burgin AM. Comparison of deterministic and stochastic SIS and SIR models in discrete time. Math Biosci. 2000 Jan 1;163(1):1–33.
149. Carnell R. Triangle: Provides the Standard Distribution Functions for the Triangle Distribution [Internet]. 2017. Available from: https://cran.r-project.org/package=triangle
150. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing; 2019.
151. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. Science (80-). 1983;220(4598):671–80.
152. Gillespie CS. Fitting heavy tailed distributions: The powerlaw package. J Stat Softw. 2015 Feb 1;64(2):1–16.
153. Thaler DS, Head MG, Horsley A. Precision public health to inhibit the contagion of disease and move toward a future in which microbes spread health. BMC Infect Dis. 2019 Feb 6;19(1).
154. McMichael AJ. Extreme weather events and infectious disease outbreaks. Virulence. 2015;6(6):543–7.
155. Radchuk V, Johst K, Groeneveld J, Turlure C, Grimm V, Schtickzelle N. Appropriate resolution in time and model structure for population viability analysis: Insights from a butterfly metapopulation. Biol Conserv. 2014 Jan 1;169:345–54.
156. Postlethwaite CM, Dennis TE. Effects of Temporal Resolution on an Inferential Model of Animal Movement. PLoS One. 2013 May 6;8(5).
157. Jaques C, Pignat E, Calinon S, Liebling M. Temporal super-resolution microscopy using a hue-encoded shutter. Biomed Opt Express. 2019 Sep 1;10(9):4727.
158. Petrov AY, Herbst M, Andrew Stenger V. Improving temporal resolution in fMRI using a 3D spiral acquisition and low rank plus sparse (L+S) reconstruction. Neuroimage. 2017 Aug 15;157:660–74.
159. Kinsley AC, Patterson G, VanderWaal KL, Craft ME, Perez AM. Parameter values for epidemiological models of foot-and-mouth disease in Swine. Front Vet Sci. 2016 Jun 1;3(JUN).
160. Chowell G, Viboud C, Simonsen L, Moghadas SM. Characterizing the reproduction number of epidemics with early subexponential growth dynamics. J R Soc Interface. 2016 Oct 1;13(123).
161. Champredon D, Dushoff J. Intrinsic and realized generation intervals in infectious-disease transmission. Proc R Soc B Biol Sci. 2015 Dec 16;282(1821).
162. Wearing HJ, Rohani P, Keeling MJ. Appropriate Models for the Management of Infectious Diseases. Ellner SP, editor. PLoS Med. 2005 Jul 26;2(7):e174.
163. Machado G, Alvarez J, Bakka HC, Perez A, Donato LE, de Ferreira Lima Júnior FE, et al. Revisiting area risk classification of visceral leishmaniasis in Brazil. BMC Infect Dis. 2019 Dec 3;19(1):2.
164. Belikov A V. The number of key carcinogenic events can be predicted from cancer incidence. Sci Rep. 2017 Dec 1;7(1).

165. Aksoy H. Use of Gamma Distribution in Hydrological Analysis.
166. Anonymous. Yellow fever in 1988. Relev Epidemiol Hebd [Internet]. 1990;65(28):213–9. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=2386708
167. Vasconcelos PFC, Monath TP. Yellow Fever Remains a Potential Threat to Public Health. Vector-Borne Zoonotic Dis [Internet]. 2016 Aug 1 [cited 2020 Apr 14];16(8):566–7. Available from: https://www.liebertpub.com/doi/10.1089/vbz.2016.2031
168. Convertino M, Muneepeerakul R, Azaele S, Bertuzzo E, Rinaldo A, Rodriguez-Iturbe I. On neutral metacommunity patterns of river basins at different scales of aggregation. Water Resour Res [Internet]. 2009 Aug 1 [cited 2020 Jun 18];45(8). Available from: http://doi.wiley.com/10.1029/2009WR007799
169. Ajelli M, Gonçalves B, Balcan D, Colizza V, Hu H, Ramasco JJ, et al. Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. BMC Infect Dis. 2010 Jun 29;10.
170. Wong JY, Kelly H, Ip DKM, Wu JT, Leung GM, Cowling BJ. Case fatality risk of influenza a (H1N1pdm09): A systematic review. Epidemiology [Internet]. 2013 Nov [cited 2020 Mar 4];24(6):830–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24045719
171. Kelly H, Cowling BJ. Case Fatality. Epidemiology [Internet]. 2013 Jul [cited 2020 Mar 4];24(4):622–3. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00001648-201307000-00023
172. Salama M, Amitai Z, Lustig Y, Mor Z, Weiberger M, Chowers M, et al. Outbreak of West Nile Virus disease in Israel (2015): A retrospective analysis of notified cases. Travel Med Infect Dis [Internet]. 2019 Mar 1 [cited 2020 Apr 18];28:41–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30016649
173. Kim KH, Tandi TE, Choi JW, Moon JM, Kim MS. Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in South Korea, 2015: epidemiology, characteristics and public health implications. J Hosp Infect. 2017 Feb 1;95(2):207–13.
174. Self JL, Conrad A, Stroika S, Jackson A, Whitlock L, Jackson KA, et al. Multistate outbreak of listeriosis associated with packaged leafy green salads, united states and canada, 2015–2016. Emerg Infect Dis. 2019 Aug 1;25(8):1461–8.
175. Cromer D, Van Hoek AJ, Jit M, Edmunds WJ, Fleming D, Miller E. The burden of influenza in England by age and clinical risk group: A statistical analysis to inform vaccine policy. J Infect. 2014 Apr 1;68(4):363–71.
176. Yu AT, Amin N, Rahman MW, Gurley ES, Rahman KM, Luby SP. Case-Fatality Ratio of Blood Culture-Confirmed Typhoid Fever in Dhaka, Bangladesh. J Infect Dis [Internet]. 2018 [cited 2020 Mar 4];218(suppl_4):S222–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30304448
177. Reich NG, Lessler J, Cummings DAT, Brookmeyer R. Estimating Absolute and

