A MIXED-METHODS APPROACH FOR VECTOR-BORNE DISEASE SURVEILLANCE IN COLOMBIA

by

Michael Richard Desjardins

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Geography

Charlotte

2019

Approved by:

_____
Dr. Eric Delmelle

_____
Dr. Irene Casas

_____
Dr. Colleen Hammelman

_____
Dr. Rajib Paul

Abstract

MICHAEL RICHARD DESJARDINS. A Mixed-Methods Approach for Vector-Borne Disease Surveillance in Colombia
(Under the direction of DR. ERIC DELMELLE)

Vector-borne diseases (VBDs) affect more than 1 billion people a year worldwide, cause over 1 million deaths, and result in hundreds of billions of dollars in societal costs. Dengue fever (DENF), chikungunya (CHIK), and Zika are three emerging VBDs that are transmitted by the Aedes mosquito. A combination of increased urbanization, globalization, climate change, and decreases in vector control have resulted in global increases in VBD epidemics, especially in previously unaffected regions. In Colombia, the co-circulation of DENF, CHIK, and Zika have resulted in severe epidemics where hundreds of thousands of people have been infected during the past decade. DENF has been endemic in Colombia for decades, and CHIK and Zika first appeared in 2013. It is critical to implement surveillance strategies that can improve the understanding of VBD transmission. Integrating mixed-method approaches in VBD surveillance are important because solely using quantitative approaches will not capture the experiences and behaviors of individuals who are susceptible to disease; and those responsible for policy-making and public health interventions.

This dissertation combines spatial and space-time statistical models, surveys, and semi-structured interviews to understand the socioeconomic, environmental, political, and institutional factors that influence the transmission of DENF, CHIK, and Zika in Colombia – one national level study; and three studies in the city of Cali. First, I detect and visualize space-time clusters of both DENF and CHIK at the national level between 2015 and 2016; and compute relative risk for each municipality that belongs to a cluster. Second, space-time conditional autoregressive (ST-CAR) models are developed to identify significant predictors of DENF, CHIK, and Zika at the neighborhood level in Cali, Colombia; and the models also include meteorological variables that are temporally lagged to predict VBD outbreaks (early warning system). Third, I administer 327 Knowledge, Attitude, and Practice (KAP) surveys to individuals in healthcare centers and select neighborhoods in Cali, Colombia in June 2019. KAP surveys are used to shed

light on at-risk communities' understanding of the vector, the pathogen, prevention and treatment strategies. I utilize Generalized Linear Models (GLMs) to identify significant predictors of KAP regarding DENF, CHIK, and Zika. The findings suggest that knowledge is related to community characteristics, while attitudes and practices are more related to individual-level factors. Access to healthcare also forms significant predictor of residents participating in preventative practices. Finally, I conduct six semi-structured interviews with high-ranking public health officials about their experiences regarding DENF, CHIK, and Zika prevention, treatment, and surveillance.

Overall, the results can be leveraged to inform public health officials and communities to motivate at-risk neighborhoods to take an active role in vector surveillance and control, and improving educational and surveillance resources in Cali, Colombia.

DEDICATION

To my late father, Richard Anthony Desjardins, Jr.

ACKNOWLEDGEMENTS

The journey of completing my PhD was the most fulfilling, fun, emotionally taxing, and challenging experience of my life so far. I never imagined that I would be able to finish in 3.5 years (if at all), considering the unexpected loss of my father and grandfather in August and October of 2017; and the loss of my other grandfather in November 2019. Graduate school is already the upper bound of emotional rollercoasters and I am eternally grateful to my friends and family in Connecticut and Charlotte for supporting me during my lowest point, and ensuring that I accomplish the dream that my father wanted more than anything for me to realize (whether he was physically here or not).

My interest in public health and epidemiology stemmed from reading *Lab 257* and *The Hot Zone* right around the time of the first Ebola epidemics in West Africa. I also wanted to combine my love for travel with my research, then the very strong partnership with my advisor (Dr. Eric Delmelle) formed. Well, he actually guided me through a master's here at Charlotte on spatial optimization of nature reserves; but his lengthy history working on dengue in Cali, Colombia with Dr. Irene Casas allowed me to accelerate into a research topic that would allow me to kill two birds with one stone (I think I accidently hit myself a few times along the way).

I am very grateful for my wonderful committee (Drs. Eric Delmelle, Irene Casas, Colleen Hammelman, and Rajib Paul) for their constant support and constructive criticism of my dissertation. I would also like to thank Dr. Matthew Eastin of UNC Charlotte for his assistance with analyzing the meterological components of my study.

Drs. Wenwu Tang and Doug Shoemaker have also been an important part of my professional growth during my time in the Center for Applied GIScience (CAGIS). I would also like to acknowledge other faculty and staff at UNC Charlotte that have played various supportive roles during my doctoral studies: Drs. Heather Smith, Elizabeth Delmelle, Joseph Kangmennaang, Wei-Ning Xiang, Harrison Campbell, Gary Silverman, and Deb Thomas; and Stephen Bowman, Teena Power, Teresa Cleveland, Patrick Jones, Elie Saliba, and Adrian Jelley.

I would like to acknowledge several people in Colombia that helped make this study possible: Drs. Dávalos and Victoria from Universidad Icesi and Dr. Dayana Carbonell from Universidad Libre for their collaboration, fieldwork support, hospitality, and Colombian adventures. Drs. Alejandro Varela and Angelica Lopez of DIME Clínica Neuro-Cardiovascular – Dr.Varela was especially critical due to his connections with literally everyone I needed to contact (his family parties are great too); the students of Universidad Libre for participating in a pilot study; Jaydiber Garcia Vaca for her help with the door-to-door surveys and treating me like family during my month in Cali (best cook in Colombia!); Robinson Pacheco of Universidad Libre for his help organizing the collaboration between UNC-Charlotte, Universidad Libre, and Universidad Icesi; the directors of the five healthcare centers where I conducted surveys; the public health officials for finding time for me to interview them; and all of the participants who volunteered their time to fill out the surveys.

I am eternally grateful for the extremely bright and caring friends I've made here at Charlotte: Claudio Kwamena Adams Owusu, Dr. Alexander Hohl, Eugene Keteni, Sara Tornabene, Abel Ayon, Yu Lan, Minrui Zheng, Dr. Adam Griffith, Adam Miner,

Dr. Coline Dony, Dr. Wenpeng Feng, Jiyang Shi, Sam Nickerson, Shawn Cummins, Matt Brooks, Dr. Claire Shuch, Dr. Ari Whiteman, Danny Yonto, Titiksha Fernandes, and Paul Jung. They've all mitigated the unrelenting perils of graduate school in various ways, preferably cracking jokes over a nice pint of craft beer and comfort food; or spending hours on the soccer pitch, scoring goals against students almost a decade younger than me.

I am also grateful for my life-long friends in Connecticut, for always making it seem like I was never gone when we got back together. Their spontaneous visits to Charlotte and summer vacations always cheered me up and reminded me that I will always be the maturely immature dude from Waterbury before "Dr. Desjardins from [insert future employers]". I love the following men like brothers: Adam Lawley, Michael Strumski, Jordan Coutu, JR Viglione, and Frank Pisani. We will forever be the Spicy Boys and Juicers!

To my family for their constant unconditional love, financial support, and becoming even stronger after experiencing the journey to hell and back since 2017. To my sister, Alexandra, for her sense of humor and being my teammate for 25 years. To my mother, Stephanie, for being the greatest mom in the world. You have always said that I got my "brains" from you (thankfully I have Dad's patience!), but I am pretty sure it was a team effort! No matter what we experienced as a family or you individually, she always made sure I had the means to an end, or the means to embark on my countless adventures.

To my favorite person and best friend, Vani Singh, for her unconditional love and support during the last year. My life has been better with you in it and I am grateful for

our numerous road trips, bookstore and coffee shop adventures, weirdness in the best way possible, countless heart-to-heart conversations, and endless hugs and affection.

Finally, to my late father, Richard Anthony Desjardins Jr., I am honored to carry your name (my middle) and was privileged to have you as a father. Despite your unexpected loss, you have made me the man I am today and your sacrifices during your life that was cut too short was and still is critical to my successes today. I did it Dad. After seeing the 800+ people at your funeral, I can only hope to have the same impact on the amount of lives that you did. I will forever be your “bud” and “dude”.

December 2019

Charlotte, NC

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Epidemiology

The World Health Organization (WHO) defines epidemiology as the "study of the distribution and determinants of health-related states or events (including disease), and the application of this study to the control of diseases and other health problems" (WHO 2018). Those who study and practice traditional epidemiology typically belong to a department or school of public health sciences. Hippocrates is widely known as the father of medicine and the first epidemiologist, being the first individual to study the link between environmental exposures and disease; for example, he observed that malaria and yellow fever occurred in swampy areas (Merrill 2015, p.24). Since then, epidemiology has greatly advanced as a scientific discipline and commonly employs the following types of approaches: (1) cross-sectional – sample or survey of individuals that examine disease exposure and status within a particular time period; (2) cohort – observing exposed and unexposed populations over time to examine particular health outcomes; (3) case-control – comparing individuals with and without a specific disease regarding their exposures; (4) and clinical trials – studying the outcomes of vaccines or drugs by randomly allocating individuals to a treatment and placebo group (Rainwater-Lovett et al. 2016). A variety of quantitative and qualitative methods (influenced by many ontologies and epistemological frameworks) can be employed in each of the four abovementioned study types.

1.1.1 Limitations of Traditional Epidemiology

Although traditional epidemiology has vastly improved our understanding of disease transmission and preventative measures, studies lack spatial methods, analysis,

and visualization, which can create uncertainty and result in a theory-practice gap. For example, Zhang et al. (2019) found a significant relationship between cardiovascular disease and mild and moderate/severe depression among U.S. adults. However, Zhang et al. do not acknowledge the geographic variations of cardiovascular disease and depression, which significantly influences an individual's risk of the two aforementioned diseases (Pedigo and Aldrich 2011; Trgovac et al. 2015). Integrating spatial analysis in an epidemiological study can facilitate targeted interventions and improve public health policy and decision-making by identifying specifically where at-risk populations are located and what is influencing disease risk and exposure.

1.1.2 Spatial Epidemiology

Elliot and Wartenberg (2004) define spatial epidemiology as "the description and analysis of geographically indexed health data with respect to demographic, environmental, behavioral, socioeconomic, genetic, and infectious risk factors". Research in spatial epidemiology is primarily concerned with disease mapping, geographical correlation studies, risk assessment in relation to point or line sources, and cluster detection and disease clustering (Elliot et al. 2000). Others have defined spatial epidemiology as: "the analysis of the spatial/geographical distribution of the incidence of disease" (Lawson 2013) and "the spatial perspective into the design and analysis of the distribution, determinants, and outcomes of all aspects of health and well-being across the continuum from prevention to treatment" (Kirby et al. 2017).

The cholera map produced by Dr. John Snow in 1854 was the first documented case of applying spatial analysis to facilitate the understanding of a disease outbreak (Snow 1856), while demonstrating that the concept of place and public health outcomes

are inherently related. Essentially, Snow mapped individual cholera deaths in London, which allowed him to determine that the vast majority of the cases were located closest to one particular well compared to every other drinking source in the city. The field of spatial epidemiology has since played an important role in improving the understanding of the processes responsible for the spread of disease (Ostfeld et al. 2005). Overall, a holistic approach to spatial epidemiology should integrate exploratory approaches to detect significant outbreaks of a disease to facilitate targeted interventions; conduct local-level analyses to examine the factors that influence disease transmission; and implement qualitative approaches to understand the perspectives and behaviors of individuals susceptible to disease and their access to healthcare resources.

1.1.3 Advances in Spatial Epidemiology

Current research in spatial epidemiology has been greatly influenced by advancements in geographic information science (GIScience), especially geographic information systems (GIS; Kirby et al. 2017). For example, mapping the spatial variation in disease rates and risk is vital for and formulating aetiological hypotheses. Identifying and mapping disease outbreaks in space and time can provide information that can be useful in planning public health interventions (Hay et al. 2013; Peterson 2014; Pigott et al. 2015). Monitoring and analyzing outbreaks under critical space-time conditions can increase the efficiency of public health responses (Kitron 1998; Kitron 2000; Jacquez et al. 2005; Rogerson and Yamada 2008; Eisen and Eisen 2011). Spatial and space-time analytics are particularly salient to estimate the dynamics of disease (Eisen and Lozano-Fuentes 2009), such as the rate of disease spread, cyclic pattern, direction, intensity, and risk of diffusion to new regions. As novel technologies emerge and data becomes

available, new epidemiological questions will arise requiring to investigate additional facets of spatial and space-time analytics. For instance, population data become increasingly detailed with respect to their spatial and temporal resolutions, which will enable methods to adjust for spatially and temporally inhomogeneous background populations (Hohl 2018). In addition, as techniques for tracking or inferring individual people's location are already available at large scales, research about spatial and space-time disease studies may shift focus from the point- and polygon-based paradigms to trajectory-based methods. High performance computing technologies (e.g. parallel computing) has enabled big geospatial health data processing and analysis (Saule et al. 2017).

1.2 Vector-borne diseases

Vector-borne diseases (e.g. malaria, dengue fever, Zika, chikungunya, etc.) affect more than 1 billion people a year worldwide, cause over 1 million deaths, and cost hundreds of billions of dollars in societal costs (World Health Organization 2014b). Mosquitoes are the most common vectors, responsible for transmitting a variety of arboviruses. The diversity of mosquito-borne pathogens (Beckham and Tyler 2015), a growing geographical vector range (Benedict et al. 2007; Rochlin et al. 2013), and a mid-century decline in vector control efforts (Floore 2006), have led to global increases in arbovirus disease transmission in recent decades (World Health Organization 2014a). In regions where multiple pathogens are widespread, comprehensive efforts to understand the interaction of multiple disease outbreaks across a heterogeneous landscape can aid in outbreak prevention and response (Ochieng et al. 2013).

Recently, several particular pathogens, most notably dengue fever (DENF) and chikungunya (CHIK), have been responsible for the majority of the burden caused by mosquito-borne diseases (Wang et al. 2016). Both diseases are caused by viruses that are primarily transmitted by the *Aedes aegypti* mosquito (Harrington et al. 2001), with *A. albopictus* acting as a secondary vector (Gratz 2004; Tsetsarkin et al. 2007; Paupy et al. 2010; Chouin-Carneiro et al. 2016). These two species are peridomestic container-breeding mosquitoes that have become prolific in urban areas due to the widespread availability of breeding habitats (Tauil 2001; Powell and Tabachnick 2013). A combination of climate change (Liu-Helmersson et al. 2014), rapid urbanization (Tauil 2001), and globalization (Charrel et al. 2007) have expanded the vector's range and caused the two infectious diseases to emerge in novel regions.

DENF, which can be caused by any of the five known viral serotypes (DEN 1 to 5), is the most rapidly spreading arbovirus on Earth (Wilder-Smith et al. 2017). Over 40% of humans are at risk of transmission, with incidence rising 30-fold in the last 50 years; and it is estimated that there are approximately 390 million DENF infections annually (Bhatt et al. 2013). Additionally, the viruses have recently spread geographically to include novel outbreaks in Europe in 2010 and the United States in 2013 (Bhatt et al. 2013). CHIK had been restricted to Africa, Southeast Asia, and India prior to 2013 when it was first detected in the Americas and Caribbean. In the subsequent 2013-2015 CHIK epidemic in the Americas, an estimated 39.5 million people were infected and an estimated societal cost of US$185 billion was incurred (Bloch 2016; Shepard 2010). Zika has also recently appeared in novel regions and was first discovered in 1947 in the Zika Forest of Uganda and was relatively rare until the 2014-2016

outbreaks in the South Pacific and Brazil (Dick et al. 1952; Duffy et al. 2009; Campos et al. 2015; Hennessey et al. 2016). Since 2015, over 90 countries around the world are at risk of Zika transmission (CDC 2018). Zika is also spread by the peridomestic container-breeding *Aedes aegypti* and *Aedes albopictus* mosquitoes, which also transmit yellow fever. Worries remain about potential novel regions becoming infected as well as the risk of antibody-dependent enhancements among populations with a history of hosting other arboviruses (Dejnirattisai et al. 2016; Kawiecki and Christofferson 2016; Durbin 2016).

1.2.1 DENF Characteristics and Clinical Manifestations

Dengue is a flavivirus that causes DENF, and there are five known serotypes of the virus, but the fifth variant follows the sylvatic cycle and the first four follow the human cycle (Mustafa et al. 2015). The incubation period ranges from 3-14 days after being bit by an infected mosquito, and symptoms can last from 2-7 days (WHO 2018), however, approximately 80% of infected individuals are asymptomatic. Symptoms of non-severe dengue can include: headache, malaise, nausea/vomiting, abdominal pain, rash, retro-orbital pain, myalgia, arthralgia, and mild bleeding (Kalayanarooj 2011). Infection from one serotype will result in lifelong immunity to that serotype, however, secondary infection with another serotype can lead to severe forms of dengue (WHO 2011), such as dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS). Both DHF and DSS can have the same symptoms of non-severe dengue, but may also include plasma leakage, skin hemorrhages, lesions, gastrointestinal bleeding, and multi-organ failure (Srikiatkhachorn et al. 2010); while DSS is DHF with circulatory failure (Rajapakse 2011). DHF and DSS primarily affects pediatric patients, but it has also been

found in adults (especially the elderly); and mortality from dengue is highest among children and those who experienced DSS (Gubler 1998).

1.2.2 CHIK Characteristics and Clinical Manifestations

Chikungunya is an alphavirus that causes CHIK, which is both an emerging and neglected tropical disease. Unlike DENF and Zika where the majority of infected individuals will be asymptomatic, between 50-97% of those infected with CHIK will be symptomatic (Nakkhara et al. 2013). The incubation period ranges from 1-12 days, and the subsequent symptoms may include: fever, arthralgia, backpain, headache, lesions, gastrointestinal issues, rash, fatigue, and conjunctivitis (Thiberville et al. 2013). Furthermore, the symptoms can last up to a week (Weaver and Lecuit 2015), however, chronic complications after the acute infection subsides is common. Symptoms of chronic CHIK may include polyarthritis and polyarthritis, which can persist for months and even several years, resulting in mobility issues (Sissoko et al. 2009; Hoarau et al. 2010; Simon et al. 2011). In rare cases, an infected individual may develop Guillain–Barré Syndrome or other atypical clinical manifestations (Oehler et al. 2015; Cunha and Trinta 2017). Maternal-fetal vertical transmission of CHIK is also possible (Lyra et al. 2016), which may result in congenital fever, apnea, encephalitis, hemorrhage, lesions, and other symptoms (Gopakumar and Ramachandran 2012).

1.2.3 Zika Characteristics and Clinical Manifestations

Zika is a flavivirus that causes Zika fever, and is closely related to the viruses that cause DENF, yellow fever, West Nile, and Japanese encephalitis (Chen and Hamer 2016). A variety of symptoms may develop after an individual is bit by an infected mosquito (incubation period of 3-12 days), including: mild fever, rash, arthralgia,

arthritis, myalgia, headache, conjunctivitis, and edema (Paixão et al. 2016), and the symptoms may last 2-7 days. However, an estimated 80% of individuals infected with the Zika virus are asymptomatic (Duffy et al. 2009), while the virus will remain in the blood for approximately one week (CDC 2018). Non-vector-borne modes of transmission are also possible for Zika, including sexual contact (vaginal, anal, and oral), blood transfusions, and vertically via pregnant mother to child (Calvet et al. 2016). Furthermore, congenital disorders are associated with Zika, especially microcephaly and Congenital Zika Syndrome (Guilland 2016). In rare cases, Zika has been linked to neurological disorders, including Guillain–Barré Syndrome, acute myelitis, and meningoencephalitis (Araujo et al. 2016).

1.2.4 DENF, CHIK, and Zika Treatment and Control Strategies

It is critical to implement surveillance strategies that can improve the understanding of VBD transmission. VBD surveillance may involve the examination of disease incidence in human populations, including the (spatial) variations among socioeconomic groups, age, and sex; the geographic distribution of vector populations capable of transmitting various VBDs, especially identifying suitable habitats (e.g. environmental variables); and analyzing human movement and interaction with their environment that may facilitate disease transmission (Palaniyandi et al. 2017). Improving vector-borne disease surveillance can facilitate the timely reporting of disease cases, reduce underreporting, inform policy-makers, increase disease awareness, define funding and research priorities (Toan et al. 2015); ultimately reducing the economic and public health burden in at-risk locations around the world (Shepard et al. 2016). VBD control surveillance and prevention programs should consider holistic studies that also

address the political and social forces that will influence decision-making (Tedesco et al. 2010). Since there is no available vaccine or medication to cure or prevent DENF, Zika, and CHIK, surveillance and control strategies should target locations with the highest risk and reported cases/rates of the disease; then inform the community about mosquito prevention techniques, such as changing water from containers, examining the interior and exterior of property for mosquito larvae and adults, and wearing proper clothing and using mosquito spray during a local epidemic.

1.3 Research Objectives

This dissertation is articulated around three objectives. First, an exploratory univariate and multivariate space-time scan statistic identifies significant space-time clusters of vector-borne diseases in Colombia. Since *Aedes* can transmit a variety of VBDs, the multivariate space-time scan statistic can identify co-occurrence of disease in space and time; and this approach has not been found in the literature in the context of VBDs. The resulting space-time clusters are visualized in a 3D-environment, which is especially novel and can improve the understanding of space-time dynamics of disease clusters. Second, more fine-scale spatial analyses in epidemiology are needed to facilitate targeted interventions and prioritize the allocation of resources. Space-time autoregressive modeling can examine the relationships between a variety of covariates and disease risk and rates. Furthermore, conducting this type of analysis at the neighborhood-level can identify neighborhoods with the highest risk of VBD transmission and the associated factors that can increase an individual's risk of contracting a VBD; and also predicting disease outbreaks using temporally lagged weather and climate variables. Finally, models and statistics do not capture the behaviors

and perspectives of those who are susceptible to disease. Integrating surveys and semi-structured interviews into an epidemiological study can be used to validate and inform modeling, and further improving targeted interventions and public health strategies to mitigate epidemics. Combining exploratory clustering and fine-scale space-time autoregression models with surveys and interviews can be effective, informative, and fall under the category of 'holistic' epidemiology. The contributions stem from three major objectives, which are further discussed below:

1.3.1 Objective 1

I detect univariate and multivariate space-time clusters of vector-borne diseases (DENF and CHIK) in Colombia using space-time scan statistics. I subsequently visualize the resulting space-time clusters using a variety of two- and three-dimensional geovisualization techniques. Next, I compare the space-time clusters of DENF and CHIK, and examine the co-occurrences (multivariate) in space and time. Finally, I compute relative risk for municipalities that belong to a significant cluster to facilitate local-level analysis and targeted interventions.

1.3.2 Objective 2

I examine the influence of socioeconomic, environmental, weather and climate, and spatial variables on vector-borne disease outbreaks in Cali, Colombia, a city that was found within a significant space-time cluster in objective 1. Next, I develop a variety of space-time conditional autoregressive models to identify significant socioeconomic, environmental, weather and climate, and spatial covariates of vector-borne disease outbreaks at the neighborhood-level. I determine if DENF, CHIK, and Zika rates and covariates in one neighborhood are influenced by rates and covariates in surrounding

neighborhoods and time periods. I also predict disease outbreaks using temporally lagged weather and climate variables that were selected via cross-correlations.

1.3.3 Objective 3

I evaluate local familiarity with vector-borne disease transmission and intervention strategies in Cali, Colombia. To do so, I administered Knowledge, Attitude, and Practice (KAP) surveys to residents in a variety of healthcare centers, universities, and door-to-door in certain neighborhoods. I examine how KAP may vary by disease (e.g. endemic vs. new), and by socioeconomic status (low, middle, and high). I also conducted six semi-structured interviews with public health officials, regarding the intervention strategies and educational campaigns in Cali. I analyze the interviews using content analysis and compared the perceptions of the residents from the surveys and the stakeholders from the interviews; essentially examining VBD awareness and the effectiveness of public health interventions and policy in Cali. Objective 3 collects both community and stakeholder information, which can be utilized to facilitate the explanation of the modeling results of Objectives 1 and 2.

1.3.4 Innovation

This is the first dissertation of its kind that combines expoloratory cluster analysis, predicitive and explanatory modeling, and qualitative approaches (i.e. surveys and interviews) to study VBDs in a holistic framework (see "holistic spatial epidemiology" in section 2.2). The three abovementioned objectives inform and supplement each other – filling in knowledge gaps that would be apparent if each study was treated separately. Objective 1 identifies Cali as a persistent high-risk location of VBDs. Objective 2 then presents the complex predictors of VBD outbreaks in Cali at the

neighborhood and weekly levels. Objective 3 presents the results of community surveys and public health official interviews that highlight the human, cultural, and political challenges of VBD surveillance and control, which can not be captured in Objectives 1 and 2. My dissertation improves VBD surveillance my suggesting a holistic framework that combines quantitative and qualitative techniques to uncover key information that are typically mentioned as "future research" or "limitations" in the literature (due to the disjointedness of single research papers in medical geography and public health).

1.4 Road Map

This remainder of this proposal is organized as follows: Chapter 2 reviews literature for the three main objectives: exploratory spatial and spatiotemporal clustering approaches (including 2D and 3D visualization techniques); modeling approaches to examine place-based determinants of VBDs; and qualitative approaches in VBD surveillance, with a focus on KAP surveys and semi-structured interviews. Chapter 3 provides the study that addresses objective 1, which was published in *Acta Tropica* in 2018 (Desjardins et al. 2018b); Chapter 4 discusses the study that addresses objective 2; and chapters 5 (submitted to *Health & Place*) and 6 discusses the studies that address objective 3. Overall, this dissertation contributes to the domains of medical geography, vector-borne disease surveillance, and epidemiology.

CHAPTER 2: LITERATURE REVIEW

This section provides a literature review pertinent to the mixed-methods approach utilized in this dissertation; and the three main research objectives, including the knowledge gaps that my research is attempting to bridge. The following topics will be discussed: mixed-methods research, which is the main purpose of this dissertation; (Objective 1) – disaggregate vs. aggregate data, spatial and space-time clustering approaches, and visualization techniques; (Objective 2) – place-based determinants of VBDs and a variety of spatial regression models; and (Objective 3) – qualitative approaches for VBD surveillance with an emphasis on KAP surveys and semi-structured interviews.

2.1 Mixed-Methods Research

Mixed-methods research is a well-respected and effective research paradigm that can address multifaceted problems that cannot solely be answered using exclusively quantitative or qualitative approaches. Despite the numerous descriptions of mixed-methods research found in the literature by leading scholars (see Johnson et al. 2007, Table 1), the general consensus is that mixed-methods integrates both quantitative and qualitative methods of data collection and analysis into a single study (Creswell 1999). Mixed-method approaches can better understand the "how" and "why" questions by integrating humanistic perspectives, such as interviews and surveys (Yin 2013). Quantitative methods typically provide generalizations about the phenomena being studied, which may not account for fine-level variation that may also produce new knowledge; and qualitative methods can consider the human element in research, which may reduce uncertainty and strengthen local-level analyses; while objective research and

the subsequent generalizations are necessary to implement and manage broader policies, because it is virtually impossible to satisfy every individual. Applying a mixed-methods framework can provide a complete and comprehensive view of what is being examined, stronger inferences and improved explanations can be obtained while offsetting the weaknesses of both quantitative and qualitative approaches, answering researchers questions that cannot be answered with a single method, and facilitate hypothesis testing and modeling development (Doyle et al. 2009). Within the context of geographic research, knowledge obtained from mixed-method approaches can provide detailed insight about the spatial heterogeneity or homogeneity of the phenomena being studied.

For example, Shay et al. (2016) examined transportation disadvantages (TD) in five rural counties in North Carolina. A mixed-methods approach was employed that combined quantitative and geographic information systems (GIS) data with data collected from focus groups and interviews. The results of the qualitative data collection were used to revise the original maps, essentially identifying the TD populations at the census-tract level within each county. As the authors stated: "interviewers and focus groups provided valuable feedback on how well the GIS maps accurately reflected local conditions" (p. 136). Therefore, this paper epitomizes the advantage of validating findings from GIS maps with ground-truthing and public-participatory mapping, identifying more accurate information that objective measures can fail to capture (human behavior). Chen et al. (2010) examined the potential ambient air pollution exposure misclassification for mothers who moved during pregnancy in New York State from 1995-2002. The authors surveyed 1,324 mothers to obtain data regarding addresses before pregnancy, during pregnancy, and at birth; and other key factors including

maternal age, education, BMI, race/ethnicity, smoking status, alcohol consumption, etc. The surveys were combined with GIS, statistics, and air pollution models to assess the exposure levels and distance moved for the participating mothers. The authors did not identify significant evidence of exposure misclassification by using the maternal address on a birth certificate rather than the addresses by gestational age. They also highlight that misclassification "may be a function of both mobility and the size of the exposure regions". Therefore, the combination of the surveys (humanistic) and quantitative analysis (positivistic) more accurately evaluates and validates the nexus between residential mobility during pregnancy, air pollution exposure, and birth outcomes.

Whether qualitative or quantitative approaches are prioritized in a mixed-methods design, it is critical that the researcher(s)' results and conclusions from both type are drawn together to form a general conclusion, rather than presenting them separately which can create disjointedness and defeat the purpose of mixed-methods (Bazely 2004). Bazely (2004) also mentions that mixed-methods studies should acknowledge the assumptions and implications of both qualitative and quantitative approaches; for example, researchers may only integrate a small number of qualitative samples (e.g. surveys) to supplement the quantitative results. As a result, the conclusions drawn may not actually broaden the perspectives and mitigate the weaknesses of solely using one method, which is one of the goals of mixed-methods research. Plausible rival hypotheses (Yin 2013, p. xvii) should also be considered and emphasizes the possibility that other factors affect a study's results that are not considered during the research process. For example, a study in medical geography may claim that living within close proximity to green spaces (e.g. parks) reduces the risk of noncommunicable diseases (NCD), such as

obesity and heart disease. However, being aware of the health benefits of physical activity, socioeconomic status, access to healthy food, age, and disabilities will also affect an individual's risk of developing a NCD. Considering plausible alternative hypotheses will improve the understanding of a particular research topic, and addressing the inherent uncertainty of studies in general.

However, using a mixed-methods approach does not necessarily make a study more valid than single-method studies. Validation procedures must be conducted for both methods to ensure the significance, reliability, and ability to improve the understanding of real-world problems (Venkatesh et al. 2013). For example, validation methods in quantitative studies (e.g. modeling), use a variety of metrics to quantify accuracy (error metrics). For example, root-mean-square-error (RMSE) quantifies the difference between predicted values and observed values (Willmott and Matsuura, 2005). Finally, mixed-methods research requires knowledge and training with various methods, which may be challenging due to time, resources, and an individual's background. The lack of expertise in one research paradigm (quantitative or qualitative) may require collaboration across multiple disciplines when a mixed-methods approach is desired. Finally, the qualitative component of mixed-methods research contains ethical concerns and will likely require approval from an institutional review board (IRB), or similar ethics committee. For example, participants in qualitative research must volunteer, understand the study's purpose, risks must be minimized, while confidentiality of the participants must be preserved (Terrell 2012). Despite the complexity and multifaceted nature of mixed-methods research, effectively utilized this third research paradigm can improve our understanding of the real-world by pursuing a "golden mean".

Johnson et al.'s (2007) general definition of mixed methods research is arguably the most concise and clear description found in the literature: "Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration" (p. 123). Mixed-methods research should not just combine multiple methods, rather, it should also include multidisciplinary and multi-level collaboration between academics, professionals, government, and communities to address the complex problems in geography and social sciences. As a result, the mixed-methods and multidisciplinary approach can facilitate the establishment and improvement of research proposals, research labs, community and government partnerships, and student training; ultimately ensuring that research conducted in geography is theoretically sound, practical, and contains short-term and long-term goals that directly address real-world problems and improve policy and human well-being. In conclusion, a true "golden mean" and mixed-methods approach should incorporate both quantitative and qualitative methods from multiple epistemological frameworks; adopting different approaches to validate research projects; and informing policymakers and community leaders about the improved understanding about the phenomena under study.

2.2 "Holistic Spatial Epidemiology"

For this dissertation, I propose a new theoretical framework called "holistic spatial epidemiology" that motivates my research philosophy. Holistic Spatial Epidemiology (HSE) is a theory that can facilitate the bridging of the theory-practice gap

in spatial epidemiology and public health research, in general. HSE addresses the wickedness of public health problems by considering the social, environmental, political, and geographic determinants of health outcomes; and also producing knowledge that can be effectively utilized by public health stakeholders and policymakers to reduce the burden of disease. Furthermore, HSE maximize practicality by including the following components:

1. Exploratory cluster analysis – detect significant outbreaks of a disease to facilitate targeted interventions.

2. Confirmatory approaches – conduct local-level analyses to examine the factors that influence disease transmission.

3. Qualitative methods - understand the perspectives and behaviors of individuals susceptible to disease and their access to healthcare resources.

4. Policy implications – ensuring that a study's findings can inform and improve public health policy.

5. Community education and outreach – inform at-risk communities about research findings.

Each component will be introduced and some examples from the literature will be provided in the subsequent subsections. It is important to note that it is virtually impossible to include all five of the HSE components in a single study. A good scholar-practitioner in spatial epidemiology should collaborate with scholars and professionals across multiple disciplines; essentially producing a body of research (i.e. multiple studies) which studies a particular disease that incorporates the five components of HSE. The holistic approach can help scholars in spatial epidemiology reach a "golden mean" by integrating quantitative and qualitative techniques, incorporating multiple epistemological frameworks; adopting different approaches to validate research projects;

and informing policy-makers and community leaders about the improved understanding about disease transmission and surveillance.

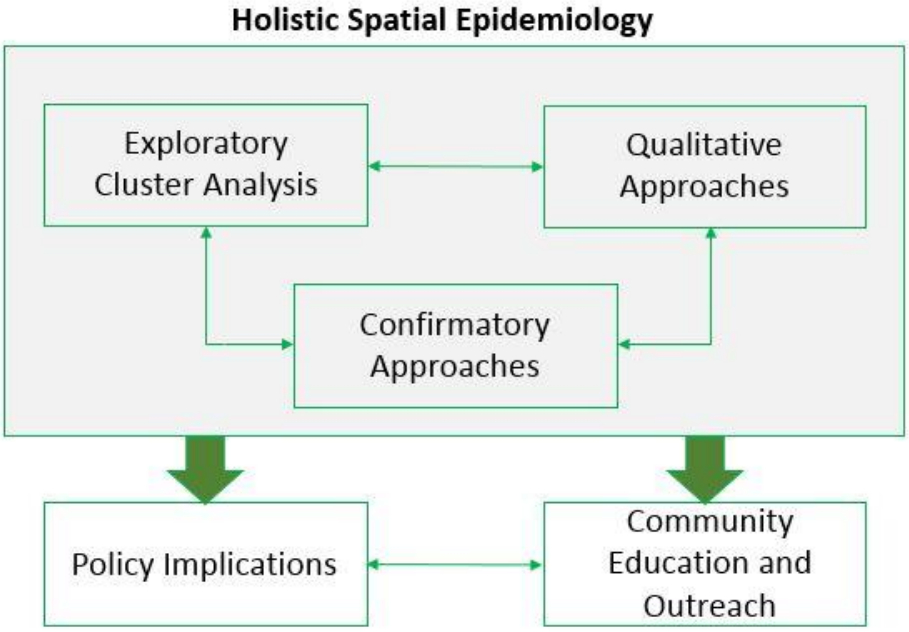


Figure 1: Conceptual Framework of Holistic Spatial Epidemiology

A conceptual framework of HSE is illustrated in Figure 1, which includes the five abovementioned components. Figure 1 suggests that a combination of methods (components 1-3) that are influenced by multiple epistemological frameworks should be utilized to study diseases. There is no particular order for components 1-3; for example, both confirmatory and qualitative approaches can occur simultaneously to shed light on the variety of factors that are responsible for disease clusters and transmission. Conversely, exploratory cluster analysis could be the first step in a spatial epidemiological analysis – identifying significant outbreak areas, and finer-level qualitative and confirmatory approaches can investigate the transmission dynamics of the epidemic. A combination of quantitative and qualitative approaches in a single research paper/project could be considered as a mixed-method approach for spatial epidemiology

(discussed in the previous section). Researchers should always be aware of the policy implications (component 4) of a study, ensuring that the methods and results can produce new knowledge that improves public health decision-making. Finally, community education and outreach can facilitate public participation and understanding of diseases that are prevalent in their community; and the public feedback can be utilized to refine research goals and also inform policymakers as well. The conceptual framework of HSE suggests that scholars, decision-makers, and at-risk communities can collaborate to mitigate the burden of disease.

2.2.1 HSE and COMBI

The World Health Organization's Communication for Behavioural Impact (COMBI) framework (WHO 2012) addresses disease outbreak control and prevention in community settings; supporting the notion that transmission dynamics are complex, context depedent, and fine-level analysis is necessary to improve health outcomes for at-risk individuals and communities. My HSE framework contains similar components as COMBI, however, COMBI's main focus is community outreach and planning; relying on many qualitative techniques to imrove epidemiological surveillance. The HSE framework can supplement COMBI by adding the mixed-methods component, especially the exploratory cluster analysis and confirmatory approaches. HSE and COMBI can be utilized together to ensure community needs are addressed before, during, and after and outbreak of disease.

2.3 Spatial and Spatiotemporal Clustering Statistics to Examine VBDs

Identifying disease clusters in space and time is typically the first step in VBD surveillance. Exploratory approaches in GIScience and spatial epidemiology can identify

high-risk areas, essentially where the observed/reported disease cases exceed the expected cases under baseline conditions. After identifying clusters, researchers can then begin to examine the place-based and local-level factors that influence VBD risk and incidence. There are a number of different methods available that can quantify the spatial and spatiotemporal clustering of VBD case data at individual or aggregated levels.

2.3.1 Individual-Level Data

The availability of individual-level data in health-related research has substantially increased due to geocoding tools available in geographic information system (GIS) software packages (Zandbergen 2014). Geocoding is the process of converting addresses to coordinates that can be displayed on a map as points (Owusu et al. 2017). The points typically represent the address of sick (cases) or at-risk individuals. Although cluster detection can be more accurate when individual-level data is utilized, privacy laws often prohibit health-related maps from being published that can theoretically identify someone's location. In the United States, the Health Insurance Portability and Accountability Act (HIPAA) was enacted in 1996 to protect individually identifiable health information (Centers for Disease Control and Prevention 2003). For geographical analyses, HIPAA states that any subdivisions smaller than state (e.g. geocodes) must eliminate identifiable information before publication (Tellman et al. 2010), essentially aggregating the data to a larger areal unit. However, the first three digits of a zip-code can be published if it contains a population greater than 20,000 people. Geomasking is another alternative, which essentially alters the coordinates to protect the true location of individuals (Armstrong et al. 1999). However, there are a variety of methods that can

quantify the spatial and spatiotemporal clustering of individual-level disease cases without the need for aggregation or geomasking.

2.3.2 Clustering Approaches for Individual Case Data

The spatial and space-time Ripley's K function evaluates the distribution of events to determine the magnitude of clustering at different spatial and temporal distances (Bailey and Gatrell 1995). The spatial and space-time Knox test identifies statistically significant case clusters at defined spatial and temporal distances (Kulldorff and Hjalmars 1999). The Mantel Index (Mantel 1967) addresses the limitations of the Knox test by assigning higher weight to nearby events both spatially and temporally. Using information from these previous tests, both spatial and space-time patterns of diseases can then be visualized. For example, space-time kernel density estimation (STKDE), which is an extension of the traditional kernel density estimation (Silverman 1986), can help visualize disease "hotspots" in both space and time, resulting in heat volume instead of heat maps (Brunsdon et al. 2007; Nakaya and Yano 2010). Since STKDE produces a heat map, the specific location of individuals is usually hidden. The spatial and space-time Knox test, Mantel Index, spatial and space-time Ripley's K function, and STKDE have all been used to map and monitor vector-borne disease outbreaks, including Dengue fever (Delmelle et al. 2013; Delmelle et al. 2014; Hohl et al. 2016), West Nile (Theophilides et al. 2003), Rift Valley Fever (Métras et al. 2012), and Chikungunya (Nsoesie et al. 2015).

2.3.3 Aggregated Data

Due to privacy concerns and data availability, health-related data is often aggregated at a variety of areal units, such as zip code, census tract, municipality, etc.

Aggregated data can suffer from the modifiable areal unit problem (MAUP), which is especially apparent spatial epidemiological studies. For example, the MAUP (Openshaw 1984) can hide local variation and can distort or exaggerate spatial patterns. The MAUP is especially an issue at coarse geographic scales (e.g. county, state, country). Aggregating data is still very useful in spatial epidemiological studies, especially when examining disease rates per population for each unit in the study area. There are a variety of approaches that can identify statistically significant clusters of disease using aggregated data, such as autocorrelation and scan statistics.

2.3.4 Spatial Autocorrelation Statistics

Spatial autocorrelation statistics are based on the first law of geography, which states that "everything is related to everything else, but near things are more related than distant things" (Tobler 1970). Positive spatial autocorrelation indicates that objects (e.g. data values) are very similar compared to distant objects within a study area, and negative spatial autocorrelation indicates that nearby objects are dissimilar. For spatial epidemiological studies, global and local autocorrelation approaches can determine the degree of disease clustering, dispersion, or spatial randomness.

2.3.5 Global Autocorrelation Methods

Global autocorrelation methods produce one statistic for the entire study area, and the most commonly used approach is Global Moran's I (Moran 1950). The Moran's I Index ranges from -1 to 1, while -1 indicates strong negative spatial autocorrelation, 0 indicates complete spatial randomness, and 1 indicates strong positive spatial autocorrelation. The Geary's C statistic (Geary 1954) is inversely related to Global Moran's I and is more sensitive to local spatial autocorrelation. Values between 0 and 1

indicate positive spatial autocorrelation, and values greater than 1 indicate negative spatial autocorrelation. Global autocorrelation approaches have been used widely in spatial epidemiology, such as examining Lyme disease in Wisconsin (Kitron and Kazmierczak 1997), influenza in Vellore, India (Lopez et al. 2014), and dengue fever in northern Thailand (Nakhapakorn and Jirakajohnkool 2006). However, global autocorrelation assumes homogeneity and does not identify where the clustering or dispersion of disease is occurring in the study area.

2.3.6 Local Autocorrelation Methods

Local autocorrelation methods identify statistically significant clusters of high values and low values, which can then be visualized on a map. Two of the most common local autocorrelation statistics are Local Indicators of Spatial Association (LISA; Anselin 1995) and the Getis-Ord Gi* statistic (Getis and Ord 1992). The LISA statistic identifies features (e.g. census tracts that have higher or lower rates than is to be expected by chance). Furthermore, a feature can be classified as one of five categories: not significant, high values surrounded by high values, high values surrounded by low values, low values surrounded by high values, and low values surrounded by low values. The Getis Ord Gi* statistic identifies significant spatial clusters of high values (hot spots) and low values (cold spots). The literature has been mainly concerned with identifying hot spots, which would indicate that a disease outbreak has occurred in that area. For example, local autocorrelation approaches have been used to identify and visualize clusters of cervix cancer in the United States (Goovaerts and Jacquez 2005), different causes of death in Hamilton, Ontario (Burra et al. 2002), and Chagas disease in Brazil (Martins-Melo et al. 2012).

2.3.7 Scan Statistics

Scan statistics are commonly used in spatial epidemiology to identify and evaluate spatial, temporal, or spatiotemporal clustering of disease cases (Kulldorff 1997). While local autocorrelation statistics detect hotspots by testing for spatial dependence of the data values, scan statistics detect clusters that are outliers (e.g. unexpected clustering given baseline conditions). Scan statistics can evaluate clustering of disaggregated data but are most commonly used to examine aggregated data. Essentially, scan statistics determine if the number of disease cases in a defined area is greater than the expected number of cases, such as the underlying population contained in the study area. The statistic utilizes circles or ellipses (scanning window) that are centered on grid points and move (scan) systematically across a study area to identify clusters of cases (each window counts number of aggregated cases per geographic unit). Each scanning window is expanded in space to include neighboring regions until a user-defined maximum radius is reached, and the number of observed cases within each window are compared to the expected cases. Before statistical inference is computed, a potential cluster is characterized when a scanning window contains more observed than expected cases. The spatial scan statistic does not consider the temporal dynamics of disease outbreaks, such as the duration of each significant cluster.

Space-time scan statistics (Kulldorff et al. 2005) incorporate a temporal dimension, where the scanning window is defined as a cylinder or three-dimensional ellipse, and the height represents the temporal dimension (e.g. time interval). The location, size, and duration of statistically significant clusters of disease cases are subsequently reported. Multivariate space-time scan statistics can identify space-time

clusters of multiple diseases that occur simultaneously (Jonsson et al., 2010; Perez et al., 2011; Greene et al., 2012; Amin et al., 2014). For example, Amin et al. (2014) used multivariate STSS to identify simultaneous clusters of three different pediatric cancers in Florida. Multivariate STSS has greater statistical power when analyzing multiple diseases, essentially ranking which disease is more prevalent and severe within a multivariate cluster. While univariate STSS have been used to examine VBD outbreaks, such as West Nile (Lian et al., 2007; Mulatti et al., 2015), malaria (Gaudart et al., 2006; Coleman et al., 2009), Lyme Disease (Li et al., 2014), chikungunya (Nsoesie et al., 2015), and dengue fever (Schmidt et al., 2011; de Melo et al., 2012; Li et al., 2012; Banu et al., 2014), the literature lacks a study that utilizes multivariate STSS to examine simultaneous outbreaks of mosquito-borne diseases.

2.3.8 Visualizing Space-Time Scan Statistics

Many studies that utilize space-time scan statistics generally visualize the results in 2D using small multiples (Naish et al. 2011; Banu et al. 2012; Uittenbogaard and Ceccato 2012; Wang et al. 2014; Mulatti et al. 2015), failing to visualize the true space-time patterns of the clusters (Bleisch 2012). Despite the complexity of displaying three-dimensional geographic space, computational improvements have facilitated 3D geovisualizations, such as graphical rendering and computer processing. Many scholars have developed 3D geovisualization techniques that incorporates a z-axis to depict the temporal dimension (Hägerstraand 1970; Kwan 2004; Miller 2005). The space-time cube framework has been widely used in studies using continuous data (Andrienko et al. 2010; Demšar and Virrantaus 2010; Nakaya and Yano 2010; Fang and Lu 2011; Sagl et al. 2013; Delmelle et al. 2014; Desjardins et al. 2018a). For example, Desjardins et al.

(2018a) conducted a space-time interpolation of pollen counts in the eastern United States and visualized the results in a space-time cube to identify the seasonality of pollen. Delmelle et al. (2014) used space-time kernel density to identify space-time clusters of dengue fever cases in Cali, Colombia, and visualized the space-clusters and the extent of spatial and temporal uncertainty in a 3D environment. The space-time cube approach has also been applied to discrete data, for example, Thakur and Hanson (2010) provided 3D visualizations of food stamp recipients, unemployment rates, and alcohol-related accidents in North Carolina.

Despite the effectiveness of 3D visualization when working with space-time data, there are some challenges and limitations worth mentioning. First, space-time layers can result in cluttered visualizations (Fang and Lu 2011). Second, static 3D visualizations can occlude data behind the main area of focus. Third, the true spatial and temporal proximity of the visualized entities may lead to biased estimates (St. John et al. 2001). If possible, the 3D visualizations should be interactive or at least multiple angles should be provided to facilitate the detection of key space-time patterns otherwise hidden by a single static image. Despite the aforementioned limitations, the main strength of 3D visualization is the scalability, as hundreds or thousands of time slices can be displayed in one graph, whereas doing so using small multiples can be cognitively too challenging when analyzing space-time data (Desjardins et al. 2018).

The literature is scarce regarding the 3D visualization of space-time clusters reported from space-time scan statistics. Nakaya and Yano (2010) and Cheng and Williams (2012) visualized the space-time clusters of crime in Kyoto, Japan and London within a 3D-environment; and Cheng and Wicks (2014) visualized space-time clusters of

tweets related to multiple events in London in a space-time cube (3D). These papers demonstrated the effectiveness of integrating 3D visualization, showing the temporal variation of clusters that 2D techniques inhibit, for example. Conversely, visualizing space-time clusters of disease in a 3D environment can improve the understanding of the space-time dynamics of an epidemic, such as variations in duration and size, and how clusters move through time, the co-occurrence of multiple VBDs, and reoccurrence of significant clusters.

2.4 Place-based determinants of VBDs

After an exploratory analysis identifies clusters of high-risk areas of VBDs, finer-level approaches can examine the factors that influence disease transmission (Rosenberg 1998) across different landscapes and areal units (e.g. census tracts and neighborhoods). In other words, disease risks and rates will vary by place and covariate data are needed to identify significant variables responsible for observable spatial patterns (Auchincloss et al. 2012). Therefore, it is critical to examine the social, economic, environmental, biological, and institutional factors that may affect VBD prevalence in a particular area. Urban regions are highly complex, and neighborhoods are the scale that public health departments most effectively operate (Whiteman 2018). Therefore, more small-area studies in spatial epidemiology are required to effectively uncover the spatial and temporal heterogeneity of disease rates across urban landscapes at these fine-levels of granularity. For example, the dynamics of temperature, precipitation, and humidity (Semenza et al. 2012); education, income, age, access to care, and quality of prevention strategies are known to strongly influence an individual's susceptibility to VBDs (Bates

et al. 2004a, 2004b). The following sections will review the place-based determinants of VBDs that are found in the literature.

2.4.1 Urbanization and Population

Since *Aedes* is a peridomestic, container-breeding mosquito, a global increase in urbanization and overpopulation has exacerbated mosquito-borne disease outbreaks (Tauil 2001). The urban heat island effect can facilitate rapid larval development by maintaining high temperatures (LaDeau et al. 2013). Both urbanization and overpopulation will increase suitable breeding habitats for developing mosquito larvae and eggs, such as open sewers, artificial containers (e.g. flowerpots), and gutters (Powell and Tabachnick 2013). Population increases, overpopulation, and human migration will increase the risk of VBD transmission due to the potential increase in pathogen (e.g. dengue virus and chikungunya virus) prevalence (Juliano and Lounibos 2005). In other words, as population increases, the number of potential human hosts also increases; and increasing population density can increase the risk of VBD transmission due to the close proximity and interaction between potential human hosts (Wilcox and Gubler 2005).

2.4.2 Weather and Climate

The impact that weather and climate have on the transmission of disease agents has been acknowledged since the beginnings of epidemiological research (Geller 2011). Understanding the dynamics of weather and climate can shed light on the cyclical nature of VBD outbreaks, the risk of transmission for susceptible human populations, and facilitate early warning surveillance systems. Climate change has especially affected VBD incidence around the world, resulting in four major observable changes (Mills et al. 2010): (1) increased vector range in nonendemic regions; (2) vector population density;

(3) pathogen prevalence in the host or vector; (4) and rate of pathogen development, reproduction, and replication.