Relative Case Fatality Ratios from Infectious Disease Surveillance Data. Biometrics. 2012 Jun;68(2):598–606.
178. Elandt-Johnson RC. Definition of rates: Some remarks on their use and misuse. Am J Epidemiol. 1975;102(4):267–71.
179. Debnath F, Ponnaiah M, Acharya P. Dengue fever in a municipality of West Bengal, India, 2015: An outbreak investigation. Indian J Public Health [Internet]. 2017 Oct 1 [cited 2020 May 1];61(4):239–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29219127
180. Jia N, Feng D, Fang LQ, Richardus JH, Han XN, Cao WC, et al. Case fatality of SARS in mainland China and associated risk factors. Trop Med Int Heal [Internet]. 2009 Nov [cited 2020 Mar 4];14(SUPPL. 1):21–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19508439
181. Yu H, Cowling BJ, Feng L, Lau EHY, Liao Q, Tsang TK, et al. Human infection with avian influenza A H7N9 virus: An assessment of clinical severity. Lancet [Internet]. 2013 Jul 13 [cited 2020 Mar 4];382(9887):138–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23803487
182. Gao Z, Parhar A, Gallant V, Heffernan C, Ahmed R, Egedahl ML, et al. Apopulation-based study of tuberculosis case fatality in Canada: Do aboriginal peoples fare less well? Int J Tuberc Lung Dis [Internet]. 2015 Jul 1 [cited 2020 Mar 4];19(7):772–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26056100
183. Uche IV, MacLennan CA, Saul A. A Systematic Review of the Incidence, Risk Factors and Case Fatality Rates of Invasive Nontyphoidal Salmonella (iNTS) Disease in Africa (1966 to 2014). PLoS Negl Trop Dis [Internet]. 2017 Jan 5 [cited 2020 Mar 4];11(1):e0005118. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28056035
184. Tan B, Wong JJM, Sultana R, Koh JCJW, Jit M, Mok YH, et al. Global Case-Fatality Rates in Pediatric Severe Sepsis and Septic Shock: A Systematic Review and Meta-analysis [Internet]. Vol. 173, JAMA Pediatrics. American Medical Association; 2019 [cited 2020 Mar 4]. p. 352–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30742207
185. Kenmoe S, Demanou M, Bigna JJ, Nde Kengne C, Fatawou Modiyinji A, Simo FBN, et al. Case fatality rate and risk factors for Nipah virus encephalitis: A systematic review and meta-analysis [Internet]. Vol. 117, Journal of Clinical Virology. Elsevier B.V.; 2019 [cited 2020 Mar 4]. p. 19–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31132674
186. Woldeamanuel YW, Andemeskel AT, Kyei K, Woldeamanuel MW, Woldeamanuel W. Case fatality of adult tetanus in Africa: Systematic review and meta-analysis [Internet]. Vol. 368, Journal of the Neurological Sciences. Elsevier B.V.; 2016 [cited 2020 Jan 14]. p. 292–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27538652
187. Portnoy A, Jit M, Ferrari M, Hanson M, Brenzel L, Verguet S. Estimates of case-fatality ratios of measles in low-income and middle-income countries: a systematic review and modelling analysis. Lancet Glob Heal. 2019 Apr

1;7(4):e472–81.
188. Szabo SM, Samp JC, Walker DR, Lane S, Cline SK, Gooch KL, et al. Liver-specific case fatality due to chronic hepatitis C virus infection: a systematic review. Ann Hepatol [Internet]. [cited 2020 Mar 4];14(5):618–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26256890
189. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. In: Journal of clinical epidemiology. 2009. p. e1–34.
190. PRISMA Checklist [Internet]. 2015 [cited 2020 May 1]. Available from: http://prisma-statement.org/PRISMAStatement/Checklist.aspx
191. Devito NJ, Goldacre B. Catalogue of bias: Publication bias. BMJ Evidence-Based Med. 2019 Apr 1;24(2):53–4.
192. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: A practical tutorial. Evid Based Ment Health. 2019 Nov 1;22(4):153–60.
193. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med [Internet]. 2002 Jun 15 [cited 2020 Apr 23];21(11):1539–58. Available from: http://doi.wiley.com/10.1002/sim.1186
194. World Health Organization. Yellow fever in Africa and South America, 2006. Relev Epidemiol Hebd. 2008 Feb;83(8):60–76.
195. Wamala JF, Malimbo M, Okot CL, Atai-Omoruto AD, Tenywa E, Miller JR, et al. Epidemiological and laboratory characterization of a yellow fever outbreak in northern Uganda, October 2010-January 2011. Int J Infect Dis. 2012 Jul;16(7):e536-42.
196. de Filippis AMB, Nogueira RMR, Schatzmayr HG, Tavares DS, Jabor A V, Diniz SCM, et al. Outbreak of jaundice and hemorrhagic fever in the Southeast of Brazil in 2001: detection and molecular characterization of yellow fever virus. J Med Virol. 2002 Dec;68(4):620–7.
197. Tuboi SH, Costa ZGA, da Costa Vasconcelos PF, Hatch D. Clinical and epidemiological characteristics of yellow fever in Brazil: analysis of reported cases 1998-2002. Trans R Soc Trop Med Hyg [Internet]. 2007 Feb 1 [cited 2018 Jul 21];101(2):169–75. Available from: https://academic.oup.com/trstmh/article-lookup/doi/10.1016/j.trstmh.2006.04.001
198. Agadzi VK, Boatin BA, Appawu MA, Mingle JA, Addy PA. Yellow fever in Ghana, 1977-80. Bull World Health Organ. 1984;62(4):577–83.
199. Jones EM, Wilson DC. Clinical features of yellow fever cases at Vom Christian Hospital during the 1969 epidemic on the Jos Plateau, Nigeria. Bull World Health Organ. 1972;46(5):653–7.
200. World Health Organization. WEEKLY EPIDEMIOLOGICAL RECORD. 1993;68(11):77–8.
201. Ingelbeen B, Weregemere NA, Noel H, Tshapenda GP, Mossoko M, Nsio J, et al. Urban yellow fever outbreak-Democratic Republic of the Congo, 2016: Towards more rapid case detection. PLoS Negl Trop Dis. 2018

Dec;12(12):e0007029.
202. De Cock KM, Monath TP, Nasidi A, Tukei PM, Enriquez J, Lichfield P, et al. Epidemic yellow fever in eastern Nigeria, 1986. Lancet (London, England). 1988 Mar;1(8586):630–3.
203. Monath TP, Craven RB, Adjukiewicz A, Germain M, Francy DB, Ferrara L, et al. Yellow fever in the Gambia, 1978-1979: Epidemiologic aspects with observations on the occurrence of Orungo virus infections. Am J Trop Med Hyg. 1980;29(5).
204. Nasidi A, Monath TP, DeCock K, Tomori O, Cordellier R, Olaleye OD, et al. Urban yellow fever epidemic in western Nigeria, 1987. Trans R Soc Trop Med Hyg [Internet]. [cited 2019 Nov 5];83(3):401–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2617590
205. Zhao YY, Jin H, Zhang XF, Wang B. Case-fatality of hand, foot and mouth disease associated with EV71: A systematic review and meta-analysis. Epidemiol Infect [Internet]. 2015 Oct [cited 2020 Apr 17];143(14):3094–102. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25721492
206. Pieters Z, Saad NJ, Antillón M, Pitzer VE, Bilcke J. Case Fatality Rate of Enteric Fever in Endemic Countries: A Systematic Review and Meta-analysis. Clin Infect Dis [Internet]. 2018 [cited 2020 Apr 17];67(4):628–38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29522159
207. Mutebi JP, Barrett ADT. The epidemiology of yellow fever in Africa. Vol. 4, Microbes and Infection. Elsevier Masson; 2002. p. 1459–68.
208. Auguste AJ, Lemey P, Pybus OG, Suchard MA, Salas RA, Adesiyun AA, et al. Yellow Fever Virus Maintenance in Trinidad and Its Dispersal throughout the Americas. J Virol. 2010 Oct 1;84(19):9967–77.
209. Farrar J, Hotez PJ, Junghanss T, Kang G, Lalloo D, White NJ. Manson's Tropical Diseases: Twenty-Third Edition. Manson's Tropical Diseases: Twenty-Third Edition. Elsevier Ltd; 2013. 1–1337 p.
210. Vainio J, Cutts F. Yellow fever [Internet]. World Health Organization Geneva. 1998 [cited 2020 May 1]. Available from: http://www.who.ch/gpv-documents/
211. Vygen S, Tiffany A, Rull M, Ventura A, Wolz A, Jambai A, et al. Changes in health-seeking behavior did not result in increased all-cause mortality during the Ebola outbreak in Western Area, Sierra Leone. Am J Trop Med Hyg. 2016 Oct 1;95(4):897–901.
212. Jin H, Zhao Y, Zhang X, Wang B, Liu P. Case-fatality risk of pregnant women with acute viral hepatitis type E: A systematic review and meta-analysis. Epidemiol Infect [Internet]. 2016 Jul 1 [cited 2020 Apr 17];144(10):2098–106. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26939626
213. Nyakarahuka L, Kankya C, Krontveit R, Mayer B, Mwiine FN, Lutwama J, et al. How severe and prevalent are Ebola and Marburg viruses? A systematic review and meta-analysis of the case fatality rates and seroprevalence. BMC Infect Dis [Internet]. 2016 Nov 25 [cited 2020 Apr 17];16(1):708. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27887599