Within the context of dengue fever and chikungunya, climate change affected the ranges of *Aedes aegypti* and *Aedes albopictus*, as well as the viruses that cause both aforementioned VBDs (Hopp and Foley 2001; Rochlin et al. 2013). *Aedes* requires warm temperatures for immature development, lifespan, and fecundity, while the optimal temperature range is between 22°C/71.6° F and 32°C/89.6° F (Marinho et al. 2016). Higher temperatures in the aforementioned range may promote faster development larvae and shorter gonotrophic cycles (Delatte et al. 2009). Eastin et al. (2014) developed models to predict dengue fever incidence rates in Cali, Colombia between 2000 and 2011, using eleven years of weather and climate data (e.g. precipitation, temperature, and humidity). The authors found that dengue outbreaks generally occurring between 18°C and 32°C, especially during warm-dry periods and extreme daily temperatures.

2.4.3 Socioeconomic Status

The risk and rates of VBDs can vary within human environments (especially urban areas) due to a variety of socioeconomic factors. Impoverished areas and neighborhoods have seen a disproportionate amount of VBDs transmitted by *Aedes* because of inadequate housing and poor water infrastructure and sanitation (Costa et al. 2017). For example, open sewers and overcrowding can facilitate the transmission of DENF, CHIK, and Zika because the abundance of suitable habitats for *Aedes* (Brasil et al. 2016). Poorer neighborhoods tend to have more open land, dilapidated housing and buildings, trash, and less resources, which may increase standing water (Becker et al. 2014). Hagenlocher et al. (2013) found that the neighborhoods in Cali, Colombia most

vulnerable to DENF were classified as having low socioeconomic status (SES), specifically high proportions of illiterate, young, and unemployed individuals, and also containing poor or absent water infrastructure and high population density. A variety of other studies found similar findings regarding the relationship between VBDs and low SES neighborhoods (Thommapalo et al. 2008; David et al. 2009; Braga et al. 2010; Delmelle et al. 2016; Krystosik et al. 2017; Farinelli et al. 2018). Despite the general consensus that low SES neighborhoods have an increased risk of *Aedes* abundance, no studies have directly compared low, medium, and high-income neighborhoods within an entire city during an epidemic.

2.4.4 Access to Healthcare

Effective healthcare utilization can substantially mitigate the burden of VBDs before, during, and after an epidemic. Healthcare facilities can provide effective treatment plans to reduce severe outcomes and provide resources and educational services to reduce the chance of future transmission (Chu et al. 2016). Infected individuals with limited financial resources may delay seeking medical care during the latter stages of a disease, which may result in worse outcomes and increase the risk of death (Ruger and Kim 2007; Cissé et al. 2007; LaBeaud 2008). For example, Khun and Manderson (2007) found that poverty, little to no cash, and skepticism towards public health care quality deterred women in Cambodia from immediately seeking care for their children with DENF symptoms. Abello et al. (2016) also found that late admission (i.e. not seeking care after onset of symptoms) for DENF was more common in public hospitals in the Philippines between 2008 and 2014. Therefore, improving the resources at healthcare facilities is necessary, besides improving access and subsidizing healthcare

for low-income individuals and families. Areas with lower access to healthcare facilities may result in an underreporting of disease cases due to individuals who are unable or refuse to seek care, which creates uncertainty and an underestimation of morbidity and mortality rates (Gibbons et al. 2014).

2.4.5 Other Place-Based Determinants of DENF, CHIK, and Zika

Individual and community risk to DENF, CHIK, and Zika are highly correlated with the abovementioned place-based determinants of disease. However, there are other significant covariates of risk found in the literature that should also be considered in a mixed-methods study in spatial epidemiology. For example, Delmelle et al. (2016) and Chiu et al. (2014) found that proximity to canals, sewers, and ditches are related to arbovirus risk. Delmelle et al. (2016) also found that proximity to tire shops and plant nurseries was also related to DENF risk in Cali, Colombia. Krystosik et al. (2018) found that neighborhoods in Cali, Colombia with higher risks of homicide experienced a higher risk of DENF, suggesting that violent crime can be a barrier to accessing preventative services and educational resources. Wu et al. (2009) found that locations with high proportions of elderly populations may be at higher risk of DENF. Mueller et al. (2016) highlights that exposure to ecosystem services (e.g. greenspaces) and disparities in infrastructure quality may affect arbovirus risk. Finally, Yen and Syme (1999) note that institutional racism may also increase the risk of disease by reducing or preventing access to preventative resources and healthcare.

2.5 Spatial Autoregressive Models

Regression-based methods can quantify the significance and strength of independent variables (predictors) on an outcome measure (e.g. disease rates), however,

spatial models assume that a disease rate occurring in one location is dependent on surrounding locations based on the first law of geography. In a spatial model, an outcome or disease rate is modeled in terms of large-scale variations and small-scale variations. Large scale variations take into account the systematic variations, and small-scale variations characterize the local variability. Specifically, small scale variations are assumed to be a random process and statistical distributions, for example, anormal distribution with aspatial autocorrelation is imposed on them. When spatial autocorrelation is present in the data, Ordinary Least Squares (OLS) may fail to capture the variability in the data and its estimates can be biased and incorrect (Matthews 2006). For example, in spatial epidemiology, the transmission dynamics of dengue fever rates in a particular location may be influenced by dengue fever rates and the explanatory variables (e.g. socioeconomic status, land use, etc.) contained in surrounding locations (spatial spillover/diffusion effects).

Simultaneous autoregressive (SAR) and conditional autoregressive (CAR) models can capture the influence of surrounding locations. Rushworth et al. (2014) highlight that disease and crime data typically exhibit spatial and spatiotemporal autocorrelations, and independence assumption is inadequate for such data. The authors also state that integrating autocorrelated random effects to the linear predictor can address the aforementioned issue (i.e. model parameters are random variables where their values depend outcomes of probability distributions and functions). Whittle (1954) first introduced the SAR model; and Besag (1974) first introduced the CAR model. Both SAR and CAR models in geography are common statistical models for spatially aggregated data on lattices, that is, regularly spaced sampled points (Carlin et al. 2014).

SAR and CAR models are similar to standard regression models but introduce an autoregressive parameter which is multiplied by a spatial weight matrix (de Smith et al. 2015). The autoregressive parameter stipulates that the resulting values for the current time period will depend on neighboring values (Anselin 2013), and the spatial weight matrix measures the degree of spatial autocorrelation between each areal unit in the study region.

SAR models are typically used for secondary order dependency and global spatial autocorrelation. Secondary order dependency describes the influence that observations have on each other (Gimond 2019). Autoregressive and SAR models have been used for VBD surveillance in a variety of study areas (Hu et al. 2010; Eastin et al. 2014; Laguna et al. 2017). For example, Eastin et al. (2014) developed an aspatial temporal autoregressive model to predict dengue fever incidence rates in Colombia using eleven years of weather and climate data (e.g. precipitation, temperature, and humidity). Eastin et al.'s model predicted dengue outbreaks two weeks to six months in advance, but failed to consider local sociodemographic and institutional variables that also increase the probability of dengue transmission. Hu et al. (2012) also utilized a SAR model to predict dengue fever cases in Queensland, Australia. Considering socioecological factors, they found that local dengue cases increased with increases in average maximum temperature, average rainfall; and overseas-acquired cases increased with increasing socioeconomic status (more likely to travel abroad).

CAR models are similar to SAR, but CAR models are "conditional" because each observation of a random process is specified conditionally on the values of the neighboring locations, which is influenced by Markov Chain properties (i.e. probability

of an event occurring depends only on previous observations; Ver Hoef et al. 2018). CAR models can be fitted to data under Bayesian paradigm (i.e. relying on prior beliefs/borrowing information to inform future estimations) using – Bayesian hierarchical models (BHM), which are widely used in techniques in geography and public health to model spatial and spatio-temporal data (Cressie and Wikle 2015). In short, BHMs model complicated space-time processes by conditionally modeling the variations in data, the process, and unknown parameters (Wang 2018). CAR models are appropriate for datasets with first order dependency and local spatial autocorrelation. First order dependency describes the variation of observations across a study area (Gimond 2019).

The temporal extension – ST-CAR can predict the value of a variable (e.g. disease rates) at a particular location and time, which will be related to current and past values of the surrounding locations; essentially testing for spatiotemporal interactions. The formulation and the mechanisms of a ST-CAR model will be provided in section 4.2.4. ST-CAR models have been used to study the effect of air pollution on human health (Lee et al. 2018); substance abuse and its relationship with child abuse (Freisthler and Weiss 2008); influenza (Lawson 2006); and Aswi et al. (2019) reviewed three papers that utilize a ST-CAR model to study dengue fever (Sani et al. 2015; Mukhsar et al. 2016a; Mukhsar et al. 2016b). For example, Freisthler and Weiss (2008) found that welfare benefits and drug arrests are positively related to Child Protective Services (CPS) referrals in California; and their ST-CAR model found that counties in California with similar rates of CPS referrals are clustered spatially and temporally.

2.5.1 Bayesian Theory

Bayesian theory is an important tool in spatial epidemiology due to its ability to combine data with prior information on the disease processes. Additionally, Bayesian models help in compartmental modeling of complex disease processes hierarchically. Since ST-CAR models are often fitted to data using Bayesian techniques, it is important to understand Bayesian theory. It is common for researchers to calculate and utilize *crude rates* when examining the distribution and risk of disease. A crude rate is simply the observed count of cases (*C*) within a location *i* divided by the population (P) at risk in the same location i, - that is C_i/P_i. However, Waller et al. (1997) mention that crude rates are unstable for regions with small at-risk populations; and particularly for rare events like cancers. They also mention that in general, crude rates can be misleading due to the potential high degree of variation across the study area. A slightly better solution to crude rates is calculating standard mortality or morbidity rate (SMR); which is essentially computing relative risk (RR) of disease by comparing the disease rate in a target location to the total disease rate in the study area (Waller and Gotway 2004).

SMR or RR better reflects the disease risk of locations in a study area by computing expected cases and comparing them to the observed cases. SMR is derived using 3 main equations:

$$\hat{\pi} = \frac{\sum_{i \in I} O_i}{\sum_{i \in I} P_i} \quad (1)$$

Equation 1 is the reference rate $\hat{\pi}$- which is simply the sum of the observed cases for all locations in the study area (O_i) divided by the sum of the total population in the study area (P_i). Once the reference rate is computed, then the expected number disease cases can be derived in Equation 2:

$$E_i = \hat{\pi} * P_i \quad (2)$$

where E_i is the expected number of cases in location *i*, which is derived by multiplying the reference rate $\hat{\pi}$ by the population in location *i* (P_i). Finally, SMR can then be derived, which is defined in Equation 3:

$$SMR_i = \frac{O_i}{E_i} \quad (3)$$

where O_i is the observed cases in location *i*, which is divided by the expected cases in location i (see Equation 2). Despite SMR being an improvement over crude rates, it still contains limitations. The main issue will come from locations with "small" populations, which is widely known as the small denominator problem (Diehr 1984). Julious et al. (2001) also point out the paradox of SMR – "the aim of SMR is to allow for different population structures, but two districts can only be compared via their SMRs if they have identical population structures!" (p. 42). For example, a location with a large population could have an expected count of 100 and an observed count of 200 – which would result in an SMR of 2; while a location with a small population could have an expected count of 3 and an observed count of 6 – also resulting in a SMR of 2. Therefore, the risk in both locations appear to be identical, however, the location with the higher population and more observed cases should be prioritized in disease surveillance. Furthermore, if the location with a smaller population experiences one more observed count (i.e. 7), then the SMR would be 2.33. As a result, the SMR is now larger than the location with a higher population and much higher observed cases. Overall, SMR can result in unreliable and "extreme" risk estimates, which can complicate disease surveillance strategies and targeted interventions.

The most popular and widely used solution to mitigating the limitations of crude rates and SMR is by using Bayesian based techniques (Lawson 2013). Bayesian estimations can stabilize risk estimates by borrowing information (i.e. data) from neighboring locations (Devine et al. 1994). Further, Bayesian disease mapping incorporates prior public health knowledge on disease rates through prior distributions. Bayes Law states that the distribution of a random variable is updated after observing data – in other words, updating a parameter based on a *prior* distribution to derive a *posterior* distribution (Gelman et al. 2014). In other words, Bayes Law can compute the probability of an event *A*, assuming that event *B* has been witnessed. Bayes Law is defined in Equation (4):

$$P(A|B) = \frac{(P(A) * P(B|A))}{(P(B))} \quad (4)$$

where P(A|B) is the probability of an event *A* occurring assuming that event B has been witnessed; P(A) is the probability of event *A* occurring; P(B|A) is the probability of event *B* occurring assuming that event *A* has been witnessed; and P(B) is the probability of event *B* occurring. Essentially, empirical Bayes estimates parameters from real data, which are the weighted average between the raw rate and global average; and the weights are proportional to at-risk population (Lawson 2013). Therefore, locations with smaller populations will have their rates adjusted substantially, and locations with high populations will barely change (Anselin et al. 2006). Calculating Empirical Bayes (EB) rate in a location *i* is defined as:

$$\pi_i^{EB} = w_i r_i + (1 - w_i)\theta \quad (5)$$

where $r_i$is the crude rate for location *i;* θ is the prior estimate – typically estimated as the reference rate (e.g. global average); and w_i is the weight for location *i,* defined as:

$$w_i = \frac{\sigma^2}{(\sigma^2 + \frac{\mu}{P_i})} \quad (6)$$

where $P_i$is the population in location i; μ is the mean of the prior (i.e. reference rate); and σ^2 is the variance, which is defined as:

$$\sigma^2 = \frac{\sum_{i=1}^{i=n} P_i(r_i - \mu)^2}{\sum_{i=1}^{i=n} P_i} - \frac{\mu}{\sum_{i=1}^{i=n} P_i / n} \quad (7)$$

The mean and variance determine the scale and shape of the parameters of the Gamma distribution, which are estimated from the data. The combination of a Gamma prior for the disease risk parameter with a Poisson distribution for the count of events yields the posterior distribution as Gamma; while the new disease risk estimate adjusts the crude rate with parameters from the prior Gamma distribution (Anselin 2006). The difference between a Poisson and Gamma distribution can be thought as follows:

(1) **Poisson**: Answers the question - "How many disease cases can we expect to see in a year"? A starting point is Poisson distribution, with the (rate of disease cases) times (the length in between reported cases) included as a parameter (Poisson distribution is defined as integers).

(2) **Gamma**: Answers the question - "How long will it take to observe 100 disease cases"? The answer is Gamma distributed, with 100 and the rate each included as parameters. The answer to this question is a real number.

Spatial Empirical Bayes (SEB) is similar to EB, but the reference rate in the equation is derived from a set of neighboring locations and their observed cases surrounding the target location. The reference rate for SEB is defined as:

$$u_i = \frac{\sum_j w_{ij} O_j}{\sum_j w_{ij} P_j} \quad (8)$$

where u_i is the reference rate for location *i*; $O_j$is the observed cases in location *j;* $P_j$is the population in location *j;* and $w_{ij}$is the binary spatial weights between location *i* and *j* – where w_{ij} is 1 if *i* and *j* are neighboring locations defined by a spatial weights matrix (e.g. queen contiguity). The variance calculation is similar to EB, but the variance is a local estimate and defined as:

$$\sigma_i^2 = \frac{\sum_j w_{ij}[P_j(r_i - \mu_i)^2]}{\sum_j w_{ij} P_j} - \frac{\mu_i}{\sum_j w_{ij} P_i/(k_i + 1)} \quad (9)$$

where $\sigma_i^2$is the variance for location *i;* $P_j$is the population of all of the neighboring locations; and $k_i$is the total number of neighbors for target location *i*. SEB can produce smoother rate estimates than EB by correcting the instability of localized variance.

2.5.2 Geographically Weighted Regression

Fotheringham et al. (1998) developed Geographically Weighted Regression (GWR) which quantifies the relationships between a dependent (e.g. disease rates) and predictor variables at different locations in a study area. GWR computes a local regression for each areal unit (e.g. neighborhood, census tract), and data from surrounding locations are used to calibrate the coefficients. The "surrounding locations" are derived from computing a spatial weight matrix. Furthermore, the distances between the target location and surrounding locations are weighted, where nearby locations are given higher importance in each local regression equation. The local coefficients and residuals can then be visualized on a map. In spatial epidemiology, GWR has been used to identify significant relationships between disease rates and a variety of the abovementioned factors that may influence an outbreak (Nakaya et al. 2005; Hu et al.

2012; Dewan et al. 2013; Delmelle et al. 2016; Wu et al. 2016; Kala et al. 2017). For example, Delmelle et al. (2016) used a GWR model which identified six significant socioeconomic and environmental predictor variables (e.g. proximity to tire shops and population density) of dengue fever rates in Cali, Colombia at the neighborhood-level; and their GWR model performed better than traditional OLS regression (R^2 = 64% vs. 29.5%). Dewan et al. (2013) also used GWR to examine the environmental factors (e.g. precipitation and temperature) that increased typhoid fever incidence in Dhaka, Bangladesh; and their GWR model had substantially better predictive power than their OLS model (R^2 = 47.5% vs. 5.2%).

Traditional GWR has recently been extended to include a temporal dimension, coined Geographical and Temporal Weighted Regression (GTWR – Huang et al. 2010; Fotheringham et al. 2015). The authors were aware that temporal dimensions can better reflect the spatial and space-time dynamics of diffusion and transmission processes of disease. GTWR is similar to the traditional GWR, but accounts for local effects in both space and time. As a result, the original GTWR computes a time-decay spatiotemporal bandwidth, where time periods within the temporal bandwidth are given less weight. Other spatiotemporal bandwidths (e.g. adaptive bandwidths) are possible and should be optimized before running the model, but an extensive discussion is beyond the scope of this paper. The results of GTWR can account for nonstationary processes in both space and time (Fotheringham et al. 2015); and modeling spatial nonstationarity with the traditional GWR may be inadequate for certain data sets, and GTWR can be more accurate if temporal variation information is added to GWR (Huang et al. 2010). GTWR has been applied to study the determinants of land use change in Baltimore, Maryland

(Wrenn and Sam 2014); PM 2.5 concentrations in Beijing, China (Guo et al. 2017); and transit ridership in Beijing, China (Ma et al. 2018).

2.5.3 Comparing ST-CAR to GTWR

Both ST-CAR and GTWR are similar in the sense that they model data generated from non-stationary processes in space and time (Finley 2011). They can both measure the degree that statistically significant predictors of disease vary spatially and temporally. Waller et al. (2007) highlight that CAR models produce model-based estimates and inference derived from varying effects via spatial random fields – e.g. borrowing strength from neighborhood spatial and temporal neighbors (i.e. Bayesian inference); while GWR models allow the covariates to vary in space (and time in the case of GTWR), but inference is ad hoc. In other words, ST-CAR models can estimate spatially and temporally varying associations between the dependent (e.g. disease rates) and predictor variables based on locally weighted regressions in both geographic AND attribute space; while GTWR can only produce local estimates in geographic space. Furthermore, Waller et al. (2007) provided an articulate comparison between GWR and CAR models; and the same comparison can be made between ST-CAR and GTWR, which add a temporal component:

> "Unlike GWR, the data do not need to be independent to define inference; the model [CAR] incorporates spatial correlation in the observations and associations through the spatial random effect (prior) distributions and these associations are automatically included in any posterior inference (e.g., point and interval estimation of parameter values, prediction of future outcomes), based on the broad substantive theory for estimation and inference underlying generalized

linear mixed models (i.e., generalized linear models with random effects)" (p. 579).

As such, GTWR models cannot define a priori spatial and spatiotemporal correlations between different coefficients, which results in posterior inference – providing stronger statistical evidence of space-time relationships between the dependent and independent variables (Chu et al. 2015). Although there is no paper that explicitly compares the results of ST-CAR and GTWR models, there are a variety of papers that strongly suggest that CAR (spatial only) models outperform GWR in terms of more accurate estimates of regression coefficients, and GWR is less robust regarding the collinearity among covariates (Waller et al. 2007). Finley (2011) suggests that GWR models should generally be utilized as tools for exploratory and descriptive data analysis due to its limited inferential capabilities.

2.6 Qualitative Approaches for VBD Surveillance

Despite GIScience and spatial epidemiology's powerful role in improving the understanding of epidemics, there is inherent uncertainty in using solely quantitative methods to inform decisions in public health. Points, lines, polygons, and statistical models fail to capture the behaviors and perspectives of those who seek healthcare and are susceptible to disease (Rosenberg 1998). Qualitative approaches should be integrated in spatial epidemiological research, which can result in a more holistic framework for disease surveillance. Quantitative geographers can greatly benefit from qualitative fieldwork approaches that are commonplace in the domains of public health, epidemiology, and human geography. Furthermore, community and stakeholder knowledge can provide detailed insight about the spatial heterogeneity of disease risk and

rates within a particular region. Chapters 5 and 6 are fieldwork-based studies that I conducted in Cali, Colombia to better understand the results presented in chapters 3 and 4.

Qualitative methods emphasize the human aspect and groups of people who make decisions that alter political, environmental and economic systems, infrastructure, etc. Three fundamental aspects of qualitative research examine human environments, social structures, and human experiences; and individual behavior and experiences may be highly influenced by their social structure position (Winchester and Rofe 2016, p. 5). Furthermore, qualitative methods have been used to shed light on what is unknown using quantitative approaches: feelings, emotions, attitudes, perceptions, and cognition; "while verifying, analyzing, interpreting, and understanding human behavior of all types" (Winchester and Rofe 2015, p. 24). General data gaps can also be filled, such as the undercounting of marginalized groups and the challenges of disease surveillance and reporting. Qualitative research may include the following approaches: (1) structured, semi-structured, or unstructured interviews (Dunn 2016); (2) oral histories (George and Stratford 2016); (3) focus groups (Cameron 2016); (4) historical/archival research (Roche 2016); (5) questionnaires/surveys (McGuirk and O'Neill 2016); (6) visual methodology (Craine and Gardner 2016); (7) discourse analysis (Waitt 2016); (8) participant observation (Kearns 2016); (9) new media (Winders 2016); and (10) participatory action research (Kindon 2016).

The ten abovementioned qualitative approaches attempt to acquire similar kinds of information – human expression as data. These data will either come from pre-existing documents, such as oral histories, historical documents, archived transcripts, or

digital sources (e.g. social media); or materials created by the researchers, such as interviews, surveys, focus groups, and participant observation (Cope and Kurtz 2016, p. 649). Any source of qualitative data will require careful reflection and it is important to essentially let the data speak for itself. Describing each of the 10 abovementioned qualitative approaches is beyond the scope of this dissertation, but it is important to note that the choice of approach will determine the type of subsequent analytical/coding techniques to extract knowledge from the data. Questionnaires/surveys can be analyzed using quantitative techniques to identify general patterns in the data, such as regression-based techniques (Polit and Beck 2010). Other less structured and more open-ended approaches, such as interviews, focus groups, and participant observation require tedious transcription, textual analysis, and careful reflection about the context (e.g. culture, time, social structure) when the data was collected.

2.6.1 Knowledge, Attitude, and Practice (KAP) Surveys

Objective 2 examines how geographic, socio-economic, demographic, environmental, and institutional factors influence VBD transmission and incidence, but three other main factors are missing from the analyses – the knowledge, attitudes, and practices (KAP) of Colombian citizens exposed to VBDs. The KAP survey approach was first used in 1950s in the domains of family planning and population studies and are now widely used to study public health (Launiala 2009). KAP studies are widely regarded as easily conductible, measureable, and interpretable (Raina 2013). Within the context of VBD research, KAP approaches are used to shed light on at-risk communities' understanding of the vector, the pathogen, and prevention and treatment strategies (Potter et al. 2016; Udayanga et al. 2018). Social, cultural, political, and economic factors affect

human behavior, which in turn can influence the transmission of VBDs (Heintze et al. 2007). The results of KAP studies can be used to improve vector control and management strategies by understanding how human knowledge, behavior, and decisions may influence VBD risk (Alobuia et al. 2015). For example, KAP surveys that were administered in endemic regions of malaria have found that both income and education levels can predict malaria risk (Launiala 2009; Bashar et al. 2012; Dawaki et al. 2016).

Besides malaria, KAP studies have also examined other VBDs in a variety of locations. Samuel et al. (2018) administered a KAP survey in New York City, which has seen over 900 travel-associated cases of Zika. They found that the majority of the 224 respondents had a poor understanding of Zika transmission, complications, and prevention practices. Udayanga et al. (2018) found that dengue free communities in Sri Lanka were characterized by individuals with fair knowledge of dengue, willingness to participate in community-based vector prevention, and water disposal practices. Corrin et al. (2017) provided a systematic review of KAP studies that examined chikungunya; finding that KAP of chikungunya among the public and health professionals varies across populations. Furthermore, they found that although knowledge of chikungunya is higher in affected areas, the vast majority of the populations do not understand the disease enough to protect themselves and their communities.

Although the KAP approach is commonly utilized in vector epidemiology, it is not well understood how KAP will vary across diseases (endemic vs. emerging), and across different neighborhoods within a municipality. Understanding KAP variation among concurrent VBDs, especially the differences between endemic and novel diseases, can improve targeted interventions, education programs, and health policy. Comparing

KAP results to VBD cases per neighborhood from a spatial perspective can be a novel approach for VBD surveillance and prevention. Colombia is a prime candidate for understanding KAP variation due to the presence of three VBDs transmitted by *Ae. aegypti* and *Ae. albopictus*: dengue (endemic), chikungunya (novel), and Zika (novel). Whiteman et al. (2018) is the only study to-date that compares DENF, CHIK, and Zika KAP across two low- and two high SES neighborhoods in Panama. They found that low income, low education, and elderly residents should be prioritized when improving education and vector surveillance programs; and participants had better knowledge of DENF than CHIK and Zika. However, Whiteman et al. (2018) did not sample residents from medium SES neighborhoods; and their sample only included residents from four neighborhoods, which may not encompass KAP variation across an entire administrative area (e.g. Panama City). They also did not compare the KAP results to the reported disease cases within their sample neighborhoods. Furthermore, KAP results may not be necessarily generalizable to other regions due to sociocultural, socioeconomic, and institutional factors.

2.6.2 Semi-structured interviews

The final approach in the holistic spatial epidemiology framework is understanding the perspectives and initiatives of policy-makers (i.e. public health officials and other relevant stakeholders). Interviewing public health officials can understand what is being done to address community needs for disease prevention, treatment, surveillance (Baum et al. 2011). Semi-structured interviews is one popular qualitative data collection method that is in-between structured and unstructured. Semi-structured interviews relies on a guide that facilitates discussion and state the questions

that must be answered (Harrell and Bradley 2009). Since it is "semi-structured", the interview can feel like a conversation and the respondents can thoroughly elaborate on the subject matter included in the guide. The interviewer may also ask relevant questions that are not on the guide, but are a result of the conversation that may improve the overall understanding of the research.

Semi-structured interviews are the most commonly used interviews in qualitative research and public health studies (Stuckley 2013). Semi-structured interviews have been used in VBD research to understand the policy-maker/public health official perspective (Deroeck et al. 2003; Hanh et al. 2009; Nagpal et al. 2012; Luo et al. 2019; Miranda et al. 2019; Odongo et al. 2019). For example, Deroeck et al. (2003) interviewed professionals in Cambodia, Indonesia, the Philippines, and Vietnam regarding their experiences with DENF in their respective countries. The authors found that the respondents supported funding for vaccine development, increased access for low-income populations in endemic regions, and combatting DENF as a top public health priority. Miranda et al. (2019) found that pharmacists in Campo Grande, Brazil lacked knowledge and training to effectively mitigate Zika. A notable discovery is that only one pharmacist suggested birth control as a preventative strategy against Zika transmission. Odongo et al. (2019) suggested the need to improve and increased open-source VBD surveillance tools. Nagpal et al. (2012) found that stakeholders in India should allocate more time and resources to CHIK outbreak management and prevention.

In Colombia, Suarez et al. (2005) found that government officials in Villavicencio (city southeast of Bogota) felt that the community does not help the public health officials combat DENF by reducing *Aedes* habitats because they are reactive to the outbreaks

rather than proactive – that is, the community tends to wait until an outbreak is affecting the population until preventative measures are practiced. Douglas et al. (2013) determined that senior public health officials consider DENF as a top priority (e.g. state of emergency in Cali in 2010) and that Colombia would consider a vaccine if it was affordable (although no vaccine is available to-date). García-Betancourt et al. (2015) investigated the social practices of storing water in Girardot (town in western Colombia) by interviewing both community members and local VBD program technicians. The technicians stated that cultural tradition and service interruptions are common factors of storing water in containers that may become breeding sites for *Aedes*, resulting in DENF risk for the household and community. Finally, Camacho et al. (2004) noted that health workers in Ibagué (city in western Colombia) suggested that administration problems (e.g. funding, leadership, health reform) were major factors to dengue control programs in the 1990s.

Currently, there lacks a study that reflect public health officials' views on vector surveillance and control; and educational campaigns regarding DENF, CHIK, and Zika. Since Cali has been a hotspot of the three aforementioned VBDs, understanding the perspectives and public health campaigns from high-ranking officials can provide key insight about the trials, tribulations, and effectiveness of current and planned policy regarding co-circulating VBDs. The information collected from the interviews can be compared with community knowledge (e.g. KAP surveys) to understand "both sides of the story". Therefore, this combination of community and policy-maker information can be utilized to address key issues to ultimately improve VBD surveillance and mitigate current and future outbreaks.

CHAPTER 3: SPACE-TIME CLUSTERS AND CO-OCCURRENCE OF CHIKUNGUNYA AND DENGUE FEVER IN COLOMBIA FROM 2015 TO 2016

Vector-borne diseases (VBDs) infect over one billion people and are responsible for over one million deaths each year, globally. Chikungunya (CHIK) and dengue Fever (DENF) are emerging VBDs due to overpopulation, increases in urbanization, climate change, and other factors. Colombia has recently experienced severe outbreaks of CHIK AND DENF. Both viruses are transmitted by the *Aedes* mosquitoes and are preventable with a variety of surveillance and vector control measures (e.g. insecticides, reduction of open containers, etc.). Spatiotemporal statistics can facilitate the surveillance of VBD outbreaks by informing public health officials where to allocate resources to mitigate future outbreaks. To fulfill objective 1 of this dissertation, I utilize the univariate Kulldorff space-time scan statistic (STSS) to identify and compare statistically significant space-time clusters of CHIK and DENF in Colombia during the outbreaks of 2015 and 2016. I also utilize the multivariate STSS to examine co-occurrences (simultaneous excess incidences) of DENF and CHIK, which is critical to identify regions that may have experienced the greatest burden of VBDs. The relative risk of CHIK and DENF for each Colombian municipality belonging to a univariate and multivariate cluster is reported to facilitate targeted interventions. Finally, I visualize the results in a three-dimensional environment to examine the size and duration of the clusters. My approach is the first of its kind to examine multiple VBDs in Colombia simultaneously, and the 3D visualizations are a novel way of illustrating the dynamics of space-time clusters of disease.

3.1 Data & Study Area

Colombia (Repùblica de Colombia) is a country in Northwestern South America; bordering Panama to the northwest, Venezuela to the northeast, Brazil to the southeast, Ecuador to the southwest, and Peru to the south (see reference map in Appendix 19). Colombia has the second-highest population in South America and third highest in Latin America with an estimated 49,954,000 people (DANE 2018). As of 2018, 80.8% of Colombia's population lives in urban areas, which has increased from 31% in 1938 (United Nations 2018). Bogotá is the capital and most populous city with ~8 million people, and four other cities have a population over 1,000,000: Medellín (~2.5 million), Cali (~2.4 million), Barranquilla (1.2 million), and Cartagena (~1 million). Colombia has not released an official census since 2005, therefore, the population estimates above are subject to uncertainty. Furthermore, Colombia is comprised of 32 departments, which are subdivided into 1,122 municipalities. Figure 2 shows the population density is Colombia for each municipality for 2015 and 2016; and also showing the change in population density between the two years of data. The figure shows that most of the population is concentrated in the northern, west-central, and south-western portions of the country; and many of the cities are very densely populated. The figure also shows that many of the major cities experienced an influx of people between 2015 and 2016.

The rapid urbanization, population growth, and population density have increased Colombia's incidence and risk of arboviruses transmitted by the peridomestic Aedes aegypti and Aedes albopictus mosquitoes (specifically dengue, chikungunya, and Zika). Between the 1950s and 1960s, fumigation and other mosquito eradication programs were carried out to eliminate *Aedes* due to the prevalence of DENF in Colombia (Dick et al.

2012). Despite some eradication success, DENF re-appeared in the 1970s and became endemic in many municipalities in Colombia since then (Messina et al. 2015). Delmelle et al. (2016) note that DENF exhibits an endemo-epidemic pattern that repeats every 2-3 years; and Eastin et al. (2014) highlight that the El Niño oscillation influences the temporal patterns. The spatiotemporal patterns of CHIK and Zika in Colombia are less known, due the diseases first appearing in the Americas in 2013.

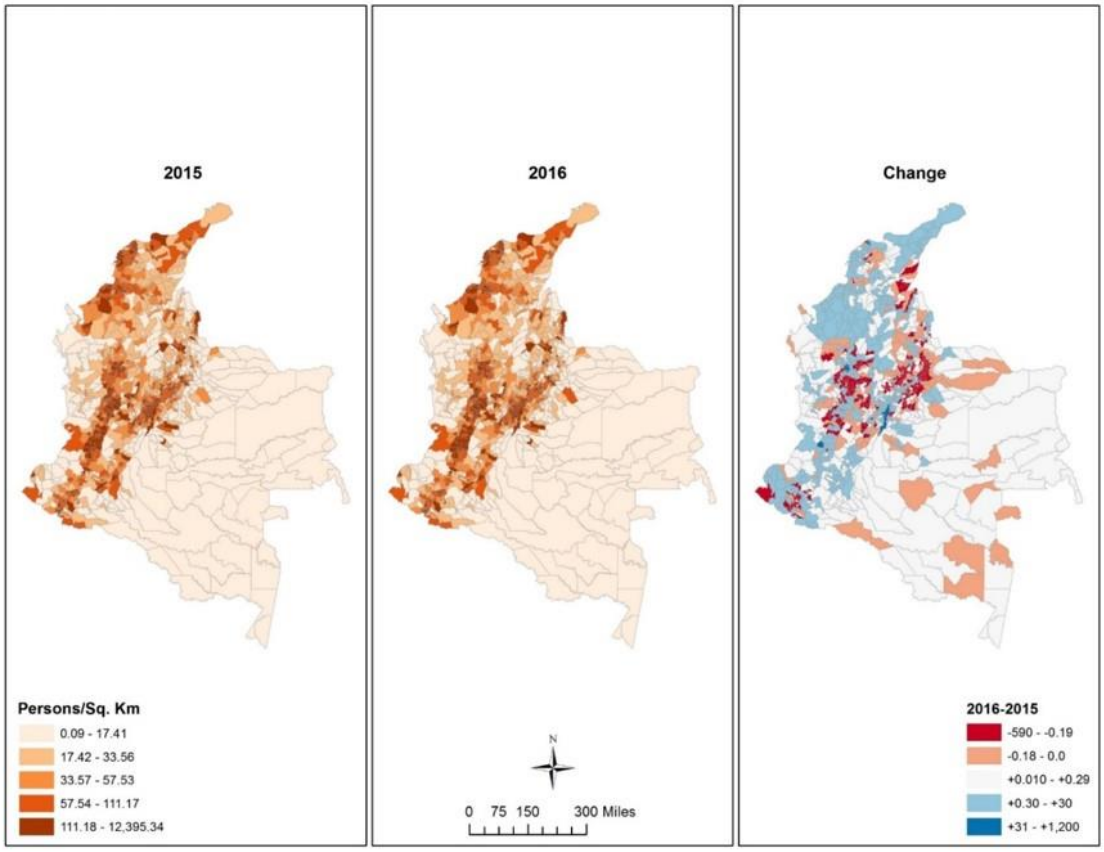


Figure 2: Population density of Colombian municipalities for 2015 and 2016; and the change between the two years (source: SIVIGILA).

Data for the case study corresponds to the number of DENF and CHIK cases per municipality in Colombia for 2015 and 2016. The data was obtained from SIVIGILA (Sistema Nacional de Vigilancia en Salud Pública – National Public Health Surveillance

System). SIVIGILA is a system administered by the National Institute of Health of Colombia (Instituto Nacional de Salud – INS) which has as a primary goal to provide information regarding events that can affect the health of the Colombian population in a timely manner. Data is uploaded into the system by the UPGDs (from their acronym in Spanish: Unidades Primarias Generadoras de Datos – Data Generating Primary Units) on a weekly basis. UPGDs are defined as any private or public entity that diagnoses the occurrence of a public health event of interest (INS 2018a).

The INS makes SIVIGILA data available at the aggregate level through their Routine Surveillance webpage (INS 2018b). Figure 3 shows the CHIK and DENF rates per 1,000 for each municipality in Colombia from 2015 to 2016.

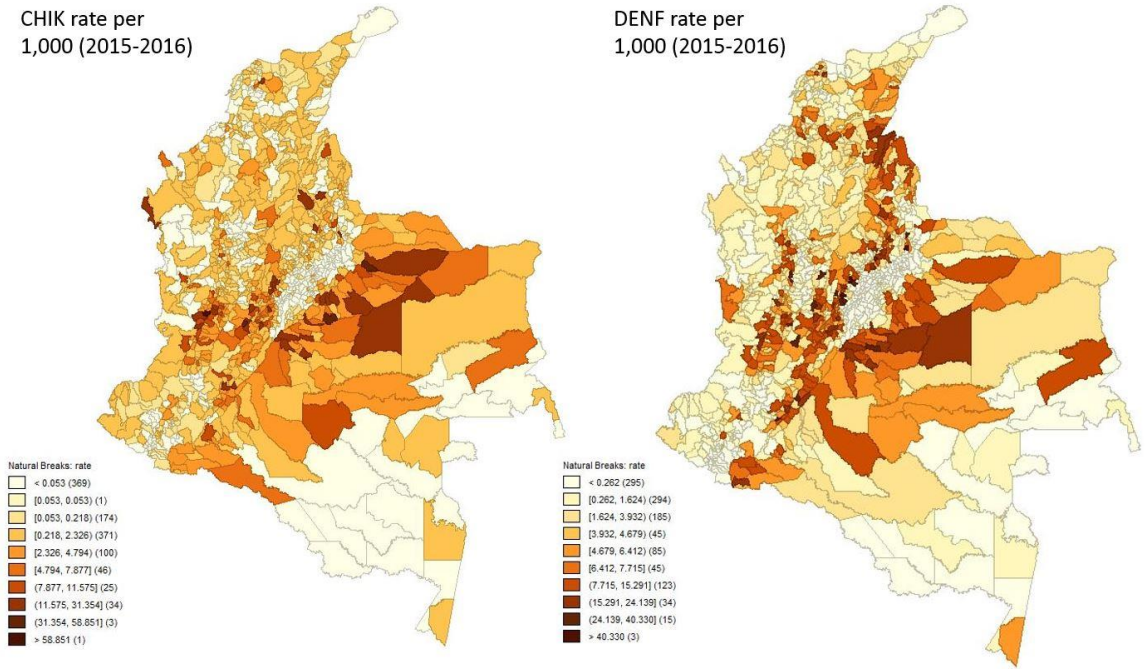


Figure 3: CHIK and DENF rates per 1,000 for each municipality in Colombia, 2015-2016

The aggregate summaries contain weekly disease cases for each municipality including suspicious, probable, and confirmed cases. A probable DENF case is identified

as exhibiting fever with two or more of the following symptoms: headache, retroocular pain, myalgia, arthralgia, and rash (INS, 2018c). A suspicious CHIK case is identified as a patient residing or visiting a healthcare facility 8–15 days prior to the onset of symptoms in a municipality where there have not been laboratory cases of CHIK confirmed; including the following symptoms: running a fever over 38°C, arthralgia or arthritis, uniform erythema, or symptoms that cannot be explained by other medical conditions (INS, 2018d). The data stored in the SIVIGILA system is described as dynamic, subject to analysis, and adjustment (this means data is revised as more information becomes available). Population data was obtained from the Geographic Information System for Planning and Land Use Ordering of Colombia (SIGOT: Sistema de Información Geográfica para la Planeación y Ordenamiento Territorial). The data contains population totals for each municipality in 2015 and 2016. The dataset includes 43,452 CHIK cases in 2015 and 11,964 CHIK cases in 2016. For DENF, there are 94,856 cases in 2015 and 99,703 cases in 2016.

3.2 Methodology

For the purpose of objective 1, I utilize a univariate and multivariate space-time scan statistic (Kulldorff et al. 2005; Kulldorff et al. 2007), which are both implemented in SaTScan™. The univariate space-time scan statistic (STSS) employs moving cylinders that scan the study area for potential space-time clusters of cases, and the base of the cylinder is the spatial dimension and the height reflects the temporal dimension. The center of the cylinder is defined as the centroid of each Colombian municipality and each cylinder is expanded until a maximum spatial and temporal size is reached (each cylinder is a potential cluster). A minimum reported temporal length of the clusters (e.g. 2

weeks), and maximum spatial and temporal size of the cylindrical scanning windows (e.g. 25% of the population at-risk and 25% of the study period) is required. STSS use a variety of probably models, depending on the characteristics of the dataset. Table 1 provides the appropriate model that should be employed for different types of data.

Table 1: Probability Models for STSS (Kulldorff 2018)

Data	Example	Model	Reference
Count	(1) case only (2) cases/controls (3) cases/total population	(1) Space-Time Permutation (2) Bernoulli (3) Discrete Poisson	(1) Kulldorff et al. (2005) (2) and (3) Kulldorff (1997)
Categorical	Disease with multiple types (dengue)	Multinomial	Jung et al. (2010)
Ordered Categorical	Cancer stages	Ordinal	Jung et al. (2007)
Survival Time	AIDS cases over a ten-year period	Exponential	Huang et al. (2007)
Other - Continuous	Body Weight	Normal	Kulldorff et al. (2009)

The data used for objective 1 requires a discrete Poisson probability model, which used cylindrical scanning windows. The VBD cases for each disease are assumed to be Poisson distributed according to the at-risk population in Colombia. Our null hypothesis stipulates that the model reflects an inhomogeneous Poisson process with an intensity μ, which is proportional to the at-risk population. The alternative hypothesis is that the number of VBD cases exceeds the number of expected cases derived from the null

model. A maximum likelihood ratio test is utilized to evaluate the null and alternative hypotheses, which is defined in Equation 10:

$$\frac{L(Z)}{L_0} = \frac{\left(\frac{n_Z}{\mu(Z)}\right)^{n_Z}\left(\frac{N-n_Z}{N-\mu_{(Z)}}\right)^{N-n_Z}}{\left(\frac{N}{\mu(A)}\right)^{N}} \quad (10)$$

Where $L(Z)$ is the likelihood function for cylinder Z, and L_0 is the likelihood function for the null hypotheses for cylinder Z. Essentially, the number of observed VBD cases in a cylinder n_Z is divided by the number of expected cases in a cylinder $\mu(Z)$ to the power of the observed n_Z, multiplied by the observed cases divided by the expected cases outside of the cylinder. The numerator is then divided by the quotient of dividing the total number of observed cases for the entire study area N across all time periods $\mu(A)$, to the power of the total number of observed cases. The cylinder will have an elevated risk if the likelihood ratio is greater than 1, that is $\frac{n_Z}{\mu(Z)} > \frac{N-n_Z}{N-\mu_{(Z)}}$. Furthermore, the space-time scan statistic uses different cylinder sizes, and the cylinder with the highest likelihood ratio (maximum) is the most likely cluster.

Since many VBDs can be transmitted by the same vector (e.g. dengue and chikungunya – *A. aegypt*i & *A. albopictus*), identifying space-time clusters where multiple VBDs co-occur can be beneficial for targeted intervention programs. To identify the co-occurrence (simultaneous excess incidence) of multiple VBD outbreaks, I utilized the multivariate space-time Poisson scan-statistic (Kulldorff et al. 2007). As with the univariate statistic, the log likelihood ratio (LLR) for each disease is computed within each cylinder. The multivariate statistic sums the LLRs for each VBD in a particular cylinder, producing a new LLR for that cylinder. Finally, the maximum summed LLR is

reported as the most likely cluster, which is defined in Equation 11:

$$T = \max_{z} \sum_{i} LLR_i(Z) \tag{11}$$

Where *T* is the most likely cluster and *i* is the dataset for a particular disease within cylinder *Z*. Secondary clusters are also reported if they are statistically significant. Monte Carlo testing is utilized to assess the statistical significance of the potential univariate and multivariate space-time clusters. Monte Carlo testing computes a *p*-value for each candidate cluster, essentially comparing simulated data sets to the real data set (recommended minimum of 999 simulations).

The majority of the literature pertaining to STSS only report the locations that belong to a significant space-time cluster. However, this approach assumes that the risk of infection is homogenous throughout the cluster. Conversely, some locations within a cluster may contain zero cases of a particular disease, due to the scanning nature of the STSS. To reduce uncertainty by identifying the municipalities that are the highest risk locations in a cluster (rather than assuming the risk of disease is homogenous throughout a cluster), I report the relative risk for each Colombian municipality belonging to a univariate and multivariate space-time cluster, which facilitates targeted interventions. Relative risk quantifies the risk of becoming infected with a disease in one location compared to all other locations, which is defined in Equation 12:

$$RR = \frac{c/e}{(C-c)/(C-e)} \tag{12}$$

Where *c* is the total number of observed cases in a municipality, e is the total number of expected cases in a municipality (proportional to at-risk population), and C is the total number of observed cases in Colombia.

The statistically significant space-time clusters detected by the univariate and multivariate STSS are visualized in both 2D and 3D. Space-time clusters of disease reported by the STSS are typically visualized solely in 2D, essentially only depicting the location and size (radius) of the clusters. Space-time (3D) visualizations are required to examine and conceptualize the space-time dynamics of disease clusters. For this study, the resulting univariate and multivariate space-time clusters of DENF and CHIK are visualized in a space-time cube using ArcScene™. The space-time clusters will have a cylindrical shape due to the scanning window utilized for the STSS. Black rings were added to the clusters to improve the conceptualization of the temporal dimension, where each ring represent a particular week during 2015 or 2016. The resulting 3D visualizations are especially effective at showing the space-time dynamics of the clusters, including the size, duration, and movement over time.

Objective 1 of this dissertation has already been published and I will share the results and discussion from Desjardins et al. (2018b) and Desjardins et al. (2019). However, the main results are extracted from Desjardins et al. (2019), which used an updated and more accurate dataset of DENF and CHIK cases in Colombia during 2015 and 2016. The discrepancy between the total cases reported in both of the aforementioned papers can be explained by the dynamic reporting of cases by the national health surveillance system (SIVIGILA). For example, CHIK, DENF, and Zika have similar symptomology, which can make initial diagnosis difficult. Furthermore, there can be uncertainty and delays in reporting to SIVIGILA from the UPGDs, especially since CHIK was a novel disease in Colombia at the time of the epidemic. The data for Desjardins et al. (2019) was retrieved in December of 2017, a year after the

epidemic was over; while the data in Desjardins et al. (2018b) was retrieved shortly after the year was over for 2015 and 2016.

3.3 Results

3.3.1 Univariate: CHIK

During the CHIK outbreaks in Colombia between 2015 and 2016, a total of n=43,452 and n=11,964 cases were reported, respectively (n=55,416 combined). The temporal distribution of CHIK cases is shown in Figure 4. May 2015 recorded the highest number of cases (n=7,934), while the first half of the year contained the majority of cases, with a sharp decline after July 2015. The CHIK epidemic in 2016 was not as severe as 2016 and although there was an increase in cases at the start of the year in January, April experienced the maximum in 2016 with n=2,180 cases. The number of cases remained elevated until July 2016 when it started to decline each month until the end of the year.

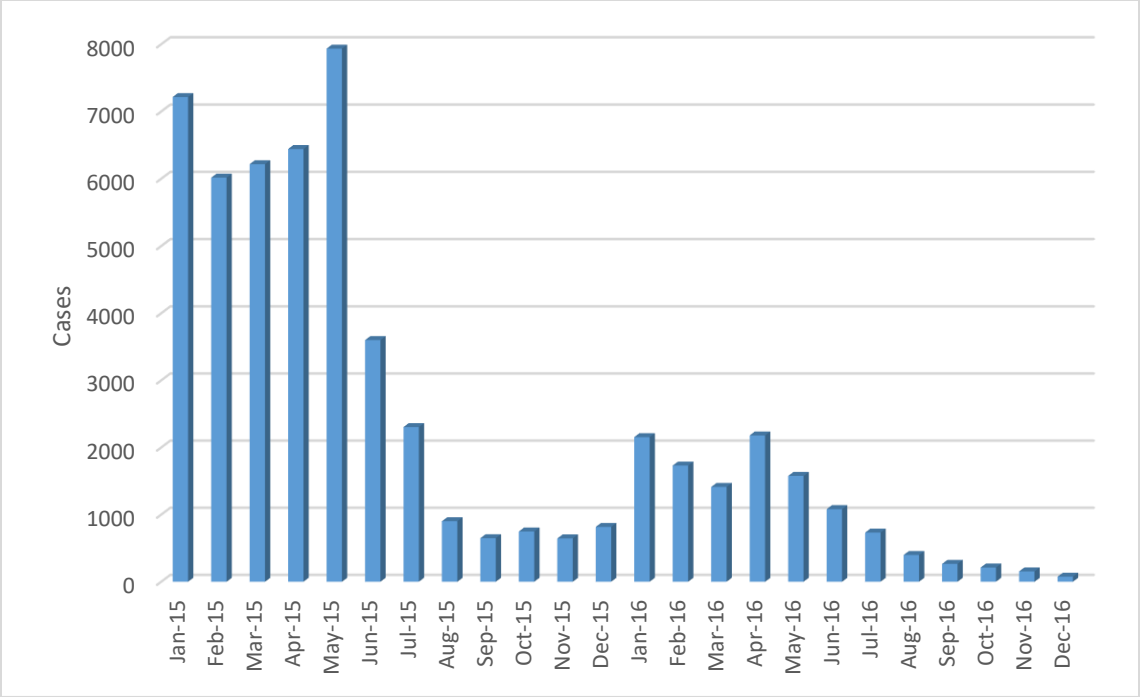


Figure 4: Temporal distribution of CHIK cases in Colombia for 2015 and 2016

Five clusters affecting 490 of the 1,125 municipalities were reported for CHIK (Figure 5A) and the results are summarized in Table 2. All five clusters were reported in 2015 with centers in (1) Ataco – Tolima Department: weeks 1–26, (2) Mapiripana – Guainía Department: weeks 15-24, (3) La Peña – Cundinamarca Department: weeks 4-29 (4) San Jacinto – Bolivar Department: weeks 1-5, and (5) Puerto Nare – Antioquia Department: weeks 7-10. Clusters 1 (most likely cluster) and 3 had the longest duration of twenty-five weeks each, andcluster 5 had the highest relative risk of 65.78. It is more informative to report the relative risk of the locations that belong to a cluster, and Figure 6A shows the relative risk of CHIK for the 490 selected municipalities. Notably, 194 of the 490 municipalities reported a relative risk greater than 1, which indicates that there were more CHIK cases than expected. Conversely, 296 municipalities had more expected than observed cases (RR between 0 and 1); and 73 municipalities had no

observed cases of CHIK (RR = 0). Cali – Valle del Cauca Department (cluster 1) had the most observed cases (4,178) during the study period with a relative risk of 1.59. Roldanillo – Valle del Cauca Department (cluster 1) had the highest relative risk of 87.6, with 3,078 observed and 37.17 expected cases.

Table 2: Space-Time Clusters of CHIK (RR: relative risk)

Cluster	Duration (weeks)	*p*	Observed	Expected	*RR*	Municipalities	Cluster pop
1	1-26	<0.01	20,407	2,172.03	14.29	144	7,671,766.9
2	15-24	<0.01	5,520	341.45	17.84	74	1,575,080.1
3	4-29	<0.01	984	23.99	41.75	10	84,561.4
4	1-5	<0.01	2,292	660.00	3.58	255	12,151,474.5
5	7-10	<0.01	328	5.02	65.78	5	115,174.4

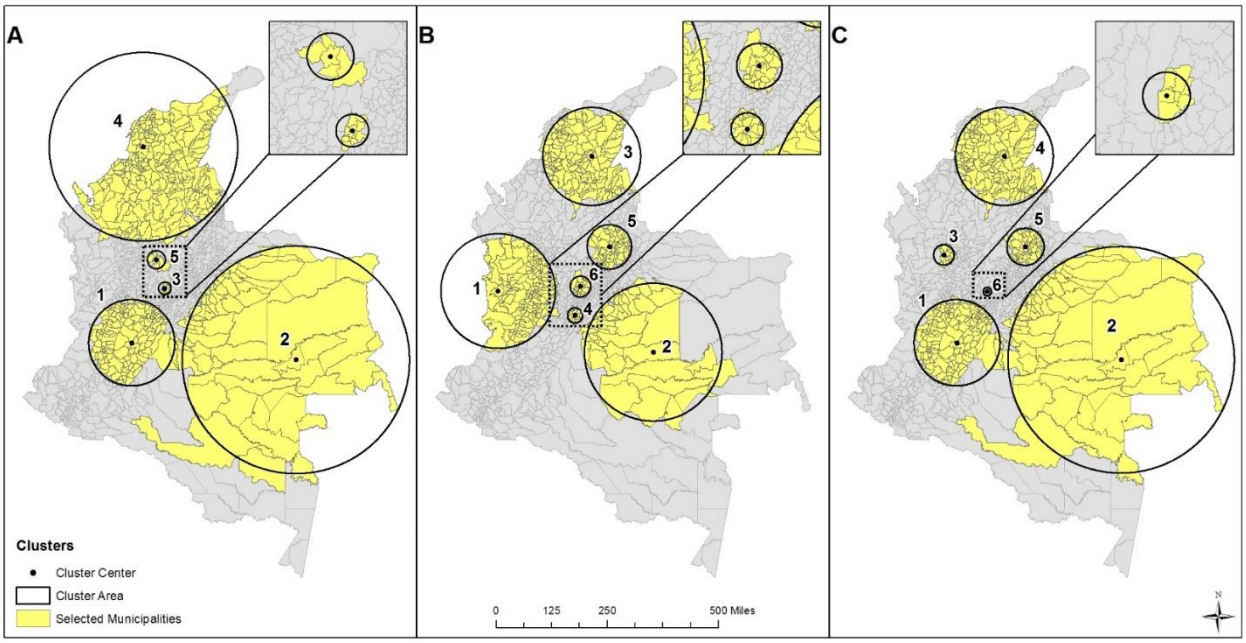


Figure 5. Significant space-time clusters of A.) CHIK, B.) DENF, and C.) Multivariate

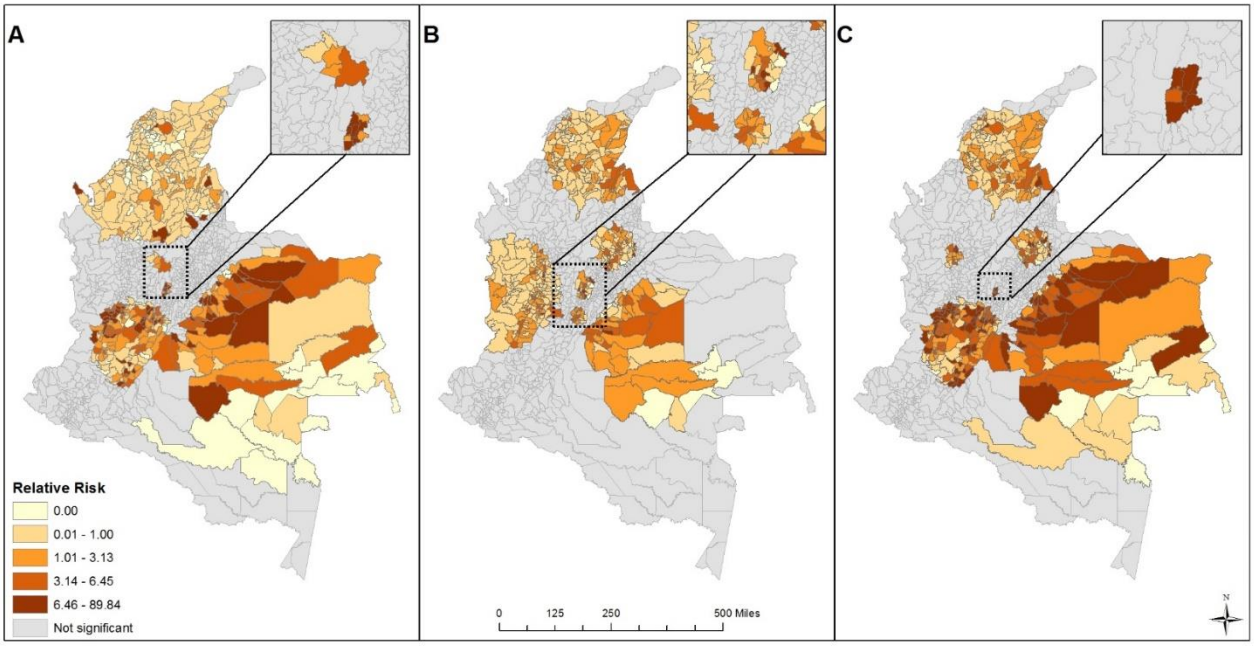


Figure 6. Relative risk per municipality for A.) CHIK, B.) DENF, and C.) Multivariate

3.3.2 Univariate: DENF

During the DENF outbreaks in Colombia between 2015 and 2016, a total of n=94,856 and n=99,703 cases were reported, respectively (n=194,559 combined). The temporal distribution of DENF cases is shown in Figure 7. In 2015, January experienced the highest number of cases (n=10,107), and the remainder of the year saw over n=5,000 cases each month. At the start of 2016, there was another sharp increase in DENF cases, with the peak of the epidemic occurring in April (n=12,907 cases). The number of cases remained high in the rainy season months, followed by a sharp decline starting in August 2016.

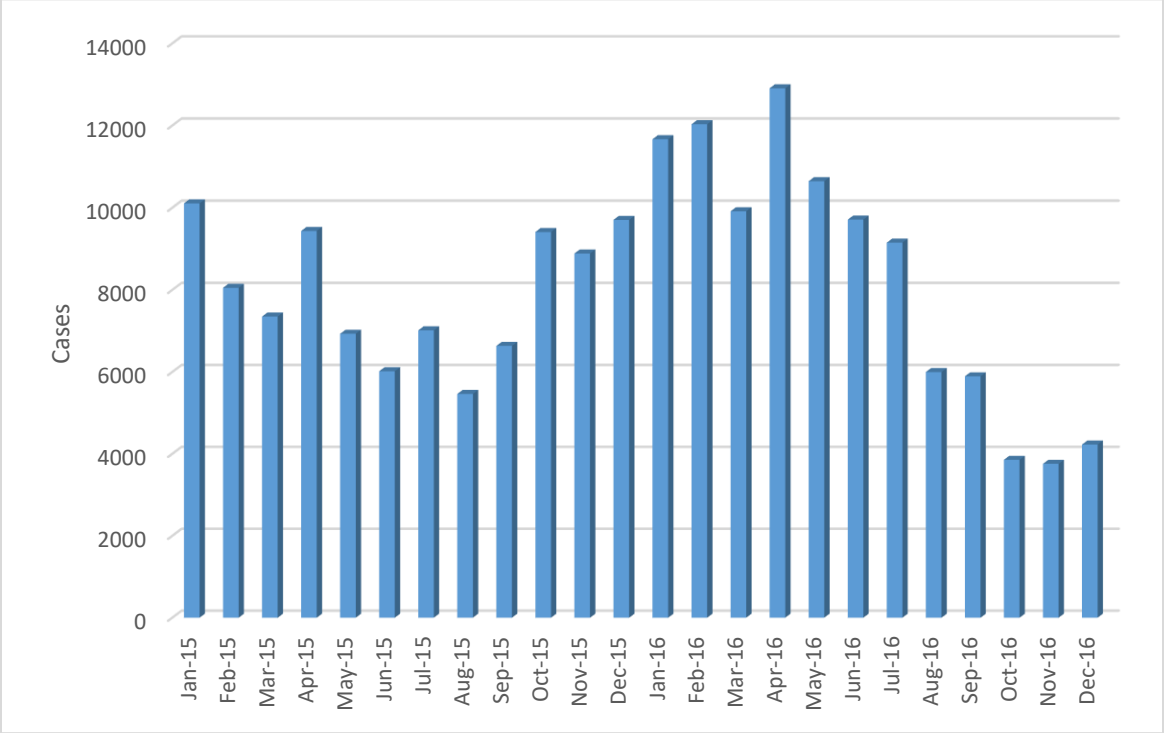


Figure 7: Temporal distribution of DENF cases in Colombia for 2015 and 2016

Six space-time clusters of DENF were reported, which included 474 of 1,125 Colombian municipalities (Figure 5B). Two clusters were reported in 2015 with centers

in (2) Mapiripán – Meta Department: weeks 1–24, and (3) Astrea – Cesar Department: weeks 32-53. The last week of cluster 3 occurred during the first week of January 2016. In 2016, four clusters were reported with centers in (1) Medio Baudó – Chocó Department: weeks 54–79, (4) Tibacuy – Cundinamarca Department: weeks 52-57, (5) Hato – Santander Department: weeks 52-77, and (6) El Peñón - Antioquia Department: weeks 52-74. Table 3 provides detailed characteristics of the six space-time DENF clusters.

Table 3: Space-Time Clusters of DENF (RR: relative risk)

Cluster	Duration (weeks)	*p*	Observe	Expect	RR	Municipalities	Cluster Pop
1	54-79	<0.01	39,363	11,638.9	3.99	165	11,610,411
2	1-24	<0.01	5,903	1,239.48	4.88	52	1,357,165
3	35-53	<0.01	8,971	3,169.72	2.92	140	4,362,901
4	52-77	<0.01	2,452	356.42	6.95	16	355,191
5	52-77	<0.01	5,527	1,894.43	2.97	74	1,893,281
6	52-74	<0.01	1,118	208.43	5.39	25	235,570.56

Clusters 1 (most likely cluster), 4, and 5 had the longest duration of twenty-five weeks each, and cluster 4 had the highest relative risk of 6.95. Figure 6B depicts the relative risk for each of the 474 municipalities belonging to a space-time DENF cluster. Out of the 474 municipalities, 211 contained a relative risk greater than 1, while 263 had more expected that observed cases (RR between 0 and 1), and 18 had no observed DENF cases (RR =0). Cali – Valle del Cauca Department (cluster 1) contained the most observed cases of DENF during the study period (n = 33,748), with a relative risk of 4.08. Soatá – Boyacá Department had the highest relative risk (RR = 17.53) with 500 observed and 28.59 expected cases. Notably, Medellín – Antioquia Department belongs to cluster 1 with 20,990 observed cases (RR = 2.25).

3.3.3 Multivariate Clusters

Fig. 5C shows multivariate space-time clusters of CHIK and DENF in Colombia during the 2015 and 2016 epidemics. Six clusters affecting 440 municipalities were reported and Table 4 summarizes the results. Four clusters were reported in 2015 with centers in (1) Ataco – Tolima Department: weeks 1-26, (2) Mapiripana - Guainía Department, (4) Astrea – Cesar Department: weeks 35-53, and (6) Quebradanegra – Cundinamarca Department: weeks 2-27. In 2016, two clusters were reported with centers in (3) Anzá – Antioquia Department: weeks 65-90, and (5) Cabrera – Cundinamarca Department: weeks 54-77. Furthermore, the last week of cluster 4 occurred in this first week of January 2016.

Table 4: Multivariate Space-Time Clusters of CHIK and DENF (RR: relative risk)

Cluster	Duration (weeks)	Municipalities	*p*	VBD	Observe	Expect	RR
1	1-26	144	<0.01	CHIK	20,407	2,172.03	14.3
				DENF	21,808	7,625.74	3
2	4-24	71	<0.01	CHIK	5,569	355.84	17.3
				DENF	5,320	1,249.30	4.3
3	65-90	18	<0.01	CHIK	0	0	0
				DENF	15,739	3,479.14	4.8
4	35-53	140	<0.01	CHIK	0	0	0
				DENF	8,971	3,169.73	2.9
5	54-77	59	<0.01	CHIK	1,517	404.54	3.8
				DENF	4,608	1,420.3	3.3
6	2-27	6	<0.01	CHIK	877	16.10	55.3
				DENF	406	56.52	7.2

The multivariate STSS will report significant clustering for one or more datasets, therefore, the multivariate results may include clusters that only contain CHIK or DENF. For this study, four of the clusters included simultaneous clustering of both CHIK and DENF (clusters 1, 2, 5, and 6), and two of the clusters only included significant clustering

of DENF (clusters 3 and 4). Clusters 1, 3, and 6 had the longest duration of twenty-five weeks each. Cluster 1 had the highest observed cases for both CHIK (n = 20,407) and DENF (n = 21,808); and cluster 6 had the highest relative risk of CHIK (RR = 55.33) and DENF (RR = 7.19). Cali – Valle del Cauca Department (cluster 1) contained the most combined observed cases with n = 37,926 (CHIK = 4,178; DENF = 33,748; combined RR = 5.67). Figure 6C shows the relative risk for the 440 municipalities belonging to the multivariate clusters. Furthermore, 13 municipalities had no observed cases of CHIK nor DENF (RR = 0); 145 municipalities had less observed than expected cases (RR between 0 and 1); and 295 had more observed than expected cases (RR > 1). Roldanillo – Valle del Cauca Department (cluster 1) had the highest combined relative risk (RR = 89.84; CHIK = 87.6 & DENF = 2.23). Notably, Medellín – Antioquia Department (cluster 1) had a combined relative risk of 2.49, with 675 observed cases of CHIK and 20,990 observed cases of DENF.

3.3.4 3D Visualizations

Figures 8-10 visualize the space-time clusters of CHIK, DENF, and co-occurrence of CHIK and DENF in a 3D-enviornment, respectively. The design of the 3D visualizations include the following elements: (1) cylinders representing the size, location, and the duration of the cluster; (2) black rings around each cluster represent a particular week during the study period; (3) a 2D layer of the municipalities belonging to a cluster, which is superimposed on Colombia; (4) a 2D layer of the radii of the clusters superimposed on Colombia; (5) labels that denote a cluster's ID; and (6) two temporal axis with labels to denote the start and end dates of each cluster. The 3D visualizations improve the conceptualization of the space-time dynamics of the reported clusters.

For example, Figure 8 shows that the five space-time clusters of CHIK began and ended during the first half of 2015. Clusters 1-3 lasted the longest, while affecting the south-central portions of Colombia. Clusters 4 and 5 occurred in the north-central portions of the country, and they had very short durations between January and March of 2015, respectively. Figure 9 shows that two DENF clusters occurred in 2015 (2 and 3), and four (1, 4-6) occurred in 2016. Cluster 2 in the central region of Colombia began in January 2015 and lasted until late June 2015. The next cluster (3) appeared in August 2015, which lasted until January 2016. The four clusters of DENF in 2016 all began in January and lasted until June and July; and they affected the central and western portions of the country. Figure 10 clearly indicates that four out of the six multivariate clusters occurred during 2015, with two occurring in 2016. Again, clusters 3 and 4 only include significant clustering of DENF, not significant co-occurrence of both DENF and CHIK (Table 4). Therefore, 2015 was a more severe epidemic year regarding the co-occurrence of DENF and CHIK, since clusters 1, 2, and 6 occurred in the first half of 2015; and cluster 5 was the only cluster displaying significant co-occurrence in 2016.

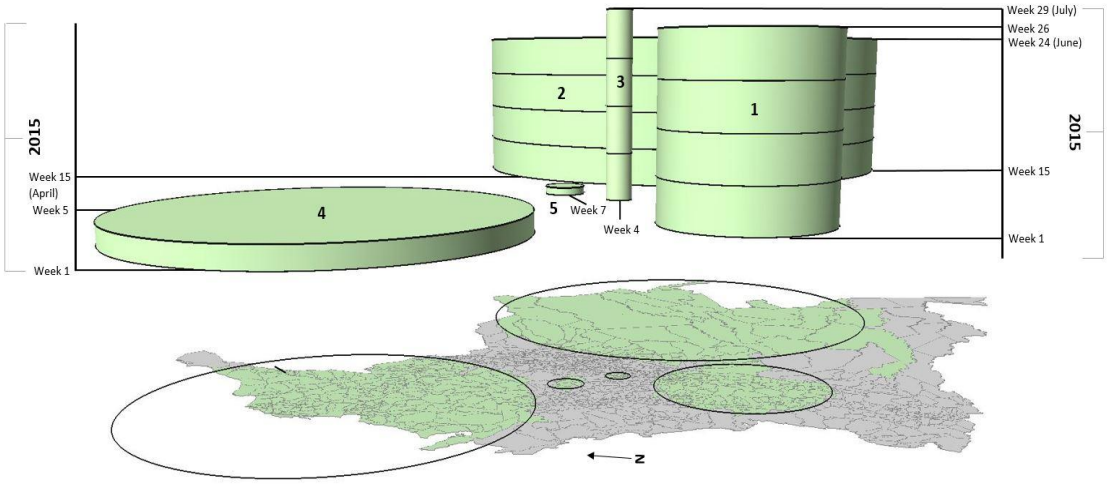


Figure 8: 3D visualization of the CHIK space-time clusters in Colombia (2015-2016)

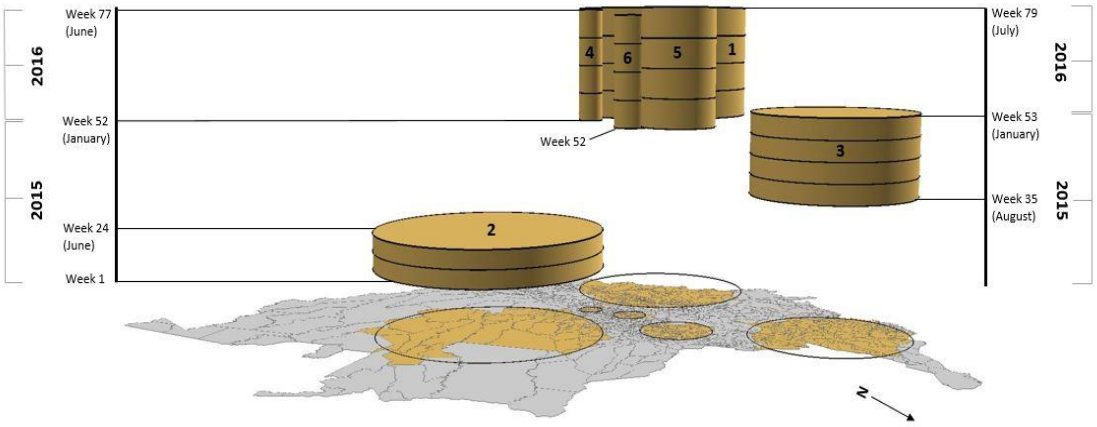


Figure 9: 3D visualization of the DENF space-time clusters in Colombia (2015-2016)

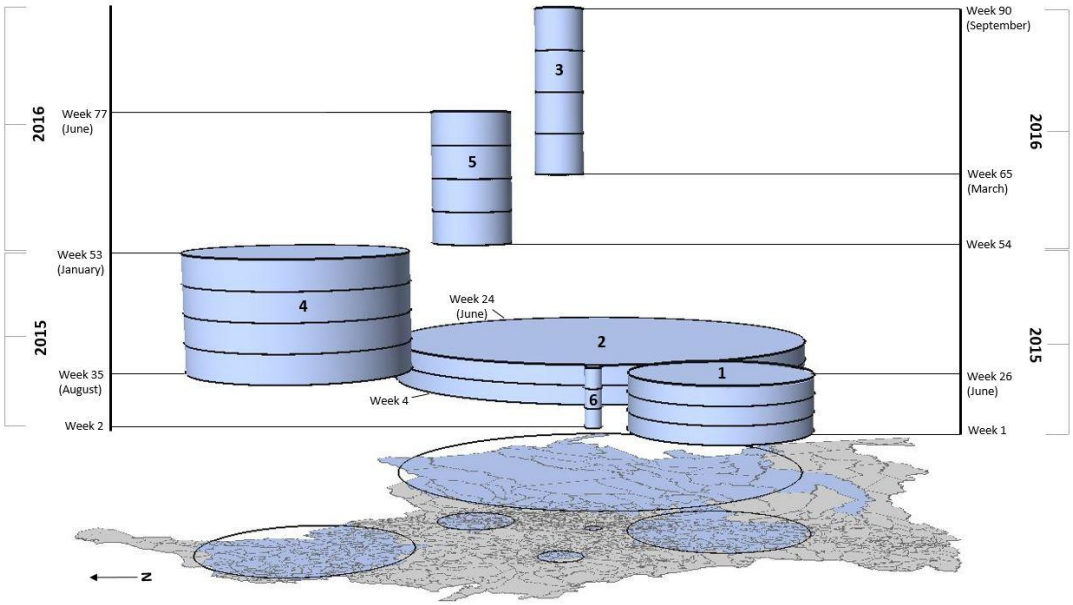


Figure 10: 3D visualization of the multivariate space-time clusters in Colombia (2015-2016)

3.4 Discussion

The results for objective 1 highlight statistically significant space-time clusters of DENF, CHIK, and regions of simultaneous excess incidence of both diseases (multivariate clusters). Further studies will be needed to determine the reasons for the differences between years, although climatic shifts (Eastin et al., 2014; Hii et al., 2016), changes in vector control (Ooi et al., 2006), and changes in reporting (Shepard et al., 2014; Silva et al., 2016) have been known to cause fluctuations in year to year infection rates. It has also been noted that DENF follows a cyclical pattern manifesting as an epidemic every 2–3 years (Cali 2010). Clusters for both diseases coincided with regions of expected vector presence of *Ae. aegypti* and *Ae. albopictus*. The clusters are partially a consequence of the geography of Colombia. Of the 1,123 municipalities, 829 have climatic conditions that are favorable for *Aedes* to reproduce (EPS C, 2015).

All clusters are centered on regions within the known precipitation, temperature, and elevation ranges of the vector. Aedes are rarely found above 1,700m, and thus, municipalities within a cluster above the aforementioned elevation reflects a consequence of the Kulldorff method rather than the occurrence of local transmission. This results from cylinders being the base shape for the scanning operation, meaning that the reported clusters can sometimes include municipalities with low rates or where transmission did not actually occur. In general, it is unlikely that the municipalities above 1,700m hosted local transmission, yet they may host travel-related cases. Unfortunately, the dataset does not indicate whether cases were suspected of being transmitted elsewhere other than the reporting municipality. However, we circumvent the elevation issue by reporting relative risk for each municipality belonging to a cluster. Many of the municipalities with a

relative risk of 0 are found in regions with an elevation greater than 1,700m. Reporting and visualizing the relative risk for each municipality also facilitates targeted interventions by identifying the municipalities that have statistically significant excess cases of each disease (i.e. RR > 1), reducing the uncertainty of solely reporting the space-time clusters.

The multivariate STSS reported four clusters of space-time co-occurrence of both DENF and CHIK. Since CHIK just recently appeared in Colombia, it is important to identify areas of co-circulation with DENF, which is hyperendemic in many regions of the country. Since the clinical manifestations of DENF and CHIK (also Zika) are similar, identifying the correct disease via clinical diagnosis is challenging in regions of co-circulation (Silva Jr. et al. 2018). Unlike DENF, chronic complications following a CHIK infection are common (de Andrade et al. 2010), which may last for weeks, months, and even years. Therefore, it is critical to implement timely and effective diagnostic methods (e.g. laboratory testing) to confirm the viral etiology between DENF, CHIK, and Zika. Reducing misdiagnosis is especially important in areas of co-circulation, and identifying areas that experience simultaneous outbreaks of DENF, CHIK, or Zika (e.g. via multivariate STSS) can facilitate targeted interventions. Co-infection of DENF and CHIK is also possible, however, there has not been any observable clinical significance, such as exacerbated symptoms (Furuya-Kanamori et al. 2016).

Despite the lack of observable clinical significance of co-infection, there is concern that antibody-dependent enhancement may increase the chance of contracting one disease after having previously been exposed to the other (Dejnirattisai et al., 2016). Both diseases are considered to be emerging VBDs. Thus, with low levels of immunity in

newly exposed regions, the capacity for co-infection remains high (Furuya-Kanamori et al., 2016). DENF and CHIK represent a challenge to health authorities given that their clinical symptomatology presents very similarly and can be difficult to identify (Alvis-Guzmán et al., 2017). While DENF has been in the Americas for over 200 years, CHIK was first reported in the Americas in 2014 (Leparc-Goffart et al., 2014). By the end of that year, 82,588 cases were eventually recorded in Colombia. The dataset represents the two years after that initial year, when immunity was first present in the population, and thus the rate of infection may be more representative of the long term. Although the objective of this study was to identify the existence of case clusters in space and time, consequent studies should investigate epidemiological conditions within each cluster to determine root causes for their occurrence. Efforts should include examinations of both vectorial capacity as well as host characteristics and variations in vector control.

The 3D visualizations (Figures 8-10) can improve the understanding of the size, duration, and movement of space-time clusters of disease (Desjardins et al. 2018b). 3D visualizations should supplement traditional 2D approaches (Desjardins et al. 2018a), especially for space-time analyses that include a large number of temporal observations. Otherwise, key space-time patterns can be masked by solely using 2D techniques. However, the 3D visualizations provided are static; crowding and occlusion could have been an issue if there were a larger number of reported space-time clusters. Integrating the 3D visualizations in an interactive environment (e.g. web-GIS platform) can improve their effectiveness by allowing the user to move around the image, for example.

3.5 Conclusion

This chapter successfully identified significant univariate and multivariate space-time clusters of DENF and CHIK in Colombia during the outbreaks in 2015 and 2016, and discussed some of the processes which were responsible for the epidemics. The reported clusters and the relative risk of each disease for each municipality were effectively visualized in both 2D and 3D, which complement each other and should be integrated in a space-time analysis of infectious diseases, improving the understanding of an epidemic. The 3D visualizations are especially effective at showing the space-time dynamics of the clusters, including the size, duration, and movement over time. Identifying the space-time co-occurrence of DENF and CHIK via multivariate STSS is another main contribution of this chapter, which sheds light on the regions in Colombia that were hotspots for simultaneous VBD outbreaks. The approaches described in this paper can be utilized to identify high-risk areas, facilitating the implementation of targeted interventions to mitigate future outbreaks, such as reducing the number of potential breeding sites (e.g. open containers) for *Aedes*, and allocating improved public health resources. The areas of co-occurrence particularly require closer examination about why clusters of DENF and CHIK intersect in space and time. The approaches are also easily transferable to other study areas, acting as a universal toolset to study space-time clustering of infectious disease, and other phenomena such as crime.

Despite the study's strengths, there are a variety of limitations and avenues for future research. First, a cluster's relative risk reported by the scan statistic does not consider sociodemographic variations within the radius. For example, children and the elderly have the highest proportion of CHIK and DENF related deaths, and relative

abundance may be related to socioeconomic variation. Therefore, the statistic could consider key sociodemographic factors (e.g. adjusted rates based on socioeconomic status), potentially generating more accurate relative risk estimates.Second, the cylindrical shape of the clusters is likely not the true boundary of outbreaks (Kulldorff et al., 2005), and some of the reported clusters extendbeyond the boundary of Colombia. I suggest that future investigations utilize other scanning methods (e.g. irregular shapes; Duczmal and Assuncao 2004; Tango and Takahashi, 2005; Ullah et al., 2017) to mitigate this concern. Finally, the data used is constantly revised by the National Institute of Health based on information reported by municipalities. Some municipalities do not submit timely information which results in necessary data adjustments. There is also under-reporting to be considered as well as potential misdiagnosis given the similarity in clinical symptoms of the viruses (Alvis-Guzmán et al., 2017).

Overall, exploratory space-time cluster approaches should be used to shed light on the space-time dynamics of epidemics and outbreaks and highlight the areas that experienced the greatest burden of disease. Subsequent research is necessary to understand the factors that influence disease transmission, while fine-level analysis (e.g. neighborhoods) can uncover local variations of disease incidence within at-risk areas. More research efforts should focus on evaluating the effectiveness of 3D visualization approaches for space-time clusters, such as user studies. 3D visualizations can also benefit from interactive environments that allow the user to navigate freely, rather that static images (such as Figures 8-10). Software that specializes in space-time clustering techniques, such as SaTScan, may not allow visualization of the results, which requires familiarity and training with a GIS and other visualization software. Future

developments in software should integrate visualization functionality to streamline subsequent analysis. As novel technologies emerge and data becomes available, new epidemiological questions will arise requiring to investigate additional facets of space-time analytics. For instance, population data become increasingly detailed with respect to their spatial and temporal resolutions, which will enable us to adjust clustering methods for spatially and temporally inhomogeneous background populations. In addition, as techniques for tracking or inferring individual people's location are already available at large scales, research about space-time disease clustering may shift focus from the point- and polygon-based paradigms to trajectory-based methods.

CHAPTER 4: SPATIO-TEMPORAL MODELING OF NEIGHBORHOOD LEVEL RISKS FOR DENGUE, CHIKUNGUNYA, AND ZIKA IN CALI, COLOMBIA

Vector-borne diseases (VBDs) affect more than 1 billion people a year worldwide, cause over 1 million deaths, and cost hundreds of billions of dollars in societal costs (WHO 2014b). Mosquitoes are the most common vectors, responsible for transmitting a variety of arboviruses. Recently, dengue fever, chikungunya, and Zika have been responsible for the majority of the burden caused by mosquito-borne diseases. These three diseases are primarily transmitted by the *Aedes Aegypti* and *Aedes Albopictus.* Since both *Aedes* species are peridomestic and container-breeding mosquitoes, vector surveillance should begin at the neighborhood level - where a variety of local factors may increase the risk of transmission.

Dengue has been endemic in Colombia for decades, and chikungunya and Zika first appeared in 2013. For this study, we examine weekly cases of three VBDs in Cali, Colombia from 2015-2016. Space-time conditional autoregressive models have been developed to quantify how disease risk is influenced by socioeconomic, environmental, and accessibility risk factors, and predicting outbreaks using lagged weather variables. Our model is also capable of identifying regions with high risk clusters at the neighborhood-level. Statistical inference is drawn under Bayesian paradigm using Markov Chain Monte Carlo techniques. The results provide detailed insight about the spatial heterogeneity of disease risk and the associated risk factors at a fine-level, coupled with an early warning surveillance system; informing public health officials to motivate at-risk neighborhoods to take an active role in vector surveillance and control, and improving educational and surveillance resources throughout the city of Cali.

4.1 Data

Cali was located in both univariate and multivariate space-time clusters detected in Chapter 3 and serves as the study area for the remainder of this dissertation. Cali is the second-largest city in Colombia and third most populous with an estimated 2010 population of 2.3 million. Cali is very densely populated with an average density of 4,000 people per square kilometer. The city is comprised of 340 neighborhoods (barrios), which are classified by socioeconomic stratum. Similar to Chapter 1, individual cases of DENF, CHIK, and Zika for the years of 2015 and 2016 are used (see Figure 11), which were provided by Colombia's National Institute of Health. Zika data was not available for Chapter 3 but was available during the data collection process here in Chapter 4. The cases were geocoded to the neighborhood level using each neighborhood's name as the address locator in the geocoder algorithm in ArcGIS 10.6. In other words, each VBD case record contained a neighborhood where the infected individual lived (individual addresses were not available), then the geocoder essentially aggregated the cases to a particular neighborhood after a successful match.

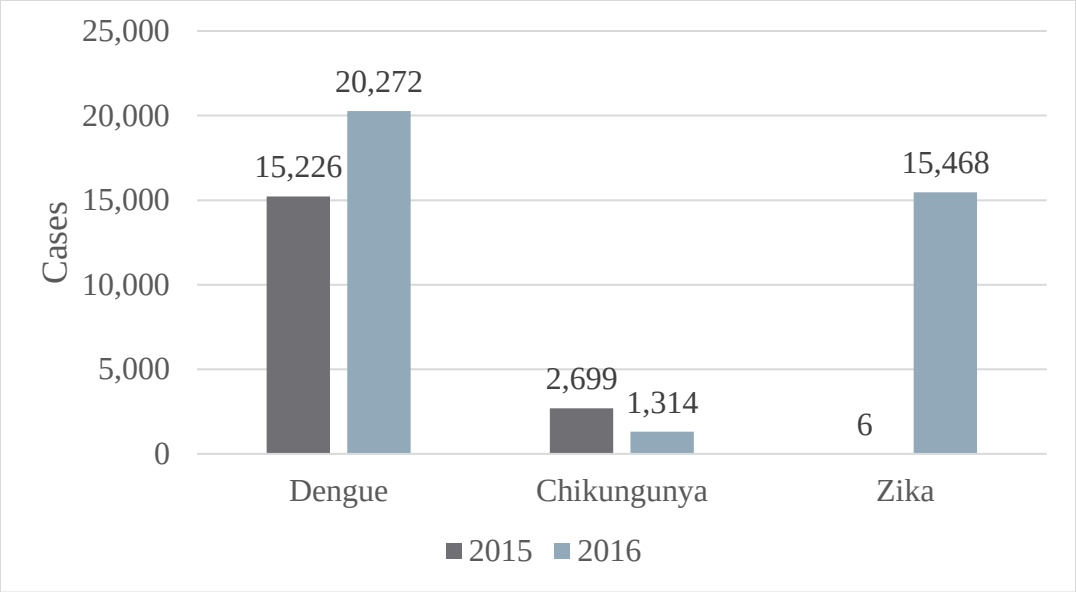


Figure 11: Total DENF, CHIK, and Zika cases in Cali, Colombia from 2015-2016

As a result, 26,503 out of 35,498 DENF cases (74.6%); 2,723 out of 4,099 CHIK cases (66.4%); and 12,247 out of 15,474 (79.1%) of Zika cases were successfully geocoded and aggregated to the neighborhoods in Cali. Cases that were not geocoded did not have an address nor a neighborhood, therefore, it was impossible to assign coordinates to the unmatched cases.

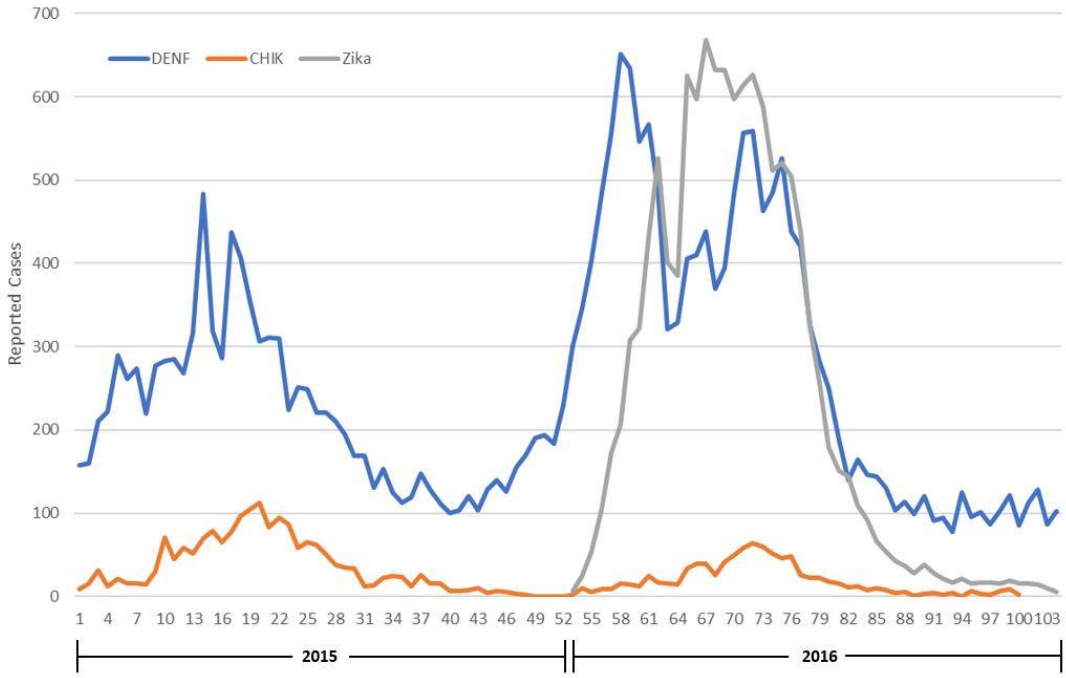


Figure 12: Temporal distribution of DENF, CHIK, and Zika cases by week in Cali from 2015 to 2016

Figure 12 shows the temporal distributions of DENF, CHIK, and Zika Cases in Cali between 2015 and 2016. Between 2015 and 2016, DENF experienced three major outbreaks: March to mid-May 2015; February to early-April 2016; and mid-June 2016 to early August 2016 (represented by the peaks in Figure 12). For CHIK, two major epidemics occurred between late February 2015 and early July 2015; and early April 2016 to early June 2016 (Figure 12). For Zika, a major epidemic occurred between late February 2016 and mid-June of the same year.

Next, socioeconomic and demographic data were provided by the Colombian census (either 2005 or 2010 estimates provided by the City of Cali), including population density, age, race, households with sewer and water access, educational attainment, employment status, socioeconomic stratum, among others. The last national census occurred in 2005, while the new 2018 census has yet to be released. The location of

healthcare centers and the environmental variables were provided by the city of Cali (2010 data) – green zones, rivers, tire shops, water pumps, cemeteries, and plant nurseries, which were geocoded as point layers with the exception of green zones (area - polygons). The environmental variables are included as potential *Aedes* habitats. For green zones, the area of the green zones for each neighborhood was computed in square-kilometers. Similar to Delmelle et al. (2016), relative proximity to rivers, tire shops, water pumps, cemeteries, plant nurseries, and healthcare centers was computed by using kernel density estimation (KDE) – representing the density of points for each layer. KDE was also computed to produce the density of trees. Zonal statistics in ArcGIS 10.6 was used to summarize the average KDE for each neighborhood in Cali.

Finally, the 11 weather variables are weekly observations that were collected from the Global Historical Climate Network (GHCN) archive (Menne et al. 2012), maintained by the National Climate Data Center (NCDC; https://www.ncdc.noaa.gov).

Due to the difference in units of measurements, each variable was normalized between 0 and 1 for subsequent analysis; that is $\hat{x} = \frac{(x_i - x_{min})}{(x_{max} - x_{min})}$. Table 5 lists the candidate independent/predictor variables of DENF, CHIK, and Zika (n=49). Appendix 11 provides a correlation matrix for the variables in Table 5 (excluding the weather variables).

Table 5: Candidate predictor variables of DENF, CHIK, and Zika in Cali

	Year	Source
Environmental/*Aedes* Habitats		City of Cali
Area of green zones (km^2)	2010	City of Cali
Relative proximity to river	2010	City of Cali
Relative proximity to tire shops	2010	City of Cali
Relative proximity to water pumps	2010	City of Cali
Relative proximity to cemeteries	2010	City of Cali
Relative proximity to plant nurseries	2010	City of Cali
Density of trees	2010	City of Cali
Healthcare Accessibility		
Relative proximity to a healthcare center	2010	City of Cali
Mean healthcare center density (km^2)	2010	City of Cali
Socioeconomic & Demographic		
Neighborhood Stratum	2010 (Est)	Census
Population density (km^2)	2005	Census
Density of occupied households (km^2)	2005	Census
Density of unoccupied households (km^2)	2005	Census
Households with sewer (%)	2005	Census
Households with water (%)	2005	Census
Individual age 0-4 years old (%)	2005	Census
Individual age 5-14 years old (%)	2005	Census
Individual age 15-24 years old (%)	2005	Census
Individual age 25-39 years old (%)	2005	Census
Individual age 40-64 years old (%)	2005	Census
Individual age 65 years old or more (%)	2005	Census
Female population (%)	2005	Census
White population (%)	2005	Census
Black population (%)	2005	Census
Indigenous population (%)	2005	Census
Individuals with disabilities (%)	2005	Census
Individuals who cannot read/write (%)	2005	Census
Individuals with no education (%)	2005	Census
Individuals with low education (%)	2005	Census
Individuals with medium education (%)	2005	Census
Individuals with high education (%)	2005	Census
Employed individuals (%)	2005	Census
Unemployed individuals (%)	2005	Census
Retired individuals (%)	2005	Census
Individuals doing housework (%)	2005	Census
Students (%)	2005	Census
Married individuals (%)	2005	Census
Single individuals (%)	2005	Census
Weather (Weekly Observations)		

Mean Temperature (°C)	2014-2016	City of Cali
Mean Maximum Temperature (°C)	2014-2016	City of Cali
Mean Minimum Temperature (°C)	2014-2016	City of Cali
Mean Daily Temperature Range (°C)	2014-2016	City of Cali
Maximum Daily Temperature Range (°C)	2014-2016	City of Cali
Mean Relative Humidity (°C)	2014-2016	City of Cali
Relative Humidity Range (°C)	2014-2016	City of Cali
Total Rain (mm)	2014-2016	City of Cali
Total days with measurable rainfall	2014-2016	City of Cali
Days with minimum temp. < 18 °C	2014-2016	City of Cali
Days with maximum temp. > 32 °C	2014-2016	City of Cali

4.2 Methodology

4.2.1 Principal component analysis

Due to the large number of socioeconomic and environmental variables (n = 29), a principal component analysis (PCA) was conducted to reduce and simplify the variables into new variable (components) that explain a large degree of variation without collinearity between the components. A PCA essentially summarizes the data by reducing the observations (variables) into a set of components, and the first component has the largest variance (Lever et al. 2017). The PCA analysis for this chapter was completed in Stata. Three main results of a PCA analysis were utilized to select the components to include in subsequent modeling: (1) scree plot – visualizes the eigenvalues (i.e. magnitude of the covariance matrix, where an eigenvector indictates the direction of the data in vector space) of each component in a line graph; (2) a table that displays the total variance explained by each component; and (3) a table that includes the rotated component loadings (direction and magnitude of each variable belonging to each principal component) for each variable in the components (essentially the relative importance of each variable in the components). The general rule of thumb is to keep component loadings with a coefficient value > 0.40 (Ventura et al. 2000).

4.2.2 Multicollinearity Testing

The remaining variables (including the weather variables) that were not included in the PCA analysis were assessed for multicollinearity using variance inflation factor (VIF) testing. VIF testing quantifies the degree of multicollinearity between variables in a regression-based analysis (Allison 1999). For this dissertation, variables with a VIF score < 3 (general rule of thumb) were included in the final modeling phase. VIF scores for each variable is defined as: $VIF_i = \frac{1}{1-R_i^2}$, where R_i^2 is the R^2 value for variable i. The VIF testing was completed in R.

4.2.3 Computing temporal lags for weather variables

One of the contributions of the subsequent space-time modeling will predict DENF, CHIK, and Zika outbreaks by including temporally lagged weather variables. The eleven weather variables in Table 5 were first assessed for multicollinearity using Variance Inflation Factor (VIF) tests. The remaining weather variables with a VIF < 3 were selected for subsequent cross-correlation analysis to select significant lags at the weekly level. Cross-correlation analysis quantifies the similarity between two random variables. For this chapter, cross-correlation was computed between disease rates and each weather variable at a particular weekly lag. The "best" temporal lag for each climate variable for each disease (DENF, CHIK, and Zika) was determined using the following process:

(1) Identify significant correlation coefficients.

(2) Identify the strongest (most positive or most negative) significant correlation coefficient.

(3) Determine if the strongest correlation coefficient is within a weekly lag within the *Aedes* life cycle (egg to adult in 2 weeks; adult *Aedes* can live up to one month).

(4) Select weekly lag for each weather variable that meets criteria 1-3 above.

4.2.4 Space-time conditional autoregressive modeling

An example of a simple SAR model (aka spatial lag model) is defined as:

$$y = \rho W y + X\beta + \varepsilon \quad (13)$$

Where ρ is a scalar parameter quantifying the effect of the dependent variable y (e.g. DENF rates) in the neighbors on the dependent variable in the target location. Wy is an N*1 vector/contiguity matrix for the dependent variable – 1 if *i* and *j* are neighbors and 0 otherwise; $X\beta$ is an N*K matrix of observations on the independent variables multiplied by a K*1 vector of regression coefficients β for each independent variable *X*; and ε is a N*1 vector of independently and identically distributed normal random white noise process with zero mean and constant variance σ^2.

SAR models have been extended to integrate a temporal dimension – called spatiotemporal autoregressive (STAR) models, and can be conceptualized as:

$$y = \rho_s W_s + \rho_t W_t + X\beta + \varepsilon \quad (14)$$

Where ρ_s is a scalar parameter for spatial autocorrelation; ρ_t is a scalar parameter for temporal autocorrelation; W_s is the spatial weight matrix for the dependent variable *y;* and W_t is the temporal contiguity matrix for the dependent variable *y*. The STAR model can also be written to incorporate space-time interaction – i.e. $W_t W_s$ (Banerjee 2018).

Abellan et al. (2008) point out that including a temporal dimension in space-time models, especially Bayesian hierarchical models (e.g. ST-CAR) considerably strengthens

epidemiologic interpretations of disease risk patterns and detecting localized excess risk. Since epidemics will vary both spatially and temporally, using solely spatial models that aggregate spatiotemporal data is both naïve and will result in unstable risk estimates. This section describe the mechanisms of a space-time conditional autoregressive model (ST-CAR), which draws statistical inference under a Bayesian paradigm using Markov Chain Monte Carlo (MCMC) techniques. ST-CAR models can determine significant spatial and spatiotemporal effects between a dependent variable and a variety of independent predictor variables. ST-CAR models are especially useful for small-area analysis of disease rates.

First, a Poisson generalized linear model (GLM) can be computed to detect significant effects of predictor variables on a dependent variable (typically disease risk); and the presence of spatiotemporal autocorrelation in the residuals. The Poisson GLM is defined as:

$$Y_{ij} \sim Poisson\ (E_{ij}R_{ij}) \tag{15}$$

$$\log(R_{ij}) = \beta_0 + \beta_1 X1_{ij} \tag{16}$$

where Y_{ij} is the observed disease count in neighborhood *i* at week *j*; E_{ij} is the expected disease count in neighborhood *i* at week *j*; and R_{ij} is the disease risk in neighborhood *i* at week *j*. β_0 is the regression intercept; $\beta_1$is the regression coefficient and $X1_{ij}$is an independent predictor variable observation in neighborhood *i* at week *j*.

Global Moran's I was then computed to detect spatial autocorrelation of the Poisson GLM residuals for each time period. Essentially, the Global Moran's I test

determines if there is evidence of unexplained spatial autocorrelation in the residuals; and if positive spatial autocorrelation is detected, then the assumption of independence is not valid for the data, and spatiotemporal autocorrelation should be considered when estimating covariate effects on the dependent variable. The global Moran's I index ranges from -1 to 1, while -1 indicates strong negative spatial autocorrelation, 0 indicates complete spatial randomness, and 1 indicates strong positive spatial autocorrelation; and the statistic is defined as:

$$I = \frac{n \sum_i \sum_l w_{il}(x_i - \bar{x})(x_l - \bar{x})}{\sum_i \sum_l w_{il} \sum (x_i - \bar{x})^2} \quad (17)$$

where *n* is the total number of neighborhoods, w_{il} is the spatial weight between neighborhood *i* and *l*, $\bar{x}$ is the mean of residuals for all neighborhoods, $x_i$is the residual value in neighborhood *i*, and $x_l$is the residual in neighborhood *l*. The Moran's I tests were conducted in R.

Next, a Bayesian hierarchical model proposed by Rushworth et al. (2014) is defined based on the formulations provided by Lee et al. (2018) using a Poisson data model (for case/population data). The model "represents the spatio-temporal pattern in the mean response with a single set of spatially and temporally autocorrelated random effects. The effects follow a multivariate autoregressive process of order 1" (Lee et al. 2018, p. 6). In other words, when going from one week to another (e.g. *t* +1), it will yield an effect on the dependent variable (disease risk). Therefore, this model examines linear trends which can be interpreted as how disease risk is influenced across time. It is assumed that the estimated effect on disease risk in the ST-CAR model is not specific to a particular week, but a process that is influenced by the covariate data across the weeks (temporal unit). Suppose a study region is divided into a collection of *N* non-overlapping

areal units (e.g. neighborhoods) indexed by $i \in \{1\ldots,N\}$; and the data is observed for multiple time periods, that is: $j \in \{1\ldots,T\}$. As suggested before, ST-CAR models utilize prior distributions, where the CAR distributions state that adjacent variables in space or time are autocorrelated, and non-adjacent variables are independent. The spatial weight matrix $W = (w_{ik})$, where a value of 1 indicates that *i* and *j* are spatially adjacent and 0 otherwise. Since we do not know where a person was infected (bit by the mosquito) or have vector surveillance data, the abovementioned adjacency matrix is valid since *Aedes* do not fly more than 400 meters from where they emerged as larvae (WHO 2019). A temporal weight matrix can also be defined as $D = (d_{jt})$, where a value of 1 is given if $t - j = 1$, and 0 otherwise. The first part of the model is defined as:

$$Y_{ij}|E_{ij,}R_{ij} \sim Poisson(E_{ij}R_{ij}) \tag{18}$$

$$\ln(R_{ij}) = X_{ij}^{T}\beta + O_{ij} + \phi_{ij} \tag{19}$$

$$\beta_k \sim N(0, 1000)\ k \in \{1, \ldots, p\}, \tag{20}$$

where Y_{ij} is the observed disease count in neighborhood *i* at week *j*; E_{ij} is the expected disease count in neighborhood *i* at week *j*; and R_{ij} is the disease risk in neighborhood *i* at week *j*. X_{ij}^{T} $(x_{ij1\ldots\ldots}, x_{ijp})$ is a vector of known covariates *p* for neighborhood *i* and week *j*. The parameter β is an associated *p* x 1 vector of regression parameters, which can come from the initial Poisson GLM in Equations 15 and 16. O is a vector of known offsets $(O_1, \ldots\ldots, O_N)_{K*N}$, where $O_j$is a $K * 1$ column vector of offsets for week *j* $(O_{1j}, \ldots\ldots, O_{Kj})$. An offset variable is used to scale the modeling of the mean in Poisson regression with a log link, which is the case in the above model (McCulloch et al. 2008). For example, since the dependent variable is rates, the offset can enforce that 10 cases of DENF in one week is not the same magnitude as 10 cases of DENF in 6 weeks. The

parameter ϕ_{ij} denotes spatiotemporally autocorrelated random effects for neighborhood *i* and week *j*. A variety of spatiotemporal structures can be fit for ϕ_{ij}, and Rushworth et al.'s (2014) ST-CAR model is used for this dissertation. In other words, "this model is appropriate if one wishes to estimate the evolution of the spatial response surface over time without forcing it to be the same for each time period" (Lee et al. 2018, p.8).

$$f(\phi_{1,\ldots,}\phi_T) \sim f(\phi_1)\prod_{j=2}^{T} f(\phi_j|\phi_{j-1}) \tag{21}$$

Where $\phi_j = \left(\phi_{1j,\ldots,}\phi_{Nj}\right)$ is a vector of random effects for week *j*. Temporal autocorrelation is enforced because ϕ_j depends on ϕ_{j-1}. $f(\phi_1)$ enforces spatial autocorrelation in the random effects, where the spatial structure is defined in the CAR prior in Equation 22:

$$\phi_{i1}|\phi_{-i} \sim N\left(\frac{\rho\sum_{k=1}^{N} w_{ik}\phi_{k1}}{\rho\sum_{k=1}^{N} w_{ik} + 1 - \rho}, \frac{\tau^2}{\rho\sum_{k=1}^{N} w_{ik} + 1 - \rho}\right) \tag{22}$$

Where ρ controls the spatial autocorrelation with ρ = 1 indicating strong spatial autocorrelation, which is conditional upon the mean random effects of adjacent neighborhoods. ρ = 0 represents independent random effects with a constant mean and constant variance. The conditional precision is controlled by τ, where precision is higher when more prior information (e.g. adjacent neighborhoods) is borrowed to determine the posterior estimates.