214. Rota PA, Moss WJ, Takeda M, De Swart RL, Thompson KM, Goodson JL. Measles. Nat Rev Dis Prim. 2016 Jul 14;2(1):1–16.
215. Woldeamanuel YW, Girma B. A 43-year systematic review and meta-analysis: Case-fatality and risk of death among adults with tuberculous meningitis in Africa. Vol. 261, Journal of Neurology. Dr. Dietrich Steinkopff Verlag GmbH and Co. KG; 2014. p. 851–65.
216. Karlberg J, Chong DSY, Lai WYY. Do Men Have a Higher Case Fatality Rate of Severe Acute Respiratory Syndrome than Women Do? Am J Epidemiol. 2004 Feb 1;159(3):229–31.
217. Sakamoto Y, Yamaguchi T, Yamamoto N, Nishiura H. Modeling the elevated risk of yellow fever among travelers visiting Brazil, 2018. Theor Biol Med Model [Internet]. 2018 Jul 2 [cited 2018 Jul 23];15(1):9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29961429
218. Tomori O. Yellow Fever: The Recurring Plague. Crit Rev Clin Lab Sci [Internet]. 2004 Jan 19 [cited 2018 Jul 21];41(4):391–427. Available from: http://www.tandfonline.com/doi/full/10.1080/10408360490497474
219. Bryant J, Wang H, Cabezas C, Ramirez G, Watts D, Russell K, et al. Enzootic Transmission of Yellow Fever Virus in Peru. Emerg Infect Dis [Internet]. 2003 Aug [cited 2018 Jul 23];9(8):926–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12967489
220. Beck AS, Barrett AD. Current status and future prospects of yellow fever vaccines. Expert Rev Vaccines [Internet]. 2015 Nov 2 [cited 2018 Jul 21];14(11):1479–92. Available from: http://www.tandfonline.com/doi/full/10.1586/14760584.2015.1083430
221. Dyer O. Yellow fever stalks Brazil in Zika's wake. BMJ [Internet]. 2017 Feb 8 [cited 2018 Jul 21];356:j707. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28179231
222. Balogun EO, Nok AJ, Kita K. Global warming and the possible globalization of vector-borne diseases: a call for increased awareness and action. Trop Med Health [Internet]. 2016 Dec 24 [cited 2018 Jul 21];44(1):38. Available from: http://tropmedhealth.biomedcentral.com/articles/10.1186/s41182-016-0039-0
223. Aspvik NP, Viken H, Ingebrigtsen JE, Zisko N, Mehus I, Wisløff U, et al. Do weather changes influence physical activity level among older adults? – The Generation 100 study. PLoS One. 2018 Jul 1;13(7).
224. Aristizabal JF, Lévêque L, Chapman CA, Serio-Silva JC. Impacts of Temperature on Behaviour of the Mexican Endangered Black Howler Monkey Alouatta pigra Lawrence, 1933 (Primates: Atelidae) in a Fragmented Landscape. Acta Zool Bulg. 2018;70(3):377–82.
225. Gershman MD, Staples JE. Yellow Fever. In: Netter's Infectious Diseases. Philadelphia: Elsevier Saunders; 2012. p. 383–9.
226. Laxminarayan R, Kakkar M, Horby P, Malavige GN, Basnyat B. Emerging and re-emerging infectious disease threats in South Asia: Status, vulnerability, preparedness, and outlook. BMJ [Internet]. 2017 [cited 2020 Mar

9];357:j1447. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28400386
227. Gilbert M, Pullano G, Pinotti F, Valdano E, Poletto C, Boëlle P-Y, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. Lancet Infect Dis [Internet]. 2020 Feb 20 [cited 2020 Mar 9];6736(20):1–7. Available from: https://doi.org/10.1016/S0140-6736
228. Moore M, Gelfeld B, Okunogbe A, Paul C. Identifying Future Disease Hot Spots: Infectious Disease Vulnerability Index. Vol. 6, Identifying Future Disease Hot Spots: Infectious Disease Vulnerability Index. RAND Corporation; 2016.
229. Quintero J, Carrasquilla G, Suárez R, González C, Olano VA. An ecosystemic approach to evaluating ecological, socioeconomic and group dynamics affecting the prevalence of Aedes aegypti in two Colombian towns. Cad Saude Publica. 2009;25(SUPPL. 1).
230. Causey OR, Kumm HW, Laemmert HW. DISPERSION OF FOREST MOSQUITOES IN BRAZIL: FURTHER STUDIES'.
231. Diallo D, Sall AA, Diagne CT, Faye O, Hanley KA, Buenemann M, et al. Patterns of a sylvatic yellow fever virus amplification in southeastern Senegal, 2010. Am J Trop Med Hyg. 2014 Jun;90(6):1003–13.
232. Obenauer JF, Andrew Joyner T, Harris JB. The importance of human population characteristics in modeling Aedes aegypti distributions and assessing risk of mosquito-borne infectious diseases. Trop Med Health. 2017;45(1).
233. Database of Global Administrative Areas. GADM data, version 3.6 [Internet]. 2018 [cited 2020 Jul 5]. Available from: https://gadm.org/download_country_v3.html
234. Zhao S, Stone L, Gao D, He D. Modelling the large-scale yellow fever outbreak in Luanda, Angola, and the impact of vaccination. PLoS Negl Trop Dis [Internet]. 2018 [cited 2018 Jul 21];12(1):e0006158. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29338001
235. Centers for Disease Control and Prevention. Mosquito life cycle [Internet]. [cited 2020 Jul 1]. Available from: https://www.cdc.gov/dengue/resources/factSheets/MosquitoLifecycleFINAL.pdf
236. Environmental Systems Research Institute (ESRI). ArcGIS Pro. Redlands, CA; 2018.
237. Hayes RO, Francy DB, Lazuick JS, Smith GC, Jones RH. Arbovirus surveillance in six states during 1972. Am J Trop Med Hyg [Internet]. 1976 [cited 2020 Jun 22];25(3):463–76. Available from: https://pubmed.ncbi.nlm.nih.gov/7148/
238. Souza RL, Mugabe VA, Paploski IAD, Rodrigues MS, Moreira PSDS, Nascimento LCJ, et al. Effect of an intervention in storm drains to prevent Aedes aegypti reproduction in Salvador, Brazil. Parasites and Vectors [Internet]. 2017 Jul 11 [cited 2020 Jun 22];10(1). Available from: https://pubmed.ncbi.nlm.nih.gov/28697811/