Equation 23 (CAR Prior) enforces temporal autocorrelation in the random effects and is defined as:

$$\phi_j|\phi_{j-1} \sim N(\alpha\phi_{j-1}, \tau^2 \mathrm{Q}(\rho, W)^{-1}) j \in \{2,\ldots,T\}, \tag{23}$$

Where Q(ρ, W) is a precision matrix, that is defined as ρ(diagonal(W1) – W) + (1 – ρ)I, where I is a N x N identity matrix and the '1' is a vector of ones (N x 1). The α controls the temporal autocorrelation, where 0 is temporally independent and 1 is strong temporal dependence. The CAR priors also include weakly informative hyperpriors (i.e. probability distribution from priors to inform/update posterior values), which are the three parameters defined below:

$$\tau \sim Uniform\ [0,1000],$$
$$\alpha \sim Uniform\ [0,1],$$
$$\rho \sim Uniform\ [0,1]$$

The values of the hyperpriors are informed by the data. For example, a non-stationary spatial process would occur when $\rho = 1$; and non-stationary temporal process would occur if $\alpha = 1$. Overall, the ST-CAR model utilized in this chapter states that when going from one week to another (t + 1), it yields an effect on the dependent variable (DENF, CHIK, or Zika risk), which is influenced by spatially and temporally dependent covariates. In other words, disease risk in a target neighborhood is influenced by current and past values of disease rates and covariates at surrounding spatial and temporal (which is a process that evolves over time). Conceptually, a spatial example would suggest that a neighborhood with low rates of DENF would have an increased risk of DENF if an adjacent neighborhood reported high rates of DENF (interaction effect).

Statistical inference is derived from Markov Chain Monte Carlo (MCMC) simulations. Monte Carlo simulation is a very common approach in statistics that estimates parameters by randomly generating values (probability distribution) for each iteration (Dwass 1957). Monte Carlo simulation compares the observed data to a large number of simulated data to derive statistical significance. The second component of

MCMC is Markov chains, which is a stochastic (random) model that describes a sequence of potential events where the probability of an event occurring depends on the previous state (i.e. the former) of the event (Gilks et al. 1995). Clearly explained by Shaver (2017), MCMC methods pick a random parameter value to consider and continues to generate random values. For a pair of parameter values, MCMC computes how well the parameter value will explain the data, influenced by prior beliefs. The generated parameter value will be added to the Markov chain if it is "better" than the previous value. Therefore, the main purpose of MCMC is to estimate the best posterior distributions that best explain observed data for ST-CAR models.

The user must select how many samples to generate (iterations) and how many samples to remove in the beginning to find a good starting point for the MCMC process (burn-ins). For example, the number of generations can be set to 100,000 with 10,000 burn-ins, essentially forcing the MCMC to not consider the first 10,000 iterations. After a model run is completed, a variety of results can be produced and analyzed, including posterior quantities, degree of spatial and temporal dependence of the data after adjusting for covariate effects, MCMC samples and fitted values for the parameters, posterior relative risk distributions, posterior medians and 95% credible intervals for relative risk of disease (dependent variable).

4.3 Results

4.3.1 PCA Analysis

Figure 13 and Tables 6 and 7 provide the results of the PCA analysis using most of the socioeconomic variables listed in Table 5. The Scree Plot in Figure 13 suggests that the first two principal components should be kept for further analysis because this is

where the “elbow” forms; essentially when the eigenvalues begin to decrease, and the remaining components explain little variance. Table 6 describes the variance explained by the top 3 principal components. The first two that were selected for inclusion in the space-time modeling explain 60.8% of the variance (42% and 18.8%, respectively).

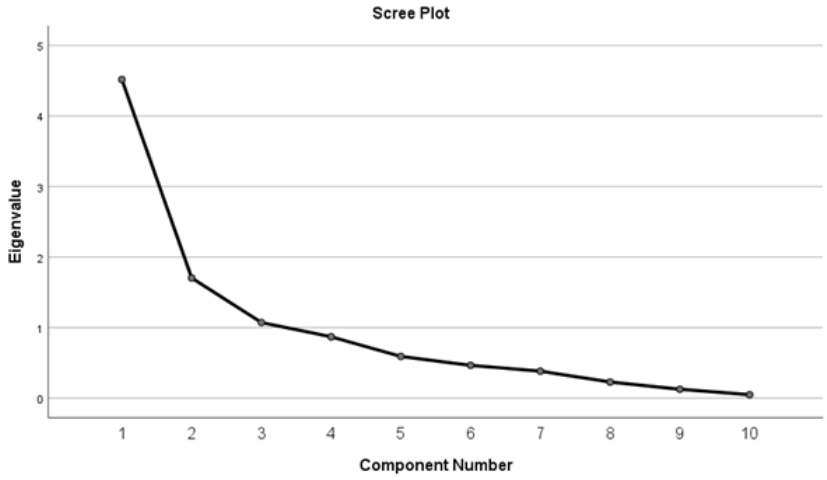


Figure 13: Scree Plot of the PCA Analysis of Socioeconomic variables in Cali, Colombia

Table 6: Explained variance by top 3 principal components

Component	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings	
	Total	Variance	Cumulative %	Total	% of Variance
1	4.516	45.163	45.163	4.203	42.032
2	1.703	17.031	62.194	1.882	18.817
3	1.072	10.719	72.912	1.206	12.064

Table 7: Rotated component loadings with coefficient > 0.40

Variable	Component		
	1	2	3
Strata	.856		
% OHH		.749	
% HW		.657	
% Work	.784		
% Retired	.781		
% Disabled			.782
% Student		.873	
% Married	.941		
% EmptyHH	.484		-.514
% HighEduc	.940		

Table 7 shows the variables that belong to each component. PC1 includes neighborhood strata, individuals who are employed, retired persons, married individuals, empty households, and individuals with high education. PC2 includes occupied households, individuals who do housework, and individuals who are students. PC3 includes disabled individuals, but was excluded from further analysis and it was included as a separate variable in further VIF testing with the remaining covariates. PC1 can be interpreted as employed, higher-income people who are likely to be older and married due to the large coefficients of retired and married individuals. PC2 can be interpreted as people who are likely to spend more time at home than those in PC1 because they are either students or do housework for a living. PC2 probably also captures younger individuals due to the inclusion of students.

4.3.2 Selecting Lagged Weather Variables

First, VIF testing was used to assess multicollinearity between the 11 candidate weather variables. As a result, 7 variables were selected for the subsequent cross-correlation analysis (VIF values < 4) - Mean Temperature (Tavg), Maximum Daily Temperature Range (DTRMax), Relative Humidity Range (RHRng), Total Rain (Train),

Total days with measurable rainfall (RainD), Days with minimum temp. < 18 °C (CoolD), and Days with maximum temp. > 32 °C (WarmD). Lagged cross-correlations were computed between average weekly disease rates and each of the seven weekly weather variables; and cross-correlations were computed from a weekly lag of 0 weeks to a lag of 15 weeks. Table 8 provides the results for DENF for each of the 16 lags and 7 covariates. The following lags were selected for DENF: 5 weeks for Tavg, 3 weeks for Train, 3 weeks for RHRng, 5 weeks for DTRMax, 5 weeks for RainD, 2 weeks for CoolD, and 5 weeks for WarmD. For CHIK (Table 9), the following lags were selected: 6 weeks for Tavg, 5 weeks for Train, 5 weeks for RHRng, 8 weeks for DTRMax, 6 weeks for RainD, 1 week for CoolD, and 6 weeks for WarmD. For Zika (Table 10), the following lags were selected: 8 weeks for Tavg, 5 weeks for Train, 7 weeks for RHRng, 3 weeks for DTRMax, 5 weeks for RainD, 3 weeks for CoolD, and 3 weeks for WarmD.

Table 8: Lagged cross-correlations between DENF rates and weekly weather variables during 2015-2016 in Cali, Colombia.

Lag (Weekly)	Tavg	Train	RHRng	DTRMax	RainD	CoolD	WarmD
0	0.419*	-0.2069*	-0.012	0.4635**	-0.5079*	0.2023*	0.5072*
-1	0.2851*	-0.4134*	-0.0574	0.1496	0.0102	-0.1097	0.3600*
-2	0.0324	0.0982	0.2118*	0.0769	0.3374*	-0.3536***	0.0945
-3	0.0739	-0.4575***	-0.4482***	0.0294	-0.0119	-0.1658	-0.012
-4	0.2369*	-0.0923	-0.1657	0.1052	-0.4879*	0.0812	0.2001*
-5	0.3691***	-0.2221*	-0.3620*	0.2062***	-0.5417***	0.2007*	0.2169***
-6	0.2545*	-0.3164*	-0.5568**	-0.1042	-0.3589*	0.3037*	0.0136
-7	-0.2201*	-0.01	-0.4728*	0.0793	-0.3376*	0.4826**	-0.0473
-8	0.2424*	-0.5149**	-0.0978	-0.0875	-0.3089*	0.1807	-0.1565
-9	0.2592*	-0.1905*	0.0667	-0.0741	-0.1939*	0.1793	-0.3166*
-10	0.2129*	-0.1208	-0.0304	0.0563	-0.0661	0.0098	0.0157
-11	0.0669	-0.4182*	-0.2715*	-0.0389	0.1658	0.1477	-0.3867*
-12	-0.44*	-0.1085	0.2702*	-0.1213	0.123	0.072	-0.4075*
-13	-0.7975**	0.1964*	0.4152*	0.2302*	0.4596*	-0.0098	-0.5279**
-14	-0.3657*	0.1404	0.4852*	0.1951*	0.4770*	0.0086	-0.3961*
-15	-0.0267	0.3656*	0.2678*	0.4587*	0.2170*	0.2420*	0.0058

*Statistically significant at the 5% level; **Maximum coefficient significant at the 5% level ***Selected lag for ST-CAR modeling

Table 9: Lagged cross-correlations between CHIK rates and weekly weather variables during 2015-2016 in Cali, Colombia.

Lag (Weeks)	Tavg	Train	RHRng	DTRMax	RainD	CoolD	WarmD
0	0.2719*	-0.3569*	-0.2131*	0.2366*	0.0899	0.1638	0.2156*
-1	-0.2550*	-0.0835	0.4816*	0.3814*	-0.3712*	0.6467***	0.1627
-2	0.0556	-0.0653	0.0247	-0.1309	-0.1949*	-0.0544	-0.1232
-3	0.0569	0.1351	-0.1902*	-0.0403	-0.0154	-0.0068	0.2203*
-4	0.0953	-0.178	-0.3161*	0.1502	0.0638	-0.1907*	0.0509
-5	0.3223*	-0.4736***	-0.4316***	0.1433	-0.3877*	0.0716	0.1411
-6	0.3912***	0.2560*	0.2426*	0.2366*	-0.4385***	-0.1717	0.3696***
-7	-0.0313	-0.3434*	-0.4371*	0.0661	-0.0097	-0.1377	-0.0048
-8	0.0439	-0.1234	-0.123	-0.4368***	0.0715	-0.1281	-0.3022*
-9	-0.0552	-0.1848*	-0.2302*	-0.2853*	-0.2733*	0.0084	-0.2143*
-10	-0.3205*	-0.3197*	-0.0275	-0.3509*	0.1021	-0.015	-0.4696**
-11	-0.0193	0.2043*	0.0018	0.2532*	-0.2317*	0.3411*	-0.2174*
-12	0.028	-0.0924	0.0099	0.1970*	0.007	0.2569*	0.0487
-13	-0.2116*	-0.0614	0	0.0258	0.2006*	0.4650*	-0.3193*
-14	0.0529	-0.0284	0.4937**	0.2367*	0.119	-0.1005	0.0663
-15	-0.4418**	-0.0451	0.3657*	0.1037	0.4364*	-0.1895*	-0.3802*

*Statistically significant at the 5% level; **Maximum coefficient significant at the 5% level ***Selected lag for ST-CAR modeling

Table 10: Lagged cross-correlations between Zika rates and weekly weather variables during 2016 in Cali, Colombia.

Lag (Weeks)	Tavg	Train	RHRng	DTRMax	RainD	CoolD	WarmD
0	-0.5412*	0.5020*	-0.0702	-0.6461*	0.3288*	-0.4841*	-0.6184*
-1	-0.4628*	0.4801*	-0.3961*	-0.6906*	0.2465*	-0.4490*	-0.5838*
-2	-0.4957*	0.3462*	-0.4049*	-0.6409*	0.2122*	-0.4755*	-0.5873*
-3	-0.4227*	0.4265*	0.0133	-0.7271***	0.4301*	-0.5677***	-0.5945***
-4	-0.4181*	0.4395*	0.2266*	-0.6577*	0.4853*	-0.4822*	-0.5205*
-5	-0.4100*	0.5257***	0.2914*	-0.6761*	0.5592***	-0.4046*	-0.4955*
-6	-0.4673*	0.4103*	0.4584*	-0.5464*	0.5394*	-0.3045*	-0.3437*
-7	-0.4846*	0.4073*	0.6216***	-0.3673*	0.6187*	-0.3109*	-0.4138*
-8	-0.6891***	0.4499*	0.5566*	-0.1743	0.6765*	-0.169	-0.3049*
-9	-0.6639*	0.5335**	0.5836*	-0.0983	0.6835**	-0.131	-0.1005
-10	-0.3379*	0.4225*	0.5437*	0.1724	0.5010*	-0.0594	0.1301
-11	-0.5529*	0.4859*	0.5236*	0.114	0.3932*	-0.1278	0.5650*
-12	0.12	0.3409*	0.2232*	0.2361*	0.2490*	-0.2381*	0.6689*
-13	0.2762*	0.1971*	-0.0414	0.5193*	-0.036	-0.3354*	0.7262**
-14	0.1414	0.2617*	-0.1357	0.5043*	-0.0703	-0.4433*	0.5313*
-15	0.3052*	0.0372	-0.3034*	0.5597*	-0.2725*	-0.5424*	0.5326*

*Statistically significant at the 5% level; **Maximum coefficient significant at the 5% level ***Selected lag for ST-CAR modeling

4.3.3 Generalized Linear Modeling Results

Equations 24 and 25 below define the Poisson GLM to detect significant effects of the covariates on DENF risk in Cali between January 2015 and December 2016.

$$\begin{aligned}\log(R_{ij}) = {} & \beta_0 + \beta_1 PC1_i + \beta_2 PC2_i + \beta_3 PNurseries_i + \beta_4 Tires_i \\ & + \beta_5 popdens_i + \beta_6 rivers_i + \beta_7 trees_i + \beta_8 TavgL5_i \\ & + \beta_9 DTRMaxL4_{ij} + \beta_{10} RelHRngL3_{ij} + \beta_{11} RainTL3_{ij} \\ & + \beta_{12} RainDL5_{ij} + \beta_{13} CoolDL2_{ij} + \beta_{14} WarmDL5_{ij}\end{aligned} \quad (24)$$

Since GLM coefficients show the multiplicative change in odds ratio, the coefficients were converted to a relative risk (Grant 2014), that is:

$$C = \exp(\beta_0)/(1 + \exp(\beta_0)) \quad (25)$$

$$RR = \frac{\exp(\beta)}{1 - C + (C * \exp(\beta))} \quad (26)$$

Table 11: DENF GLM Results

	Coefficient	Relative Risk	Std. Error	p
Intercept	-0.7722	NA	0.0471	<0.001
PC1	0.0916	1.0636	0.0097	<0.001
PC2	-0.1865	0.8298	0.0161	<0.001
Pnurseries	0.6651	1.9448	0.0340	<0.001
Tires	0.1155	1.1225	0.0497	<0.001
Popdens	-1.1714	0.3099	0.0596	<0.001
Rivers	-0.0106	0.9894	0.0363	<0.001
Trees	0.3765	1.4571	0.0672	<0.001
TavgL5	1.5640	4.7779	0.0438	<0.001
DTRMaxL4	0.7221	2.0588	0.0405	<0.001
RelHRngL3	0.8401	2.3167	0.0312	<0.001
RainTL3	-1.9794	0.1382	0.0512	<0.001
RainDL5	0.0129	1.0130	0.0332	<0.001
CoolDL2	-0.3471	0.7067	0.0388	<0.001
WarmL5	-0.8104	0.4447	0.0334	<0.001

Table 11 provides the results for the DENF Poisson GLM. All fourteen potential covariates of the subsequent ST-CAR modeling are statistically significant. PC2, population density, proximity to rivers, lagged total rain, lagged cool days, and lagged

warms days all have a negative relationship with DENF rates, suggesting a decreased risk of DENF transmission. PC1, proximity to plant nurseries, proximity to tire shops, tree density, lagged average temperature, lagged maximum daily temperature range, lagged relative humidity range, and lagged rain days all have a positive relationship with DENF rates, suggesting an increased risk of DENF transmission. Since there is evidence of significant relationships between the fourteen covariates and DENF rates, they were all included in the ST-CAR modeling. Furthermore, there was statistically significant evidence of unexplained spatial autocorrelation in the residuals after conducting a Moran's I test for a few weeks with a high number of DENF cases. Suvsequently, there was strong evidence of spatial autocorrelation of the residuals. For example, week 1 (January 2015) had a Moran's I value of 0.11 ($p<0.01$); and week 2 had a Moran's I value of 0.19 ($p<0.01$).

Equations 28 and 29 below define the Poisson GLM to detect significant effects of the covariates on CHIK risk in Cali between January 2015 and November 2016.

$$\begin{aligned}\log(R_{ij}) = {} & \beta_0 + \beta_1 PC1_i + \beta_2 PC2_i + \beta_3 PNurseries_i + \beta_4 Tires_i \\ & + \beta_5 popdens_i + \beta_6 rivers_i + \beta_7 trees_i + \beta_8 TavgL6_{ij} \\ & + \beta_9 DTRMaxL8_{ij} + \beta_{10} RelHRngL5_{ij} + \beta_{11} RainTL5_{ij} \\ & + \beta_{12} RainDL6_{ij} + \beta_{13} CoolDL1_{ij} + \beta_{14} WarmDL6_{ij}\end{aligned} \quad (27)$$

Table 12: CHIK GLM Results

	Coefficient	Relative Risk	Std. Error	p
Intercept	1.5761	NA	0.1594	<0.001
PC1	-0.0690	0.9879	0.0323	<0.001
PC2	-0.1320	0.8763	0.0415	<0.001
Pnurseries	0.5612	1.7528	0.1055	<0.001
Tires	0.2886	1.3345	0.1407	<0.001
Popdens	-1.5351	0.2154	0.1652	<0.001
Rivers	-0.1837	0.8322	0.1122	<0.001
Trees	0.2224	1.2490	0.2112	<0.001
TavgL6	-0.2434	0.7839	0.1417	<0.001
DTRMaxL8	0.1303	1.1392	0.1161	<0.001
RelHRngL5	-0.3466	0.7071	0.1118	<0.001
RainTL5	-0.9790	0.3757	0.1557	<0.001
RainDL6	-1.1744	0.3090	0.1177	<0.001
CoolDL1	1.1641	3.2031	0.0986	<0.001
WarmL6	-0.0045	0.9955	0.1034	<0.001

Table 12 provides the results for the CHIK Poisson GLM. All fourteen potential covariates of the subsequent ST-CAR modeling are statistically significant. PC1, PC2, population density, proximity to rivers, lagged average temperature, lagged relative humidity range, lagged total rain, lagged rain days, and lagged warm days all have a negative relationship with CHIK rates, suggesting a decreased risk of CHIK transmission. Proximity to plant nurseries, proximity to tire shops, tree density, lagged maximum daily temperature range, and lagged cool days all have a positive relationship with CHIK rates, suggesting an increased risk of CHIK transmission. Since there is evidence of significant relationships between the fourteen covariates and CHIK rates, they were all included in the ST-CAR modeling. Furthermore, there was statistically significant evidence of unexplained spatial autocorrelation in the residuals after conducting a Moran's I test for a few weeks with a high number of CHIK cases. For example, week 3 (January 2015) had

a Moran's I value of 0.18 ($p<0.01$); and week 15 (April 2015) had a Moran's I value of 0.11 ($p<0.01$).

Equations 30 and 31 below define the Poisson GLM to detect significant effects of the covariates on Zika risk in Cali between January 2016 and December 2016.

$$\begin{aligned}\log(R_{ij}) = {} & \beta_0 + \beta_1 PC1_i + \beta_2 PC2_i + \beta_3 PNurseries_i + \beta_4 Tires_i \\ & + \beta_5 popdens_i + \beta_6 rivers_i + \beta_7 trees_i + \beta_8 TavgL8_{ij} \\ & + \beta_9 DTRMaxL3_{ij} + \beta_{10} RelHRngL7_{ij} + \beta_{11} RainTL5_{ij} \\ & + \beta_{12} RainDL5_{ij} + \beta_{13} CoolDL3_{ij} + \beta_{14} WarmDL3_{ij}\end{aligned} \quad (28)$$

Table 13: Zika GLM Results

	Coefficient	Relative Risk	Std. Error	p
Intercept	0.4203	NA	0.0745	<0.001
PC1	0.0640	1.0252	0.0174	<0.001
PC2	-0.1337	0.8749	0.0196	<0.001
Pnurseries	-0.4481	0.6389	0.0594	<0.001
Tires	0.1099	1.1162	0.0702	<0.001
Popdens	-0.9121	0.4017	0.0799	<0.001
Rivers	-0.5372	0.5844	0.0586	<0.001
Trees	0.2168	1.2421	0.1078	<0.001
TavgL8	-0.2658	0.7666	0.0561	<0.001
DTRMaxL3	2.4696	11.8174	0.0762	<0.001
RelHRngL7	-0.1269	0.8808	0.0568	<0.001
RainTL5	0.6637	1.9419	0.0718	<0.001
RainDL5	0.0258	1.0261	0.0497	<0.001
CoolDL3	0.7121	2.0382	0.0830	<0.001
WarmL3	-1.3880	0.2496	0.0523	<0.001

Table 13 provides the results for the Zika Poisson GLM. All fourteen potential covariates of the subsequent ST-CAR modeling are statistically significant. PC2, proximity to plant nurseries, population density, proximity to rivers, lagged average temperature, lagged relative humidity range, and lagged warm days all have a negative relationship with Zika rates, suggesting a decreased risk of Zika transmission. PC1,

proximity to tire shops, tree density, lagged maximum daily temperature range, lagged total rain, lagged rain days, and lagged cool days all have a positive relationship with Zika rates, suggesting an increased risk of Zika transmission. Since there is evidence of significant relationships between the fourteen covariates and Zika rates, they were all included in the ST-CAR modeling. Furthermore, there was statistically significant evidence of unexplained spatial autocorrelation in the residuals after conducting a Moran's I test for a few weeks with a high number of Zika cases. For example, week 2 (January 2016) had a Moran's I value of 0.21 ($p<0.01$); and week 15 (April 2016) had a Moran's I value of 0.59 ($p<0.01$).

4.3.4 ST-CAR Modeling Results

Using the CARBayesST (Lee et al. 2018) in R, the three models above (Poisson log-linear GLM) were each fitted to the ST-CAR model described in section 4.2.4. The neighborhood matrix W is binary where $w_{ik} = 1$ if two neighborhoods in Cali share a common border and 0 otherwise. Using the defaults described in Lee et al. (2018), the model is run for 220,000 MCMC samples; 20,000 samples are removed by the initial burn-in period; and the thinning parameter is set to 10 (keeping every 10^{th} value and removing all others), which thins the samples to reduce autocorrelation of the Markov Chain – which results in 20,000 samples used for statistical inference. Thinning reduces the computational burden of long Markov chains, since it requires a large amount of memory. The deviance information criterion (DIC) – generalization of the Akaike information criterion (AIC), which essentially measures how well the model fits the data, where lower values of DIC indicate a better model fit. The shrinking of the confidence intervals for each covariate can also provide evidence of better model fit. The results of

six models are presented: Model 1 (DENF with no lags); Model 2 (DENF with lagged weather variables); Model 3 (CHIK with no lags); Model 4 (CHIK with lagged weather variables); Model 5 (Zika with no lags); Model 6 (Zika with lagged weather variables).

4.3.5 Model 1 – DENF with no lags

Table 14 summarizes the results of Model 1. The 95% credible intervals do not contain a value of 0, indicating that all of the covariates exhibit statistically significant relationships with DENF risk at the neighborhood level in Cali between January 2015 and December 2016. The spatial autocorrelation value of 0.98 indicates that there is very strong spatial dependence of the data after adjusting for covariate effects. The temporal autocorrelation value of 0.11 suggests that there is some presence of temporal dependence of the data after adjusting for covariate effects. Therefore, DENF risk in Cali at the neighborhood level is influenced by DENF rates and covariates in surrounding neighborhoods and time periods (weeks).

Table 14: Model 1 Results

DIC: 67,055.98	**2.50%**	**Median**	**97.50%**
Intercept	-2.0296	-1.6427	-1.3183
PC1	0.0159	0.0463	0.0785
PC2	-0.0885	-0.0475	-0.0068
Plant Nursery	0.0203	0.1213	0.2215
Tire Shops	-0.0838	0.0531	0.1902
Pop Dens (km2)	-0.173	-0.0146	0.1482
River/Ravine	-0.2473	-0.1233	-0.0029
Tree Density	0.1254	0.3071	0.4909
Avg Temp	1.5864	1.9947	2.3546
Days Temp Max	-0.24	0.2361	0.9014
Rel Humid Range	0.4641	0.8713	1.342
Rain Total	-1.3716	-0.613	0.003
Rain Days	-0.5249	-0.1837	0.1929
Cool Days	0.0013	0.5678	1.1938
Warm Days	-1.7085	-1.3392	-0.9888
Spatial Autocorrelation	0.977	**0.9829**	0.9872
Temporal Autocorrelation	0.0736	**0.1189**	0.1636

Table 15: Model 1 Results – relative risk of DENF for each covariate

Variable	RR (0.025)	RR (0.5)	RR(0.975)	RR %
PC1	1.016	1.047	1.082	**4.7**
PC2	0.915	0.954	0.993	**-4.6**
Plant Nursery	1.02	1.129	1.248	**12.9**
Tire Shops	0.92	1.055	1.209	**5.5**
Pop Dens (km2)	0.841	0.985	1.16	**-1.5**
River/Ravine	0.781	0.884	0.997	**-11.6**
Tree Density	1.134	1.36	1.634	**36**
Avg Temp	4.886	7.35	10.534	**635**
Days Temp Max	0.787	1.266	2.463	**26.6**
Rel Humid Range	1.591	2.39	3.827	**139**
Rain Total	0.254	0.542	1.003	**-45.8**
Rain Days	0.592	0.832	1.213	**-16.8**
Cool Days	1.001	1.764	3.3	**76.4**
Warm Days	0.181	0.262	0.372	**-73.8**

Table 15 shows the posterior medians and 95% credible intervals for relative risk of DENF in Cali between January 2015 and December 2016. The results suggest a 4.7% increase in DENF risk for PC1; a 4.7% decrease for PC2; a 12.9% increase for proximity to plant nurseries; a 5.5% increase for proximity to tire shops; a 1.5% decrease for population density; a 11.6% decrease for proximity to rivers/ravines; a 36% increase for tree density; a 635% increase for average temperature; a 26.6% increase for days with maximum temp. > 32 °C; a 139% increase for relative humidity range; a 45.8% decrease for total rainfall; a 16.8% decrease for total rain days; a 76.4% increase for cool days; and a 73.6% decrease for warm days. The results will be further explained in the discussion.

4.3.6 Model 2 – DENF lagged weather variables

Table 16 summarizes the results of Model 2. The 95% credible intervals do not contain a value of 0, indicating that all of the covariates exhibit statistically significant relationships with DENF risk at the neighborhood level in Cali between January 2015 and December 2016. The spatial autocorrelation value of 0.98 indicates that there is very strong spatial dependence of the data after adjusting for covariate effects. The temporal autocorrelation value of 0.11 suggests that there is some presence of temporal dependence of the data after adjusting for covariate effects. Therefore, DENF risk in Cali at the neighborhood level is influenced by DENF rates and covariates in surrounding neighborhoods and time periods (weeks). The DIC (66,964.16) is slightly lower than Model 1 (DENF with no lags – DIC = 67,055.98). In general, the lagged weather variables in model 2 shrunk the confidence intervals of the coefficients (most notably for average temperature and Days Temp Max).

Table 16: Model 2 Results

DIC: 66,954.16	2.50%	Median	97.50%
Intercept	-0.8195	-0.5423	-0.3055
PC1	0.0459	0.0732	0.1
PC2	-0.0042	0.0382	0.0798
Plant Nursery	0.0841	0.1806	0.2787
Tire Shops	-0.023	0.1118	0.2469
Pop Dens (km2)	-0.3296	-0.1749	-0.0176
River/Ravine	-0.2532	-0.1381	-0.0227
Tree Density	0.0901	0.2626	0.4404
Avg Temp (L5)	-0.6022	-0.2908	0.0415
Days Temp Max (L4)	0.4579	0.8115	1.2709
Rel Humid Range (L3)	0.4317	0.615	0.8213
Rain Total (L3)	-2.8222	-2.4343	-2.0671
Rain Days (L5)	-0.9178	-0.7063	-0.4845
Cool Days (L2)	-0.8337	-0.2914	0.1285
Warm Days (L5)	0.2861	0.6383	0.9196
Spatial Autocorrelation	0.977	**0.9824**	0.9867
Temporal Autocorrelation	0.0737	**0.1193**	0.1651

Table 17: Model 2 Results – relative risk of DENF for each covariate

Variable	RR (0.025)	RR (0.5)	RR (0.975)	RR %
PC1	1.047	1.076	1.105	**7.6**
PC2	0.996	1.039	1.083	**3.9**
Plant Nursery	1.088	1.198	1.321	**19.8**
Tire Shops	0.977	1.118	1.28	**11.8**
Pop Dens (km2)	0.719	0.84	0.983	**-16**
River/Ravine	0.776	0.871	0.978	**-12.9**
Tree Density	1.094	1.3	1.553	**30**
Avg Temp (L5)	0.548	0.748	1.042	**-25.2**
Days Temp Max (L4)	1.581	2.251	3.564	**125.1**
Rel Humid Range (L3)	1.54	1.85	2.273	**85**
Rain Total (L3)	0.059	0.088	0.127	**-91.2**
Rain Days (L5)	0.399	0.493	0.616	**-50.7**
Cool Days (L2)	0.434	0.747	1.137	**-25.3**
Warm Days (L5)	1.331	1.893	2.508	**89.3**

Table 17 shows the posterior medians and 95% credible intervals for relative risk of DENF in Cali between January 2015 and December 2016. The results suggest a 7.6% increase in DENF risk for PC1; a 3.9% increase for PC2 (negative relationship [decreased risk] in Model 1); a 19.8% increase for proximity to plant nurseries; a 11.8% increase for proximity to tire shops; a -16% decrease for population density; a 12.9% decrease for proximity to rivers/ravines; a 30% increase for tree density; a 25.2% decrease for average temperature; a 125.1% increase for days with maximum temp. > 32 °C; an 85% increase for relative humidity range; a 91.2% decrease for total rainfall; a 50.7% decrease for total rain days; a 25.3% decrease for cool days; and a 89.3% increase for warm days.

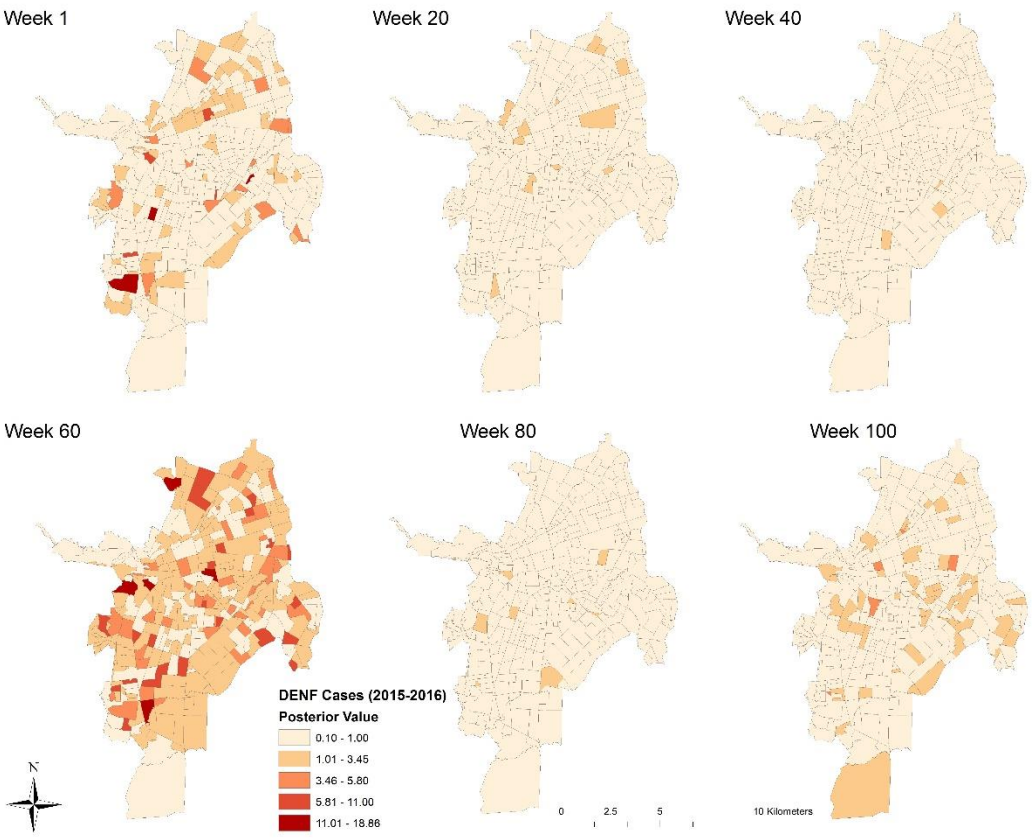


Figure 14: Temporal cross-sections of model 2 posterior values for each neighborhood of DENF Cases in Cali.

The negative to positive relationship between DENF and PC2 observed when comparing Models 1 and 2 is difficult to interpret. This could be due to the lagged weather variables affecting the posterior estimates. The magnitude of PC2 (low RR) is much lower than the other predictor variables, therefore, I can hypothesize that people that spend more time at home (PC2) may or may not be more susceptible to DENF and further investigation is required.

Figure 14 provides maps of the temporal cross-sections of Model 2 posterior values for each neighborhood of DENF cases in Cali between 2015 and 2016 (Weeks 1, 20, 40, 60, 80, and 100). When comparing the six temporal cross-sections, week 60 (late February 2016) experienced the highest DENF risk (predicted posterior mean values) after accounting for the 14 covariates (including the lagged weather variables). Interestingly, some of the highest risk neighborhoods for DENF during week 60 (for example) are classified as high strata (5 or 6). These high strata neighborhoods are adjacent to low strata neighborhoods (1 or 2), which suggests that there is spatial-temporal interaction between them. In other words, there is evidence that high strata neighborhoods are at higher risk when surrounded by lower strata neighborhoods due to the covariates in the model. Appendix 16 provides population adjusted rates of DENF for temporal cross-section in Figure 14.

4.3.7 Model 3 – CHIK with no lags

Table 18 summarizes the results of Model 3. The 95% credible intervals do not contain a value of 0, indicating that all of the covariates exhibit statistically significant relationships with CHIK risk at the neighborhood level in Cali between January 2015 and November 2016. The spatial autocorrelation value of 0.99 indicates that there is very

strong spatial dependence of the data after adjusting for covariate effects. The temporal autocorrelation value of 0.93 suggests that there is very strong temporal dependence of the data after adjusting for covariate effects. Therefore, CHIK risk in Cali at the neighborhood level is influenced by CHIK rates and covariates in surrounding neighborhoods and time periods (weeks).

Table 18: Model 3 Results

DIC: 17,001.25	2.50%	Median	97.50%
Intercept	-1.1347	-0.3113	0.5116
PC1	-0.0498	0.0171	0.0842
PC2	0.0132	0.1054	0.1955
Plant Nursery	-0.0132	-0.061	0.151
Tire Shops	0.0382	0.3257	0.6117
Pop Dens (km2)	-0.2969	0.0388	0.3775
River/Ravine	-0.1966	0.065	0.322
Tree Density	-0.8125	-0.3986	0.0125
Avg Temp	-0.8697	0.2283	1.3227
Days Temp Max	-1.2522	-0.2425	0.7938
Rel Humid Range	-0.7962	-0.1744	0.4913
Rain Total	-1.3439	-0.4354	0.6048
Rain Days	-0.3772	0.2286	0.8125
Cool Days	-1.5331	-0.7108	0.2846
Warm Days	-0.7404	0.0719	0.9503
Spatial Autocorrelation	0.9943	**0.9965**	0.9978
Temporal Autocorrelation	0.9049	**0.9287**	0.9454

Table 19: Model 3 Results – relative risk of CHIK for each covariate

Variable	RR (0.025)	RR (0.5)	RR(0.975)	RR %
PC1	0.951	1.017	1.088	**1.7**
PC2	1.013	1.111	1.216	**11.1**
Plant Nursery	0.754	0.941	1.163	**-5.9**
Tire Shops	1.039	1.385	1.844	**38.5**
Pop Dens (km2)	0.743	1.04	1.459	**4**
River/Ravine	0.822	1.067	1.38	**6.7**
Tree Density	0.444	0.671	1.013	**-32.9**
Avg Temp	0.419	1.256	3.754	**25.6**
Days Temp Max	0.286	0.785	2.212	**-21.5**
Rel Humid Range	0.451	0.84	1.634	**-16**
Rain Total	0.261	0.647	1.831	**-35.3**
Rain Days	0.686	1.257	2.254	**25.7**
Cool Days	0.216	0.491	1.329	**-50.9**
Warm Days	0.477	1.075	2.587	**7.5**

Table 19 shows the posterior medians and 95% credible intervals for relative risk of CHIK in Cali between January 2015 and November 2016. The results suggest a 1.7% increase in DENF risk for PC1; a 11.1% increase for PC2; a 5.9% decrease for proximity to plant nurseries; a 38.5% increase for proximity to tire shops; a 4% decrease for population density; a 6.7% decrease for proximity to rivers/ravines; a 32.9% decrease for tree density; a 25.6% increase for average temperature; a 21.5% decrease for days with maximum temp. > 32 °C; a 16% decrease for relative humidity range; a 35.3% decrease

for total rainfall; a 25.7% increase for total rain days; a 50.9% decrease for cool days; and a 7.5% increase for warm days.

4.3.8 Model 4 – CHIK with lagged weather variables

Table 20 summarizes the results of Model 4. The 95% credible intervals do not contain a value of 0, indicating that all of the covariates exhibit statistically significant relationships with CHIK risk at the neighborhood level in Cali between January 2015 and November 2016. The spatial autocorrelation value of 0.99 indicates that there is very strong spatial dependence of the data after adjusting for covariate effects. The temporal autocorrelation value of 0.93 suggests very strong presence of temporal dependence of the data after adjusting for covariate effects. Therefore, CHIK risk in Cali at the neighborhood level is influenced by CHIK rates and covariates in surrounding neighborhoods and time periods (weeks). The DIC (17,009.32) is virtually the same as Model 3 (CHIK with no lags – DIC = 17,001.25). Comparing the confidence intervals, Model 3 already displayed "tight" CIs for most of the covariates; whereas Model 4 with the lagged weather variables either shifted the CIs more positive or negative with minimal shrinking. This suggests that the lags are beneficial, but are more valuable for predicting DENF outbreaks in Cali than CHIK (likely due to the co-occurrence of outbreaks).

Table 20: Model 4 Results

DIC: 17,009.32	2.50%	Median	97.50%
Intercept	-0.6661	0.1475	1.0507
PC1	-0.0498	0.0159	0.0831
PC2	0.0128	0.1022	0.1944
Plant Nursery	-0.2808	-0.0641	0.1526
Tire Shops	0.0349	0.3231	0.6023
Pop Dens (km2)	-0.3047	0.0419	0.3907
River/Ravine	-0.1959	0.0687	0.3306
Tree Density	-0.8103	-0.3957	0.003
Avg Temp (L6)	-1.7351	-0.5251	0.5561
Days Temp Max (L8)	-1.527	-0.6548	0.103
Rel Humid Range (L5)	-1.2451	-0.5716	0.1836
Rain Total (L5)	-1.5518	-0.6055	0.3071
Rain Days (L6)	-0.7366	0.0471	0.7073
Cool Days (L1)	-0.0321	0.6058	1.3367
Warm Days (L6)	-0.228	0.7587	1.5168
Spatial Autocorrelation	0.9939	**0.9964**	0.9978
Temporal Autocorrelation	0.9047	**0.9304**	0.9504

Table 21: Model 4 Results – relative risk of CHIK for each covariate

Variable	RR (0.025)	RR (0.5)	RR(0.975)	RR %
PC1	0.951	1.016	1.087	**1.6**
PC2	1.013	1.108	1.215	**10.8**
Plant Nursery	0.755	0.938	1.165	**-6.2**
Tire Shops	1.036	1.381	1.826	**38.1**
Pop Dens (km2)	0.737	1.043	1.478	**4.3**
River/Ravine	0.822	1.071	1.392	**7.1**
Tree Density	0.445	0.673	1.003	**-32.7**
Avg Temp (L6)	0.176	0.592	1.744	**-40.8**
Days Temp Max (L8)	0.217	0.52	1.108	**-48**
Rel Humid Max (L5)	0.288	0.565	1.202	**-43.5**
Rain Total (L5)	0.212	0.546	1.359	**-45.4**
Rain Days (L6)	0.479	1.048	2.029	**4.8**
Cool Days (L1)	0.968	1.833	3.807	**83.3**
Warm Days (L6)	0.796	2.135	4.558	**113.5**

Table 21 shows the posterior medians and 95% credible intervals for relative risk of CHIK in Cali between January 2015 and November 2016. The results suggest a 1.6% increase in DENF risk for PC1; a 10.8% increase for PC2; a 6.2% decrease for proximity to plant nurseries; a 38.1% increase for proximity to tire shops; a 4.3% decrease for population density; a 7.1% increase for proximity to rivers/ravines; a 32.7% decrease for tree density; a 40.8% decrease for average temperature; a 48% decrease for days with maximum temp. > 32 °C; a 43.5% decrease for relative humidity range; a 45.4% decrease for total rainfall; a 4.8% increase for total rain days; a 83.3% increase for cool days; and a 113.5% increase for warm days.

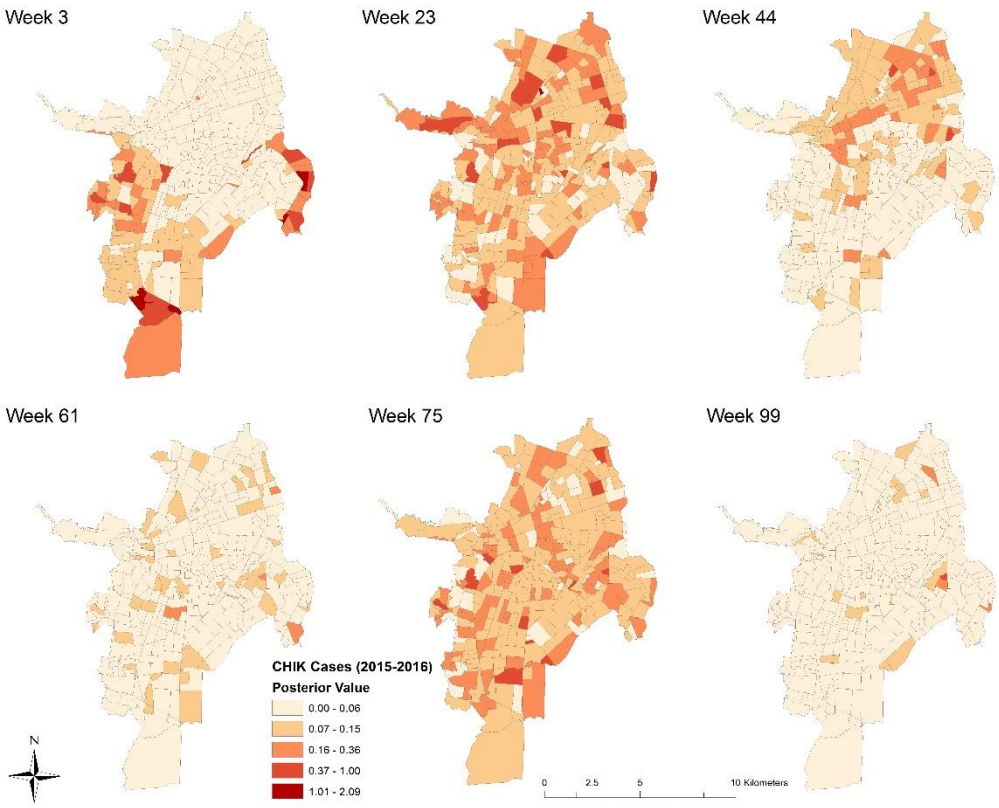


Figure 15: Temporal cross-sections of model 4 posterior values for each neighborhood of CHIK Cases in Cali.

Figure 15 provides maps of the temporal cross-sections of Model 4 posterior values for each neighborhood of CHIK cases in Cali between 2015 and 2016 (Weeks 3, 23, 44, 61, 75, and 99). When comparing the six temporal cross-sections, weeks 3 and 23 (early December 2015 and late May 2015) experienced the number of CHIK cases (predicted posterior mean values) after accounting for the 14 covariates (including the lagged weather variables). The highest risk neighborhoods for the cross-sections were predominately in low and middle strata neighborhoods in the west, east, and southern parts of Cali. The high strata neighborhoods with a high posterior value (high cases) are adjacent to low- or middle-income neighborhoods, suggesting that the risk is influenced by surrounding locations. Appendix 17 provides population adjusted rates of CHIK for temporal cross-section in Figure 15.

4.3.9 Model 5 – Zika with no lags

Table 22 summarizes the results of Model 5. The 95% credible intervals do not contain a value of 0, indicating that all of the covariates exhibit statistically significant relationships with Zika risk at the neighborhood level in Cali between January 2016 and December 2016. The spatial autocorrelation value of 0.96 indicates that there is very strong spatial dependence of the data after adjusting for covariate effects. The temporal autocorrelation value of 0.76 suggests that there is strong temporal dependence of the data after adjusting for covariate effects. Therefore, Zika risk in Cali at the neighborhood level is influenced by Zika rates and covariates in surrounding neighborhoods and time periods (weeks).

Table 22: Model 5 Results

DIC: 23,360.51	2.50%	Median	97.50%
Intercept	-2.583	-2.0151	-1.5298
PC1	0.0813	0.1369	0.1955
PC2	-0.3595	-0.2931	-0.225
Plant Nursery	0.623	0.8281	1.0353
Tire Shops	0.1719	0.4231	0.6688
Pop Dens (km2)	-1.687	-1.4065	-1.1237
River/Ravine	-0.3309	-0.1208	0.0791
Tree Density	0.4899	0.8671	1.2417
Avg Temp	-0.4081	0.1814	0.743
Days Temp Max	-0.2054	0.2422	0.6673
Rel Humid Range	-0.9696	-0.5983	-0.3179
Rain Total	-1	-0.5969	-0.1883
Rain Days	0.5037	0.7834	1.0769
Cool Days	-0.7366	-0.264	0.4405
Warm Days	-1.5433	-1.0104	-0.2438
Spatial Autocorrelation	0.9467	**0.9637**	0.9766
Temporal Autocorrelation	0.733	**0.7582**	0.7842

Table 23: Model 5 Results – relative risk of Zika for each covariate

Variable	RR (0.025)	RR (0.5)	RR(0.975)	RR %
PC1	1.085	1.147	1.216	**14.7**
PC2	0.698	0.746	0.799	**-25.4**
Plant Nursery	1.865	2.289	2.816	**128.9**
Tire Shops	1.188	1.527	1.952	**52.7**
Pop Dens (km2)	0.185	0.245	0.325	**-75.5**
River/Ravine	0.718	0.886	1.082	**-11.4**
Tree Density	1.632	2.38	3.462	**138**
Avg Temp	0.665	1.199	2.102	**19.9**
Days Temp Max	0.814	1.274	1.949	**27.4**
Rel Humid Range	0.379	0.55	0.728	**-45**
Rain Total	0.368	0.551	0.828	**-44.9**
Rain Days	1.655	2.189	2.936	**118.9**
Cool Days	0.479	0.768	1.553	**-23.2**
Warm Days	0.214	0.364	0.784	**-63.6**

Table 23 shows the posterior medians and 95% credible intervals for relative risk of Zika in Cali between January 2016 and December 2016. The results suggest a 14.7% increase in Zika risk for PC1; a 25.4% decrease for PC2; a 128.9% increase for proximity to plant nurseries; a 52.7% increase for proximity to tire shops; a 75.5% decrease for population density; a 11.4% decrease for proximity to rivers/ravines; a 138% increase for tree density; a 19.9% increase for average temperature; a 27.4% increase for days with maximum temp. > 32 °C; a 45% decrease for relative humidity range; a 44.9% decrease for total rainfall; a 118.9% increase for total rain days; a 23.2% decrease for cool days; and a 63.6% increase for warm days.

4.3.10 Model 6 – Zika with lagged weather variables

Table 24 summarizes the results of Model 6. The 95% credible intervals do not contain a value of 0, indicating that all of the covariates exhibit statistically significant relationships with Zika risk at the neighborhood level in Cali between January 2016 and December 2016. The spatial autocorrelation value of 0.96 indicates that there is very strong spatial dependence of the data after adjusting for covariate effects. The temporal autocorrelation value of 0.76 suggests a strong presence of temporal dependence of the data after adjusting for covariate effects. Therefore, Zika risk in Cali at the neighborhood level is influenced by Zika rates and covariates in surrounding neighborhoods and time periods (weeks). The DIC (23,371.51) is virtually the same as Model 5 (Zika with no lags – DIC = 23,360.51). Comparing the confidence intervals, Model 5 already displayed "tight" CIs for most of the covariates; whereas Model 6 with the lagged weather variables either shifted the CIs more positive or negative with minimal shrinking. Similar to Model 4 (lagged CHIK), this also suggests that the lags are beneficial, but are more valuable for predicting DENF outbreaks in Cali than Zika (likely due to the co-occurrence of outbreaks).

Table 24: Model 6 Results

DIC: 23,371.51	2.50%	Median	97.50%
Intercept	-2.3297	-1.9437	-1.5579
PC1	0.0736	0.1339	0.1987
PC2	-0.4073	-0.3427	-0.2772
Plant Nursery	0.4467	0.6681	0.8841
Tire Shops	0.0938	0.353	0.6139
Pop Dens (km2)	-1.5873	-1.3043	-1.029
River/Ravine	-0.2207	-0.0233	0.1833
Tree Density	0.5991	0.9761	1.3599
Avg Temp (L8)	-0.7613	-0.4615	-0.1825
Days Temp Max (L3)	-0.2413	0.2053	0.6557
Rel Humid Range (L7)	0.1297	0.4259	0.7139
Rain Total (L5)	0.5616	0.9567	1.5309
Rain Days (L5)	-0.3435	0.026	0.4304
Cool Days (L3)	-0.5225	0.2538	0.9768
Warm Days (L3)	-0.8464	-0.5494	-0.2569
Spatial Autocorrelation	0.9443	**0.9621**	0.9756
Temporal Autocorrelation	0.7331	**0.7609**	0.7873

Table 25: Model 6 Results – relative risk of Zika for each covariate

Variable	RR (0.025)	RR (0.5)	RR(0.975)	RR %
PC1	1.076	1.143	1.22	**14.3**
PC2	0.665	0.71	0.758	**-29**
Plant Nursery	1.563	1.951	2.421	**95.1**
Tire Shops	1.098	1.423	1.848	**42.3**
Pop Dens (km2)	0.204	0.271	0.357	**-72.9**
River/Ravine	0.802	0.977	1.201	**-2.3**
Tree Density	1.82	2.654	3.896	**165.4**
Avg Temp (L8)	0.467	0.63	0.833	**-37**
Days Temp Max (L3)	0.786	1.228	1.927	**22.8**
Rel Humid Max (L7)	1.138	1.531	2.042	**53.1**
Rain Total (L5)	1.753	2.603	4.622	**160.3**
Rain Days (L5)	0.709	1.026	1.538	**2.6**
Cool Days (L3)	0.593	1.289	2.656	**28.9**
Warm Days (L3)	0.429	0.577	0.773	**-42.3**

Table 25 shows the posterior medians and 95% credible intervals for relative risk of Zika in Cali between January 2016 and December 2016. The results suggest a 14.3% increase in Zika risk for PC1; a 29% decrease for PC2; a 95.1% increase for proximity to plant nurseries; a 42.3% increase for proximity to tire shops; a 72.9% decrease for population density; a 2.3% decrease for proximity to rivers/ravines; a 165.4% increase for tree density; a 37% decrease for average temperature; a 22.8% increase for days with maximum temp. > 32 °C; a 53.1% decrease for relative humidity range; a 160.3% increase for total rainfall; a 2.6% increase for total rain days; a 28.9% increase for cool days; and a 42.3% decrease for warm days.

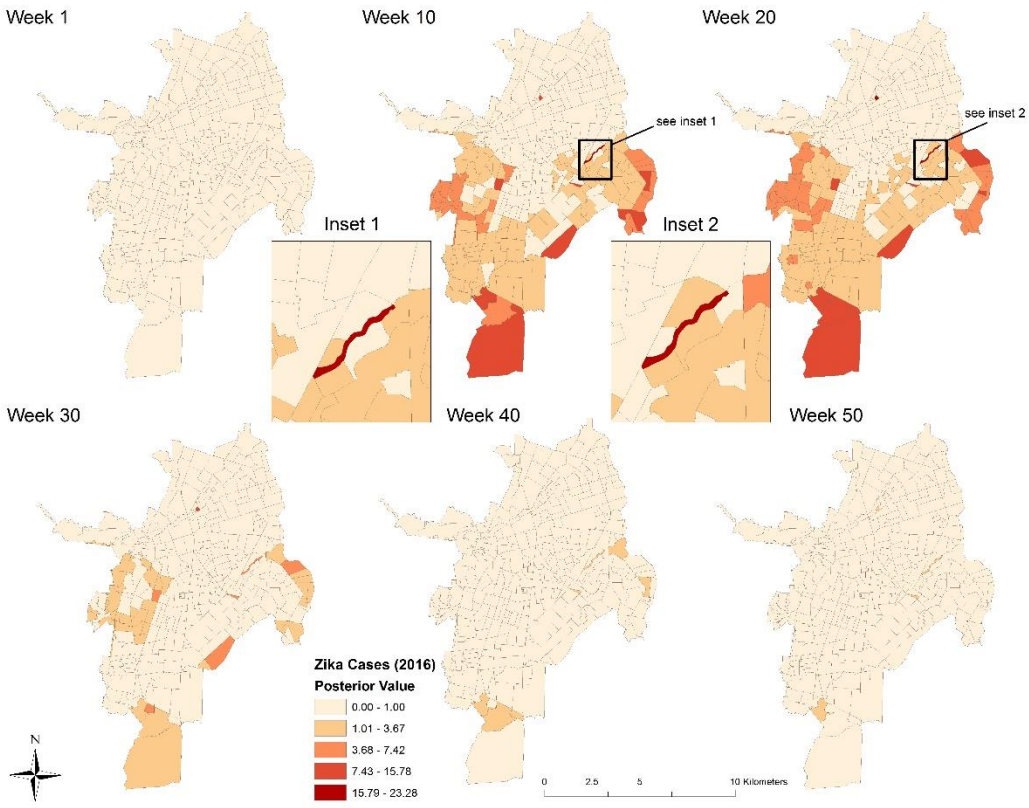


Figure 16: Temporal cross-sections of model 6 posterior values for each neighborhood of Zika Cases in Cali.

Figure 16 provides maps of the temporal cross-sections of Model 6 posterior values for each neighborhood of Zika cases in Cali (Weeks 1, 10, 20, 30, 40, and 50 in 2016). When comparing the six temporal cross-sections, weeks 10 and 20 (mid-March 2016 and early May 2016) experienced the highest number of Zika cases (predicted posterior mean values) after accounting for the 14 covariates (including the lagged weather variables). The highest risk neighborhoods for the cross-sections were predominately in low and middle strata neighborhoods in the west, east, and southern parts of Cali. However, there is evidence of high number of cases of Zika in high strata neighborhoods, especially in the southernmost part of the City. This finding is somewhat similar to DENF and CHIK, where spatial interaction between adjacent neighborhoods affects risk; however, there were a high number of reported Zika cases and rates in the high strata neighborhoods in the southernmost portion, partially explaining the high posterior values in that region. Finally, the elongated neighborhood in insets 1-2 in Figure 16 is Lleras - Cinta Larga (cases were reported there, but no population data is available as seen in the rate maps in Appendix 18). Appendix 18 provides population adjusted rates of Zika for temporal cross-section in Figure 16.

4.4 Discussion

This study is the first of its kind to model space-time risk at the neighborhood-level of three co-occurring VBDs at the weekly level across two years of disease surveillance data. This study is also the first that incorporates temporally lagged weather variables in a ST-CAR approach that can predict outbreaks in advance. Coupling the lagged weather variables with the spatial covariates of DENF, CHIK, and Zika risk, the models can serve as early warning systems with neighborhood-level effects explaining

where and why certain locations are more at-risk than others. There are many key findings that warrant further investigation and explanation.

First, there is very strong evidence that there is both spatial and temporal interaction between disease risk and the significant covariates for adjacent neighborhoods in Cali. Although DENF had a much lower temporal autocorrelation (value of 0.12 for both Models 1 and 2), this can be explained by the distribution of cases between 2015 and 2016 – there were three distinct peaks, but DENF remains a persistent threat due to the four serotypes of the virus. The very strong temporal autocorrelation values for both CHIK and Zika are due to the massive spike in cases during 2015 and 2016, followed by very sharp declines after the epidemics subsided. Both CHIK and Zika were still new to Cali during that time, and lifelong immunity to both viruses after infection can result in herd immunity, which is partially responsible for the decrease in overall cases of CHIK in 2016. The very strong spatial autocorrelation values (although extremely high) for all three VBDs suggest that DENF, CHIK, and Zika outbreaks in adjacent neighborhoods are strongly related (i.e. living next to a neighborhood with high cases will strongly influence disease risk and cases in your neighborhood of residence).

When examining the results of the three socioeconomic covariates (PC1, PC2, and population density), the increased risk of DENF, CHIK, and Zika for PC1 is an interesting finding. This corroborates with Delmelle et al. (2016), whom suggest that in the southern part of Cali, houses are typically bigger with relatively larger yards, which may provide a suitable habitat *Aedes* to breed. Neighborhoods with a high proportion of people in the PC1 category (e.g. employed, older, more educated) are also typically adjacent to middle or lower strata neighborhoods (with the exception of the extreme

South), which also may increase the risk of disease due to the very strong evidence space-time interaction effect between the locations as suggested by the models. An increased risk of DENF and CHIK was reported for neighborhoods with higher proportion of individuals in the PC2 category (i.e. work from home and students). This finding corroborates with strong evidence that *Aedes* proliferates in and around homes (Baldacchino et al. 2015; Lindsay et al. 2017; Wilson and Sevarkodiyone 2017); and it has also been found that cases can substantially decline if *Aedes* trap interventions are put in place in at-risk communities (Lorenzi 2016). The negative relationship between Zika and PC2 is likely due to the spatial and temporal distribution of cases and Zika.

Another unexpected result was the negative relationship between population density and DENF and Zika risk. One explanation can be that some of the densely populated neighborhoods in the eastern part of Cali have a high concentration of Afro-Colombian population, which has been suggested to be less susceptible to the viruses (Chacón-Duque et al. 2014). Furthermore, there is evidence that shows that low density areas with poor infrastructure may have increased *Aedes* presence (Maciel-de-Freitas and Lourenço-de-Oliveira 2009); and DENF's complex immunology and the herd immunity resulting after infection from the viruses (Schmidt et al. 2011) may have contributed to these patterns. Amongst other factors, high density of populations may not necessarily be a main risk factor of VBD transmission (Feldstein et al. 2015). These findings regarding population density dispute evidence found in other studies described in section 2.4.1. While increases in population and urbanization will undoubtedly increase risk of VBD transmission, the true affect of population density may vary at fine spatial levels (e.g. neighborhoods). My findings suggest that the infrastructure and availability of breeding

habitats for *Aedes* within densely populated areas influence risk more than close proximity of hosts alone.

Closer proximity to plant nurseries, tire shops, and higher tree density all exhibited an increased risk of DENF and Zika. Plant nurseries and tire shops are common breeding grounds of *Aedes*, thus neighborhoods within close proximity are generally at a higher risk of disease transmission (Delmelle et al. 2016). Tree density may also be a significant risk factor of *Aedes* presence since studies have shown that high tree shade density stimulates breeding; and tree holes are suitable water containers where *Aedes* have been found in abundance (Lian et al. 2006; Mangudo et al. 2015). Although CHIK exhibited a negative relationship with tree density and plant nurseries, this is likely due to the spatial distribution of the cases at the neighborhood level. CHIK cases were more concentrated in middle strata neighborhoods, away from plant nurseries and green areas with a higher tree density.

Closer proximity to rivers and ravines (i.e. any moving bodies of water) resulted in significantly lower risk of DENF and Zika; and there was a higher risk for CHIK. *Aedes* require stagnant water as a breeding ground, therefore, the flowing water of a river or ravine would prove to be an unsuitable habitat for the mosquitoes. Although flooding events during the rainy season could create stagnant water sources surrounding the rivers, further research using remote sensing techniques (such as flow analysis) could provide insight to areas prone to stagnant water. The positive relationship between CHIK and water bodies may also be due to the distribution of the cases, since the two main rivers form the eastern and western boundaries of Cali's core urban area.

In general, adding significantly lagged weather variables shrunk the confidence intervals of the models' coefficients and relative risk estimates. Although the significant weekly lags chosen were different for each weather covariate and for each disease, they were all within the range of the *Aedes* life cycle and the differences are due to the weekly variations in observed cases. The first notable finding is that for each lagged weather covariate, the relationship with disease risk was either positive or negative (with the exception of average temperature). This highly suggests that DENF, CHIK, and Zika forecasting is dependent on a complex combination of lagged weather variables, where the direction of the association (increased or decreased risk) corresponds to certain weather conditions before an outbreak occurred. El Niño (ENO) also occurred between 2015 and 2016 (Null 2019); and there is strong evidence that ENO years correspond with major epidemics of DENF, CHIK, and Zika (Eastin et al. 2014; Vincenti-Gonzalez 2018). Table 26 facilitates the interpretation of the lagged weather variables, which correspond to particular stages of the *Aedes* life cycle.

4.4.1 DENF & Weather

The variables at a 5-week lag in Model 2 is most likely corresponding to larval development: Tavg, DTRMax, RainD, and WarmD. Despite Tavg having an unexpected negative relationship with DENF, its overall predictive magnitude is relatively small compared to the other three. DTRMax and WarmD exhibit expected positive relationships with DENF; and RainD exhibits an expected negative relationship. At 5 weeks before a DENF outbreak, the weather is often characterized by multiple days with short-lived rain showers (i.e., RainD is above average, its correlation with DENF is negative, and the regression coefficient is negative); and each day experiences sufficient

sunshine to allow the daily maximum temperature to exceed 32°C (i.e. WarmD is above average, its correlation with DENF is positive, and the regression coefficient is positive), but then the short-lived rain showers also induce evaporational cooling that significantly decreases the daily minimum temperature (i.e. DTRMax is above average, its correlation with DENF is positive, and the regression coefficient is positive) and leads to a slightly cooler, but above-average daily mean temperature (i.e. the Tavg remain above average, its correlation with DENF remains positive, but the regression coefficient is slightly negative and the relative risk is small). Overall, the regular rainfall combined with average to above-average temperatures produce numerous stagnant pools within a favorable thermal environment for prolific larval development.

Next, two variables contain a 3 week lag (Train and RHRng), which is mostly likely related to the gonotrophic cycle. Both Train and RHRng exhibit relationships to DENF that are consistent with expectations (negative and positive, respectively), and both show similar predictive importance based on their relative risk estimations. The weather 3 weeks before a DENF outbreak is as follows: the week is often characterized by relatively below-average rainfall (i.e. Train is below average, its correlation with DENF is negative, and the regression coefficient is negative). Less rainfall coincides with clear skies, which allows the relative humidity to fluctuate between small daytime values and large nighttime values (i.e. RHRng is above average, its correlation with DENF is positive, and the regression coefficient is positive). Overall, the relatively dry conditions (clear skies and below-average rainfall) maximize solar heating, minimize evaporational cooling, and allow the warm temperatures most favorable for *Aedes* feeding to occur.

One variable exhibited a strong correlation at a 2-week lag (CoolD), which most likely is related to the extrinsic incubation period and gonotrophic cycle. CoolD exhibits a negative relationship to DENF that is consistent with expectations, but its relative importance is small and roughly equivalent to Tavg. The most probable weather occurring 2 weeks before a DENF outbreak is as follows: the week is characterized by above average temperatures (i.e. CoolD is below average, its correlation with DENF is negative, and the regression coefficient is negative). Overall, warmer temperatures will accelerate both viral replication within *Aedes* and *Aedes* feeding, which increases viral transmission to humans.

4.4.2 CHIK & Weather

Eight weeks before a CHIK outbreak, there were multiple days of below average temperature (i.e. DTRMax correlation with CHIK is negative, regression coefficient is negative, and relative risk is negative). Next, three variables contain a 6-week lag: Tavg, RainD, and WarmD. At six weeks before a CHIK outbreak, the weather is often as follows: multiple days with short-lived rain showers (i.e., RainD is above average, its correlation with CHIK is negative, and the regression coefficient is positive, and the relative risk is positive); and each day experiences sufficient sunshine to allow the daily maximum temperature to exceed 32°C (i.e. WarmD is above average, its correlation with CHIK is positive, and the regression coefficient and relative risk are positive). Despite the negative relative risk and coefficient of Tavg, the predictive magnitude is much higher for WarmD. Therefore, the short-lived rain showers and numerous days of warm weather promote and accelerate larval development and viral development in mature mosquitoes at 6 weeks before a CHIK outbreak. In other words, the cooler temperatures

8 weeks in advance would delay larval development; followed by increasing larval and mosquito density in subsequent weeks (6-week lag) due to the more favorable conditions.

Two variables contain a 5-week lag: RHRng and Train. The weather 5 weeks before a CHIK outbreak is characterized as follows: multiple days with a small relative humidity range (i.e. correlation with CHIK is negative, regression coefficient and relative risk are negative) and dryer conditions due to the lack of measureable rainfall (i.e. Train correlation with CHIK is negative, regression coefficient and relative risk are negative). The low fluctuating relative humidity and low rainfall totals can result in fewer active hours for *Aedes* to feed. Finally, CoolD contained a one-week lag; furthermore, one week before a CHIK outbreak is characterized by below average temperatures (i.e. positive correlation, negative coefficient and relative risk). This was an unexpected result and could be an artifact of the modeling process. Moreover, a one week lag often corresponds to the instrinsic incubation period when weather conditions are not believed to play a role. As such, we can disregard this result and place more emphasis on the six other lagged variables, which correspond to larval development, extrinsic incubation period, and gonotrophic cycle.

4.4.3 Zika & Weather

Four variables are related to larval development: Tavg lagged at 8 weeks, and RHRng at 7 weeks, and Train and RainD at 5 weeks. Despite Tavg and RHRng having an unexpected negative and positive relationship with Zika, respectively; their overall predictive magnitude is smaller than Train. Train and RainD both exhibited a positive relationship with Zika. At 8 weeks before a Zika outbreak, the weather is often characterized by multiple days with below average daily mean temperature (i.e. the Tavg

remains below average, its correlation with Zika is negative, and the regression coefficient is negative and the relative risk is moderate). At 7 weeks before, the below average temperatures lead to multiple days of relative humidity fluctuating between small daytime values and large nighttime values (i.e. RHRng is above average, its correlation with Zika is positive, the regression coefficient is positive, and the relative risk is high). At 5 weeks before, there are multiple days of heavy rainfall (i.e. Train and RainD are above average, their correlation with Zika are positive, the regression coefficients are positive, and the relative risk is very high for total rainfall). The unexpected relationships for Tavg and RHRng at 7 weeks may be indicative of a delay of larval development – until more optimal conditions with sufficient pools of stagnant water arrive (in week 5) to catalyze the larval development process and leading to a greater larval densities. Overall, a few weeks of cooler weather and a large range of relative humidity was followed by a week a heavy rain, which produced numerous stagnant pools of water for *Aedes* larval development.

Three variables are related to both the gonotrophic cycle and extrinsic incubation period: DTRMax, CoolD, and WarmD at 3-week lags. The weather 3 weeks before a Zika outbreak is as follows: the week is characterized by cooler, but still above average temperatures (i.e. CoolD is above average, its correlation with Zika is negative, regression coefficient is negative, and relative risk is positive; and WarmD is below average, its correlation with Zika is negative, regression coefficient is negative, and relative risk is negative). Finally, DTRMax is above average, its correlation with Zika is negative, the regression coefficient is positive, and the relative risk is positive. Overall, slightly cooler, but multiple days of above average temperature favor and accelerate both

viral replication within *Aedes* and *Aedes* feeding, which increases viral transmission to humans.

Table 26: Temporally-lagged weather variales and correlation with *Aedes* life cycle

Developmental Stage	Total Days **	Total Weeks **	LAG Range ##	Variable	Expected Relation	Rationale
Larval/Pupa Development	10-21	1-3	3-8	RAINT	Positive	More total rain produces more stagnant pools and promotes larval development
(vector grows in water)				RAIND	Positive	More frequent rain produces more stagnant pools and promotes larval development
				TAVG	Positive	Warmer temperatures promote larval development
				DTR	Negative	Smaller daily temperature ranges imply fewer hours with cold temperatures (and greater larval survivial)
				DRH	Negative	Smaller ranges in relative humidity imply fewer dry days (when stagnant pools would more quickly evaporate)
				WARMD	Positive	Higher larval densities in warmer weather
				COOLD	Negative	Fewer cold days would promote more larval survival
Gonotrophic Cycle	3-7	0-1	2-5	RAINT	Negative	Rainfall tends to coincide with cooler temperatures -- less vector feeding (viral transmission)
(vectors feed on humans)				RAIND	Negative	Rainfall tends to coincide with cooler temperatures -- less vector feeding (viral transmission)
				TAVG	Positive	Vector feeding (viral transmission) more frequent at warmer temperatures
				DTR	Negative	Vector feeding (viral transmission) more frequent at warmer temperatures -- smaller DTR implies fewer cold hours
				DRH	Positive	Vector feeding (viral transmission) more frequent during dry periods -- larger DRH implies more dry days
				WARMD	Positive	Vector feeding (viral transmission) more frequent at warmer temperatures
				COOLD	Negative	Vector feeding (viral transmission) more frequent at warmer temperatures -- fewer cold hours
Extrinsic Incubation	7-15	1-2	1-4	RAINT	Unknown	No direct link between rainfall and extrinsic incubation
(virus matures in vector)				RAIND	Unknown	No direct link between rainfall and extrinsic incubation
				TAVG	Positive	Viral development more rapid at warmer temperatures
				DTR	Negative	Viral development more rapid at warmer temperatures -- smaller DTR implies fewer cold hours
				DRH	Positive	Decreased incubation period - rapid replication of virus
				WARMD	Positive	Viral development more rapid at warmer temperatures
				COOLD	Negative	Viral development more rapid at warmer temperatures -- fewer cold hours
Total	20-43	3-8	0-8			

**These ranges are based on literature cited in Eastin et al. (2014).

These ranges in lag (since the DENF virus was reported – or lab confirmed) correspond to the typical vector-DENF transmission cycle outlined above (including multiple vector feedings)

4.4.4 Limitations & Future Research

Despite the strengths and contributions of this research, there are notable limitations and areas of future work that is worth discussing. First, the underreporting of cases and unmatched addresses during the geocoding process likely undermines the true burden of all three VBDs examined in this study. Second, further validation is needed to determine the uncertainty and predictive power of the lagged weather variables by comparing our findings to observed VBD cases during the study period. Third, the socioeconomic and demographic data was a mix from the Colombian National Census and 2010 population estimates. Colombia recently administered a new national census (the first since 2005), but is currently unavailable. Using 2005 and 2010 data for this study will bias the results, but the neighborhood classifications (strata) mostly remained unchanged. The uncertainty resulting from using outdated census data is a common limitation found in many studies in Latin America and developing countries. Fourth, including vector surveillance data (presence/absence) in each neighborhood would improve the accuracy of the relative risk estimates (Whiteman et al. 2019). Fifth, the spatial weight matrix only considered adjacent neighborhoods as "neighbors", we recognize that individual activity spaces expand far beyond locations nearby their home.

Future research can implement different spatial and temporal weight matrices for sensitivity analysis purposes. Sixth, the weather conditions and severity of outbreaks for all three diseases varied between 2015 and 2016, which may have affected the model results. Future work can disaggregate the years and run two separate models for further examination. Seventh, future research can examine the possibility of underestimation or overestimation (according to the models' residuals). Eighth, the space-time patterns of

the VBD outbreaks did not exhibit much seasonality, which is likely due to only using two years of data. Further work can utilize 5-10 years of VBD and weather data to detect potential seasonal patterns of the epidemics. Finally, the three diseases are transmitted by the same vector (*Aedes*), and the 6 univariate space-time models are compared as if the diseases occurred independently from each other. Future research can develop a multivariate space-time CAR model to examine which neighborhoods are at higher risk for one disease, two diseases, or all three concurrently. Lee et al. (2018) note that the multivariate space-time (MVST) approach in CAR modeling is still in its infancy stages, and the development of such models can more accurately examine and compare the co-occurrence of diseases transmitted by the same vector. After running Pearson's correlation tests in Stata, there is strong evidence that DENF, CHIK, and Zika cases per week exhibit strong and positive correlation: DENF and CHIK with a coefficient of 0.75 in 2015 ($p < 0.5$); DENF and CHIK with a coefficient of 0.65 ($p < 0.5$) in 2016; DENF and Zika with a coefficient of 0.73 ($p < 0.5$) in 2016; and CHIK and Zika with a coefficient of 0.88 ($p < 0.5$) in 2016. These significant results support the notion of MVST approaches in future work.

4.5 Conclusion

A space-time CAR modeling approach was utilized to examine significant socioeconomic, demographic, environmental, and meteorological predictors of DENF, CHIK, and Zika in Cali, Colombia during the 2015 and 2016. The temporally lagged weather covariates can significantly estimate when risk of transmission is highest, and the spatial covariates can help explain the differences in disease risk at the neighborhood-level. Adding weather and climate data to a space-time model can improve disease

surveillance, especially for VBDs that require specific conditions for transmission to occur. This study demonstrated that there was strong spatial and temporal dependence between adjacent neighborhoods and time periods, which provides strong evidence that disease transmission can be influenced by characteristics and phenomena occurring in surrounding locations. We also provide evidence that DENF, CHIK, and Zika are not just diseases of the poor; although risk factors may be higher in neighborhoods of lower socioeconomic status, we have shown that the transmission dynamics of all three VBDs are place- and temporally based. Despite this study being retrospective in nature, the modeling approach can be applied in a contemporary surveillance setting when significant outbreaks have not yet occurred, serving as an early warning system to help promote proactive community health, improve public health educational campaigns, targeted interventions. We hope that this research influences further small area space-time analysis, since we support the notion that disease prevention should start at the neighborhood and community level. Chapters 5 and 6 will capture the perspectives and behaviors of those susceptible to DENF, CHIK, and Zika in Cali and those responsible for VBD surveillance, education, treatment, and control.

CHAPTER 5: KNOWLEDGE, ATTITUDES, AND PRACTICES REGARDING DENGUE, CHIKUNGUNYA, AND ZIKA IN CALI, COLOMBIA.

Dengue fever (DENF), chikungunya (CHIK), and Zika are responsible for the majority of the burden caused by vector-borne diseases (VBDs); which are caused by viruses primarily transmitted by the Aedes mosquito. *Aedes* have become prolific in urban areas due to a combination of climate change, rapid urbanization, increased human mobility, and globalization, causing the three VBDs to emerge in novel regions. Community knowledge can provide detailed insights about the spatial heterogeneity of disease risk and rates within a particular region, improving public health interventions. Knowledge, Attitude, and Practice (KAP) surveys are used to shed light on at-risk communities' understanding of the vector, the pathogen, prevention and treatment strategies. Little is known how KAP varies across diseases, and across neighborhoods within a city. Understanding KAP variation among concurrent VBDs at a fine-level, especially differences between endemic and emerging diseases, may improve targeted interventions, education programs, and health policy. I administered KAP surveys to 327 individuals in healthcare centers and select neighborhoods in Cali, Colombia in June 2019. I utilized GLMs to identify significant predictors of KAP. The results can be leveraged to inform public health officials and communities to motivate at-risk neighborhoods to take an active role in vector surveillance and control, and improving educational and surveillance resources in Cali, Colombia.

5.1 Methodology

I developed a KAP survey that is influenced by similar questionnaires found in studies discussed in chapter 2 – literature review (e.g. Whiteman et al. 2018). Many of the questions were also influenced by feedback by faculty members of Universidad Libre

and Universidad Icesi in Cali, Colombia, who provided assistance during the fieldwork portion in June of 2019. My KAP survey can easily be modified to study other infectious diseases in any location across the globe. Informed verbal consent was obtained from each participant before the survey was administered by myself or a trained researcher. The survey and all of the components of chapter 3 are approved by the University of North Carolina at Charlotte's IRB board (UNC-Charlotte Case No. 18-0399). The full survey (both English and Spanish versions) is attached in the appendix 1 and appendix 2, respectively. The surveys were filled out by pen and paper by participants in selected healthcare centers, universities, and door-to-door in Cali.

The questionnaire took approximately 15 minutes to complete and is comprised of six sections: (1) sociodemographic information, (2) general questions, (3) knowledge, (4) attitudes, (5) practices 1, and (6) practices 2. Section 1 collected the participant's gender, age, address, neighborhood, education level, monthly income, occupation, civil status, number of children living at home, household size, and race. Section 2 contained questions on previous diagnoses and familiarity with DENF, CHIK, or Zika; additionally, we did ask respondents about their accessibility to healthcare services. Section 3 was organized around 12 multiple choice and fill-in questions about a participant's knowledge of DENF, CHIK, and Zika. Section 4 contained twelve statements and utilizes a Likert scale to capture a participant's fear of a disease, prevention responsibility (government or individual), belief of fumigation effectiveness, and necessity to seek treatment if infected with DENF, CHIK, or Zika. Section 5 contained five statements and questions regarding the potential risk that sexual intercourse may have on disease transmission, motivation to learn about preventative techniques, concern about mosquitoes in their neighborhood and

contacting authorities when they and/or larvae are observed, and how often their home is sprayed with insecticides or larvicides. Finally, section 6 asked participants: "how often do you practice the following measures to prevent mosquitoes in your home / neighborhood?" Nine preventative measures were listed, and a Likert scale was provided to capture how frequently each individual participates in each of the preventative measures.

5.2 Data Coding

Survey answers were coded and entered into an Excel database. For section 1 (sociodemographic information), answers were entered exactly as they are answered – e.g. participant 1 is male, has a postgraduate degree, is married, etc. The address and neighborhood of their home were used for geocoding purposes that allowed us to link their residence to a neighborhood stratum (Figure 17) and were recoded accordingly. For section 2 (general questions), a '1' was entered if a participant previously had DENF, CHIK, or Zika – each in three separate columns; a value between '0' and '3' was entered depending on the number of diseases they are familiar with; rating access to health services were recoded between '1' (poor) and '5' (excellent); average travel time to a health facility was entered categorically, from "less than 5 minutes" to " 1 hour or more"; and the questions about stopping and refusing medical treatment were given a '1' for "yes" and '0' for "no".

For section 3 (knowledge), a value between '0' and '3' was entered, depending on the number of VBDs participants were "familiar with"; a value between '0' and '3' was entered, depending on the number of VBDs the participant thinks they know how to prevent; with separate columns for each disease ('1' if they know how to prevent,

‘0’otherwise). Correct answers were given a value of ‘1’, that is – each disease are transmitted by mosquitoes; Zika can be transmitted via sexual contact; there are currently no vaccines available for the three VBDs; a baby is at risk of birth defects if a pregnant woman is infected with each disease; female *Aedes* bite in the morning and afternoon; and people are at a higher risk of contracting the three VBDs during the Colombian winter (wet) seasons. A value from 0-9 was entered depending on how many information sources were checked – which was completed for each disease. For section 4 (attitudes), the responses were coded from 1-6 (not sure to completely agree). For section 5 (practices 1), a value of ‘1’ was entered for “yes” responses and ‘0’ otherwise. The frequency of spraying was entered from ‘0’ (never/don’t know) to ‘5’ (daily). For section 6 (practices 2), the frequency of individuals participating in mosquito prevention measures were assessed using a range from ‘0’ (never) to ‘5’ (daily).

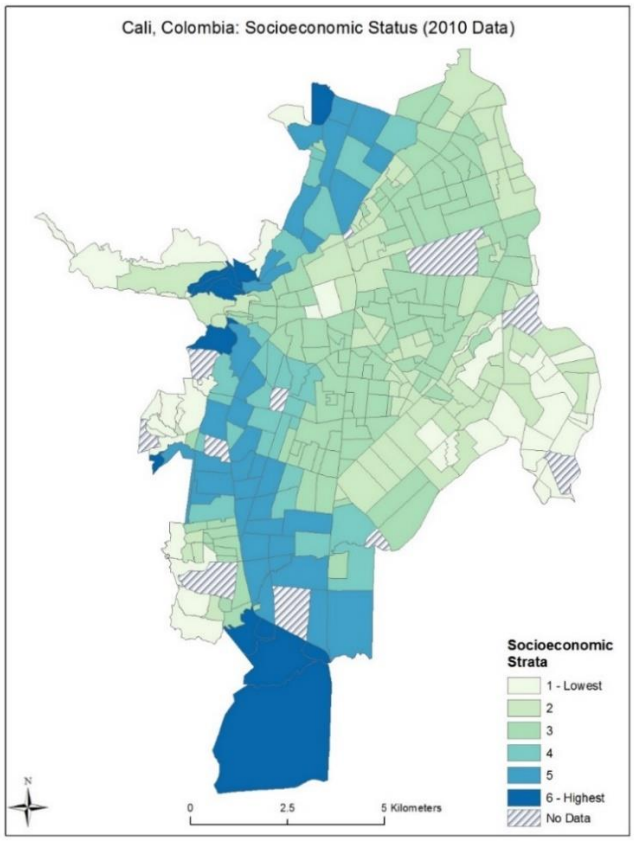


Figure 17: Socioeconomic strata for neighborhoods in Cali, Colombia

5.3 Sampling Design

Participants were primarily identified at healthcare centers in Cali. The surveys that were administered in healthcare centers belonged to each strata group (low, medium, high). The surveys were also administered in neighborhoods that belong to disease clusters AND non-clusters. The following six queries were used to identify healthcare centers where the surveys were administered (neighborhoods must have a healthcare center within boundary):

1. low-socioeconomic status neighborhoods (strata 1 and 2) that belong to a cluster of dengue, chikungunya, or Zika.
2. middle-socioeconomic status neighborhoods (strata 3 and 4) that belong to a cluster of dengue, chikungunya, or Zika.
3. high-socioeconomic status neighborhoods (strata 5 and 6) that belong to a cluster of dengue, chikungunya, or Zika.
4. low-socioeconomic status neighborhoods (strata 1 and 2) that do not belong to a cluster of dengue, chikungunya, or Zika.
5. middle-socioeconomic status neighborhoods (strata 3 and 4) that do not belong to a cluster of dengue, chikungunya, or Zika.
6. high-socioeconomic status neighborhoods (strata 5 and 6) that do not belong to a cluster of dengue, chikungunya, or Zika.

Disease clusters were determined by the local Moran's I autocorrelation statistic. Essentially, this statistic identifies statistically significant clusters of high and low values. Disease cases per neighborhood were used as the variable in the autocorrelation statistic. Clusters were computed for DENF, which has the highest number of cases between 2015 and 2016. Figure 18 (left) provides the results of DENF cases in Cali for 2015 and 2016; and the resulting clusters after computing local Moran's I. The local Moran's I statistic identified four types of clusters: high number of dengue cases surrounded by neighborhoods with a high number of dengue cases (high-high); high number of dengue cases surrounded by neighborhoods with a low number of dengue cases (high-low); low number of dengue cases surrounded by neighborhoods with a low number of dengue cases (low-low); and low number of dengue cases surrounded by a high number of

dengue cases (low-high). Figure 18 (right) also includes the location of health centers in Cali (n=494).

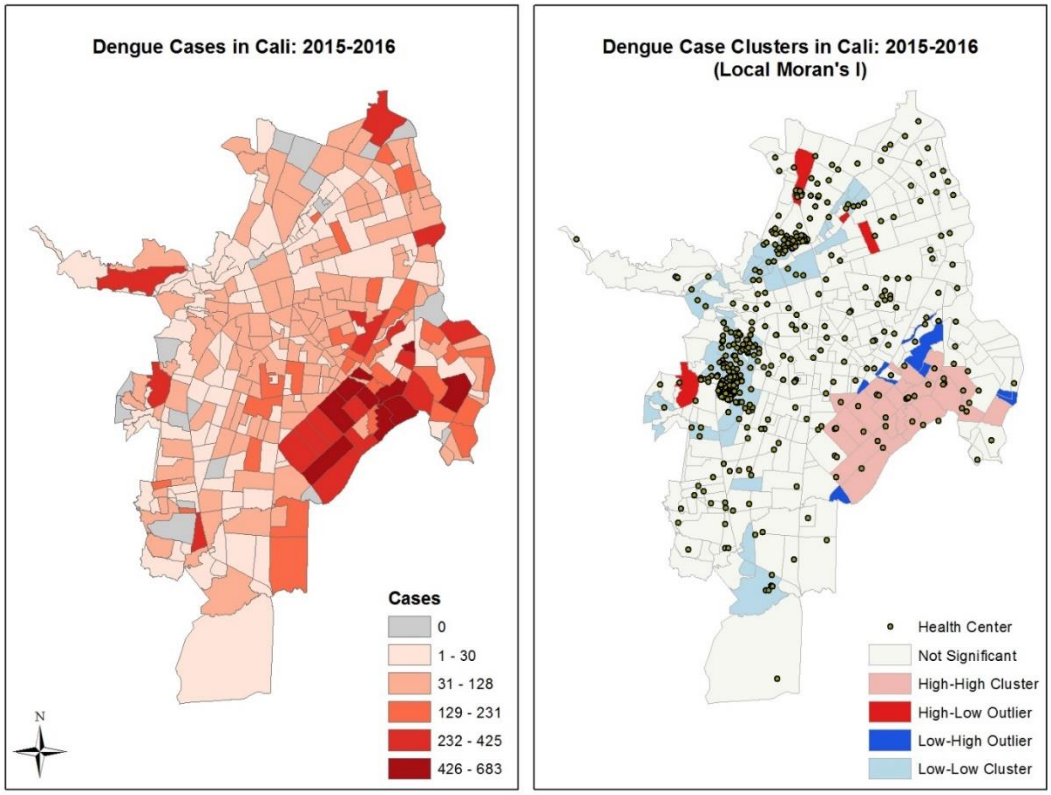


Figure 18: dengue cases (left) and case clusters (right) in Cali, Colombia for 2015 and 2016.

Of course, the location of the healthcare center may not be indicative of the socioeconomic status of the individual that is surveyed. For example, surveys from healthcare centers in high-income neighborhoods can include participants from middle- or low-income neighborhoods. The main idea of the abovementioned sample design was to capture as much variation as possible. To maximize sample size, surveys were also conducted at Universidad Icesi, Universidad Libre, and door-to-door in several neighborhoods throughout Cali. Therefore, the original sampling method proposed above

was not execited completely since opportunities were presented to survey at the universities and neighborhoods during the fieldwork portion in June of 2019.

Figure 19 shows the selected neighborhoods and health centers in Cali, which was derived from the sampling methods/queries in the previous section. The results are solely based on DENF cases, however, the selected neighborhoods and health centers also treat patients for CHIK and Zika too.

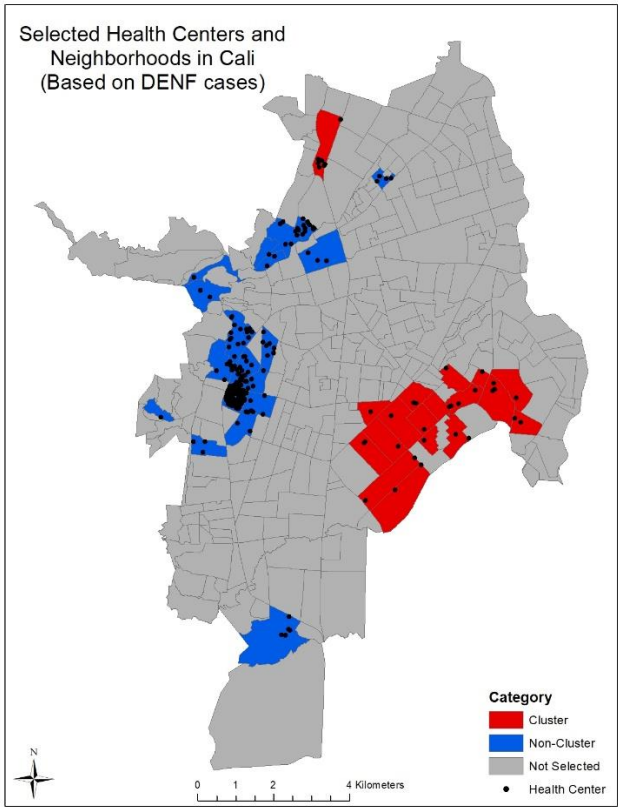


Figure 19: Selected neighborhoods and health centers in Cali, Colombia.

The table in Appendix 3 includes the 37 selected neighborhoods, the stratum, the type, and number of health centers within each neighborhood. After further consultation with colleagues in Colombia, the table in Appendix 4 includes the 21 healthcare centers that were candidates for the surveys. The final list was refined mainly after speaking with Dr. Alejandro Varela (General Manager of DIME clinics), who is the former

secretary of health in Cali. Dr. Varela identified the 21 places in Appendix 4 as facilities that he could help secure permission to survey. The forms that request permission to survey are found in Appendix 8. Figure 20 provides a map that shows the spatial locations of the final refined sample of the 21 candidate healthcare centers in Cali. The final survey sites are provided in the results section of this chapter.

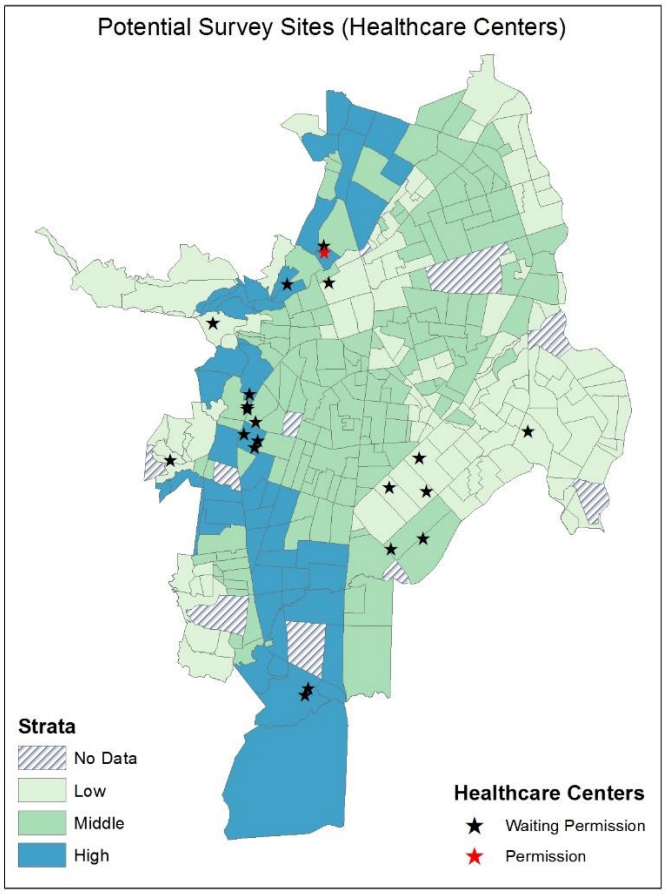


Figure 20: Final selection of potential health centers to survey in Cali, Colombia.

5.4 Data Analysis

Generalized linear models (GLM) were utilized to examine the effects between KAP scores and socioeconomic, sociodemographic, and accessibility variables. Three separate GLMs were used for knowledge, attitudes, and practices, respectively (each acting as the dependent variable). Overall knowledge, attitude, and practice scores were

determined by adding the total responses for each section (maximum value of 57, 60, and 60, respectively); The independent variables come from section 1 of the survey (A1), and the accessibility variables also serve as potential predictors in the GLMs. Variable inflation factor (VIF) tests were conducted to assess potential collinearity. Chi-square (X^2) tests were employed to compare the independent variables between the three strata groups and to test for significant differences regarding the answers to individual questions. Data analysis was conducted in R.

5.5 Results

5.5.1 Sociodemographic Information

A total of 327 surveys were collected during the one-month fieldwork window (211 female – 64.5% and 116 male – 35.5%) – 64 (19.6%) from homes in various neighborhoods across Cali, 66 (20.2%) from Universidad Icesi and Libre, and 197 (60.2%) from five healthcare centers. Figure 21 shows the locations where the surveys were conducted. Figure 22 shows the place of residence for the survey participants.

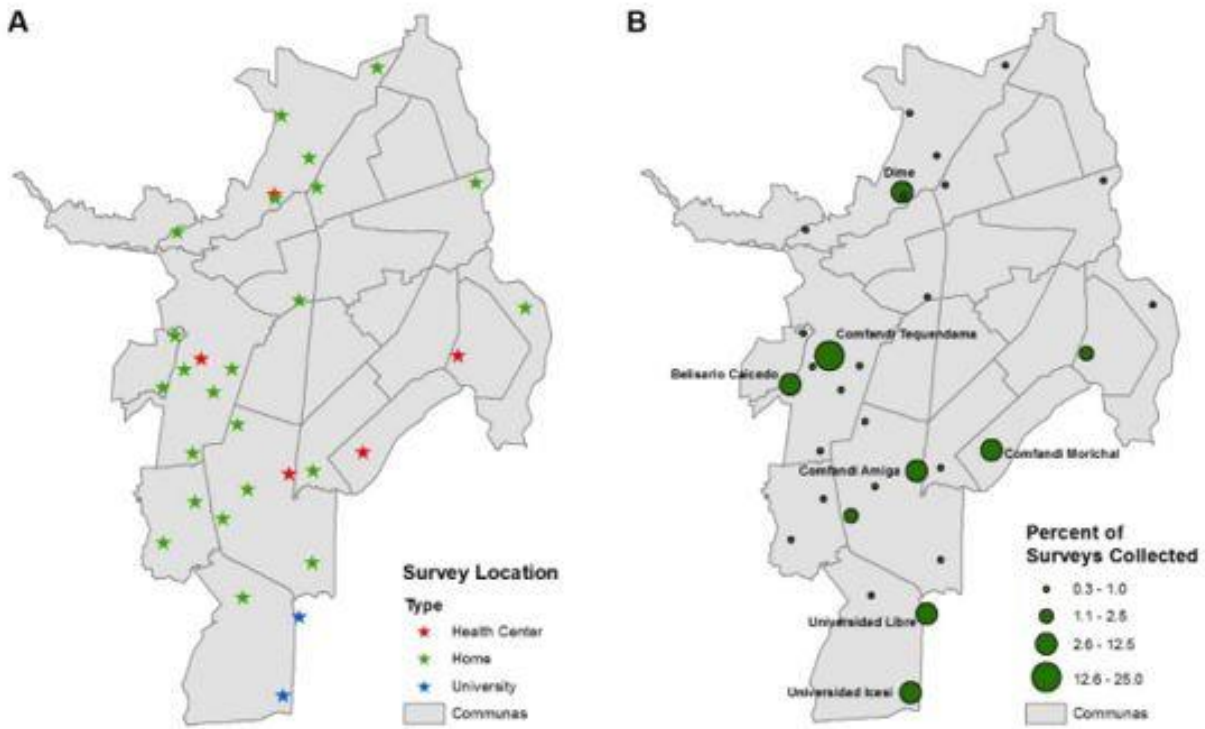


Figure 21: Location of surveys – homes indicate centroid of neighborhood where survey was collected (A); and percent of surveys collected from each location (B).

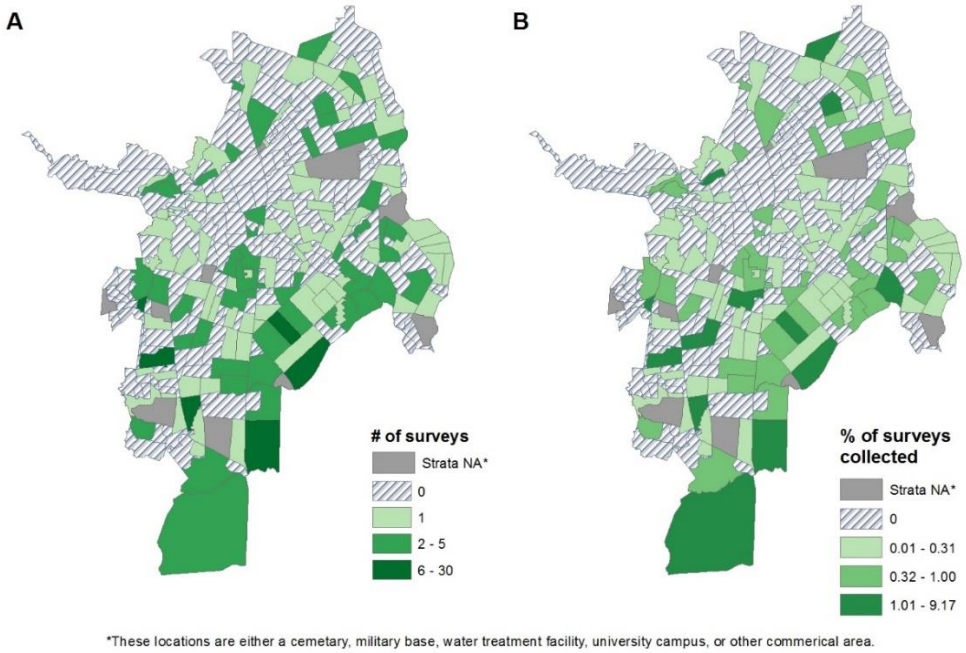


Figure 22: KAP participants' place of residence per neighborhood – 21 participants live outside of Cali's boundary, but live in Valle del Cauca. Total number of surveys per neighborhood in (A); and percent of surveys collected per neighborhood in (B).

Table 27 provides the socioeconomic and demographic characteristics of the 327 participants in our study, with 96 (29.4%) surveys collected in the low strata, 159 (48.6%) in the middle strata, and 72 (22%) in the high strata. A total of 143 (43.7%) participants were between the ages of 18 and 35; 126 (38.5%) were single; 109 (33.3%) held a secondary education; 120 (36.7%) worked full-time; 163 (49.9%) had children living in their household; 152 (46.5%) lived in a household with 3-4 individuals; and 158 (48.3%) reported their race as Mestizo.

Chi-squared tests were conducted to compare the sociodemographic attributes between the low, middle, and high strata neighborhoods, suggesting significant differences between the 3 strata groups for 6 out the 9 sociodemographic attributes.

There was no significant variation in gender, educational attainment, and income. However, we identified significant variation in the age groups of respondents ($p < 0.05$); occupation status ($p < 0.01$); civil status ($p < 0.01$); children in the household with a higher proportion of children in middle and low strata ($p < 0.01$); household size ($p < 0.01$); and race with a higher proportion of Afro-Colombians and Mestizos in low strata neighborhoods ($p < 0.01$).