239. R S, J L, M S, A M, H K, E K, et al. Effects of Irrigation and Rainfall on the Population Dynamics of Rift Valley Fever and Other Arbovirus Mosquito Vectors in the Epidemic-Prone Tana River County, Kenya. J Med Entomol [Internet]. 2017 [cited 2020 Jun 22];54(2). Available from: https://pubmed.ncbi.nlm.nih.gov/28011732/
240. Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth. Bioscience. 2001 Nov 1;51(11):933–8.
241. Venables WN, Ripley BD. Modern Applied Statistics with S [Internet]. 4th ed. New York: Springer; 2002 [cited 2020 Mar 13]. Available from: http://link.springer.com/10.1007/978-0-387-21706-2_7
242. Fox J, Weisberg S. An R companion to applied regression. 3rd ed. Thousand Oaks, CA: SAGE Publications; 2019.
243. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. Global Sensitivity Analysis. The Primer [Internet]. Global Sensitivity Analysis. The Primer. Chichester, UK: John Wiley and Sons; 2008 [cited 2020 Apr 29]. 1–292 p. Available from: http://doi.wiley.com/10.1002/9780470725184
244. Wu J, Dhingra R, Gambhir M, Remais J V. Sensitivity analysis of infectious disease models: methods, advances and their application. J R Soc Interface. 2013 Sep 6;10(86):20121018.
245. Sobol I. Sensitivity analysis for non-linear mathematical model. Math Model Comput Exp. 1993;1:407–14.
246. Lu R, Wang D, Wang M, Rempala GA. Estimation of Sobol's sensitivity indices under generalized linear models. Commun Stat - Theory Methods. 2018 Nov 2;47(21):5163–95.
247. Funk S, Bansal S, Bauch CT, Eames KTD, Edmunds WJ, Galvani AP, et al. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. Epidemics. 2015 Mar 1;10:21–5.
248. Figueiredo PO, Silva ATS, Oliveira JS, Marinho PE, Rocha FT, Domingos GP, et al. Detection and Molecular Characterization of Yellow Fever Virus , 2017, Brazil. EcoHealth 2018 154. 2018 Aug 16;15(4):864–70.
249. Griffiths S, Zhou X-N. Why research infectious diseases of poverty? [Internet]. 2012 [cited 2020 Apr 29]. Available from: https://www.who.int/tdr/stewardship/global_report/2012/chapitre1_web.pdf
250. Grief SN, Miller JP. Infectious Disease Issues in Underserved Populations. Vol. 44, Primary Care - Clinics in Office Practice. W.B. Saunders; 2017. p. 67–85.
251. Johansson, Michael A, Arana-Vizcarrondo, Neysari, Biggerstaff, Brad J, Staples JE, Johansson MA, Arana-Vizcarrondo N, Biggerstaff BJ, Staples JE. Incubation periods of Yellow fever virus. Am J Trop Med Hyg [Internet]. 2010 Jul 1 [cited 2018 Jul 21];83(1):183–8. Available from: http://www.ajtmh.org/content/journals/10.4269/ajtmh.2010.09-0782
252. Vasconcelos PFC, Costa ZG, Travassos da Rosa ES, Luna E, Rodrigues SG, Barros VLRS, et al. Epidemic of jungle yellow fever in Brazil, 2000:

Implications of climatic alterations in disease spread. J Med Virol [Internet]. 2001 Nov 1 [cited 2018 Jul 21];65(3):598–604. Available from: http://doi.wiley.com/10.1002/jmv.2078
253. Hoshen MB, Morse AP. A weather-driven model of malaria transmission. Malar J. 2004 Sep 6;3.
254. Craig M, Le Sueur D, Snow B. A climate-based distribution model of malaria transmission in sub-Saharan Africa. Vol. 15, Parasitology Today. Elsevier; 1999. p. 105–11.
255. Sirisena P, Noordeen F, Kurukulasuriya H, Romesh TA, Fernando LK. Effect of climatic factors and population density on the distribution of dengue in Sri Lanka: A GIS based evaluation for prediction of outbreaks. PLoS One. 2017 Jan 1;12(1).
256. Fuller D, Troyo A, Beier J. El Niño Southern Oscillation and vegetation dynamics as predictors of dengue fever cases in Costa Rica. - PubMed - NCBI. Env Res Lett. 2009 Mar;4(4):140111–8.
257. Zhang H, Wang L, Lai S, Li Z, Sun Q, Zhang P. Surveillance and early warning systems of infectious disease in China: From 2012 to 2014. Int J Health Plann Manage. 2017;32(3):329–38.
258. Morin CW, Semenza JC, Trtanj JM, Glass GE, Boyer C, Ebi KL. Unexplored Opportunities: Use of Climate- and Weather-Driven Early Warning Systems to Reduce the Burden of Infectious Diseases. Vol. 5, Current environmental health reports. NLM (Medline); 2018. p. 430–8.
259. Douam F, Ploss A. Yellow Fever Virus: Knowledge Gaps Impeding the Fight Against an Old Foe. Vol. 26, Trends in Microbiology. Elsevier Ltd; 2018. p. 913–28.
260. Eisen L, Beaty BJ, Morrison AC, Scott TW. Proactive Vector Control Strategies and Improved Monitoring and Evaluation Practices for Dengue Prevention. J Med Entomol. 2009 Nov 1;46(6):1245–55.
261. Pan American Health Organization. Epidemiological Update: Yellow Fever [Internet]. 2019 Mar [cited 2020 May 28]. Available from: http://www.paho.org
262. Salami D, Capinha C, Do Rosário Oliveira Martins M, Sousa CA. Dengue importation into Europe: A network connectivity-based approach. PLoS One. 2020;15(3).
263. Cui Z, Lin D, Chongsuvivatwong V, Gravis EA, Chaiprasert A, Palittapongarnpim P, et al. Hot and cold spot areas of household tuberculosis transmission in southern china: Effects of socio-economic status and mycobacterium tuberculosis genotypes. Int J Environ Res Public Health. 2019 May 2;16(10).
264. Abedi-Astaneh F, Hajjaran H, Yaghoobi-Ershadi MR, Hanafi-Bojd AA, Mohebali M, Shirzadi MR, et al. Risk mapping and situational analysis of cutaneous leishmaniasis in an endemic area of Central Iran: A GIS-based survey. PLoS One. 2016 Aug 1;11(8).
265. Gou F, Liu X, Liu D, Ren X, Li J, Liu H, et al. Spatial-temporal distribution of