Table 27: Socioeconomic and demographic characteristics of the KAP participants

Variable	n	%	Variable	n	%
Strata			**Occupational Status**		
Low	96	29.4	Full-time	120	36.7
Middle	159	48.6	Part-time	20	6.1
High	72	22	Independent	66	20.2
Age			Unemployed	12	3.7
18-35	143	43.7	Student	36	11
36-55	108	33	Pension	30	9.2
56-70	60	18.3	Housewife	43	13.1
70+	15	4.6	**Children in Household**		
Marital Status			Yes	163	49.8
Single	126	38.5	No	163	49.8
Married	89	27.2	**Household Size**		
Free Union	81	24.8	1	16	4.9
Separated/Divorced	19	5.8	2	72	22
Widow	18	3	3-4	152	46.5
Other	1	0.3	5-6	68	20.8
Education Level			> 6	19	5.8
Primary	25	7.6	**Race**		
Secondary	109	33.3	White	115	35.2
Undergraduate	73	22.3	Mestizo	158	48.3
Postgraduate	46	14	Afro-Colombian	42	12.9
Technical	73	22.3	Indigenous	6	1.8
			Other	3	0.9
			Mixed-Race	2	0.6

5.5.2 General Questions

For question 1, 194 (59.3%) respondents reported never having DENF, CHIK, or Zika; 48 (14.7%) reported having CHIK; 51 (15.6%) reported having DEN; 6 (1.8%) had Zika; 6 (1.8%) had CHIK and Zika; 19 (5.8%) had both DENF and CHIK; and 3 (0.9%) individuals reported having all three of the VBDs at some time in their lives. Table 28 summarizes the responses of question 1 by strata – where 46 individuals who reported having at least of the three VBDs also lived in a low strata neighborhood.

Table 28: Percent of participants who reported having one or more of the three VBDs

Disease	High Strata	Middle Strata	Low Strata
None	14.7%	29.4%	15.3%
DENF	3.4%	9.5%	2.8%
CHIK	2.4%	5.2%	7.0%
CHIK/Zika	0.3%	0.6%	0.9%
DENF/CHIK	0.9%	2.4%	2.4%
All three	0.0%	0.3%	0.6%
Zika	0.3%	1.2%	0.3%

For question 2, 37 (11.3%) respondents reported having no knowledge of the three VBDs; 17 (5.2%) only had knowledge of CHIK; 1 (0.3%) had knowledge of CHIK and Zika; 43 (13.1%) only had knowledge of DENF; 52 (15.9%) had knowledge of both DENF and CHIK; 1 (0.3%) only had knowledge of Zika; and 176 (53.8%) reported having knowledge of all three VBDs.

For question 3, low strata neighborhoods rated their access to healthcare as 2.55, on average; and middle and high strata rated their access as 2.88 and 3.11, respectively (1-5 scale). For question 4, 4 (1.2%) individuals reported traveling less than 5 minutes to a healthcare center; 79 (24.1%) reported traveling 5-15 minutes; 129 (39.4%) travel 15-30 minutes; 75 (22.9%) travel 30-60 minutes; and 40 (12.2%) reported traveling an hour

or more. Table 29 summarizes the responses of question 4 by strata; we find that participants from low and middle strata were more likely to travel 60 minutes or more ($p < 0.05$).

Table 29: KAP participants' approximate time to a healthcare center

Time	High Strata	Middle Strata	Low Strata	Total
1 hr or more	1.2%	4.9%	6.1%	12.2%
30-60	5.5%	10.1%	7.1%	22.7%
15-30	8.9%	20.2%	11.9%	41%
5-15	6.1%	11%	5.9%	23%
< 5 min	0.3%	0.5%	0.3%	1.1%

For question 5-1, 70 (21.4%) respondents reported discontinuing medical treatment because of the distance to a healthcare center; 208 (63.6%) never discontinued treatment; and 49 (15%) individuals were not sure. For question 5-2, 24 (7.3%) individuals refused medical treatment because of the distance to a healthcare center; 245 (75%) never refused treatment; and 58 (17.7%) individuals were not sure.

5.5.3 Knowledge

For question 1, 131 (40%) individuals reported being familiar with all three VBDs; 34 (10.4%) reported both DENF and CHIK; 1 (0.3%) reported both DENF and Zika; 90 (27.5%) reported only DENF; 31 (9.5%) reported only CHIK; 3 (0.9%) reported only Zika; and 37 (11.3%) reported not being familiar with any of the three VBDs. For question 2, 80 (24.5%) individuals did not correctly state how DENF is transmitted; and 247 (75.5%) reported the correct answer (214 mosquito; 24 *Ae. aegypti*; and 9 *Aedes*). For question 3, 137 (41.9%) individuals did not correctly state how CHIK is transmitted, and 190 (58.1%) reported the correct answer (171 mosquito; 7 *Aedes*; 11 *Ae. aegypti*; and 1 *Ae. aegypti/albopictus*). For question 4, 176 (53.8%) individuals did not correctly state

how Zika is transmitted; and 151 (46.1%) reported the correct answer (136 mosquito; 4 *Aedes*, 10 *Ae. aegypti*; and 1 *Aedes*/vertical).

For question 5-DENF, individuals learned about DENF from nearly three sources ($\bar{x}$ = 2.8). Regarding the types of sources, television was the top source of knowledge for DENF with 252 (77%) selections, radio was second with 122 (37.3%), and doctor was third with 115 (35.1% - see Table 30 for the complete results). For question 5-CHIK, individuals learned about CHIK from two sources or more ($\bar{x}$ = 2.6). Regarding the types of sources, television was the main source of knowledge for CHIK with 236 (72.2%) selections, radio was second with 116 (35.5%), and doctor was third with 98 (30%). For question 5-Zika, individuals learned about Zika from exactly 2.0 sources. Regarding the types of sources, television was the main source of knowledge for Zika with 194 (59.3%) selections, radio was second with 95 (29%), and doctor was third with 82 (25%). Figure 23 shows the average number of sources that individuals learned about each of the three VBDs by strata. On average, respondents from high strata neighborhoods seem to have had the most knowledge of the three VBDs; and there was a slight wider variety information sources for DENF, followed by CHIK and Zika. Interestingly, low strata neighborhoods reported slightly more information sources about DENF and CHIK than middle strata neighborhoods (p<0.05).

Table 30: Educational sources reported by KAP participants in Cali, Colombia.

Type	DENF	CHIK	Zika
TV	252	236	194
Radio	122	116	95
Doctor	115	98	82
Social Network	92	88	65
Pamphlets	79	82	69
School	41	40	32
Newspaper	34	83	80
Billboards	25	26	28
Work	9	9	9
Internet	3	3	5
Articles	2	2	0
Patient	2	1	1
Curiosity	1	0	1
Other	1	0	1
Research	1	0	0
Presentations	1	1	1

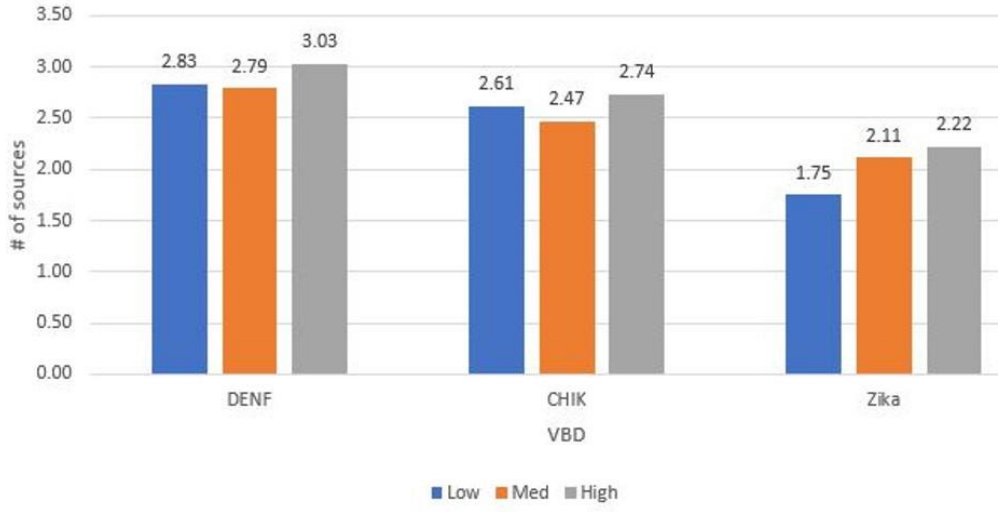


Figure 23: Average Educational Sources per Neighborhood Strata Regarding DENF, CHIK, and Zika

Question 6 asked if the three VBDs can be transmitted sexually, and 186 (56.9%) reported the correct answer for DENF (no); 179 (54.7%) reported the correct answer for CHIK (no); and only 23 (7%) reported the correct answer for Zika (yes). Question 7

asked if the three VBDs can be prevented with a vaccine, and 93 (28.4%) reported the correct answer for DENF (no); 104 (31.8%) reported the correct answer for CHIK (no); and 99 (30.3%) reported the correct answer for Zika (no). Question 8 asked if a baby is at-risk of birth defects if a pregnant mother is infected with one of the three VBDs, and 113 (34.5%) reported the correct answer for DENF (yes); 143 (43.7%) reported the correct answer for CHIK (yes); and 206 (63%) reported the correct answer for Zika (yes).

Question 9 asked which time of day are you most at-risk to be infected by the three VBDs (morning and afternoon), and 124 (38%) were completely correct for DENF (received 2 points), 63 (19.2%) were partially correct stating at least morning or afternoon (received 1 point), and 140 (42.8%) were incorrect stating at night or not knowing at all; 113 (34.5%) were completely correct for CHIK, 66 (20.2%) were partially correct, and 148 (45.2%) were incorrect; 108 (33%) were completely correct for Zika, 55 (16.8%) were partially correct, and 164 (50.1%) were incorrect.

Question 10 asked if the individual knows how to prevent one or more of the three VBDs, and 258 (78.9%) reported yes for DENF; 217 (66.7%) reported yes for CHIK; and 198 (60.5%) reported yes for Zika. Question 11 asked if they think each of the three VBDs are worse in Cali than they used to be, and 162 (49.5%) reported yes for DENF (49.5%); 141 (43.1%) reported yes for CHIK; and 138 (42.2%) reported yes for Zika (42.2%). Finally, question 12 asked which season results in a higher risk of transmission from each of the three VBDs (winter/wet seasons is correct), and 207 (63.3%) reported the correct answer for DENF; 185 (56.6%) reported the correct answer for CHIK; and 167 (51%) reported the correct answer for Zika.

When comparing the total knowledge scores for all three diseases, DENF had an average score of 9.52, CHIK with 8.60, and Zika with 6.87. When comparing total knowledge scores by neighborhood strata, low strata had an average of 26.14, middle with 27.27, and high with 29.43. Figure 24 compares the average knowledge scores by strata and disease in Cali. For DENF, low strata neighborhoods had an average score of 9.16, middle with 9.48, and high with 10.07. For CHIK, low had an average score of 8.60, middle with 8.31, and high with 9.22. For Zika, low had an average score of 6.15, middle with 6.96, and high with 7.63.

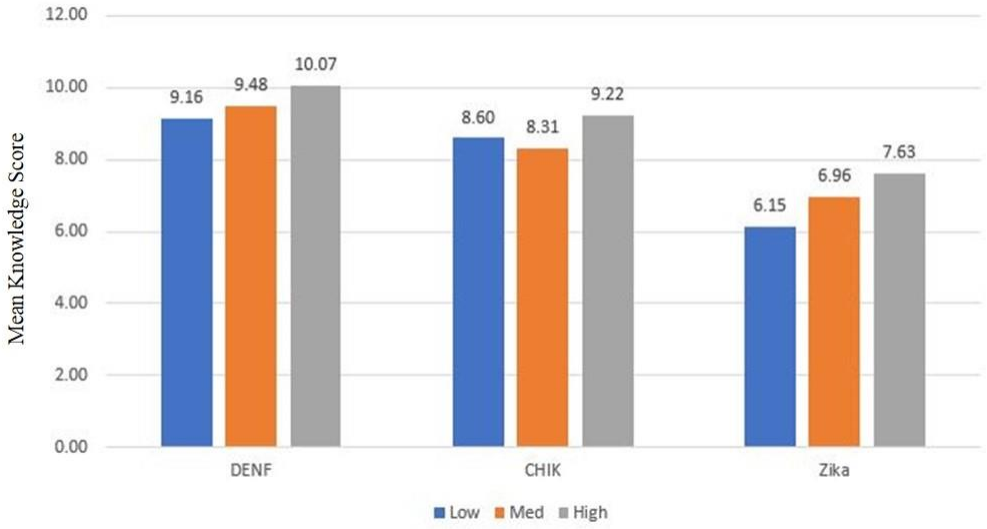


Figure 24: Average knowledge scores by strata and disease

Overall, the results suggest that residents have the greatest knowledge of DENF, followed by CHIK and Zika, and particularly so for residents in high strata neighborhoods. When comparing strata and disease, average knowledge scores are greatest for high strata, followed by middle and low for both DENF and Zika; and low strata neighborhoods have slightly higher average knowledge of CHIK than middle strata neighborhoods.

Table 31 shows the results of the GLM using total knowledge score as the dependent variable. The total knowledge scores were significantly related to strata, sex, civil status, race, and occupation. Compared to high strata neighborhoods, low and middle strata neighborhoods had significantly less knowledge of DENF, CHIK, and Zika. Notably, males had significantly less knowledge than females. Married individuals had significantly more knowledge, and separated/divorced individuals had significantly less knowledge. Compared to Afro-Colombians, Whites and Mestizos had significantly more knowledge. Finally, individuals working full-time had significantly more knowledge of the three VBDs.

Table 31: Modeling results of Knowledge - significant predictor variables

Variable	Coefficient	*p*	Compare
Intercept	3.225	<0.05	NA
Low Estrata	-0.089	<0.05	High Estrata
Middle Estrata	-0.067	<0.05	High Estrata
Males	-0.064	<0.05	Females
Married	0.105	<0.05	Free Union
Separated/Divorced	-0.158	<0.05	Free Union
White	0.125	<0.05	Afro-Colombians
Mestizo	0.135	<0.05	Afro-Colombians
Work Full-Time	0.082	<0.05	Student

5.5.4 Attitudes

The median attitude scores for each individual question were either ‘5’ (agree) or ‘6’ (totally agree). Furthermore, the results suggest that residents think DENF is the most severe disease out of the three (median = 6), followed by CHIK (median = 5), and Zika (median = 5). Next, DENF, CHIK, and Zika all had a median at-risk score of ‘5’. Residents suggested that they are more responsible to prevent the three VBDs (median = 6) than the government (median = 5). A median of ‘6’ suggests that most people think

fumigation efforts are effective. Regarding seeking immediate treatment for each VBD, all three VBDs had a median of '6'. In total, all three VBDs had a median attitude score of '31' (maximum of 48). When comparing the total attitude score by strata, all three neighborhood categories had a median score of '62' (max = 72). Figure 25 compares the median attitude scores by strata and disease. For DENF, low and middle strata neighborhoods had a median score of '32', and high strata had a score of '31'. For CHIK and Zika, low and middle had a score of '31', and high with '30'. Overall, residents seem to think that DENF is the most severe out of the three VBDs.

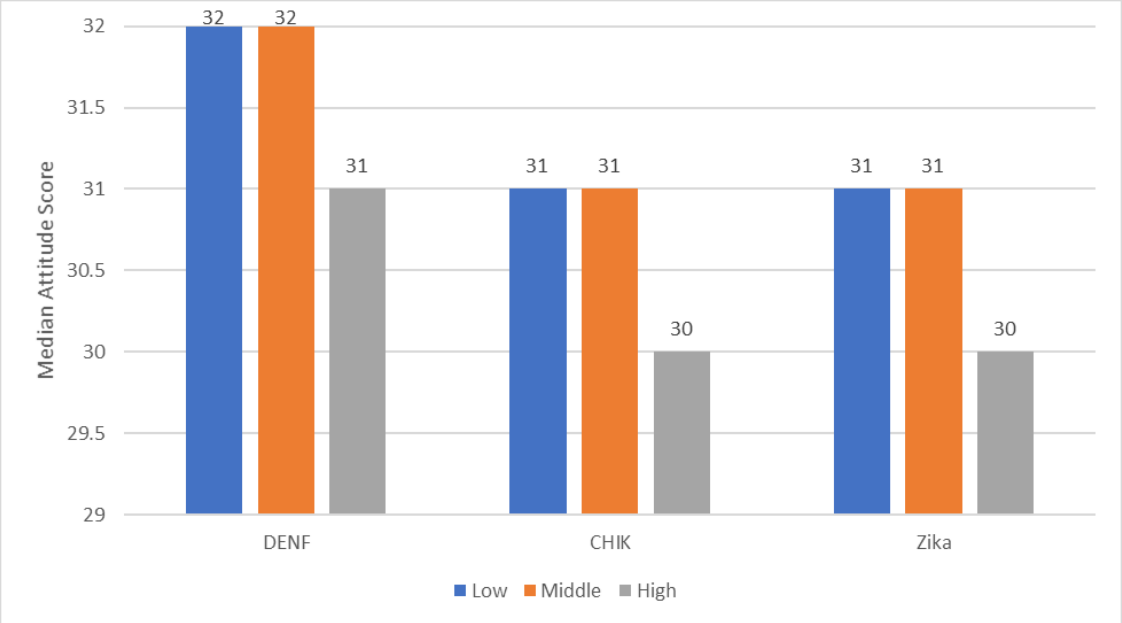


Figure 25: Median attitude scores by strata and disease

Table 32 shows the results of the GLM using total attitude score as the dependent variable. The total attitude scores were significantly related to education, race, occupation, and previous infection of one or more of the VBDs. Compared to individuals with a postgraduate education, individuals with a secondary education had a significantly lower attitude score. Compared to Afro-Colombians, Whites and Mestizos had a significantly higher attitude score, and Mixed-Race individuals had a significantly lower

attitude score. Compared to students, individuals who worked full-time, part-time, independently, collected pension, housewives, and those who were unemployed all had a significantly higher attitude score. Finally, individuals who were previously infected with one or more of the three VBDs were significantly more likely to report higher attitude scores.

Table 32: Modeling results of Attitudes - significant predictor variables

Variable	Coefficient	*p*	Compare
Intercept	3.991	<0.05	NA
Secondary Educ	-0.075	<0.05	Postgraduate
White	0.057	<0.05	Afro-Colombian
Mestizo	0.063	<0.05	Afro-Colombian
Mixed-Race	-0.221	<0.05	Afro-Colombian
Work Full-time	0.077	<0.05	Student
Housewife	0.11	<0.05	Student
Independent Work	0.065	<0.05	Student
Work Part-Time	0.087	<0.05	Student
Pension	0.086	<0.05	Student
Unemployed	0.11	<0.05	Student
Had 1 or more VBDs	0.028	<0.05	NA

5.5.5 Practices I

Question 1 asked if the individual strives to learn about the prevention of each of the three VBDs. For DENF, 205 (62.7%) said yes; 186 (56.9%) said yes for CHIK; 167 (51%) said yes for Zika (51%). Question 2 asked if the individual was worried about mosquitoes in their neighborhood, and 235 (71.8%) said yes. Question 3 asked if they have ever contacted the proper authorities if they have identified mosquitoes or mosquito larvae on their property, and 52 (15.9%) said yes. Question 4 asked how frequently does their residence get sprayed with insecticides or larvicides, and 5 (1.5%) reported daily

(1.5%), 4 (1.2%) weekly, 61 (18.6%) monthly, 42 (12.8%) twice per year, 42 (12.8%) annually, and 110 (33.6%) never/don't know. All three VBDs had a median practice I score of '3'. When comparing the median practice score by neighborhood strata, low had a score of '4', and middle and high both had a median of '5'. When comparing the Practices I scores by strata and disease, the median score was the same for each category at a value of '3'.

5.5.6 Practices II

The median Practices II scores for each individual question ranged from '0' (never) and '5' (daily). The median score for using repellent, using long-sleeved shirts and pants when working outside, and using window screens was '0' (never); '4' (weekly) for frequently changing water in materials that accumulate water, removing materials that accumulate water outside the home, and removing materials that accumulate water inside the home; '0' (never) for participating in community cleanup activities; '4' (weekly) for examining deposits of water for mosquito larvae; '0' (never) for using mosquito nets in the home; and '5' (daily) for closing the windows and doors in the home. When comparing the total score by neighborhood strata, all three had a median value of '30'. Next, the total Practices I and Practices II scores were combined by summing the two into a single practice category; and the median scores were compared by neighborhood strata. Low strata neighborhoods had a median of 36, middle with 35, and high with 34.

Table 33 shows the results of the GLM using total practice score as the dependent variable. The total practice scores were significantly related to previous infection of one or more of the VBDs, familiarity with one or more VBDs, access rating to healthcare, travel time to a healthcare facility, and refusing healthcare because of the cost.

Interestingly, no demographic or socioeconomic variables had a significant relationship with total practice score. Individuals who were previously infected with or were familiar with one or more one or more of the three VBDs were significantly more likely to participate in more preventative measures against DENF, CHIK, or Zika. Next, the positive relationship between practices and access rating suggests that individuals are significantly more likely to participate in preventative measures as their access to healthcare increases. Furthermore, compared to individuals who travel an hour or more to a healthcare facility, individuals who travel 5-15 minutes were significantly less likely to participate in preventative measures; and individuals who travel < 5 minutes were significantly more likely to participate in preventative measures.

Table 33: Modeling results of Practices - significant predictor variables

Variable	Coefficient	*p*	Compare
Intercept	3.504	<0.05	NA
Had 1 or more VBDs	0.052	<0.05	NA
Familiar with 1 or more VBDs	0.027	<0.05	NA
Rating Access to Healthcare	0.017	<0.05	NA
Travel 5-15 minutes to Healthcare	-0.069	<0.05	1 hr +
Travel < 5 minutes to Healthcare	0.295	<0.05	1 hr +
Refused Healthcare because of Cost	-0.042	<0.05	NA

5.6 Discussion

This study is the first of its kind to assess KAP that compares low, middle, and high strata neighborhoods across an entire city that is at-risk for co-occurring VBDs. Although the usefulness of studies that evaluate KAP of populations at risk of diverse diseases has been previously recognized, these studies are essential to understand individuals' prevention practices in order to assess the impact of various education

strategies and design, develop and implement community tailored programs (Hernández-Escolar et al. 2014). Our findings suggest that knowledge of DENF, CHIK, and Zika in Cali, Colombia is related to community characteristics (e.g. strata), and attitudes and practices are more related to the individual level. Individuals knew more about DENF, followed by CHIK, then Zika (although Zika had more reported cases in Chapter 4); likely due to the latter two diseases first appearing between 2013 and 2014, while DENF has been endemic since the 1970s. As we expected, knowledge scores were higher in high strata neighborhoods, then middle, followed by low. Knowledge gaps are clear barriers for the population to get empowered with the corresponding prevention plans. This issue could be approached with the proper use of media and the dissemination of diverse campaigns, detailing the origin of these diseases, the factors that perpetuate them, their transmission, necessary treatment and prevention methods. This way, each citizen acquires responsibility for their environment, the periodic inspection of the potential reservoirs and their respective elimination. As Hernández-Escolar et al. (2014) mentioned, empowerment is a key condition for the community to face complex public health problems. Even though, the constant role of public health expert authorities is essential to guarantee the transformation of good knowledge into better practices (Sarmiento-Senior et al. 2019).

On average, attitude scores were similar in both low and high strata neighborhoods, and individuals residing in the middle strata (3 and 4) seem to be less worried about the three diseases. The community is the one that exerts control over the vector and its proliferation, therefore, the greatest importance lies in the attitudes of individuals and their awareness of the problem which will ultimately create the

foundation for behaviors turned into habits over time. It is necessary to raise awareness in communities, so attention focuses on vector control through fumigation and use of repellent, but also on prevention and targeted actions in the eradication of the disease (Castañeda-Porras et al. 2017). Despite TV, radio, and doctors as the predominant source of information about the risks of the three VBDs in this study, there is a degree of ineffectiveness and educational campaigns need to be improved. The involvement of local stakeholders such as community leaders and teachers from the initial stages of education and promulgation of campaigns could improve the acceptance and adaptation of the strategies to the specific conditions of each neighborhood (which is echoed later in Chapter 6 by the public health officials). Furthermore, assessing the specific problems faced by each area of the municipality and having information on risk conducts related to vector breeding allows the design of local communication strategies that would consider concrete proposals for behavioral changes (Castro et al. 2008).

Access to health and previous infection with the disease significantly improved an individual's willingness to take preventative measures against DENF, CHIK, and Zika. Timely diagnosis a disease is critical to morbidity and mortality and mitigate current and future outbreaks (Casas et al. 2017). Furthermore, doctors were the third main source of knowledge, which suggests that prevention measures should be pursued outside of healthcare settings. These findings in the Practices section are important because Chapter 4's did find healthcare facility density to be significant; although better accessibility modeling is required to further investigate my findings on the importance of access in the KAP. Corroborating with Whiteman et al. (2018), we found that there were a low number of participants who closed their windows throughout the day. Since *Aedes* is

known to seek shelter inside during the hottest temperatures of mid-day (Dzul-Manzanilla et al. 2017), vector-control efforts should address the issues of utilization of window and door screens. Public health officials may need to communicate with property owners to install structures (i.e. screens) to reduce mosquito presence inside residences.

Although many studies typically compare high and low-income neighborhoods or locations, we have found evidence that middle-income neighborhoods may lack resources due to a variety of reasons. In this study, high strata individuals are typically more educated and have access to more healthcare resources, and the low strata individuals in Cali live in neighborhoods that are frequently targeted by public health campaigns and mosquito eradication programs. The modeling results in Chapter 4 provide strong evidence that there are many high-risk neighborhoods of DENF, CHIK, and Zika in middle strata neighborhoods. Therefore, individuals living in middle strata neighborhoods in Cali may face public health disparities regarding targeted interventions both before, during, and after outbreaks of DENF, CHIK, and Zika.

It is also worth noting that this study occurred in the summer of 2019, while major outbreaks of CHIK and Zika occurred between 2014 and 2016 (Krystosik et al. 2017; Desjardins et al. 2018). It is therefore surprising that the findings of the KAP survey suggest that CHIK and Zika are still relatively misunderstood and awareness did not match the levels of DENF in Cali. Since CHIK and Zika are transmitted by the same vector, it is critical to improve the community's understanding regarding CHIK and Zika. In practice, if an individual is participating in preventative measures against DENF, they are also protecting themselves against CHIK and Zika. However, CHIK and Zika can

also be spread vertically, and safe-sex education campaigns should be improved to combat the congenital complications (especially microcephaly resulting from vertical transmission of Zika).

Despite the strengths of this study, I acknowledge that there are several limitations which can serve as avenues for future research. First, the results may be slightly biased towards middle strata participants due to the higher proportion of surveys collected from that group. Second, as with any survey, recall bias may have affected the accuracy of the individual responses. Third, accuractely measuring individual attitudes is challenging and can be biased. The participiants may have responded to the attitude questions that they perceived to be correct (which can be influenced by the survey's context and location where is was administered). Launiala (2009) suggests that in order to improve the reliability of attitude-based questions, the researcher can "transform some of the attitude statements into direct questions in the other sections and to assess whether there is any discrepancy between the results or not (p. 5)". Fourth, follow-up studies with the highest risk neighborhoods can determine the best approaches for improving KAP. Finally, our results may not be generalizable to other at-risk populations of DENF, CHIK, and Zika outside of Cali; but should inspire other communities to assess their understanding and preventative practices regarding infectious disease.

5.7 Conclusion

I encourage public health officials and community leaders to examine the effectiveness of their educational campaigns, while improvin preventative measures against the *Aedes* mosquito before an epidemic may occur to minimize future cases. I also suggest that educational and targeted intervention campaigns could be improved by

allocating resources to community leaders, considering that certain populations have a distrust in government. These community leaders can disseminate educational materials to at-risk neighborhoods, which may increase awareness and willingness to practice VBD preventative measures. VBD prevention should also be proactive instead of reactive, that is public health officials should not wait until an epidemic begins to intervene; rather funding and stakeholders should be utilized year-round to raise awareness and understanding about the diseases that are constantly prevalent in certain communities.

This study also provides evidence that public health programs may be lacking in middle income areas, due to targeted interventions focusing on the lowest and "highest risk" neighborhoods. It is also clear that more resources need to be devoted to improving the understanding of CHIK and Zika. Therefore, educational material (e.g. TV, radio, and social media – see Giustini et al. 2018) should discuss the risks and complications of DENF, CHIK, and Zika since there are all transmitted by the same vector and require the same climatic and environmental conditions for successful infection. Overall, more fine-level studies that evaluate community and individual KAP are necessary to improve public awareness of VBD transmission.

CHAPTER 6: PERCEPTIONS OF DENF, CHIK, AND ZIKA FROM PUBLIC HEALTH OFFICIALS IN CALI, COLOMBIA

The findings from chapter 5 suggested that the at-risk communities and individuals still face knowledge gaps regarding the transmission dynamics of DENF, CHIK, and Zika. There was also evidence of disparities in accessibility and taking preventative measures against *Aedes* (reducing breeding sites, preventing bites, and practicing safe sex during outbreaks – especially for Zika). The goal of chapter 6 is to understand the other side of the story – what the public health officials are doing to mitigate VBD outbreaks, what is working, what is not working, their perceptions of the communities' role in vector control, and suggestions for improved prevention and control strategies. Collecting information from both the community and public health officials can facilitate suggestions for improved VBD surveillance; and also refine modeling approaches and data collection for quantitative studies.

6.1 Methodology

I developed a semi-structured interview that is comprised of ten main questions. The main objective was to ask high-ranking public health officials about their career experiences, their familiarity with DENF, CHIK, and Zika in Cali, the policies and protocols in place to combat and prevent outbreaks, necessary improvements, the importance of community health surveys and studies, and identifying the highest risk areas on a map (sketch maps). The sketch map question asked the participants to identify the highest-risk areas of DENF, CHIK, and Zika in Cali (circling locations with a marker). Participants were recruited for an interview if they were a professor of public health sciences at a university, director or general manager of a healthcare center, or worked for Valle del Cauca's or Cali's Secretary of Health Department. Each interview

was intended to last no longer than thirty minutes at the location of decided by each participant. Besides one interview with a team of four public health officials that lasted 1.5 hours, the duration of each did not last longer than 30 minutes.

Informed verbal consent was obtained from each participant before the interview was conducted by me and a colleague that was fluent in Spanish. The interview and all the components are approved by the University of North Carolina at Charlotte's IRB board (UNC-Charlotte Case No. 18-0399). The full interview (both English and Spanish versions) is attached in Appendix 9 and Appendix 10, respectively; and each one was recorded and later translated and transcribed. The sketch mapping completed by each participant was digitized in ArcGIS to facilitate analysis and comparison. Text-based analysis of the transcriptions was completed in NVivo by coding the interviews into categories. There are ten main categories, which are the responses from the ten questions.

6.2 Results

Six completed interviews were conducted with high ranking public health officials and researchers in Cali and Valle del Cauca. A total of nine participants were interviewed because one session included multiple high-ranking officials. Also, brief unstructured interview with a Cali resident was also included in the analysis. The following subsections discuss the main findings from each question asked in the interviews. For privacy purposes, the names and the titles of the participants have been masked. As a result, codenames were assigned to each participant: public health official 1 (PHO1, PHO2, etc.).

6.2.1 Participant job descriptions and duties

The majority of the participants were trained and licensed as medical doctors before assuming their current roles. At the time of the interviews, all participants were involved with public health education, disease surveillance and control, and community health research in Cali. Their job titles and descriptions ranged from sociologist, entomologist, sanitation manager, environmental health and vector management, professor, epidemiologist, health educator, and policymakers. The participants' experience working in Cali ranged from 5 years to 32 years; and it was clear that each participant is highly educated with both national and international experience in the public health field; and all were familiar with either researching, diagnosing, treating DENF, CHIK, and Zika. Community health and social outreach was also a central theme of all the participants. Although most of the efforts focused on the city limits of Cali, others work with the corregimientos (rural subdivisions) surrounding the urbanized region.

> "We do community medicine – we take the bus and go to rural areas and have outpatient clinics. And we look at the social determinants and all these other factors that influence health" [PHO9].

6.2.2 Challenges in identifying and diagnosing DENF, CHIK, and Zika

The public health officials seemed to have struggled with the first outbreaks of CHIK and Zika (2013-2014), which co-occurred with outbreaks of endemic DENF. The focus of combatting the diseases was reactive, and priority was given to the diseases that were novel threats (CHIK and Zika).

> "During the chikungunya epidemic, that was 4-5 years ago, it was diagnosed only by the National Institute of Health, sending samples from here to the department, and then the department sent them. There were not a large percentage of samples

> confirmed. It was a problem to confirm dengue samples because the country gave priority to chikungunya and then a year after Zika [PHO1]".

Due to the similar onset symptomologies of all three VBDs, it is very difficult to make an initial diagnosis. It is clear that there was vast underreporting and over-reporting of DENF, CHIK, and Zika due to the similar clinical manifestations during the initial acute period. Since CHIK and Zika were new in the region, the public health officials could not accurately estimate the expected morbidity and mortality rates; which may explain why the CHIK case data in chapter 4 was much lower than what was reported by the public health officialsThe co-circulation of all three VBDs during that time also put severe pressure on the healthcare system. Instead of treating each disease equally regarding the allocation of resources and educational campaigns, there was a mass panic depending on the reported cases for a particular week or month. For example, people seemed to have forgotten about DENF when CHIK arrived, despite the persistent viral circulation in Colombia and Cali. Zika also took priority once there was a high number of confirmed cases and poor pregnancy outcomes, such as microcephaly and other birth defects resulting from vertical transmission.

> "Dengue has been here forever, but still you never get dengue or someone near you never gets dengue, you actually don't see dengue. Which amazes me because it actually kills people. But then chikungunya got here and everyone got chikungunya, but no one died from chikungunya. But there was a lot of arthritis. The chronic complications have been abysmal, but still it wasn't something as severe as dengue, but it was very visible, very visible. So, I think people were actually more aware of chikungunya than dengue at that point, chikungunya and Zika than they were about dengue" [PHO9]".

Despite PHO9 suggesting that people were more ware of CHIK during the first epidemics in Cali, the results from my KAP study provide evidence that individuals have more awareness of DENF, followed by CHIK, then Zika. This could be due to the fact that this

KAP survey was conducted in 2019, which was a few years after the major epidemics of CHIK and Zika. Therefore, recall bias and awareness of DENF could have taken priority again (since we have not seen major epidemics of CHIK and Zika since 2016).

6.2.3 Treatment challenges

Despite the awareness and educational programs being improved since the initial outbreaks of CHIK and Zika, the public health officials are concerned that individuals are not seeking immediate treatment and wait too long to get diagnosed, attempt to treat the symptoms at home, or simply refuse medical treatment because they know there is no cure or vaccine. This corresponds to the KAP results which indicate that people who refused medical treatment practice less preventative measures against *Aedes* than individuals who seek care.

"Intra-hospitality problems or family delays. They take too long to go to consult. They think they can handle it with acetaminophen, water and people get sicker. Older people and children or people who have other disease, than the clinical chart gets worse [PHO5]".

Therefore, it seems that there needs to be consistent educational campaigns each month to keep the population aware of the continued risks of DENF, CHIK, and Zika. This finding also corroborates with the KAP study in Chapter 5, suggesting that access to healthcare will significantly increase preventative measures against DENF, CHIK, and Zika. The participants also acknowledge that because DENF is endemic, it will be virtually impossible to completely eradicate the disease from Cali and from Colombia. It must be instilled in the population that immediate treatment for DENF, CHIK, and Zika is critical to minimize severe complications and potentially fatal outcomes.

6.2.4 Prevention strategies and obstacles

Educational campaigns and prevention strategies carried out by public health officials in Cali is multifaceted and conducted at various spatial scales, which includes several components: laboratory, public health surveillance, social participation, and clinical attention. However, there are several obstacles that Cali faces regarding VBD surveillance, treatment, and control.

First, the dynamics of human movement and commuting to Cali from surrounding towns and Departments create a difficult environment for VBD surveillance. Also, since ~80% of individuals infected with DENF and Zika are asymptomatic (Duffy et al. 2009; WHO 2018), it is extremely challenging to track viral circulation and perform effective disease surveillance.

> "First, for me the key is to identify the area where the person is coming from. It is very important. Why? Because here in Cali we have a floating population that comes from neighboring municipalities and in those municipalities we already know that the sanitary conditions favor the growth of the mosquitos and breeding sites. And many of those cases that happen are imported. They get here. So knowing if a person goes to consult here and because they work here or study here and they come from Candelaria or Jamundi, the probability increases of it being a vector borne disease, especially dengue [PHO9]".

PHO9 highlights the need to understand individuals' activity spaces and their exposure to *Aedes* breeding sites and interaction with other human hosts – which was not capture in the models in Chapter 4 (which assumed the individuals were infected/bit in their place of residence).

Second, political elections in Cali influence disease surveillance and the allocation of public health resources. To control corruption, the government implements a "guarantee law" during an election year, essentially making it illegal for mayors and governors to spend money during the process:

> "And the thing is usually you have it [major dengue outbreak] every 4 years. And it's a situation that the year that you have it, in the last 4 years, it's the electoral year. So, the year they are working on campaigns, they are not working on disease [PHO8]".

Third, the management of funds is also a challenging situation, such as effectively allocating money towards the correct resources to mitigate and prevent outbreaks. It seems that many agree that too much funding is going towards fumigation programs (and *Wolbachia*) and not enough is being allocated towards education. The complexity of the population, policymakers and elections, and disease dynamics create a difficult environment that requires constant adaptation and awareness (community awareness needs improvement – see Chapter 5). Constant vigilance and ensuring that people take an active part in vector control is essential to minimizing outbreaks.

> "What are we going to do to prevent it? And I think most of the efforts that we've seen are when everyone is sick. We are giving more attention and resources to just treat the symptoms once they are in the hospital, but we don't take the time or the efforts to do the preventive strategies [PHO9]".

Fourth, the medical doctors' training is also not continuous, which may reduce the educational material that are disseminated to patients. There also seems to be a discrepancy between what medical doctors (primary care providers) and the epidemiologists think about what is best for preventing DENF, CHIK, and Zika. The medical doctors seem to be just treating what the epidemiologists are reporting in the city; and the providers could be missing opportunities to educate their patients about how to prevent the disease, rather than just discussing treatment plans. There is some blame being projected by the public health officials of the city, suggesting that primary care physicians need to improve their educational material disseminated to the patients, either by the facility or medical doctor themselves.

Fifth, educational campaigns seem to be more successful at the local, community level. In fact, educating and relying on community leaders to disseminate information may be more effective than relying on public health officials to educate the communities. Despite my KAP study identifying TV, Radio, and Doctors as their primary source of learning about VBDs, it was clear that mass media campaigns and primary care as a source of information are not completely effective – likely due to the complex determinants of health education and vector presence/absence that may differ at fine spatial scales. The community leaders are responsible for a certain amount of homes or families and can report disease risk or suspected cases to the program members. However, the communities may rely on the government for eradicating mosquitoes, and the migration of community leaders and normal residents complicate education. A Cali woman also stated that: "the people from the streets are not conscious, education is lacking" – corroborating with the KAP findings.

Sixth, storm sewers (sumideros) in Cali appear to be one of the main sources of *Aedes* breeding grounds and larval development. There are approximately 58,000 storm sewers in Cali, and they are constant targets for spraying programs and larval sampling. The storm sewers also seem to be the main vector breeding sites in wealthier neighborhoods (which was not examined in Chapter 4 due to unavailable data).

> "If Cali had a combined sewer system, we wouldn't have those breeding sites. I live in El Prado neighborhood in Cartago, and it is a rich neighborhood, so why are people getting sick? Living in that place, they do not go to work, they are confined in their homes, why are they getting sick? They do not have breeding sites in their homes. But the problem is that 10 meters of their house, 50 meters of their house they have a rainwater sumidero [storm sewer] and that is contributing to the proliferation of the *Culex* and *Aedes* populations. They come in through the window, through any place in your home. So, the problem in Cali is not going to be resolved [PHO6]".

Seventh, the interviews alluded to the fact that spraying is too infrequent; and does not necessarily reach the highest risk neighborhoods because public health officials may be too afraid or are not allowed in dangerous and violent areas. Quite simply, prevention against *Aedes* should start at the individual level to reduce the burden placed on the public health officials and policymakers (which was suggested by every interview participant). However, individuals and communities are not responsible for eradicating *Aedes* from the storm sewers, which is the job of the public health officials. The peridomestic, container-breeding nature of *Aedes* requires humans to create their habitats and be available as the hosts for a blood meal. Therefore, reducing available breeding sites at home is the first step of prevention. The participants also mentioned that fumigation is important for maintaining the moral and support of the communities because individuals feel that the public health officials are taking action against VBDs (despite the economic burden and ineffectiveness).

> "However, you know, all the studies show that the mosquitoes live inside of the houses. So, this is another thing that I thought, how ineffective it is to spend all this money fumigating the neighborhoods; and but, you know, if rains so hard, maybe you say, yeah it's ineffective. But if it rains not so hard, you know, the poison will stay in this little pool [PHO8]".

Another major effort to eradicate *Aedes* larvae is the use of guppy fish in stagnant pools of water, such as storm sewers, fountains, and rainwater boxes. However, the guppy fish strategy is low impact and high cost. They are often killed due to inhospitable conditions.

> "People washed their cars and bikes and killed the fish. Sand came with rain and asphyxiated the fish because of the gills. Any type of strong rain pushed them into the river. It is very expensive [PHO 4 and PHO5]".

Finally, the consensus recommended by the participants was individual awareness and education; and minimizing potential *Aedes* breeding sites and habitats inside and outside of the home. This would could reduce the economic burden of spraying and fish programs that attempt to eliminate mosquito larvae.

> "Stop the creation of breeding sites [PHO4]" and "mainly (reducing) breeding sites. Is not the spraying. It is not the nets. Each one contributes a percentage to the commitment and effectiveness. But, we need to attack the factory that are the breeding sites [PHO5]".

6.2.5 Participant suggestions for future research and interventions

All participants agreed that there is always need for better data at a finer-level, better educational campaigns, and more research, especially using spatial and space-time analysis. Tracking people's movement and defining activity spaces to examine the potential interaction between the hosts (humans), the vector (*Aedes*), and where people are getting infected. There was also a strong sense that more research and preventative efforts need to be made before an outbreak occurs.

> "Because we found here in work that we conducted in this administration that we will do an integral campaign, we compared the incidence of the disease three months before and three months after and found that there was a 60-70% reduction of the disease. This was only sustained for a month. Each month that went by the incidence will progressively increase and on the fourth month if you didn't do anything, after four months, again the incidence was the same as before the intervention [PHO7]".

Each participant expressed their great interest and support for KAP-type studies to gain a "baseline" and understand the perceptions of the at-risk individuals and communities. This should be done on a regular basis to follow-up and assess intervention campaigns. We also need to bridge the gap between academia and decision-makers. I strongly suggest that my KAP study in Chapter 5 is administered again in a year or two; determining if KAP has improved since the time of the initial surveys.

> "It is very important because the EPSs (healthcare centers) have programs on health promotion and prevention, but generally there are no resources to conduct studies on community perception and knowledge towards diseases. This is why it is so important to strengthen the relationship between academia and public institutions so the research results can be implemented [PHO2]".

Education needs to be continuous, and KAP studies can assess the improvement (or lack thereof) of the people's role in vector control. Many surveys and similar studies are conducted in Cali but are never disseminated to the public or health centers. Therefore, it is critical to share the results with everyone involved. It is also critical because of human migration and neighborhood change, which can substantially alter VBD risk and transmission.

6.2.6 Sketch maps of high-risk neighborhoods of VBDs

Figure 26 shows digitized versions of the sketch maps of at-risk areas of DENF, CHIK, and Zika. Each participant was more or less precise with their sketch maps, with some simply putting a star on the comunas, and others drawing imprecise circles around the general locations of high-risk. It is clear that the participants who completed a sketch map agree that the eastern part of Cali has the highest risk of VBDs, which has consistently had the highest rates of DENF, CHIK, and Zika.

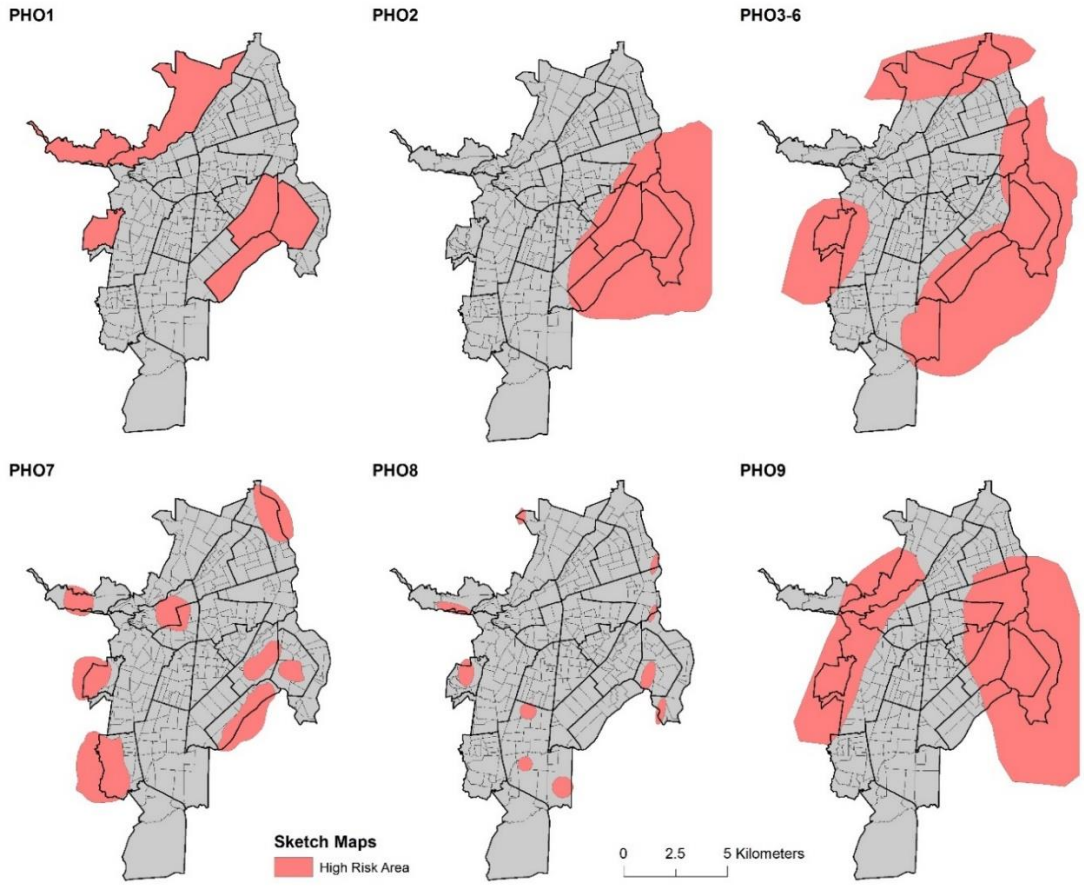


Figure 26: Participant sketch maps of perceived high-risk areas of DENF, CHIK, and Zika

Many comunas and neighborhoods along the western border were highlighted, as well as a few middle strata neighborhoods in the central part of Cali. However, the findings from the ST-CAR modeling in Chapter 4 provided strong evidence that the middle strata neighborhoods in the central location of Cali contain a high proportion of DENF, CHIK, and Zika cases. It is interesting to find that the PHO's did not prioritize those locations on the sketch map in Figure 26. Although not shown on the sketch maps, the PHO3 also mentioned that "universities are of high importance" because:

"they have very large populations. And populations that come and go. On vacation they come and go. And sometimes on the weekends. And they bring and take. And they also

get sick. And because of being so big they have green areas. For example, UniValle in Melendez, the Chiminango trees they are the best breeding site. It is magnificent. It is the one of the best breeding sites. UniValle that is located in a good area, that is attached to Ciudad Jardin, only has 200 rain water sumideros inside the university, plus all the students that come and go that is an enormous amount and that come from other municipalities that can bring different serotypes. Think of this variety of situations. Cali is super critical. And also what we know, people from neighboring municipalities come and contribute cases here".

PHO3's discussion of trees being suitable breeding sites for Aedes corroborates with Chapter 4 (higher tree density significantly increased disease risk). Furthermore, the higher strata, less urbanized neighborhoods in the very southern tip of Cali seem to be a major concern for many of the public health officials. This also highlights the importance of spatial and space-time studies to detect interaction between adjacent neighborhoods, and the mobility of the individuals who become infected with DENF, CHIK, or Zika.

Table 34 summarizes the number of neighborhoods that belonged to a high-risk area on the sketch maps in Figure 26. It is interesting to see the discrepancy between the quantity and recorded locations on the sketch maps for each of the six participants. PHO9's map included 136 total neighborhoods considered as high-risk; which also included the highest proportion of middle and high strata neighborhoods. PHO8's map contained the lowest number of neighborhoods and was also the most precise on the sketch map. PHO8's clinic does not treat DENF, CHIK, or Zika, so it makes sense that the other participants identified much more high-risk neighborhoods. The variation in the number of high-risk neighborhoods identified by the public health officials in government positions could be due to recall bias; locations that were being actively targeted for fumigation and surveillance; historically high-risk locations; and locations

with a large number of reported cases. However, the high-risk regions that were identified are all in the same general locations in Cali.

Table 34: Characteristics of neighborhoods belonging to a high risk area in the participant sketch maps

Participant	High Risk Neighborhoods (n)	Low Strata (n)	Middle Strata (n)	High Strata (n)	Not Ranked (n)
PHO9	136	73	38	18	7
PHO3-6	112	61	27	17	7
PHO2	81	62	16	0	3
PHO1	79	50	13	15	1
PHO7	59	48	10	0	1
PHO9	26	15	5	4	2

6.3 Discussion

This study developed and administered six semi-structured interviews regarding DENF, CHIK, and Zika with nine high-ranking public health officials in Cali, Colombia. Supplemental anecdotes were also provided by a woman who is a resident of the city. To my knowledge, this is the first qualitative study that interviewed major public health officials that are responsible for educational campaigns, VBD surveillance and control, and targeted interventions. The interviews provided key insight to the perceptions, surveillance strategies, and decision-making before, during, and after outbreaks of DENF, CHIK, and Zika; including the major public health burden and challenges faced when CHIK and Zika first arrived as emerging diseases in Cali between 2013 and 2014.

The first major finding is that there needs to be continuous educational programs that are available before, during, and after an outbreak of DENF, CHIK, and Zika. Bryan

et al. (1994) emphasize the importance of community participation in vector control programs, which are more cost-effective than top-down (vertical) government-based programs. Gubler and Clark (1994) and Healy et al. (2014) also provide evidence of *Aedes* habitat reduction when communities actively participate in vector control. Education programs should include lectures, community-based cleanup activities, improving mass media outreach (e.g. TV and radio campaigns); and integrating VBD awareness in primary and secondary schools can potentially promote behavioral changes (i.e. increase knowledge and prevention) in community health, since children can educate their parents and guardians about the risk factors (Deepthi et al. 2014).