hepatitis B in Gansu province, 2009-2014. Chinese J Endem [Internet]. 2016 Jan 1 [cited 2020 Apr 30];37(1):85–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26822650
266. Litvoc MN, Novaes CTGF, Lopes MIBF. Yellow fever. Rev Assoc Med Bras. 2018 Feb 1;64(2):106–13.
267. Lowe R, Barcellos C, Brasil P, Cruz OG, Honório NA, Kuper H, et al. The zika virus epidemic in brazil: From discovery to future implications. Int J Environ Res Public Health. 2018 Jan 9;15(1).
268. Chen Y, Chu CW, Chen MIC, Cook AR. The utility of LASSO-based models for real time forecasts of endemic infectious diseases: A cross country comparison. J Biomed Inform. 2018 May 1;81:16–30.
269. Ruiz D, Poveda G, Vélez ID, Quiñones ML, Rúa GL, Velásquez LE, et al. Modelling entomological-climatic interactions of Plasmodium falciparum malaria transmission in two Colombian endemic-regions: Contributions to a National Malaria Early Warning System. Malar J. 2006 Aug 1;5.
270. Tesla B, Demakovsky LR, Mordecai EA, Ryan SJ, Bonds MH, Ngonghala CN, et al. Temperature drives Zika virus transmission: Evidence from empirical and mathematical models. Proc R Soc B Biol Sci. 2018 Aug 15;285(1884).
271. World Health Organization. Sustaining the drive to overcome the global impact of neglected tropical diseases [Internet]. WHO. World Health Organization; 2013 Jan [cited 2020 May 18]. Available from: https://www.who.int/neglected_diseases/9789241564540/en/
272. George DB, Taylor W, Shaman J, Rivers C, Paul B, O'Toole T, et al. Technology to advance infectious disease forecasting for outbreak management. Nat Commun. 2019 Sep 2;10(1):1–4.
273. Chen Y, Ong JHY, Rajarethinam J, Yap G, Ng LC, Cook AR. Neighbourhoodlevel real-time forecasting of dengue cases in tropical urban Singapore. BMC Med. 2018 Aug 6;16(1).
274. Chae S, Kwon S, Lee D. Predicting infectious disease using deep learning and big data. Int J Environ Res Public Health. 2018 Aug 1;15(8).
275. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infect Dis Model. 2020 Jan 1;5:256–63.
276. Shanafelt DW, Jones G, Lima M, Perrings C, Chowell G. Forecasting the 2001 Foot-and-Mouth Disease Epidemic in the UK. Ecohealth. 2018 Jun 1;15(2):338–47.
277. Zhao D, Wang L, Cheng J, Xu J, Xu Z, Xie M, et al. Impact of weather factors on hand, foot and mouth disease, and its role in short-term incidence trend forecast in Huainan City, Anhui Province. Int J Biometeorol. 2017 Mar 1;61(3):453–61.
278. Fischer LS, Santibanez S, Hatchett RJ, Jernigan DB, Meyers LA, Thorpe PG, et al. CDC Grand Rounds: Modeling and Public Health Decision-Making. MMWR Morb Mortal Wkly Rep [Internet]. 2016 Dec 9 [cited 2020 Jun 25];65(48):1374–7. Available from:

http://www.cdc.gov/mmwr/volumes/65/wr/mm6548a4.htm
279. Dieng H, Rahman GMS, Hassan AA, Salmah MRC, Satho T, Miake F, et al. The effects of simulated rainfall on immature population dynamics of Aedes albopictus and female oviposition. Int J Biometeorol [Internet]. 2012 Jan [cited 2020 Jul 13];56(1):113–20. Available from: https://pubmed.ncbi.nlm.nih.gov/21267602/
280. Brady OJ, Johansson MA, Guerra CA, Bhatt S, Golding N, Pigott DM, et al. Modelling adult Aedes aegypti and Aedes albopictus survival at different temperatures in laboratory and field settings. Parasites and Vectors [Internet]. 2013 Dec 12 [cited 2020 Jul 13];6(1). Available from: https://pubmed.ncbi.nlm.nih.gov/24330720/
281. Instituto Brasileiro de Geografia Estastistica. Population Estimates [Internet]. 2018 [cited 2018 Aug 1]. Available from: https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9103-estimativas-de-populacao.html?edicao=17283&t=downloads
282. Imai C, Armstrong B, Chalabi Z, Mangtani P, Hashizume M. Time series regression model for infectious disease and weather. Environ Res [Internet]. 2015 Oct 1 [cited 2020 Jun 30];142:319–27. Available from: https://pubmed.ncbi.nlm.nih.gov/26188633/
283. Swaan CM, Wong A, Marinović AB, Kretzschmar MEE, Van Steenbergen JE. Timeliness of infectious disease reporting, the Netherlands, 2003 to 2017: Law change reduced reporting delay, disease identification delay is next [Internet]. Vol. 24, Eurosurveillance. European Centre for Disease Prevention and Control (ECDC); 2019 [cited 2020 Jul 13]. Available from: /pmc/articles/PMC6905299/?report=abstract
284. Opayele AV, Adekunle AJ, Ibrahim KT, Olaleye DO. Influence of Meteorological Variables on Diversity and Abundance of Mosquito Vectors in Two Livestock Farms in Ibadan, Nigeria: Public Health Implications. J Mosq Res. 2017;7(9):70.
285. Sun H, Jit M, Cook AR, Carrasco LR, Dickens BL. Determining environmental and anthropogenic factors which explain the global distribution of aedes aegypti and Ae. Albopictus. BMJ Glob Heal. 2018 Jul 1;3(4).
286. Laneri K, Cabella B, Prado PI, Coutinho RM, Kraenkel RA. Climate drivers of malaria at its southern fringe in the Americas. PLoS One. 2019;14(7).
287. Paaijmans KP, Read AF, Thomas MB. Understanding the link between malaria risk and climate. Proc Natl Acad Sci U S A. 2009 Aug 18;106(33):13844–9.
288. Choi Y, Tang CS, McIver L, Hashizume M, Chan V, Abeyasinghe RR, et al. Effects of weather factors on dengue fever incidence and implications for interventions in Cambodia. BMC Public Health. 2016 Dec 8;16(1):241.
289. Kakarla SG, Caminade C, Mutheneni SR, Morse AP, Upadhyayula SM, Kadiri MR, et al. Lag effect of climatic variables on dengue burden in India. Epidemiol Infect [Internet]. 2019 [cited 2020 Jun 25];147. Available from: /pmc/articles/PMC6518529/?report=abstract
290. Zuur AF, Ieno EN. Beginner's guide to zero-inflated models with R. Newburgh,

UK: Highland Statistics Ltd; 2016.
291. Anderson SC. Gamma Hurdle Models - Fitting and interpreting Gamma hurdle models [Internet]. 2014 [cited 2020 May 28]. Available from: https://seananderson.ca/2014/05/18/gamma-hurdle/
292. Saunders SP, Ries L, Neupane N, Isabel Ramírez M, García-Serrano E, Rendón-Salinas E, et al. Multiscale seasonal factors drive the size of winter monarch colonies. Proc Natl Acad Sci U S A. 2019;116(17):8609–14.
293. Harris M, Caldwell JM, Mordecai EA. Climate drives spatial variation in Zika epidemics in Latin America. Proc R Soc B Biol Sci. 2019 Aug 28;286(1909).
294. Imai C, Brooks WA, Chung Y, Goswami D, Anjali BA, Dewan A, et al. Tropical influenza and weather variability among children in an urban low-income population in Bangladesh. Glob Health Action [Internet]. 2014 [cited 2020 Jun 30];7(1). Available from: https://pubmed.ncbi.nlm.nih.gov/25128806/
295. Wang C, Jiang B, Fan J, Wang F, Liu Q. A study of the dengue epidemic and meteorological factors in Guangzhou, China, by using a zero-inflated poisson regression model. Asia-Pacific J Public Heal [Internet]. 2014 Jan [cited 2020 Jun 30];26(1):48–57. Available from: https://pubmed.ncbi.nlm.nih.gov/23761588/
296. Cavalcante KRLJ, Tauil PL. Epidemiological characteristics of yellow fever in Brazil, 2000-2012. Epidemiol e Serv saude Rev do Sist Unico Saude do Bras. 2016;25(1):11–20.
297. Pascual M, Bouma MJ. Do rising temperatures matter? Ecology. 2009;90(4):906–12.
298. De Paiva CA, Oliveira APDS, Muniz SS, Calijuri ML, Dos Santos VJ, Alves SDC. Determination of the spatial susceptibility to yellow fever using a multicriteria analysis. Mem Inst Oswaldo Cruz. 2019 Mar 1;114(3).
299. Akter R, Hu W, Gatton M, Bambrick H, Naish S, Tong S. Different responses of dengue to weather variability across climate zones in Queensland, Australia. Environ Res. 2020 May 1;184.
300. Ahmed T, Hyder MZ, Liaqat I, Scholz M. Climatic conditions: Conventional and nanotechnology-based methods for the control of mosquito vectors causing human health issues. Vol. 16, International Journal of Environmental Research and Public Health. MDPI AG; 2019.
301. Costa EAP de A, Santos EM de M, Correia JC, de Albuquerque CMR. Impact of small variations in temperature and humidity on the reproductive activity and survival of Aedes aegypti (Diptera, Culicidae). Rev Bras Entomol. 2010;54(3):488–93.
302. Stolerman LM, Maia PD, Nathan Kutz J. Forecasting dengue fever in Brazil: An assessment of climate conditions. PLoS One. 2019;14(8).
303. Manley D. Scale, aggregation, and the modifiable areal unit problem. In: Handbook of Regional Science. Berlin: Springer Berlin Heidelberg; 2014.
304. Johansson MA, Arana-Vizcarrondo N, Biggerstaff BJ, Staples JE. Incubation periods of yellow fever virus. Am J Trop Med Hyg. 2010 Jul;83(1):183–8.
305. Manning WG, Basu A, Mullahy J. Generalized Modeling Approaches to Risk

Adjustment of Skewed Outcomes Data. NBER Work Pap No t0293 [Internet]. 2003 Oct [cited 2020 Jun 11]; Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=459402
306. Barrera R, Felix G, Acevedo V, Amador M, Rodriguez D, Rivera L, et al. Impacts of hurricanes irma and maria on aedes aegypti populations, aquatic habitats, and mosquito infections with Dengue, Chikungunya, and Zika Viruses in Puerto Rico. Am J Trop Med Hyg. 2019 May 1;100(6):1413–20.
307. Calzolari M, Albieri A. Could drought conditions trigger Schmallenberg virus and other arboviruses circulation? Int J Health Geogr. 2013 Feb 14;12.
308. Rueda LM. Global diversity of mosquitoes (Insecta: Diptera: Culicidae) in freshwater. Vol. 595, Hydrobiologia. 2008. p. 477–87.
309. Karesh JW, Mazzoli RA, Heintz SK. Ocular Manifestations of Mosquito-Transmitted Diseases. Mil Med. 2018 Mar;183(Supp 1):450–8.