Second, dynamic populations, especially daily human mobility and migrations in and out of Cali result in complex transmission dynamics of DENF, CHIK, and Zika; including difficulties tracking viral circulation of the three VBDs. Wesolowski et al. (2015) showed that using data from approximately 40 million mobile phones in Pakistan better predicted the spatial extent and temporal duration of DENF epidemics than predictive models without human mobility data. Examining the activity spaces of at-risk individuals using social media data, GPS, cell phone data, for example, can substantially improve the understanding of virus-vector-host interactions (Paz-Soldan et al. 2010; Gomide et al. 2011; Stoddard et al. 2013; de Almeida et al. 2017), which can reduce uncertainty of identifying high risk locations.

Third, the effectiveness of *Aedes* fumigation programs have been subject to debate. A major issue is the economic burden, lack of consistent spraying, and wide availability of breeding sites inside and outside of homes, buildings, and greenspaces (as mentioned in the interviews). More importantly, insecticide resistance is a major issue,

and studies have shown that both *Ae. aegypti* and *Ae. albopictus* have developed resistance to certain insecticides which reduces the ability to control the mosquito populations (Randson et al. 2010; Vontas et al. 2012). The most cost-effective approach to vector control is community empowerment via education and participation and understanding the ecological, biological, and social determinants of DENF, CHIK, and Zika (Lima et al. 2015). Therefore, more resources need to be allocated towards an integrated vector control approach that involves community participation.

Fourth, sumideros (storm sewers) are a major source of non-residential larval development sites for *Aedes* in both wet and dry seasons (Arana-Guardia et al. 2014) and should be a priority for vector surveillance and control. Storm sewer sites and larval samples could be added to the ST-CAR modeling described in Chapter 4 to quantify the risk of being in close proximity to sewers.As previously mentioned, an integrated vector control strategy that involves community participation may be the most cost-effective approach of mitigating *Aedes* prevalance, but the storm sewer issue is solely the local government's problem. The communities can do their part and still be at high risk of DENF, CHIK, or Zika transmission if the storm sewers are not properly fumigated.

Finally, the co-circulation of DENF, CHIK, and Zika in Cali, Colombia requires educational programs and research inititivates that prioritize all three diseases. As mentioned in the interviews, the three diseases were prioritized based on the number of cases reported during a particular time period. Since all three VBDs are transmitted by *Aedes*, it is critical allocate equal weight to the diseases for maximum prevention. Co-infection is also possible, that is, an individual can be infected with two or all three of the viruses at the same time; which can exacerbate preexisting conditions such as sickle cell

anemia (Rodriguez-Morales et al. 2016; Villamil-Gómez et al. 2016). Furthermore, if an individual gets infected with one of the three VBDs,.there is evidence that antibody-dependent enhancement can increase the risk of contracting another VBD (Dejnirattisai et al. 2016). Researchers should also examine the causes of co-circulation of DENF, CHIK, and Zika, such as vectorial capacity, human characteristics and behavior, and vector control strategies (Desjardins et al. 2018).

6.4 Conclusion

Understanding the perspectives, decisions, and knowledge of high-ranking public health officials can help researchers prioritize and refine project goals, methods, and scale of analysis. It is critical to gain insight about disease surveillance and control techniques being used in high risk locations of VBDs, especially in regions where there is co-circulation of DENF, CHIK, and Zika. The information gained from the interviews in this study can be combined with community health surveys (e.g. KAP studies) to understand both sides of the story. The realities that at-risk individuals and communities face may be different than the realities that public health officials believe to be true. For example, the PHO's did not seem to prioritize middle strata neighborhoods as high-risk areas; while the ST-CAR models and KAP results provided evidence that many strata neighborhoods are at high-risk and individual awareness of the three VBDs needs improvement. The findings suggest a need to allocate more resources to studies that evaluate targeted interventions and educational campaigns (i.e. follow-up studies); as well as training primary care physicians to better educate their patients about VBD prevention. The healthcare facilities are ultimately responsible for training their doctors

and should be actively promote VBD awareness and prevention techniques. Therefore, we can determine what is effective and what needs improvement.

It was clear from the interviews that the participants believe that the communities need to improve their role in vector control by taking more precautions and practicing more preventative measures against *Aedes*. However, there remains cultural and political barriers (e.g. storing water, the guarantee law, and budget issues) that underscores the persistent outbreaks of DENF; and potential of future CHIK and Zika epidemics. Finally, it is also critical to bridge the gap between academics and public health officials that actively make decisions regarding VBD surveillance and control. Academics can facilitate VBD education and surveillance by disseminating results to the public health officials and conducting more community health studies. Also, public health officials can provide important knowledge and context to the researchers, validate findings, and give insight on how to make academic research relevant and useful in their practice. It is my hope that this study empowers academics to work more closely with both the communities and stakeholders in their study areas to ensure that their research has a direct impact on positive public health outcomes. Overall, the community, the public health officials, the healthcare providers, and academics need to work together in a holistic way to contribute to VBD surveillance and prevention.

CHAPTER 7: GENERAL DISCUSSION AND CONCLUSIONS

In this dissertation, I supported my notion of "holistic" epidemiology by implementing a mixed-methods approach across four studies that improve vector-borne disease surveillance in Colombia, but also improves our understanding of the complex determinants of three emerging and reemerging VBDs that put billions at risk of transmission, globally. Chapter 3 identified and visualized space-time clusters of DENF and CHIK in Colombia at the national level between 2015 and 2016. Multivariate clusters were also computed to examine the co-occurrence of DENF and CHIK in space and time. The relative risk of at-risk municipalities belonging to the clusters were computed to facilitate targeted interventions and prioritize the highest risk regions.

Chapter 4 identified significant predictors of DENF, CHIK, and Zika risk at the neighborhood level between 2015 and 2016 in Cali, Colombia. ST-CAR models were developed to examine how disease risk in a target neighborhood is affected by disease rates and predictor variables in surrounding neighborhoods and time periods. Lagged weather variables were added as covariates to the models, which can be used as an early warning system to predict outbreaks in advance. The results showed disease risk for each neighborhood every week in the study period, for each disease. Therefore, the neighborhood-level risk of each disease can be tracked at a weekly level. Chapter 4 is the first of its kind to combine lagged weather variables and neighborhood-level risks at a weekly level to measure disease risk spatially and temporally across an entire city (Cali). The models can be utilized to explain "when" the outbreaks could occur, "where" they will likely be a high number of cases, and answer the "what" and "why" questions regarding what is influencing VBD transmission.

In the absence of a vaccine, prevention strategies are the most effective means of control. Therefore, understanding how people learn about the strategies to prevent the disease, how they apply them to their daily lives, and how effective they consider certain preventative strategies are important to help explain the spatial patterns of the disease. Chapter 5 administered KAP surveys to individuals in healthcare centers, universities, and neighborhoods (door-to-door) in Cali, Colombia during June of 2019. The main objective was to identify significant predictors of knowledge, attitudes, and preventative practices regarding DENF, CHIK, and Zika. The results were compared by neighborhood strata (low, middle, and high) and by disease.

The results will be shared with public health officials, stakeholders, and community leaders about improving vector-borne disease surveillance and control based on this study's findings. Therefore, the results of this chapter can propose alternatives to engage the community in taking an active role in vector control. Disseminating the results of a study to the public can also allow communities to provide feedback regarding the major findings, which can help refine research goals. The results of the KAP surveys can substantially improve targeted interventions and education programs that meet the specific needs of particular neighborhoods in Cali. Data and result sharing arrangements and agreements should be a part of the research process; and "by involving local communities in the spatial nature of [health] risk, behavioral changes to reduce susceptibility are more likely to be adopted, as traditionally participation in a health (or hazard mitigation) program increases whenever those at risk believe they are stakeholders in the process" (Mills and Curtis 2008, p. 69).

Chapter 6 conducted semi-structured interviews to high-ranking public health officials in Cali, Colombia regarding their experiences with DENF, CHIK, and Zika. A plethora of information was gathered and analyzed, which provided key insight into national and local educational programs, surveillance strategies, targeted interventions, and preventative measures; including their perception of the community's role in vector control. It was clear that the best preventative measure against the three VBDs is eliminating breeding sites in and outside of homes and storm sewers (sumideros).

Chapters 3-6 can be combined into a comprehensive and holistic mixed-methods study of vector-borne disease surveillance in Colombia. I was able to meet the suggestions provided in my conceptual framework of holistic spatial epidemiology. The first objective of this dissertation was identifying high-risk municipalities of DENF and CHIK using a space-time cluster detection approach. Since Cali was found in significant space-time clusters of both DENF and CHIK; it guided local-level analysis and confirmatory approaches presented in chapters 4-6. The next objective was identifying factors that influence DENF, CHIK, and Zika transmission in Cali. The ST-CAR approach allowed me to examine at-risk neighborhoods for each disease at the weekly level for a twoyear period. Adding the lagged weather variables also allows the prediction disease outbreaks (early warning system), which influences the larval development, gonotrophic cycle, and extrinsic incubation period of the *Aedes* mosquitoes.

Combining the results gained from the KAP and interview studies can supplement the modeling results by facilitating the explanation of key findings (Objective 3). I gained perspectives from both the individuals susceptible to the VBDs and the public

health officials responsible for mitigating the outbreaks and providing educational materials to the communities to empower them to take an active role in vector control. Some of the unexpected results produced by the ST-CAR models can be explained by the KAP and interviews.

For example, the models in chapter 4 suggested that PC1 (higher education, employed, and live in high strata neighborhoods) exhibited a positive relationship with disease risk. Although some studies provide evidence that high strata neighborhoods have larger yards and potentially less vector-control efforts carried out by the government, the interviews emphasized that the presence of storm sewers are primary habitats for *Aedes* in wealthier neighborhoods. Although the expected cases are low, the mosquitoes will still feed on human hosts in their home by flying through windows or doors, especially during the hottest part of the day. Many high strata neighborhoods are also adjacent to middle and low strata neighborhoods, which typically have a higher abundance of man-made breeding sites for *Aedes*. Furthermore, it was clear that university and public areas with less population density are at high-risk due to human movement and susceptibility to *Aedes* in these areas.

The KAP surveys and semi-structured interviews also allowed me to understand further complexities and predictors of DENF, CHIK, and Zika outbreaks that would have been impossible to uncover in Chapters 3 and 4. This includes the following: (1) KAP varies by neighborhood and for each individual, which can affect VBD transmission dynamics. (2) Accessibility to healthcare significantly affects the likelihood of taking preventative steps against *Aedes*. (3) Spraying is infrequent and often ineffective; and future research can compare the space-time patterns of spraying in Cali with VBD cases

per week, possibly quantifying the effectiveness of fumigation programs in various neighborhoods. (4) Storm sewers are a primary breeding site for *Aedes*; and adding sewer locations in the ST-CAR models may have improved predictions and risk estimations. (5) University areas and public spaces are high-risk areas due to human interaction and potential availability of breeding sites. (6) The significance of trees as suitable breeding sites detected in the ST-CAR models in Chapter 4 was corroborated with the PHO interviews in Chapter 6.

Conducting both community surveys and public health official interviews allowed me to shed light on the tension between the communities at-risk of VBDs, and the public health officials in charge of decision-making, surveillance, and control programs. These findings capture the complexities of VBD surveillance and community health issues that cannot be mapped, estimated, nor quantified as provided in chapters 3 and 4. In order to truly maxmimize preventative efforts against *Aedes* and viral transmission, it is worthwhile to understand the behaviors, decisions, and perspectives that are a part of the seemingly simple agent/vector/host process of VBD transmission. The KAP surveys and interviews highlighted the cultural, political, and educational barriers that either increase or decrease VBD risk and infection.

It was clear that reducing *Aedes* breeding sites inside and outside the home, as well as the storm sewers should be the primary control strategy for Cali – and I will also suggest that reducing breeding sites should be the priority anywhere that is at-risk of DENF, CHIK, or Zika. The most cost-effective way of reducing breeding sites is improving community participation in vector control; and improving access to healthcare resources (where physicians should be constantly trained to educate their patients about

being proactive, rather than getting infected and learning about the transmission dynamics). Although fumigation can be temporarily effective against larval development, it is too expensive and infrequent to effectively minimize VBD risk. My notion of *holistic spatial epidemiology* requires community and policy-maker interaction before, during, and after research studies regarding disease surveillance, control, and treatment. Transparency between everyone involved is vital, where monthly meetings with the community and public health officials can address knowledge gaps.. Ensuring that both the community and public health officials understand each other's realities necessary to achieve my holistic approach.

Finally, I suggest two major avenues that can be taken to reduce the burden of VBDs, not only in Cali, but in other locations experiencing outbreaks of DENF, CHIK, Zika, and others. First, it is critical to bridge the knowledge gaps between the community and public health officials/policymakers. I recommend that monthly or biweekly meetings take place in multiple areas of Cali that allow the communities and stakeholders to communicate and discuss the realities each are facing regarding disease surveillance, prevention, treatment, etc. This would allow the public health officials to present key findings and allow individuals of the communities to ask questions and state any concerns (see Section 2.2. and Figure 1). These meetings can also allow public health officials to administer KAP surveys that can be used as a baseline; then follow-up studies can be conducted to ensure that educational campaigns and targeted interventions are effective and improving. The meetings can also allow the communities to discuss their realities that the public health officials may not be aware of; which can facilitate research

objectives and consider aspects of community health that may have been previously hidden (e.g. cultural barriers).

Second, empowering communities to take an active role in vector control may require community-appointed leaders to disseminate information and report potential infected individuals to public health officials. It is fiscally impossible to monitor an entire at-risk region. Furthermore, there may be certain areas of a city that are too dangerous for public health officials to examine in person. Therefore, appointing community leaders that are trained (and potentially paid) to help with disease surveillance and control efforts can increase available resources and fine-level monitoring. The meetings discussed in the previous paragraph may be one place to disseminate findings and discuss necessary inteventions. However, an epidemic requires action on a daily basis; therefore, the community leaders can be equipped with technologies (e.g. phone and tablet apps, web resources, etc.) to directly inform public health officials of sick individuals, and the presence of vectors and vector breeding habitats.

Overall, educating the communities about their role in vector surveillance and control needs improvement, however, educating the public health officials about community realities and activity spaces is also required for improving educational materials and campaigns. It is my hope that this dissertation can be utilized to facilitate research proposals, improve community and government partnerships, and student training; ultimately ensuring that research conducted in spatial epidemiology is theoretically sound, practical, and contains short-term and long-term goals that directly address and improve community health and policy.

REFERENCES

Abellan, J. J., Richardson, S., & Best, N. (2008). Use of space–time models to investigate the stability of patterns of disease. *Environmental health perspectives*, *116*(8), 1111-1119.

Aitken, S.C., & Valentine, G. (2015). Ways of knowing and ways of doing geographic research. In: Aitken SC, Valentine, G (eds) *Approaches to Human Geography*, Second Edition, Thousand Oaks: SAGE, pp. 1-13.

Allison, P. D. (1999). *Multiple regression: A primer*. Pine Forge Press.

de Almeida Marques-Toledo, C., Degener, C. M., Vinhal, L., Coelho, G., Meira, W., Codeço, C. T., & Teixeira, M. M. (2017). Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting Dengue at country and city level. *PLoS neglected tropical diseases*, *11*(7), e0005729.

Alvis-Guzmán, N., Zakzuk-Sierra, J., Vargas-Moranth, R., Alcocer-Olaciregui, A., Parra-Padilla, D. (2017). Dengue: chikunguña y zika en Colombia 2015–2016. Rev. MVZ Córdoba 22 (Suppl), 5994–6003.

de Andrade, D. C., Jean, S., Clavelou, P., Dallel, R., & Bouhassira, D. (2010). Chronic pain associated with the Chikungunya Fever: long lasting burden of an acute illness. *BMC infectious diseases*, *10*(1), 31.

Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S. I., ... & Tominski, C. (2010). Space, time and visual analytics. *International Journal of Geographical Information Science*, *24*(10), 1577-1600.

Alobuia, W. M., Missikpode, C., Aung, M., & Jolly, P. E. (2015). Knowledge, attitude, and practices regarding vector-borne diseases in Western Jamaica. *Annals of global health*, *81*(5), 654-663.

Amin, R., Hendryx, M., Shull, M., & Bohnert, A. (2014). A cluster analysis of pediatric cancer incidence rates in Florida: 2000–2010. *Statistics and Public Policy*, *1*(1), 69-77.

Anselin, L. (2013). *Spatial econometrics: methods and models* (Vol. 4). Springer Science & Business Media.

Anselin, L., Lozano, N., & Koschinsky, J. (2006). Rate transformations and smoothing. *Urbana*, *51*, 61801.

Arana-Guardia, R., Baak-Baak, C. M., Loroño-Pino, M. A., Machain-Williams, C., Beaty, B. J., Eisen, L., & García-Rejón, J. E. (2014). Stormwater drains and catch basins as sources for production of Aedes aegypti and Culex quinquefasciatus. *Acta tropica*, *134*, 33-42.

Araujo, A. Q., Silva, M. T. T., & Araujo, A. P. (2016). Zika virus-associated neurological disorders: a review. *Brain*, *139*(8), 2122-2130.

Armstrong, M. P., Rushton, G., & Zimmerman, D. L. (1999). Geographically masking health data to preserve confidentiality. *Statistics in medicine*, *18*(5), 497-525.

Aswi, A., Cramb, S. M., Moraga, P., & Mengersen, K. (2019). Bayesian spatial and spatio-temporal approaches to modelling dengue fever: a systematic review. *Epidemiology & Infection*, *147*.

Auchincloss, A. H., Gebreab, S. Y., Mair, C., & Diez Roux, A. V. (2012). A review of spatial methods in epidemiology, 2000–2010. *Annual review of public health*, *33*, 107-122.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis* (Vol. 413). Essex: Longman Scientific & Technical.

Baldacchino, F., Caputo, B., Chandre, F., Drago, A., della Torre, A., Montarsi, F., & Rizzoli, A. (2015). Control methods against invasive Aedes mosquitoes in Europe: a review. *Pest management science*, *71*(11), 1471-1485.

Banerjee, S. (2018). Areal data models. https://sudipto.biostat.ucla.edu/ph8472/Areal.pdf [last accessed 1 April 2019].

Banu, S., Hu, W., Guo, Y., Naish, S., & Tong, S. (2014). Dynamic spatiotemporal trends of dengue transmission in the Asia-Pacific region, 1955–2004. *PLoS One*, *9*(2), e89440.

Bashar, K., Al-Amin, H. M., Reza, M. S., Islam, M., & Ahmed, T. U. (2012). Socio-demographic factors influencing knowledge, attitude and practice (KAP) regarding malaria in Bangladesh. *BMC Public Health*, *12*(1), 1084.

Bates, I., et al. (2004a). Vulnerability to malaria, tuberculosis, and HIV/AIDS infection and disease. Part 1: determinants operating at individual and household level. *The Lancet infectious diseases*, *4*(5), 267-277.

———. (2004b). Vulnerability to malaria, tuberculosis, and HIV/AIDS infection and disease. Part II: determinants operating at environmental and institutional level. *The Lancet infectious diseases*, 4(6), 368-375.

Baum, N. M., DesRoches, C., Campbell, E. G., & Goold, S. D. (2011). Resource allocation in public health practice: a national survey of local public health officials. *Journal of Public Health Management and Practice, 17*(3), 265-274.

Bazeley, P. (2004). Issues in mixing qualitative and quantitative approaches to research. *Applying qualitative methods to marketing management research*, *141*, 156.

Beckham, J. D., & Tyler, K. L. (2015). Arbovirus infections. *Continuum (Minneapolis, Minn.), 21*(6 NEUROINFECTIOUS DISEASE), 1599.

Becker, B., Leisnham, P. T., & LaDeau, S. L. (2014). A tale of two city blocks: Differences in immature and adult mosquito abundances between socioeconomically different urban blocks in Baltimore (Maryland, USA). *International journal of environmental research and public health, 11*(3), 3256-3270.

Benedict, M. Q., Levine, R. S., Hawley, W. A., & Lounibos, L. P. (2007). Spread of the tiger: global risk of invasion by the mosquito Aedes albopictus. *Vector-borne and zoonotic Diseases*, *7*(1), 76-85.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion), *J. Roy. Statist. Soc. Ser. B,* 36: 192-236.

Bhatt, S., et al. (2013). The global distribution and burden of dengue. Nature, 496(7446), 504-507.

Bleisch, S. (2012). 3D Geovisualization-Definition and Structures for the Assessment of Usefulness. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 129-134.

Bloch, D. (2016). The Cost and Burden of Chikungunya in the Americas. *Public Health Theses*.http://elischolar.library.yale.edu/ysphtdl/1022?utm_source=elischolar.library.yale.edu%2Fysphtdl%2F1022&utm_medium=PDF&utm_campaign=PDFCoverPages.

Braga, C., Luna, C. F., Martelli, C. M., De Souza, W. V., Cordeiro, M. T., Alexander, N., ...& Marques, E. T. (2010). Seroprevalence and risk factors for dengue infection in socio-economically distinct areas of Recife, Brazil. *Acta tropica*, *113*(3), 234-240.

Brasil, P., Calvet, G. A., Siqueira, A. M., Wakimoto, M., de Sequeira, P. C., Nobre, A., ... & Romero, C. (2016). Zika virus outbreak in Rio de Janeiro, Brazil: clinical characterization, epidemiological and virological aspects. *PLoS neglected tropical diseases*, *10*(4), e0004636.

Brunsdon, C., Corcoran, J., & Higgs, G. (2007). Visualising space and time in crime patterns: A comparison of methods. *Computers, Environment and Urban Systems*, *31*(1), 52-75.

Bryan, R. T., Balderrama, F., Tonn, R. J., & Dias, J. C. P. (1994). Community participation in vector control: lessons from Chagas' disease. *The American journal of tropical medicine and hygiene*, *50*(6_Suppl), 61-71.

Burra, T., Jerrett, M., Burnett, R. T., & Anderson, M. (2002). Conceptual and practical issues in the detection of local disease clusters: a study of mortality in Hamilton, Ontario. *The Canadian Geographer/Le Géographe Canadien*, *46*(2), 160-171.

Cali, S., (2010). Historia del dengue en Cali. Endemia o una continua epidemia. Secretaria de Salud Publica Municipal de Cali, Cali.

Calle-Giraldo, J. P., Rojas, C. A., Hurtado, I. C., Barco, C., Libreros, D., Sánchez, P. J., ... & Pinzón, E. (2019). Outcomes of Congenital Zika Virus Infection During an Outbreak in Valle del Cauca, Colombia. *The Pediatric infectious disease journal*, *38*(7), 735-740.

Calvet, G. A., Santos, F. B. D., & Sequeira, P. C. (2016). Zika virus infection: epidemiology, clinical manifestations and diagnosis. *Current opinion in infectious diseases*, *29*(5), 459-466.

Cameron, J. (2016). Focusing on the Focus Group. In: Hay, I. (ed) *Qualitative Research Methods in Human Geography*, Fourth Edition, Oxford University Press, pp. 203-224.

Campos, G. S., Bandeira, A. C., & Sardi, S. I. (2015). Zika virus outbreak, Bahia, Brazil. *Emerging infectious diseases*, *21*(10), 1885.

Carlin, B. P., Gelfand, A. E., & Banerjee, S. (2014). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.

Camacho, T., de la Hoz, F., Cárdenas, V., Sánchez, C., de Calderón, L., Pérez, L., & Bermúdez, A. (2004). Incomplete surveillance of a dengue-2 epidemic in Ibague, Colombia, 1995-1997. *Biomedica*, *24*(2), 174-182.

Casas, I., Delmelle, E., & Delmelle, E. C. (2017). Potential versus revealed access to care during a dengue fever outbreak. *Journal of Transport & Health*, *4*, 18-29.

Castañeda-Porras O., Segura, O., Garón-Lara, E. C., Manosalva-Sánchez, C. (2017). Conocimientos, actitudes y prácticas frente al control del vector Aedes aegypti, Villanueva Casanare, Colombia, 2016. *Rev. Méd. Risaralda*, 23 (1): 14 – 22.

Castrillón, J.C., Castaño J.C., Urcuqui S. (2015). Dengue en Colombia: diez años de evolución. *Rev Chil Infectol.* 32(2):142–9.

Castro, M., Pérez, D., Pérez, K., Polo, V., López, M., Sánchez, L. (2008). Contextualización de una estrategia comunitaria integrada para la prevención del dengue. *Rev Cubana Med Trop,* v.60 n.1.

Centers for Disease Control and Prevention. (2003). HIPAA privacy rule and public health. Guidance from CDC and the US Department of Health and Human Services. *MMWR: Morbidity and mortality weekly report*, *52*(Suppl. 1), 1-17.

Centers for Disease Control and Prevention. (2018). World Map of Areas with Risk of Zika. https://wwwnc.cdc.gov/travel/page/world-map-areas-with-zika [last accessed 9 Sept 2018].

Centers for Disease Control and Prevention (2018). Zika Symptoms, Testing, & Treatment. https://www.cdc.gov/zika/symptoms/symptoms.html [last accessed 28 Sept 2018].

Chacón-Duque, J. C., Adhikari, K., Avendaño, E., Campo, O., Ramirez, R., Rojas, W., ... & Bedoya, G. (2014). African genetic ancestry is associated with a protective effect on Dengue severity in colombian populations. *Infection, Genetics and Evolution*, *27*, 89-95

Charrel, R. N., de Lamballerie, X., & Raoult, D. (2007). Chikungunya outbreaks-the globalization of vectorborne diseases. *New England Journal of Medicine*, *356*(8), 769.

Chen, L., Bell, E. M., Caton, A. R., Druschel, C. M., & Lin, S. (2010). Residential mobility during pregnancy and the potential for ambient air pollution exposure misclassification. *Environmental research*, *110*(2), 162-168.

Chen, Lin H., and Davidson H. Hamer. "Zika virus: rapid spread in the western hemisphere". *Annals of internal medicine* 164.9 (2016): 613-615.

Cheng, T., & Williams, D. (2012). Space-time analysis of crime patterns in central London. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *39*, 47-52.

Cheng, T., & Wicks, T. (2014). Event detection using Twitter: a spatio-temporal approach. *PloS one*, *9*(6), e97807.

Chiu, C. H., Wen, T. H., Chien, L. C., & Yu, H. L. (2014). A probabilistic spatial dengue fever risk assessment by a threshold-based-quantile regression method. *PloS one*, *9*(10), e106334.

Chouin-Carneiro, T., et al. (2016). Differential Susceptibilities of Aedes aegypti and Aedes albopictus from the Americas to Zika Virus. *PLoS neglected tropical diseases*, *10*(3), e0004543.

Chu, H. J., Huang, B., & Lin, C. Y. (2015). Modeling the spatio-temporal heterogeneity in the PM10-PM2. 5 relationship. *Atmospheric Environment*, *102*, 176-182.

Chu, H. J., Lin, B. C., Yu, M. R., & Chan, T. C. (2016). Minimizing spatial variability of healthcare spatial accessibility—The case of a dengue fever outbreak. *International journal of environmental research and public health*, *13*(12), 1235.

Cissé, B., Luchini, S., & Moatti, J. P. (2007). Progressivity and horizontal equity in health care finance and delivery: What about Africa?. *Health policy*, *80*(1), 51-68.

Coleman, M., Coleman, M., Mabuza, A. M., Kok, G., Coetzee, M., & Durrheim, D. N. (2009). Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. *Malaria Journal*, *8*(1), 68.

Cope, M., & Kurtz, H. (2016). Organizing, coding and analyzing qualitative data. In: Clifford, N., Cope, M., Gillespie, T. & French, S. (eds) *Key methods in geography*, *36*(3), 647-64. Sage Publications.

Corrin, T., Waddell, L., Greig, J., Young, I., Hierlihy, C., & Mascarenhas, M. (2017). Risk perceptions, attitudes, and knowledge of chikungunya among the public and health professionals: a systematic review. *Tropical medicine and health*, *45*(1), 21.

Costa, F., Carvalho-Pereira, T., Begon, M., Riley, L., & Childs, J. (2017). Zoonotic and Vector-Borne Diseases in Urban Slums: Opportunities for Intervention. *Trends in parasitology*, *33*(9), 660-662.

Craine, J., & Gardner, C. (2016). Visual Methodology. In: Hay, I. (ed) *Qualitative Research Methods in Human Geography*, Fourth Edition, Oxford University Press, pp. 274-287.

Cressie, N., & Wikle, C. K. (2015). *Statistics for spatio-temporal data*. John Wiley & Sons

Creswell, J. W. (1999). Mixed-method research: Introduction and application. In *Handbook of educational policy* (pp. 455-472). Academic Press.

Cunha, R. V., & Trinta, K. S. (2017). Chikungunya virus: clinical aspects and treatment-A Review. *Memórias do Instituto Oswaldo Cruz*, *112*(8), 523-531.

DANE. (2018). Reloj de Población. Departamento Administrativo Nacional de Estadística. http://www.dane.gov.co/reloj/ [last accessed 19 Sep. 2018].

David, M. R., Lourenço-de-Oliveira, R., & Freitas, R. M. D. (2009). Container productivity, daily survival rates and dispersal of Aedes aegypti mosquitoes in a high income dengue epidemic neighbourhood of Rio de Janeiro: presumed influence of differential urban structure on mosquito biology. *Memórias do Instituto Oswaldo Cruz*, *104*(6), 927-932.

Dawaki, S., Al-Mekhlafi, H. M., Ithoi, I., Ibrahim, J., Atroosh, W. M., Abdulsalam, A. M., ... & Ahmed, A. (2016). Is Nigeria winning the battle against malaria? Prevalence, risk factors and KAP assessment among Hausa communities in Kano State. *Malaria journal*, *15*(1), 351.

Deepthi, R., Naresh Kumar, S. J., Prasanna Kamath, B. T., & Rajeshwari, H. (2014). Participatory school health education on vector-borne diseases: engaging children as change agents. *International Journal of Health Promotion and Education*, *52*(2), 68-77.

Delatte, H., Gimonneau, G., Triboire, A., & Fontenille, D. (2009). Influence of temperature on immature development, survival, longevity, fecundity, and gonotrophic cycles of Aedes albopictus, vector of chikungunya and dengue in the Indian Ocean. *Journal of medical entomology*, *46*(1), 33-41.

Dejnirattisai, W., et al. (2016). Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus. *Nature immunology*, *17*(9), 1102-1108.

Dick, G. W. A., Kitchen, S. F., & Haddow, A. J. (1952). Zika virus (I). Isolations and serological specificity. *Transactions of the royal society of tropical medicine and hygiene*, *46*(5), 509-520.

Dick, O. B., San Martín, J. L., Montoya, R. H., del Diego, J., Zambrano, B., & Dayan, G. H. (2012). The history of dengue outbreaks in the Americas. *The American journal of tropical medicine and hygiene*, *87*(4), 584-593.

Delmelle, E., Casas, I., Rojas, J. H., & Varela, A. (2013). Spatio-temporal patterns of Dengue Fever in Cali, Colombia. *International Journal of Applied Geospatial Research (IJAGR)*, *4*(4), 58-75.

Delmelle, E., Dony, C., Casas, I., Jia, M., & Tang, W. (2014). Visualizing the impact of space-time uncertainties on dengue fever patterns. *International Journal of Geographical Information Science*, *28*(5), 1107-1127.

Delmelle, E., Hagenlocher, M., Kienberger, S., & Casas, I. (2016). A spatial model of socioeconomic and environmental determinants of dengue fever in Cali, Colombia. *Acta tropica*, *164*, 169-176.

Demšar, U., & Virrantaus, K. (2010). Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science*, 24 (10), 1527-1542.

DeRoeck, D., Deen, J., & Clemens, J. D. (2003). Policymakers' views on dengue fever/dengue haemorrhagic fever and the need for dengue vaccines in four southeast Asian countries. *Vaccine*, *22*(1), 121-129.

Desjardins, M. R., Hohl, A., Griffith, A., & Delmelle, E. (2018a). A space–time parallel framework for fine-scale visualization of pollen levels across the Eastern United States. *Cartography and Geographic Information Science*, 1-13.

Desjardins, M. R., Whiteman, A., Casas, I., & Delmelle, E. (2018b). Space-time clusters and co-occurrence of chikungunya and dengue fever in Colombia from 2015 to 2016. *Acta tropica*, *185*, 77-85.

Desjardins, M.R., Hohl, A., Delmelle, E., & Casas, I. (2019). Identifying and visualizing space-time clusters of vector-borne diseases. In Geospatial Technology for Human Well-Being and Health. Springer. Forthcoming.

Devine, O. J., Louis, T. A., & Halloran, M. E. (1994). Empirical Bayes methods for stabilizing incidence rates before mapping. *Epidemiology*, 622-630.

Dewan, A. M., Corner, R., Hashizume, M., & Ongee, E. T. (2013). Typhoid Fever and its association with environmental factors in the Dhaka Metropolitan Area of Bangladesh: a spatial and time-series approach. *PLoS neglected tropical diseases*, *7*(1), e1998.

Diehr, P. (1984). Small area statistics: large statistical problems. *American Journal of Public Health*, *74*(4), 313-314.

Dixon, D.P., & Jones III, J.P. (2015). Feminist Geographies of Difference, Relation, and Construction. In: Aitkin, S. and G. Valentine (eds) *Approaches to Human Geography*, Second Edition, Thousand Oaks: SAGE, pp. 49-63.

Douglas, D. L., DeRoeck, D. A., Mahoney, R. T., & Wichmann, O. (2013). Will dengue vaccines be used in the public sector and if so, how? Findings from an 8-country survey of policymakers and opinion leaders. *PLoS neglected tropical diseases*, *7*(3), e2127.

Doyle, L., Brady, A. M., & Byrne, G. (2009). An overview of mixed methods research. *Journal of research in nursing*, *14*(2), 175-185.

Duczmal, L., & Assuncao, R. (2004). A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistics & Data Analysis*, *45*(2), 269-286.

Duffy, M. R., Chen, T. H., Hancock, W. T., Powers, A. M., Kool, J. L., Lanciotti, R. S., ... & Guillaumot, L. (2009). Zika virus outbreak on Yap Island, federated states of Micronesia. *New England Journal of Medicine*, *360*(24), 2536-2543.

Dunn, K. (2016). Interviewing. In: Hay, I. (ed) *Qualitative Research Methods in Human Geography*, Fourth Edition, Oxford University Press, pp. 149-188.

Durbin, A. P. (2016). Dengue antibody and Zika: friend or foe?. *Trends in immunology*, *37*(10), 635-636.

Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *The Annals of Mathematical Statistics*, 181-187.

Dzul-Manzanilla, F., Ibarra-López, J., Bibiano Marín, W., Martini-Jaimes, A., Leyva, J. T., Correa-Morales, F., ... & Vazquez-Prokopec, G. M. (2016). Indoor Resting Behavior of Aedes aegypti (Diptera: Culicidae) in Acapulco, Mexico. *Journal of medical entomology*, *54*(2), 501-504.

Eastin, M. D., Delmelle, E., Casas, I., Wexler, J., & Self, C. (2014). Intra-and interseasonal autoregressive prediction of dengue outbreaks using local weather and regional climate for a tropical environment in Colombia. *The American journal of tropical medicine and hygiene*, *91*(3), 598-610.

Eisen, L., & Eisen, R. J. (2011). Using geographic information systems and decision support systems for the prediction, prevention, and control of vector-borne diseases. *Annual review of entomology*, *56*, 41-61.

Eisen, L., & Lozano-Fuentes, S. (2009). Use of mapping and spatial and space-time modeling approaches in operational control of Aedes aegypti and dengue. *PLoS neglected tropical diseases*, *3*(4), e411.

Elliot, P., Wakefield, J. C., Best, N. G., & Briggs, D. J. (2000). *Spatial epidemiology: methods and applications*. Oxford University Press.

Elliott, P., & Wartenberg, D. (2004). Spatial epidemiology: current approaches and future challenges. *Environmental health perspectives*, *112*(9), 998-1006.

EPS C, (2015). El incremento del chikunguña es normal. Asi se comparta el virus. Boletin dePrensa COOMEVA EPS. No. 005-2015-1.

Fang, T.B., & Lu, Y. (2011). Constructing a Near Real-time Space-time Cube to Depict Urban Ambient Air Pollution Scenario. *Transactions in GIS*, 15 (5), 635-649.

Farinelli, E. C., Baquero, O. S., Stephan, C., & Chiaravalloti-Neto, F. (2018). Low socioeconomic condition and the risk of dengue fever: A direct relationship. *Acta Tropica*.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160.

Feldstein, L. R., Brownstein, J. S., Brady, O. J., Hay, S. I., & Johansson, M. A. (2015). Dengue on islands: a Bayesian approach to understanding the global ecology of dengue viruses. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *109*(5), 303-312.

Finley, A. O. (2011). Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, *2*(2), 143-154.

Floore, T. G. (2006). Mosquito larval control practices: past and present. *Journal of the American Mosquito Control Association*, *22*(3), 527-533.

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and planning A*, *30*(11), 1905-1927.

Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, *47*(4), 431-452.

Freisthler, B., & Weiss, R. E. (2008). Using Bayesian space-time models to understand the substance use environment and risk for being referred to child protective services. *Substance Use & Misuse*, *43*(2), 239-251.

Furuya-Kanamori, L., Liang, S., Milinovich, G., Magalhaes, R. J. S., Clements, A. C., Hu, W., ... & Yakob, L. (2016). Co-distribution and co-infection of chikungunya and dengue viruses. *BMC infectious diseases*, *16*(1), 84.

García-Betancourt, T., Higuera-Mendieta, D. R., González-Uribe, C., Cortés, S., & Quintero, J. (2015). Understanding water storage practices of urban residents of an endemic dengue area in Colombia: Perceptions, rationale and socio-demographic characteristics. *PloS one*, *10*(6), e0129054.

Gaudart, J., Poudiougou, B., Dicko, A., Ranque, S., Toure, O., Sagara, I., ... & Doumbo, O. K. (2006). Space-time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village. *BMC public health*, *6*(1), 286.

Gibbons, C. L., Mangen, M. J. J., Plass, D., Havelaar, A. H., Brooke, R. J., Kramarz, P., ... & Kretzschmar, M. E. (2014). Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC public health*, *14*(1), 147.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC.

Gimond, M. (2019). Point Pattern Analysis. https://mgimond.github.io/Spatial/point-pattern-analysis.html#first-and-second-order-effects [last accessed 1 April 2019].

Geller, L. (2001). Under the Weather: Climate, Ecosystems, and Infectious Disease. *Emerging Infectious Diseases*, *7*(7), 606-608. https://dx.doi.org/10.3201/eid0707.017750.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

George, K., & Stratford, E. (2016). Oral History and Human Geography. In: Hay, I. (ed) *Qualitative Research Methods in Human Geography*, Fourth Edition, Oxford University Press, pp. 189-201.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical analysis*, *24*(3), 189-206.

Goovaerts, P., & Jacquez, G. M. (2005). Detection of temporal changes in the spatial distribution of cancer rates using local Moran's I and geostatistically simulated spatial neutral models. *Journal of Geographical Systems*, *7*(1), 137-159.

Gomide, J., Veloso, A., Meira Jr, W., Almeida, V., Benevenuto, F., Ferraz, F., & Teixeira, M. (2011, June). Dengue surveillance based on a computational model of spatio-temporal locality of Twitter. In *Proceedings of the 3rd international web science conference* (p. 3). ACM.

Gopakumar, H., & Ramachandran, S. (2012). Congenital chikungunya. *Journal of clinical neonatology*, *1*(3), 155.

Gratz, N. G. (2004). Critical review of the vector status of Aedes albopictus. *Medical and veterinary entomology*, *18*(3), 215-227.

Greene, S. K., Huang, J., Abrams, A. M., Gilliss, D., Reed, M., Platt, R., ... & Kulldorff, M. (2012). Gastrointestinal disease outbreak detection using multiple data streams from electronic medical records. *Foodborne pathogens and disease*, *9*(5), 431-441.

Gubler, D. J. (1998). Dengue and dengue hemorrhagic fever. *Clinical microbiology reviews*, *11*(3), 480-496.

Gubler, D. J., & Clark, G. G. (1994). Community-based integrated control of Aedes aegypti: a brief overview of current programs. *The American journal of tropical medicine and hygiene*, *50*(6_Suppl), 50-60.

Guilland, A. (2016). Zika virus is a global public health emergency, declares WHO.

Guo, Y., Tang, Q., Gong, D. Y., & Zhang, Z. (2017). Estimating ground-level PM2. 5 concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sensing of Environment*, *198*, 140-149.

Hagenlocher, M., Delmelle, E., Casas, I., & Kienberger, S. (2013). Assessing socioeconomic vulnerability to dengue fever in Cali, Colombia: statistical vs expert-based modeling. *International journal of health geographics*, *12*(1), 36.

Hägerstrand, T. (1970). What about people in regional science?. In *Papers of the Regional Science Association* (Vol. 24, No. 1, pp. 6-21). Springer-Verlag.

Hanh, T. T. T., Hill, P. S., Kay, B. H., & Quy, T. M. (2009). Development of a framework for evaluating the sustainability of community-based dengue control projects. *The American journal of tropical medicine and hygiene*, *80*(2), 312-318.

Harrell, M. C., & Bradley, M. A. (2009). *Data collection methods. Semi-structured interviews and focus groups*. Rand National Defense Research Inst santa monica ca.

Hay, S. I., et al. (2013). Global mapping of infectious disease. *Phil. Trans. R. Soc. B*, *368*(1614), 20120250.

Harrington, L. C., Edman, J. D., & Scott, T. W. (2001). Why do female Aedes aegypti (Diptera: Culicidae) feed preferentially and frequently on human blood?. *Journal of medical entomology*, *38*(3), 411-422.

Healy, K., Hamilton, G., Crepeau, T., Healy, S., Unlu, I., Farajollahi, A., & Fonseca, D. M. (2014). Integrating the public in mosquito management: active education by community peers can lead to significant reduction in peridomestic container mosquito habitats. *PLoS One*, *9*(9), e108504.

Heintze, C., Garrido, M. V., & Kroeger, A. (2007). What do community-based dengue control programmes achieve? A systematic review of published evaluations. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *101*(4), 317-325.

Hennessey, M., Fischer, M., & Staples, J. E. (2016). Zika virus spreads to new areas—region of the Americas, May 2015–January 2016. *American Journal of Transplantation*, *16*(3), 1031-1034.

Hernández-Escolar, J., Consuegra-Mayor, C., Herazo-Beltrán, Y. (2014). Conocimientos, actitudes y prácticas sobre Dengue en un barrio de la ciudad de Cartagena de Indias. *Rev. Salud Pública.* Volumen 16, Número 2, p. 281-292, 2014. Electronic ISSN 2539-3596. Printed ISSN 0124-0064.

Hii, Y. L., Zaki, R. A., Aghamohammadi, N., & Rocklöv, J. (2016). Research on climate and dengue in Malaysia: a systematic review. *Current environmental health reports*, *3*(1), 81-90.

Hoarau, J. J., Bandjee, M. C. J., Trotot, P. K., Das, T., Li-Pat-Yuen, G., Dassa, B., ... & Tallet, F. (2010). Persistent chronic inflammation and infection by Chikungunya arthritogenic alphavirus in spite of a robust host immune response. *The Journal of Immunology*, ji_0900255.

Hohl, A. (2018). *Accelerating the Detection of Space-Time Patterns under Nonstationary Background Population*(Doctoral dissertation, The University of North Carolina at Charlotte).

Hohl, A., Delmelle, E., Tang, W., & Casas, I. (2016). Accelerating the discovery of space-time patterns of infectious diseases using parallel computing. *Spatial and spatio-temporal epidemiology*, *19*, 10-20.

Hopp, M. J., & Foley, J. A. (2001). Global-scale relationships between climate and the dengue fever vector, Aedes aegypti. *Climatic change*, *48*(2-3), 441-463.

Hu, W., Clements, A., Williams, G., Tong, S., & Mengersen, K. (2010). Bayesian spatiotemporal analysis of socio-ecologic drivers of Ross River virus transmission in Queensland, Australia. *The American journal of tropical medicine and hygiene*, *83*(3), 722-728.

Hu, M., Li, Z., Wang, J., Jia, L., Liao, Y., Lai, S., ... & Yang, W. (2012). Determinants of the incidence of hand, foot and mouth disease in China using geographically weighted regression models. *PloS one*, *7*(6), e38978.

Huang, L., Kulldorff, M., & Gregorio, D. (2007). A spatial scan statistic for survival data. *Biometrics*, *63*(1), 109-118.

Jacquez, G. M., Greiling, D. A., & Kaufmann, A. M. (2005). Design and implementation of a Space-Time Intelligence System for disease surveillance. *Journal of Geographical Systems*, *7*(1), 7-23.

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, *24*(3), 383-401.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of mixed methods research*, *1*(2), 112-133.

Jonsson, M. E., Heier, B. T., Norström, M., & Hofshagen, M. (2010). Analysis of simultaneous space-time clusters of Campylobacter spp. in humans and in broiler flocks using a multiple dataset approach. *International journal of health geographics*, *9*(1), 48.

Juliano, S. A., & Philip Lounibos, L. (2005). Ecology of invasive mosquitoes: effects on resident species and on human health. *Ecology letters*, *8*(5), 558-574.

Julious, S. A., Nicholl, J., & George, S. (2001). Why do we continue to use standardized mortality ratios for small area comparisons?. *Journal of Public Health*, *23*(1), 40 46.

Jung, I., Kulldorff, M., & Klassen, A. C. (2007). A spatial scan statistic for ordinal data. *Statistics in medicine*, *26*(7), 1594-1607.

Jung, I., Kulldorff, M., & Richard, O. J. (2010). A spatial scan statistic for multinomial data. *Statistics in medicine*, *29*(18), 1910-1918.

Kala, A. K., Tiwari, C., Mikler, A. R., & Atkinson, S. F. (2017). A comparison of least squares regression and geographically weighted regression modeling of West Nile virus risk based on environmental parameters. *PeerJ*, *5*, e3070.

Kalayanarooj, S. (2011). Clinical manifestations and management of dengue/DHF/DSS. *Tropical medicine and health*, *39*(4SUPPLEMENT), S83-S87.

Kawiecki, A. B., & Christofferson, R. C. (2016). Zika Virus–Induced Antibody Response Enhances Dengue Virus Serotype 2 Replication In Vitro. *The Journal of infectious diseases*, *214*(9), 1357-1360.

Kearns, R. (2016). Place Observation in the Research Toolkit. In: Hay, I. (ed) *Qualitative Research Methods in Human Geography*, Fourth Edition, Oxford University Press, pp. 313-333.

Khun, S., & Manderson, L. (2007). Health seeking and access to care for children with suspected dengue in Cambodia: an ethnographic study. *BMC Public Health*, *7*(1), 262.

Kirby, R. S., Delmelle, E., & Eberth, J. M. (2017). Advances in spatial epidemiology and geographic information systems. *Annals of epidemiology*, *27*(1), 1-9.

Kitchin, R. (2015). Positivist Geography. In: Aitkin, S. and G. Valentine (eds)

Kindon, S. (2016). Empowering Approaches: Participatory Action Research. In: Hay, I. (ed) *Qualitative Research Methods in Human Geography*, Fourth Edition, Oxford University Press, pp. 350-370.

Kitron, U., & Kazmierczak, J. J. (1997). Spatial analysis of the distribution of Lyme disease in Wisconsin. *American journal of Epidemiology*, *145*(6), 558-566.

Kitron, U. (1998). Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. *Journal of medical entomology*, *35*(4), 435-445.

Kitron, U. (2000). Risk Maps: Transmission and Burden of Vector-borne Diseases. *Parasitology today*, *16*(8), 324-325.

Krystosik, A. R., Curtis, A., Buritica, P., Ajayakumar, J., Squires, R., Dávalos, D., ... & James, M. A. (2017). Community context and sub-neighborhood scale detail to explain dengue, chikungunya and Zika patterns in Cali, Colombia. *PloS one*, *12*(8), e0181208.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics-Theory and methods*, *26*(6), 1481-1496.

Kulldorff, M. (2018). SaTScan™ user guide for version 9.6. [https://www.satscan.org/].

Kulldorff, M., Heffernan, R., Hartman, J., Assunçao, R., & Mostashari, F. (2005). A space–time permutation scan statistic for disease outbreak detection. *PLoS medicine*, *2*(3), e59.

Kulldorff, M. & Hjalmars, U. (1999). The Knox method and other tests for space-time interaction. *Biometrics,* 55**,** 544-552.

Kulldorff, M., Huang, L., & Konty, K. (2009). A scan statistic for continuous data based on the normal probability model. *International journal of health geographics*, *8*(1), 58.

Kulldorff, M., Mostashari, F., Duczmal, L., Katherine Yih, W., Kleinman, K., & Platt, R. (2007). Multivariate scan statistics for disease surveillance. *Statistics in medicine*, *26*(8), 1824-1833.

Krystosik, A., Curtis, A., LaBeaud, A., Dávalos, D., Pacheco, R., Buritica, P., ... & James, M. (2018). Neighborhood violence impacts disease control and surveillance: case study of Cali, Colombia from 2014 to 2016. *International journal of environmental research and public health*, *15*(10), 2144.

Kwan, M. P. (2004). Gis methods in time-geographic research: geocomputation and geovisualization of human activity patterns. *Geografiska Annaler: Series B, Human Geography*, *86*(4), 267-280.

LaBeaud, A. D. (2008). Why arboviruses can be neglected tropical diseases. *PLoS neglected tropical diseases*, *2*(6), e247.

LaDeau, S. L., Leisnham, P. T., Biehler, D., & Bodner, D. (2013). Higher mosquito production in low-income neighborhoods of Baltimore and Washington, DC: understanding ecological drivers and mosquito-borne disease risk in temperate cities. *International journal of environmental research and public health*, *10*(4), 1505-1526.

Laguna, F., Grillet, M. E., León, J. R., & Ludeña, C. (2017). Modelling malaria incidence by an autoregressive distributed lag model with spatial component. *Spatial and spatio-temporal epidemiology*, *22*, 27-37.

Launiala, A. (2009). How much can a KAP survey tell us about people's knowledge, attitudes and practices? Some observations from medical anthropology research on malaria in pregnancy in Malawi. *Anthropology Matters*, *11*(1).

Lawson, A. B. (2006). Disease cluster detection: a critique and a Bayesian proposal. *Statistics in medicine*, *25*(5), 897-916.

Lawson, A. B. (2013). *Statistical methods in spatial epidemiology*. John Wiley & Sons.

Lee, D., Rushworth, A., & Napier, G. (2018). Spatio-temporal areal unit modelling in R with conditional autoregressive priors using the CARBayesST package. *Journal of Statistical Software*, *84*(9).

Leparc-Goffart, I., Nougairede, A., Cassadou, S., Prat, C., & De Lamballerie, X. (2014). Chikungunya in the Americas. *The Lancet*, *383*(9916), 514.

Lever, J., Krzywinski, M., & Altman, N. (2017). Points of significance: Principal component analysis. *Nature Methods*, *14*, 641-642.

Li, J., Kolivras, K. N., Hong, Y., Duan, Y., Seukep, S. E., Prisley, S. P., ... & Gaines, D. N. (2014). Spatial and temporal emergence pattern of Lyme disease in Virginia. *The American journal of tropical medicine and hygiene*, *91*(6), 1166-1172.