Appendix A. Supplement for Chapter 2

Table A1. Summary of distribution fitting. Shows the best fitting parameter set and associated Kolmogorov Smirnov statistic for Yellow Fever incidence for each temporal resolution for both time periods

		Jan 2000 - Nov 2016			Dec 2016 - Mar 2018		
		Daily	Weekly	Monthly	Daily	Weekly	Monthly
Normal (mean, var)	Parameters	(7.078, 6.725)	(8.280, 8.021)	(9.430, 8.760)	(4.557, 3.995)	(7.060, 6.190)	(10.532, 9.566)
	KS stat	0.148	0.153	0.142	0.128	0.127	0.136
Lognormal (mean, var)	Parameters	(7.072, 6.729)	(8.284, 8.040)	(9.432, 8.760)	(4.549, 3.995)	7.048, 6.168)	(10.514, 9.534)
	KS stat	0.148	0.153	0.142	0.128	0.127	0.135
Trunc Normal (mean, var)	Parameters	(0.006, 10.155)	(0.002, 11.556)	(0.018, 12.581)	(0.014, 5.997)	(0.000, 9.261)	(0.004, 13.826)
	KS stat	0.139	0.142	0.133	0.113	0.114	0.126
Exponential (rate)	Parameter	0.112	0.099	0.091	0.192	0.127	0.086
	KS stat	0.096	0.091	0.078	0.06	0.065	0.083
Gamma (shape, rate)	Parameters	(0.787, 0.083)	(0.749, 0.070)	(0.809, 0.068)	(0.817, 0.143)	(0.747, 0.084)	(0.609, 0.042)
	KS stat	0.078	0.068	0.054	0.035	0.023	0.034
Uniform (min, max)	Parameters	(0.004, 15.945)	(0.000, 17.823)	(0.004, 20.424)	(0.006, 10.194)	(0.044, 15.903)	(0.006, 23.416)
	KS stat	0.209	0.202	0.196	0.181	0.187	0.195
Triangle (min, max, peak)	Parameters	(0.067, 28.820, 0.708)	(0.198, 31.845, 0.339)	(0.161, 34.953, 0.910)	(0.143, 15.897, 0.145)		
	KS stat	0.135	0.13	0.112	0.113		
PowerLaw (xmin, alpha)	Parameters	(1.131, 1.575)	(1.067, 1.491)	(0.987, 1.520)		(0.659, 1.438)	(0.755, 1.375)
	KS stat	0.103	0.096	0.087		0.121	0.124

Appendix B. Supplement for Chapter 3

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	36
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	36
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	38-39
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	39
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	None
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	40-41
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	40
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	40
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	41-42
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	43
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	42-43
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	None
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	40
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	44

Figure B1a. PRISMA checklist, page 1.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	43
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	44
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	41-42
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	47-48
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	None
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	48
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	48-50
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	None
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	49-50
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	50-52
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	53-54
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	54
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	i

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

Figure B1b. PRISMA checklist, page 2

Table B1. Symptomatic definitions of severe Yellow Fever by article. Cells denoted with an X indicate that the symptom was required for inclusion in the study as a Yellow Fever case. Cells denoted with "OrX" indicate that at least one symptom among those denoted was required for inclusion as a Yellow Fever case. Empty cells indicate that a symptom was not considered.

Paper Number	Authors	Country	Lab confirmed	Fever	Jaundice	Hemorrhaging	Abdominal Pain	Organ Failure	Travel
1	WHO 2008 [194]	Cote d'Ivoire	Yes	X	X				
1	WHO 2008 [194]	Mali	Yes	X	X				
21	Wamala 2012 [195]	Uganda	Yes	X	Or	Or			
22	de Filippis 2002 [196]	Brazil	Yes	Or	Or	Or			
48	Tuboi 2007 [197]	Brazil	Yes	X	X	Or	Or		Or
48	Tuboi 2007 [197]	Brazil	Yes	X	X	X			
73	Agadzi 1984 [198]	Ghana	Not stated	Or1	Or1	Or1	Or1		Or2
73	Agadzi 1984 [198]	Ghana	Not stated	Or1	Or1	Or1	Or1		Or2
73	Agadzi 1984 [198]	Ghana	Not stated	Or1	Or1	Or1	Or1		Or2
73	Agadzi 1984 [198]	Ghana	Not stated	Or1	Or1	Or1	Or1		Or2
73	Agadzi 1984 [198]	Ghana	Not stated	Or1	Or1	Or1	Or1		Or2
73	Agadzi 1984 [198]	Ghana	Not stated	Or1	Or1	Or1	Or1		Or2
81	Jones 1972 [199]	Nigeria	Yes	Or	Or	Or	Or		

Table B1 cont.

122	WHO 1993 [200]	Kenya	Yes	Or	Or		Or	Or	
167	Ingelbeen 2018 [201]	Democratic Republic of Congo	Yes	X	X				
168	deCock 1988 [202]	Nigeria	Not stated	X	X				
170	Monath 1980 [203]	Gambia	Yes	X	X				
171	Nasidi 1989 [204]	Nigeria	No	X	X				
171	Nasidi 1989 [204]	Nigeria	No	X	X				
171	Nasidi 1989 [204]	Nigeria	No	X	X				
171	Nasidi 1989 [204]	Nigeria	No	X	X				
171	Nasidi 1989 [204]	Nigeria	No	X	X				
171	Nasidi 1989 [204]	Nigeria	No	X	X				
171	Nasidi 1989 [204]	Nigeria	No	X	X				