Li, Z., Yin, W., Clements, A., Williams, G., Lai, S., Zhou, H., ... & Hu, W. (2012). Spatiotemporal analysis of indigenous and imported dengue fever cases in Guangdong province, China. *BMC infectious diseases*, *12*(1), 132.

Lian, M., Warner, R. D., Alexander, J. L., & Dixon, K. R. (2007). Using geographic information systems and spatial and space-time scan statistics for a population-based risk analysis of the 2002 equine West Nile epidemic in six contiguous regions of Texas. *International Journal of Health Geographics*, *6*(1), 42.

Lima, E. P., Goulart, M. O. F., & Neto, M. L. R. (2015). Meta-analysis of studies on chemical, physical and biological agents in the control of Aedes aegypti. *BMC public health*, *15*(1), 858.

Lindsay, S. W., Wilson, A., Golding, N., Scott, T. W., & Takken, W. (2017). Improving the built environment in urban areas to control Aedes aegypti-borne diseases. *Bulletin of the World Health Organization*, *95*(8), 607.

Liu-Helmersson, J., Stenlund, H., Wilder-Smith, A., & Rocklöv, J. (2014). Vectorial capacity of Aedes aegypti: effects of temperature and implications for global dengue epidemic potential. *PloS one*, *9*(3), e89783.

Lofgren, K. (2013). What is Epistemology? Available via http://www.streetarticles.com/science/what-is-epistemology [Accessed 20 Mar 2019].

Lorenzi, O. D. (2016). Reduced incidence of Chikungunya virus infection in communities with ongoing Aedes Aegypti mosquito trap intervention studies—Salinas and Guayama, Puerto Rico, November 2015–February 2016. *MMWR. Morbidity and mortality weekly report*, *65*.

Luo, R., Fongwen, N., Kelly-Cirino, C., Harris, E., Wilder-Smith, A., & Peeling, R. (2019). Rapid diagnostic tests for determining dengue serostatus: a systematic review and key informant interviews. *Clinical Microbiology and Infection*.

Lyra, P. P. R., Campos, G. S., Bandeira, I. D., Sardi, S. I., de Moura Costa, L. F., Santos, F. R., ... & Moreira, L. M. O. (2016). Congenital chikungunya virus infection after an outbreak in Salvador, Bahia, Brazil. *AJP reports*, *6*(3), e299.

Ma, X., Zhang, J., Ding, C., & Wang, Y. (2018). A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. *Computers, Environment and Urban Systems*, *70*, 113-124.

Maciel-de-Freitas, R., & Lourenço-de-Oliveira, R. (2009). Presumed unconstrained dispersal of Aedes aegypti in the city of Rio de Janeiro, Brazil. *Revista de saude publica*, *43*, 8-12.

Mangudo, C., Aparicio, J. P., & Gleiser, R. M. (2015). Tree holes as larval habitats for Aedes aegypti in urban, suburban and forest habitats in a dengue affected area. *Bulletin of entomological research*, *105*(6), 679-684.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, *27*(2 Part 1), 209-220.

Marinho, R. A., Beserra, E. B., Bezerra-Gusmão, M. A., Porto, V. D. S., Olinda, R. A., & dos Santos, C. A. (2016). Effects of temperature on the life cycle, expansion, and dispersion of Aedes aegypti (Diptera: Culicidae) in three cities in Paraiba, Brazil. *Journal of Vector Ecology*, *41*(1), 1-10.

Martins-Melo, F. R., Ramos, A. N., Alencar, C. H., Lange, W., & Heukelbach, J. (2012). Mortality of Chagas' disease in Brazil: spatial patterns and definition of high-risk areas. *Tropical Medicine & International Health*, *17*(9), 1066-1075.

Matthews, S. A. (2006, June). GeoDa and spatial regression modeling. In *Center for Spatially Integrated Social Science, Population Research Institute. University of California, Santa Barbara. http://ibis. geog. ubc. ca/* brian/workshop/GeoDa_Spatial_Regression. pdf. Santa Barbara, California, USA. [*Accessed 1 April 2019] (Vol. 27, p. 2017).

McGuirk, P.M., & O'Neill, P. (2016). Using Questionnaires in Qualitative Human Geography. In: Hay, I. (ed) *Qualitative Research Methods in Human Geography*, Fourth Edition, Oxford University Press, pp. 246-272.

de Melo, D. P. O., Scherrer, L. R., & Eiras, A. E. (2012). Dengue fever occurrence and vector detection by larval survey, ovitrap and MosquiTRAP: a space-time clusters analysis. *PloS one*, *7*(7), e42125.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, *29*(7), 897-910.

Merrill, R. M. (2015). *Introduction to epidemiology*. Jones & Bartlett Publishers.

Messina, J. P., Brady, O. J., Pigott, D. M., Golding, N., Kraemer, M. U., Scott, T. W., ... & Hay, S. I. (2015). The many projected futures of dengue. *Nature Reviews Microbiology*, *13*(4), 230.

Métras, R., Porphyre, T., Pfeiffer, D. U., Kemp, A., Thompson, P. N., Collins, L. M., & White, R. G. (2012). Exploratory space-time analyses of Rift Valley fever in South Africa in 2008–2011. *PLoS neglected tropical diseases*, *6*(8), e180.

Miller, H. J. (2005). A measurement theory for time geography. *Geographical analysis*, *37*(1), 17-45.

Mills JW, Curtis A. (2008) Geospatial approaches for disease risk communication in marginalized communities. Progress in community health partnerships: *Research, education, and Action*, 2(1): 61-72

Mills, J. N., Gage, K. L., & Khan, A. S. (2010). Potential influence of climate change on vector-borne and zoonotic diseases: a review and proposed research plan. *Environmental health perspectives*, *118*(11), 1507.

Miranda, E., Santos-Pinto, C., Antunes, C., Ferreira, L., & Osorio-de-Castro, C. (2019). Primary Care Pharmacist Interventions in Risk Reduction for the Zika Virus Epidemic: A Study in Campa Grande, Mato Grosso do Sul, Brazil. *Prehospital and Disaster Medicine*, *34*(s1), s62-s63.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17-23.

Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, *10*(2), 243-251.

Mueller, N., Rojas-Rueda, D., Basagaña, X., Cirach, M., Cole-Hunter, T., Dadvand, P., ... & Tonne, C. (2016). Urban and transport planning related exposures and mortality: a health impact assessment for cities. *Environmental health perspectives*, *125*(1), 89-96.

Mukhsar, Asrul Sani, Bahriddin Abapihi, Edi Cahyono. (2016a) Construction Posterior Distribution for Bayesian Mixed ZIP Spatio-Temporal Model. *International Journal of Biology and Biomedicine, 1*, 32-39.

Mukhsar, Abapihi, B., Sani, A., Cahyono, E., Adam, P., & Aini Abdullah, F. (2016b). Extended convolution model to bayesian spatio-temporal for diagnosing the DHF endemic locations. *Journal of Interdisciplinary Mathematics*, *19*(2), 233-244.

Mulatti, P., Mazzucato, M., Montarsi, F., Ciocchetta, S., Capelli, G., Bonfanti, L., & Marangon, S. (2015). Retrospective space–time analysis methods to support West Nile virus surveillance activities. *Epidemiology & Infection*, *143*(1), 202-213.

Mustafa, M. S., Rasotgi, V., Jain, S., & Gupta, V. (2015). Discovery of fifth serotype of dengue virus (DENV-5): A new public health dilemma in dengue control. *Medical Journal Armed Forces India*, *71*(1), 67-70.

Nagpal, B. N., Saxena, R., Srivastava, A., Singh, N., Ghosh, S. K., Sharma, S. K., ... & Ojha, V. P. (2012). Retrospective study of chikungunya outbreak in urban areas of India. *The Indian journal of medical research*, *135*(3), 351.

Naish, S., Hu, W., Mengersen, K., & Tong, S. (2011). Spatial and temporal clusters of Barmah Forest virus disease in Queensland, Australia. *Tropical Medicine & International Health*, *16*(7), 884-893.

Nakaya, T., Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in medicine, 24*(17), 2695-2717.

Nakaya, T. & Yano, K. (2010). Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics. *Transactions in GIS,* 14**,** 223-239.

Nakhapakorn, K., & Jirakajohnkool, S. (2006). Temporal and spatial autocorrelation statistics of dengue fever. *Dengue Bulletin, 30,* 177-183.

Nakkhara, P., Chongsuvivatwong, V., & Thammapalo, S. (2013). Risk factors for symptomatic and asymptomatic chikungunya infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 107*(12), 789-796.

Nsoesie, E. O., Ricketts, R. P., Brown, H. E., Fish, D., Durham, D. P., Mbah, M. L. N., ... & Owers, K. (2015). Spatial and temporal clustering of chikungunya virus transmission in Dominica. *PLoS neglected tropical diseases*, *9*(8), e0003977.

INS (2014). *Informe Quincenal Epidemiológico Nacional (IQEN).* Colombia: Instituto Nacional de Salud y Ministerio de Salud y Protección Social. https://www.ins.gov.co/buscador-eventos/IQEN/IQEN%20vol%2020%202015%20num%205.pdf

INS (2017). *Protocolo de Vigilancia en Salud Pública – Enfermedad por Virus Zika.* Colombia: Instituto Nacional de Salud y Ministerio de Salud y Protección Social. https://www.ins.gov.co/buscador-eventos/Lineamientos/PRO%20Zika.pdf

INS (2019). *Boletín Epidemiológico Semanal (BES) Chikungunya.* Colombia: Instituto Nacional de Salud y Ministerio de Salud y Protección Social. https://www.ins.gov.co/buscadoreventos/BoletinEpidemiologico/2019%20Bolet%C3%ADn%20epidemiol%C3%B3gico%20semana%2018.pdf

Krystosik, A. R., Curtis, A., Buritica, P., Ajayakumar, J., Squires, R., Davalos, D., ... & James, M. A. (2017). Community context and sub-neighborhood scale detail to explain dengue, chikungunya and Zika patterns in Cali, Colombia. *PLoS One*, *12*(8), e0181208.

Krystosik, A., Curtis, A., LaBeaud, A., Dávalos, D., Pacheco, R., Buritica, P., ... & James, M. (2018). Neighborhood violence impacts disease control and surveillance: case study of Cali, Colombia from 2014 to 2016. *International journal of environmental research and public health, 15*(10), 2144.

Launiala, A. (2009). How much can a KAP survey tell us about people's knowledge, attitudes and practices? Some observations from medical anthropology research on malaria in pregnancy in Malawi. *Anthropology Matters*, *11*(1).

Mustafa, M. S., Rasotgi, V., Jain, S., & Gupta, V. (2015). Discovery of fifth serotype of dengue virus (DENV-5): A new public health dilemma in dengue control. *Medical Journal Armed Forces India*, *71*(1), 67-70.

Nakkhara, P., Chongsuvivatwong, V., & Thammapalo, S. (2013). Risk factors for symptomatic and asymptomatic chikungunya infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *107*(12), 789-796.

Null, J. (2019). El Niño and La Niña Years and Intensities. https://ggweather.com/enso/oni.htm [last accessed 10 December 2019].

Ochieng, C., et al. (2013). Mosquito-borne arbovirus surveillance at selected sites in diverse ecological zones of Kenya; 2007–2012. *Virology journal*, *10*(1), 140.

Odongo, W., Ransom, R., & Henao, O. (2019). Informatics & Surveillance in Global Health: Informatics Capacity for Zika Outbreak. *Online Journal of Public Health Informatics*, *11*(1).

Oehler, E., Fournier, E., Leparc-Goffart, I., Larre, P., Cubizolle, S., Sookhareea, C., ... & Ghawche, F. (2015). Increase in cases of Guillain-Barré syndrome during a Chikungunya outbreak, French Polynesia, 2014 to 2015. *Eurosurveillance*, *20*(48), 30079.

Ooi, E. E., Goh, K. T., & Gubler, D. J. (2006). Dengue prevention and 35 years of vector control in Singapore. *Emerging infectious diseases*, *12*(6), 887.

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment and planning A*, *16*(1), 17-31.

Owusu, C., Lan, Y., Zheng, M., Tang, W., & Delmelle, E. (2017). Geocoding Fundamentals and Associated Challenges. In *Geospatial Data Science Techniques and Applications*. Karimi, H. A., & Karimi, B. (Eds.). CRC Press. 41-62.

Ostfeld, R. S., Glass, G. E., & Keesing, F. (2005). Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in ecology & evolution*, *20*(6), 328-336.

Paixão, E. S., Barreto, F., da Glória Teixeira, M., da Conceição N. Costa, M., & Rodrigues, L. C. (2016). History, epidemiology, and clinical manifestations of Zika: a systematic review. *American journal of public health*, *106*(4), 606-612.

Palaniyandi, M., Anand, P. H., & Pavendar, T. (2017). Environmental risk factors in relation to occurrence of vector borne disease epidemics: Remote sensing and GIS for rapid assessment, picturesque, and monitoring towards sustainable health. *International Journal of Mosquito Research*, *4*(3): 09-20.

Paupy, C., et al. (2010). Comparative role of Aedes albopictus and Aedes aegypti in the emergence of Dengue and Chikungunya in central Africa. *Vector-Borne and Zoonotic Diseases*, *10*(3), 259-266.

Paz-Soldan, V. A., Stoddard, S. T., Vazquez-Prokopec, G., Morrison, A. C., Elder, J. P., Kitron, U., ... & Scott, T. W. (2010). Assessing and maximizing the acceptability of global positioning system device use for studying the role of human movement in dengue virus transmission in Iquitos, Peru. *The American journal of tropical medicine and hygiene*, *82*(4), 723-730.

Pedigo, A., & Aldrich, T. (2011). Neighborhood disparities in stroke and myocardial infarction mortality: a GIS and spatial scan statistics approach. *BMC public health*, *11*(1), 644.

Perez, A., Alkhamis, M., Carlsson, U., Brito, B., Carrasco-Medanic, R., Whedbee, Z., & Willeberg, P. (2011). Global animal disease surveillance. *Spatial and spatio-temporal epidemiology*, *2*(3), 135-145.

Peterson, A. T. (2014). *Mapping Disease Transmission Risk: Geographic and Ecological Contexts*. Johns Hopkins University Press.

Pigott, D. M., et al. (2015). Prioritising infectious disease mapping. *PLoS neglected tropical diseases*, *9*(6), e0003756.

Polit, D. F., & Beck, C. T. (2010). Generalization in quantitative and qualitative research: Myths and strategies. *International journal of nursing studies*, *47*(11), 1451-1458.

Potter, A., Jardine, A., & Neville, P. J. (2016). A survey of knowledge, attitudes, and practices in relation to Mosquitoes and Mosquito-Borne disease in Western Australia. *Frontiers in public health*, *4*, 32.

Powell, J. R., & Tabachnick, W. J. (2013). History of domestication and spread of Aedes aegypti-A Review. *Memorias do Instituto Oswaldo Cruz*, *108*, 11-17.

Raina, S. (2013). Assessment of knowledge, attitude, and practice in health care delivery. *North American journal of medical sciences*, *5*(3), 249.

Rainwater-Lovett, K., Rodriguez-Barraquer, I., & Moss, W. J. (2016). Viral Epidemiology: Tracking Viruses with Smartphones and Social Media. In *Viral Pathogenesis* (pp. 241-252). Academic Press.

Rajapakse, S. (2011). Dengue shock. *Journal of Emergencies, Trauma and Shock*, *4*(1), 120.

Ranson, H., Burhani, J., Lumjuan, N., & Black IV, W. C. (2010). Insecticide resistance in dengue vectors. *TropIKA. net [online]*, *1*(1).

Roche, M. (2016). Historical Research and Archival Sources. In: Hay, I. (ed) *Qualitative Research Methods in Human Geography*, Fourth Edition, Oxford University Press, pp. 225-245.

Rochlin, I., Ninivaggi, D. V., Hutchinson, M. L., & Farajollahi, A. (2013). Climate change and range expansion of the Asian tiger mosquito (Aedes albopictus) in Northeastern USA: implications for public health practitioners. *PloS one*, *8*(4), e60874.

Rodriguez-Morales, A. J., Villamil-Gómez, W. E., & Franco-Paredes, C. (2016). The arboviral burden of disease caused by co-circulation and co-infection of dengue, chikungunya and Zika in the Americas. *Travel medicine and infectious disease*, *14*(3), 177.

Rogerson, P., & Yamada, I. (2008). *Statistical detection and surveillance of geographic clusters*. CRC Press.

Rosenberg, M. W. (1998). Medical or health geography? Populations, peoples and places. *International Journal of Population Geography*, *4*(3), 211.

Ruger, J. P., & Kim, H. J. (2006). Global health inequalities: an international comparison. *Journal of Epidemiology & Community Health*, *60*(11), 928-936.

Rushworth, A., Lee, D., & Mitchell, R. (2014). A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. *Spatial and spatio-temporal epidemiology*, *10*, 29-38.

Sagl, G., Delmelle, E., & Delmelle, E. (2014). Mapping collective human activity in an urban environment based on mobile phone data. *Cartography and Geographic Information Science*, *41*(3), 272-285.

Samuel, G., DiBartolo-Cordovano, R., Taj, I., Merriam, A., Lopez, J. M., Torres, C., ... & Thakur, K. T. (2018). A survey of the knowledge, attitudes and practices on Zika virus in new York City. *BMC public health*, *18*(1), 98.

Sani, A., Abapihi, B., Mukhsar, M., & Kadir, K. (2015). Relative risk analysis of dengue cases using convolution extended into spatio-temporal model. *Journal of Applied Statistics*, *42*(11), 2509-2519.

Sarmiento-Senior D, Matiz M. I., Jaramillo-Gómez JF, Olano, V. A., Vargas, S. L., Alexander N., Lenhart, A., Stenström, T. H., Overgaard, H. J. (2019). Knowledge, attitudes, and practices about dengue among pupils from rural schools in an endemic area in Colombia. *Biomedica,* 39:478-90.

Saule, E., Panchananam, D., Hohl, A., Tang, W., & Delmelle, E. (2017, August). Parallel space-time kernel density estimation. In *2017 46th International Conference on Parallel Processing (ICPP)* (pp. 483-492). IEEE.

Schmidt, W. P., Suzuki, M., Thiem, V. D., White, R. G., Tsuzuki, A., Yoshida, L. M., ... & Ariyoshi, K. (2011). Population density, water supply, and the risk of dengue fever in Vietnam: cohort study and spatial analysis. *PLoS medicine*, *8*(8), e1001082.

Seamon, D. (2015). Lived Emplacement and the Locality of Being: A Return to Humanistic Geography? In: Aitkin, S. and G. Valentine (eds) *Approaches to Human Geography*, Second Edition, Thousand Oaks: SAGE, pp. 35-48.

Semenza, J. C., Suk, J. E., Estevez, V., Ebi, K. L., & Lindgren, E. (2012). Mapping climate change vulnerabilities to infectious diseases in Europe. *Environmental health perspectives*, *120*(3), 385.

Shaver, B. (2017). A Zero-Math Introduction to Markov Chain Monte Carlo Methods. *Towards Data Science*. https://towardsdatascience.com/a-zero-math-introduction-to-markov-chain-monte-carlo-methods-dcba889e0c50 [accessed 11 April 2019].

Shay, E., Combs, T. S., Findley, D., Kolosna, C., Madeley, M., & Salvesen, D. (2016). Identifying transportation disadvantage: Mixed-methods analysis combining GIS mapping with qualitative data. *Transport Policy*, *48*, 129-138.

Shepard, D. S. (2010). Cost and Burden of Dengue and Chikungunya from the Americas to Asia. *Dengue Bulletin* – 34. http://apps.who.int/iris/bitstream/10665/170991/1/db2010v34p1.pdf.

Shepard, D. S., Undurraga, E. A., Betancourt-Cravioto, M., Guzmán, M. G., Halstead, S. B., Harris, E., ... & Gubler, D. J. (2014). Approaches to refining estimates of global burden and economics of dengue. *PLoS neglected tropical diseases*, *8*(11), e3306.

Shepard, D. S., Undurraga, E. A., Halasa, Y. A., & Stanaway, J. D. (2016). The global economic burden of dengue: a systematic analysis. *The Lancet infectious diseases*, *16*(8), 935-941.

Silva Jr, J. V., Ludwig-Begall, L. F., de Oliveira-Filho, E. F., Oliveira, R. A., Durães-Carvalho, R., Lopes, T. R., ... & Gil, L. H. (2018). A scoping review of Chikungunya virus infection: epidemiology, clinical characteristics, viral co-circulation complications, and control. *Acta tropica*.

Silva, M. M., Rodrigues, M. S., Paploski, I. A., Kikuti, M., Kasper, A. M., Cruz, J. S., ... & Ko, A. I. (2016). Accuracy of dengue reporting by national surveillance system, Brazil. *Emerging infectious diseases*, *22*(2), 336.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). CRC press.

Simon, F., Javelle, E., Oliver, M., Leparc-Goffart, I., & Marimoutou, C. (2011). Chikungunya virus infection. *Current infectious disease reports*, *13*(3), 218.

Silva Jr, J. V., Ludwig-Begall, L. F., de Oliveira-Filho, E. F., Oliveira, R. A., Durães-Carvalho, R., Lopes, T. R., ... & Gil, L. H. (2018). A scoping review of Chikungunya virus infection: epidemiology, clinical characteristics, viral co-circulation complications, and control. *Acta tropica*.

Sissoko, D., Malvy, D., Ezzedine, K., Renault, P., Moscetti, F., Ledrans, M., & Pierre, V. (2009). Post-epidemic Chikungunya disease on Reunion Island: course of rheumatic manifestations and associated factors over a 15-month period. *PLoS neglected tropical diseases*, *3*(3), e389.

de Smith, M., Longley, P., and Goodchild, M. (2015). Spatial autoregressive and Bayesian modeling. In *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools , 5th Edition*. Matador: Leicester, United Kingdom.

Snow, J. (1856). On the mode of communication of cholera. *Edinburgh medical journal*, *1*(7), 668.

Srikiatkhachorn, A., Gibbons, R. V., Green, S., Libraty, D. H., Thomas, S. J., Endy, T. P., ... & Nimmannitaya, S. (2010). Dengue hemorrhagic fever: the sensitivity and specificity of the world health organization definition for identification of severe cases of dengue in Thailand, 1994–2005. *Clinical Infectious Diseases*, *50*(8), 1135-1143.

St. John, M., Cowen, M. B., Smallman, H. S., & Oonk, H. M. (2001). The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors*, *43*(1), 79-98.

Stoddard, S. T., Forshey, B. M., Morrison, A. C., Paz-Soldan, V. A., Vazquez-Prokopec, G. M., Astete, H., ... & Kochel, T. J. (2013). House-to-house human movement drives dengue virus transmission. *Proceedings of the National Academy of Sciences*, *110*(3), 994-999.

Stuckey, H. L. (2013). Three types of interviews: Qualitative research methods in social health. *Journal of Social Health and Diabetes*, *1*(02), 056-059.

Suarez, M. R., Olarte, S. M. F., Ana, M. F. A., & González, U. C. (2005). Is what I have just a cold or is it dengue? Addressing the gap between the politics of dengue control and daily life in Villavicencio-Colombia. *Social Science & Medicine*, *61*(2), 495-502.

Tango, T., & Takahashi, K. (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International journal of health geographics*, *4*(1), 11.

Tauil, P. L. (2001). Urbanization and dengue ecology. *Cadernos de saúde pública*, *17*, S99-S102.

Tedesco, C., Ruiz, M., & McLafferty, S. (2010). Mosquito politics: local vector control policies and the spread of West Nile Virus in the Chicago region. *Health & Place*, *16*(6), 1188-1195.

Tellman, N., Litt, E. R., Knapp, C., Eagan, A., Cheng, J., & Lewis Jr, J. (2010). The effects of the Health Insurance Portability and Accountability Act privacy rule on influenza research using geographical information systems. *Geospatial health*, *5*(1), 3-9.

Terrell, S. R. (2012). Mixed-methods research methodologies. *The qualitative report*, *17*(1), 254-280.

Thakur, S., & Hanson, A. J. (2010). A 3D visualization of multiple time series on maps. In *Information Visualisation (IV), 2010 14th International Conference* (pp. 336-343). IEEE.

Thammapalo, S., Chongsuvivatwong, V., Geater, A., & Dueravee, M. (2008). Environmental factors and incidence of dengue fever and dengue haemorrhagic fever in an urban area, Southern Thailand. *Epidemiology & Infection*, *136*(1), 135-143.

Theophilides, C. N., Ahearn, S. C., Grady, S., & Merlino, M. (2003). Identifying West Nile virus risk areas: The dynamic continuous-area space-time system. *American Journal of Epidemiology*, *157*(9), 843-854.

Thiberville, S. D., Moyen, N., Dupuis-Maguiraga, L., Nougairede, A., Gould, E. A., Roques, P., & de Lamballerie, X. (2013). Chikungunya fever: epidemiology, clinical syndrome, pathogenesis and therapy. *Antiviral research*, *99*(3), 345-370.

Toan, N. T., Rossi, S., Prisco, G., Nante, N., & Viviani, S. (2015). Dengue epidemiology in selected endemic countries: factors influencing expansion factors as estimates of underreporting. *Tropical Medicine & International Health*, *20*(7), 840-863.

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography* 46 (sup1):234-240.

Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E., & Higgs, S. (2007). A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS pathogens*, *3*(12), e201.

Trgovac, A. B., Kedron, P. J., & Bagchi-Sen, S. (2015). Geographic variation in male suicide rates in the United States. *Applied Geography*, *62*, 201-209.

Udayanga, L., Gunathilaka, N., Iqbal, M. C. M., Pahalagedara, K., Amarasinghe, U. S., & Abeyewickreme, W. (2018). Socio-economic, Knowledge Attitude Practices (KAP), household related and demographic based appearance of non-dengue infected individuals in high dengue risk areas of Kandy District, Sri Lanka. *BMC infectious diseases*, *18*(1), 88.

Uittenbogaard, A., & Ceccato, V. (2012). Space-time clusters of crime in Stockholm, Sweden. *Review of European Studies*, *4*(5), 148.

Ullah, S., Daud, H., Dass, S. C., Khan, H. N., & Khalil, A. (2017). Detecting space-time disease clusters with arbitrary shapes and sizes using a co-clustering approach. *Geospatial health*, *12*(2).

United Nations. (2018). 2018 Revision of World Urbanization Prospects. https://population.un.org/wup/ [last accessed 19 Sept. 2018].

Valle del Cauca S. D. d. S. (2019). *Autoridades de salud estudian probable caso de dengue grave, tras fallecimiento de niña de 5 años*. Valle del Cauca: Secretaria Departamental de Salud.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS quarterly*, 21-54.

Ventura, J., Nuechterlein, K. H., Subotnik, K. L., Gutkind, D., & Gilbert, E. A. (2000). Symptom dimensions in recent-onset schizophrenia and mania: a principal components analysis of the 24-item Brief Psychiatric Rating Scale. *Psychiatry research*, *97*(2-3), 129-135.

Ver Hoef, J. M., Hanks, E. M., & Hooten, M. B. (2018). On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. *Spatial statistics*, *25*, 68-85.

Villamil-Gómez, W. E., González-Camargo, O., Rodriguez-Ayubi, J., Zapata-Serpa, D., & Rodriguez-Morales, A. J. (2016). Dengue, chikungunya and Zika co-infection in a patient from Colombia. *Journal of infection and public health*, *9*(5), 684-686.

Vincenti-Gonzalez, M. F., Tami, A., Lizarazo, E. F., & Grillet, M. E. (2018). ENSO-driven climate variability promotes periodic major outbreaks of dengue in Venezuela. *Scientific reports*, *8*(1), 5727.

Vontas, J., Kioulos, E., Pavlidi, N., Morou, E., Della Torre, A., & Ranson, H. (2012). Insecticide resistance in the major dengue vectors Aedes albopictus and Aedes aegypti. *Pesticide Biochemistry and Physiology*, *104*(2), 126-131.

Waitt, G. (2016). Doing Foucauldian Discourse Analysis-Revealing Social Identities. In: Hay, I. (ed) *Qualitative Research Methods in Human Geography*, Fourth Edition, Oxford University Press, pp. 288-312.

Waller, L. A., Carlin, B. P., Xia, H., & Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical association*, *92*(438), 607-617.

Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data* (Vol. 368). John Wiley & Sons.

Waller, L. A., Zhu, L., Gotway, C. A., Gorman, D. M., & Gruenewald, P. J. (2007). Quantifying geographic variations in associations between alcohol distribution and violence: a comparison of geographically weighted regression and spatially varying coefficient models. *Stochastic Environmental Research and Risk Assessment*, *21*(5), 573-588.

Wang, M. (2018). Spatio-temporal Statistical Modeling: Climate Impacts due to Bioenergy Crop Expansion. *Dissertation*, Arizona State University (May 2018).

Wang, H, et al. (2016). Global, Regional, and National Life Expectancy, All-Cause Mortality, and Cause-Specific Mortality for 249 Causes of Death, 1980-2015: A Systematic Analysis for the Global Burden of Disease Study 2015. *The Lancet* 388 (10053): 1459–1544.

Wang, J., Cao, Z., Zeng, D. D., Wang, Q., Wang, X., & Qian, H. (2014). Epidemiological analysis, detection, and comparison of space-time patterns of Beijing hand-foot-mouth disease (2008–2012). *PLoS one*, *9*(3), e92745.

Weaver, S. C., & Lecuit, M. (2015). Chikungunya virus and the global spread of a mosquito-borne disease. *New England Journal of Medicine*, *372*(13), 1231-1239.

Wesolowski, A., Qureshi, T., Boni, M. F., Sundsøy, P. R., Johansson, M. A., Rasheed, S. B., ... & Buckee, C. O. (2015). Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proceedings of the National Academy of Sciences*, *112*(38), 11887-11892.

Whiteman, A., Desjardins, M.R., Eskildsen, G.A., & Loaiza, J.R. (2019). Integrating vector surveillance data to improve space-time risk estimation of dengue fever in Panama. *PLOS Neglected Tropical Diseases.* 13(9): e0007266.

Whiteman, A., Mejia, A., Hernandez, I., & Loaiza, J. R. (2018). Socioeconomic and demographic predictors of resident knowledge, attitude, and practice regarding arthropod-borne viruses in Panama. *BMC public health*, *18*(1), 1261.

Whiteman, A. (2018). *Socioeconomic Variation and Aedes Mosquitoes: An Examination of Vector-Borne Disease Risk in the Urban Environment* (Doctoral dissertation, The University of North Carolina at Charlotte).

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 434-449.

Wilcox, B.A. (2005). Disease ecology and the global emergence of zoonotic pathogens *Environmental Health and Preventive Medicine*, *10*(5), 263-272.

Wilder-Smith, A., Gubler, D. J., Weaver, S. C., Monath, T. P., Heymann, D. L., & Scott, T. W. (2017). Epidemic arboviral diseases: priorities for research and public health. *The Lancet infectious diseases*, *17*(3), e101-e106.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, *30*(1), pp. 79-82.

Wilson, J. J., & Sevarkodiyone, S. P. (2017). Household survey of dengue and chikungunya vectors (Aedes aegypti and Aedes albopictus)(Diptera: Culicidae) in Tirunelveli district, Tamil Nadu, India. *Journal of Entomological Research*, *41*(3), 317-324.

Winchester, H.P.M., & Rofe, M.W. (2016). Qualitative Research and Its Place in Human Geography. In: Hay, I. (ed) *Qualitative Research Methods in Human Geography*, Fourth Edition, Oxford University Press, pp. 1-28.

Winders, J. (2016). New Media. In: Hay, I. (ed) *Qualitative Research Methods in Human Geography*, Fourth Edition, Oxford University Press, pp. 334-349.

World Health Organization. (2011). Comprehensive guideline for prevention and control of dengue and dengue haemorrhagic fever.

World Health Organization (2012). Communication for behavioural impact (COMBI): A toolkit for behavioural and social communication in outbreak response. https://www.who.int/ihr/publications/combi_toolkit_outbreaks/en/ [last accessed 10 December 2019].

World Health Organization. (2014a). A Global Brief on Vector-Borne Diseases. *World Health Organization*, 9. doi:WHO/DCO/WHD/2014.1.

———. (2014b). "WHO Factsheet Vector-Borne Diseases." *Factsheet Number 387*, no. March: 10. http://www.who.int/kobe_centre/mediacentre/vbdfactsheet.pdf.

World Health Organization (2018). Dengue and severe dengue. http://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue [last accessed 28 Sept 2018].

Wrenn, D. H., & Sam, A. G. (2014). Geographically and temporally weighted likelihood regression: Exploring the spatiotemporal determinants of land use change. *Regional Science and Urban Economics*, *44*, 60-74.

Wu, L., Deng, F., Xie, Z., Hu, S., Shen, S., Shi, J., & Liu, D. (2016). Spatial analysis of severe fever with thrombocytopenia syndrome virus in China using a geographically weighted logistic regression model. *International journal of environmental research and public health*, *13*(11), 1125.

Wu, P. C., Lay, J. G., Guo, H. R., Lin, C. Y., Lung, S. C., & Su, H. J. (2009). Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. *Science of the total Environment*, *407*(7), 2224-2233.

Xiang, W. N. (2017). Pasteur's quadrant: an appealing ecophronetic alternative to the prevalent Bohr's quadrant in ecosystem services research. *Landscape Ecology*, 1-7.

Yen, I. H., & Syme, S. L. (1999). The social environment and health: a discussion of the epidemiologic literature. *Annual review of public health*, *20*(1), 287-308.

Yin, R. K. (2013). Chapter 1: Getting Started. In: *Case study research: Design and methods*. Sage publications. 3-25.

Zandbergen, P. A. (2014). Ensuring confidentiality of geocoded health data: assessing geographic masking strategies for individual-level data. *Advances in medicine*, *2014*.

Zhang, Z., Jackson, S., Merritt, R., Gillespie, C., & Yang, Q. (2019). Association between cardiovascular health metrics and depression among US adults: National Health and Nutrition Examination Survey, 2007–2014. *Annals of epidemiology*.

APPENDIX 1: KAP SURVEY (ENGLISH)

ID#_____ Date____/_____/_____

Knowledge, Attitudes, and Practices (KAP) of dengue, chikungunya, and Zika in Cali, Colombia

Sociodemographic Information	
Sex	☐ Male ☐ Female
Age	☐ 18-35 ☐ 36-55 ☐ 56-70 ☐ 70+
Address of home	_____
Neighborhood	_____
Education level	☐ None ☐ Primary ☐ Secondary ☐ Technical ☐ Undergraduate ☐ Postgraduate
Monthly Income	[Insert range COP] _____
Occupation	☐ Full-time ☐ Housewife ☐ Part-time ☐ Unemployed ☐ Independent ☐ Student ☐ Pension
Civil Status	☐ Married ☐ Separated/Divorced ☐ Single ☐ Widowed ☐ Free Union ☐ Other
Do you have children who live in your home?	☐ Yes ☐ No
Number of people living in your home (including yourself)	☐ 1 ☐ 2 ☐ 3-4 ☐ 5-6 ☐ 6+

Race	□ White □ Indigenous □ Mestizo □ African-Colombian □ Asian □ Other:_____

General Questions

Have you ever had the following diseases? Check all that apply.	□ Dengue □ Chikungunya □ Zika □ None
Do you know information about the following diseases? Check all that apply.	□ Dengue □ Chikungunya □ Zika □ None
How would you rate your access to health services?	□ Excellent □ Very good □ Good □ Fair □ Poor
On average, how long does it take to get to a health service?	□ less than 5 minutes □ 5-15 minutes □ 15-30 minutes □ 30-60 minutes □ 1 hour or more
Due to the distance of the health service:	1. Have you stopped seeking medical treatment or going to the doctor? □ Yes □ No □ I'm not sure 2. Have you refused medical treatment? □ Yes □ No □ I'm not sure

Knowledge

What diseases are you familiar with? Check all that apply.	□ Dengue □ Chikungunya □ Zika
Do you know how dengue is transmitted?	□ Si □ No If yes, how?: _____

Do you know how chikungunya is transmitted?	□ Si □ No If yes, how?: _____
Do you know how Zika is transmitted?	□ Si □ No If yes, how?: _____
By what means did you learn about the following diseases? Check all that apply.	Dengue □ Television □ Radio □ Newspapers/Magazines □ Pamphlets/Brochures □ Billboards □ School □ Friends/Family □ Doctors □ Other:_____ □ I do not know anything about this disease Chikungunya □ Television □ Radio □ Newspapers/Magazines □ Pamphlets/Brochures □ Billboards □ School □ Friends/Family □ Doctors □ Other:_____ □ I do not know anything about this disease Zika □ Television □ Radio □ Newspapers/Magazines □ Pamphlets/Brochures □ Billboards □ School □ Friends/Family □ Doctors □ Other:_____ □ I do not know anything about this disease
Mark which diseases you think can be transmitted through sexual contact.	Dengue □ Yes □ No □ Don't know Chikungunya □ Yes □ No □ Don't know Zika □ Yes □ No □ Don't know
Mark what diseases you think can be prevented with vaccine.	Dengue □ Yes □ No □ Don't know Chikungunya □ Yes □ No □ Don't know Zika □ Yes □ No □ Don't know
If a pregnant woman is infected with the following disease, is the baby at risk of birth defects?	Dengue □ Yes □ No □ Don't know Chikungunya □ Yes □ No □ Don't know Zika □ Yes □ No □ Don't know
At what time is it more likely that people will become infected with the following diseases? Select all that apply.	Dengue □ Morning □ Afternoon □ Night Chikungunya □ Morning □ Afternoon □ Night Zika □ Morning □ Afternoon □ Night

I know how to prevent the following diseases. Select all that apply.	□ Dengue □ Chikungunya □ Zika
The following diseases are worse than they used to be in this city.	Dengue □ Yes □ No □ Don't know Chikungunya □ Yes □ No □ Don't know Zika □ Yes □ No □ Don't know
I am at higher risk of contracting the following diseases in the following seasons. Check all that apply.	Dengue □ Winter □ Summer □ Don't know Chikungunya □ Winter □ Summer □ Don't know Zika □ Winter □ Summer □ Don't know

Attitudes

	Completely agree	Agree	Neutral	Disagree	Completely Disagree	Not sure
Dengue is a serious disease.						
Chikungunya is a serious disease.						
Zika is a serious disease.						
I'm at risk of getting dengue.						
I am at risk of getting chikungunya.						
I am at risk of contracting Zika						
The Colombian government and public health officials are responsible for preventing dengue, chikungunya and Zika.						
I am responsible for preventing dengue, chikungunya and Zika.						
Fumigating mosquito breeding sites with larvicides is the most effective way to reduce the mosquito population.						
It is necessary to seek immediate treatment for dengue fever.						
It is necessary to seek immediate treatment for chikungunya.						
It is necessary to seek immediate treatment for Zika.						

Practices I

Can the following diseases be transmitted through sexual intercourse?	Dengue □ Yes □ No □ Don't know Chikungunya □ Yes □ No □ Don't know Zika □ Yes □ No □ Don't know
I strive to learn about the prevention of the following diseases.	Dengue □ Yes □ No □ Don't know Chikungunya □ Yes □ No □ Don't know Zika □ Yes □ No □ Don't know
I'm worried about mosquitoes in my neighborhood.	□ Yes □ No □ I'm not sure
I have contacted the authorities when I have identified mosquitoes or mosquito larvae on my property.	□ Yes □ No □ I'm not sure
How often do you spray your home with insecticides / larvicides to kill mosquitoes and / or mosquito larvae?	□ Daily □ Weekly □ Monthly □ Twice a year □ Yearly □ Never □ Don't know

Practices II

How often do you practice the following measures to prevent mosquitoes in your home / neighborhood?

	Never	Daily	Weekly	Biweekly	Monthly	Yearly
Use repellent						
Wear long-sleeved shirts and pants when working outside.						
Change the water in the pots frequently.						
Remove materials / objects that may accumulate water **outside** the house.						
Remove materials / objects that may accumulate water **inside** the house.						
Participate in community clean-up activities.						
Examine water deposits to identify mosquito larvae.						
Use of mosquito nets in the house.						
Keep windows and doors closed in the house.						

APPENDIX 2: KAP SURVEY (SPANISH)

ID#_____ Fecha____/_____/_____

Conocimientos, Actitudes y Prácticas (CAP) del dengue, chikungunya y zika en Cali, Colombia

Información sociodemográfica	
Sexo	☐ Masculino ☐ Femenino
Edad	☐ 18-35 ☐ 36-55 ☐ 56-70 ☐ 70+
Dirección de la casa	_____
Barrio	_____
Formación Educativa	☐ No estudió ☐ Primaria ☐ Secundaria ☐ Técnico ☐ Pregrado ☐ Posgrado
Ingreso mensual	[insertar rango COP] _____
Ocupación	☐ Empleado de tiempo completo ☐ Empleado de tiempo parcial ☐ Independiente ☐ Pensionado ☐ Ama de casa ☐ Desempleado ☐ Estudiante
Estado Civil	☐ Casado ☐ Separado/Divorciado ☐ Soltero ☐ Viudo ☐ Unión libre ☐ Otro
¿Tiene hijos que viven en su casa?	☐ Sí ☐ No
Número de personas que viven en su hogar (debe incluirse)	☐ 1 ☐ 2 ☐ 3-4 ☐ 5-6 ☐ 6+

Raza	□ Blanco □ Indígena □ Mestizo □ Afro-Colombiano □ Asiático □ Otro:_____

Preguntas Generales

¿Alguna vez ha tenido las siguientes enfermedades? Marque todas las que aplican.	□ Dengue □ Chikungunya □ Zika □ Ninguna
¿Conoce información sobre las siguientes enfermedades? Marque todas las que aplican.	□ Dengue □ Chikungunya □ Zika □ Ninguna
¿Cómo calificaría su acceso a los servicios de salud?	□ Excelente □ Muy bueno □ Bueno □ Justo □ Pobre
En promedio, ¿cuánto tiempo demora en llegar a un servicio de salud?	□ Menos de 5 minutos □ 5-15 minutos □ 15-30 minutos □ 30-60 minutos □ 1 hora o mas
Debido a la distancia del servicio de salud:	1. Ha dejado de buscar tratamiento médico o ir al médico? □ Sí □ No □ No estoy seguro 2. Ha rechazado tratamiento médico? □ Sí □ No □ No estoy seguro

Conocimientos

¿Con qué enfermedades está familiarizado? Marque todas las que aplican.	□ Dengue □ Chikungunya □ Zika
¿Sabe como se transmite el dengue?	□ Si □ No Si, Sí cómo?: _____
¿ Sabe como se transmite el chikungunya?	□ Si □ No Si, Sí cómo?: _____
¿ Sabe como se transmite el Zika?	□ Si □ No Si, Sí cómo?: _____
¿Por qué medio conoció, sobre las siguientes enfermedades? Marque todas las que aplican.	Dengue □ Televisión □ Radio □ Periódicos/Revistas □ Panfletos/Folleto □ Vallas publicitarias □ Colegio □ Amigos/Familia □ Médico □ Otro:_____ □ No sé nada de esta enfermedad Chikungunya □ Televisión □ Radio □ Periódicos/Revistas □ Panfletos/Folleto □ Vallas publicitarias □ Colegio □ Amigos/Familia □ Médico □ Otro:_____ □ No sé nada de esta enfermedad Zika □ Televisión □ Radio □ Periódicos/Revistas □ Panfletos/Folleto □ Vallas publicitarias □ Colegio □ Amigos/Familia □ Médico □ Otro:_____ □ No sé nada de esta enfermedad
Marque cuales enfermedades cree que se pueden transmitir por contacto sexual.	Dengue □ Sí □ No □ No sabe Chikungunya □ Sí □ No □ No sabe Zika □ Sí □ No □ No sabe

Marque cuales enfermedades cree que se pueden prevenir con vacuna.	Dengue □ Sí □ No □ No sabe Chikungunya □ Sí □ No □ No sabe Zika □ Sí □ No □ No sabe
Si una mujer embarazada está infectada con la siguiente enfermedad, ¿ el bebé corre el riesgo de tener defectos al nacer?	Dengue □ Sí □ No □ No sabe Chikungunya □ Sí □ No □ No sabe Zika □ Sí □ No □ No sabe
¿A qué hora es más probable que las personas se infecten con las siguientes enfermedades? Seleccione todas las que aplican.	Dengue □ Mañana □ Tarde □ Noche Chikungunya □ Mañana □ Tarde □ Noche Zika □ Mañana □ Tarde □ Noche
Sabe cómo prevenir las siguientes enfermedades. Marque todas las que aplican.	□ Dengue □ Chikungunya □ Zika
Las siguientes enfermedades son peores de lo que solían ser en esta ciudad	Dengue □ Sí □ No □ No sabe Chikungunya □ Sí □ No □ No sabe Zika □ Sí □ No □ No sabe
Estoy en mayor riesgo de contraer las siguientes enfermedades en las siguientes temporadas. Marque todas las que aplican.	Dengue □ Invierno □ Verano □ No sabe Chikungunya □ Invierno □ Verano □ No sabe Zika □ Invierno □ Verano □ No sabe

Actitudes

	Totalmente de acuerdo	De acuerdo	Neutral	En desacuerdo	Totalmente en desacuerdo	No está seguro
El dengue es una enfermedad grave.						
El chikungunya es una enfermedad grave.						
El Zika es una enfermedad grave.						
Estoy en riesgo de contraer dengue.						
Estoy en riesgo de contraer chikungunya.						
Estoy en riesgo de contraer Zika.						
El gobierno colombiano y los funcionarios de salud pública son responsables de prevenir el dengue, el chikungunya y el Zika.						
Yo soy responsable de prevenir el dengue, el chikungunya y el Zika.						
Fumigar los criaderos de mosquitos con larvicidas es la forma más efectiva de reducir la población de mosquito.						
Es necesario buscar tratamiento inmediato para la fiebre del dengue.						
Es necesario buscar tratamiento inmediato para chikungunya.						
Es necesario buscar tratamiento inmediato para el Zika.						

Prácticas I

¿Se pueden transmitir las siguientes enfermedades a través de las relaciones sexuales?	Dengue □ Sí □ No □ No sabe Chikungunya □ Sí □ No □ No sabe Zika □ Sí □ No □ No sabe
Me esfuerzo por aprender sobre la prevención de las siguientes enfermedades.	Dengue □ Sí □ No □ No sabe Chikungunya □ Sí □ No □ No sabe Zika □ Sí □ No □ No sabe
Estoy preocupado por los mosquitos en mi vecindario.	□ Sí □ No □ No estoy seguro
Ha contactado a las autoridades cuando he identificado mosquitos o larvas de mosquitos en mi propiedad.	□ Sí □ No □ No estoy seguro
¿Con qué frecuencia fumiga su casa con insecticidas / larvicidas para matar mosquitos y / o larvas de mosquitos?	□ Diario □ Semanal □ Mensual □ Dos veces al año □ Anual □ Nunca □ No sabe

Prácticas II

¿Con qué frecuencia practica las siguientes medidas para prevenir los mosquitos en su hogar / vecindario?

	Nunca	A Diario	Semanal	Quincenal	Mensual	Anual
Uso de repelente.						
Use camisas de manga larga y pantalones cuando trabaje afuera.						
Cambiar el agua de las materas con frecuencia.						
Remover los materiales / objetos que puedan acumular agua **fuera** de la casa.						
Remover los materiales / objetos que puedan acumular agua **adentro** de la casa.						
Participar en actividades de limpieza comunitaria.						
Examinar depósitos de agua para identificar larvas de mosquitos.						
Uso de toldillos en la casa.						
Mantener las ventanas y puertas cerradas en la casa.						

APPENDIX 3: SELECTED NEIGHBORHOODS AND HEALTH CENTERS IN CALI, COLOMBIA

Barrio	Estrata	Type	Saluds
Alfonso Bonilla Aragon	1	Cluster	3
Antonio Narino	2	Cluster	1
Bajos Ciudad Cordoba	3	Cluster	2
Barrio Eucaristico	4	Non-Cluster	8
Bellavista	2	Non-Cluster	2
Brisas de Mayo	1	Non-Cluster	1
Centenario	5	Non-Cluster	2
Ciudad Cordoba	3	Cluster	2
Cuarto de Legua - Guadalupe	5	Non-Cluster	3
El Cedro	4	Non-Cluster	7
El Diamante	2	Cluster	2
El Poblado I	2	Cluster	1
El Poblado II	2	Cluster	4
El Retiro	1	Cluster	1
El Vergel	1	Cluster	1
Granada	4	Non-Cluster	6
Guillermo Valencia	3	Non-Cluster	4
Jose Manuel Marroquin I	2	Cluster	1
Juanambu	5	Non-Cluster	1
La Flora	5	Cluster	8
Las Orquideas	1	Cluster	2
Los Cambulos	4	Non-Cluster	5
Los Comuneros I	2	Cluster	2
Los Comuneros II	2	Cluster	1
Los Sauces	3	Cluster	1
Nueva Tequendama	5	Non-Cluster	4
Omar Torrijos	2	Cluster	1
Republica de Israel	2	Cluster	2

San Fernando Nuevo	5	Non-Cluster	25
San Fernando Viejo	5	Non-Cluster	17
San Nicolas	2	Non-Cluster	3
Santa Isabel	4	Non-Cluster	9
Santa Teresita	6	Non-Cluster	1
Union de Vivienda Popular	2	Cluster	1
Urbanizacion Ciudad Jardin	6	Non-Cluster	5
Urbanizacion Tequendama	5	Non-Cluster	114
Versalles	5	Non-Cluster	14

APPENDIX 4: REFINED LIST OF POTENTIAL HEALTH CENTERS TO SURVEY IN CALI, COLOMBIA

Salud	**Barrio**	**Strata**
HOSPITAL INFANTIL Niño Dios	Alfonso Bonilla Aragon	1
PUESTO DE SALUD BRISAS DE MAYO	Brisas de Mayo	1
CENTRO DE SALUD ANTONIO NARIÑO	Antonio Narino	2
HOSPITAL CARLOS CARMONA MONTOYA	Republica de Israel	2
Centro De Salud Unión De Vivienda Popular	Union de Vivienda Popular	2
PUESTO DE SALUD BELLAVISTA	Bellavista	2
HOSPITAL SAN JUAN DE DIOS	San Nicolas	2
PUESTO DE SALUD CIUDAD CORDOBA	Ciudad Cordoba	3
I.P.S. COMFANDI MORICHAL	Bajos Ciudad Cordoba	3
CLINICA SAN FERNANDO	Santa Isabel	4
Coomeva Emergencia Medica CEM	Santa Isabel	4
CLINICA CENTRO MEDICO IMBANACO	San Fernando Nuevo	4
CLINICA SEBASTIAN DE BELALCAZAR	Juanambu	5
Clinicentro Sanitas	Urbanizacion Tequendama	5
HOSPITAL UNIVERSITARIO DEL VALLE	San Fernando Viejo	5
CLINICA COMFANDI TEQUENDAMA	Urbanizacion Tequendama	5
DIME	Versalles	5
Dinámica	Urbanizacion Tequendama	5
Eps CafeSalud Cl22norte	Versalles	5
Clinisanitas Ciudad Jardin	Urbanizacion Ciudad Jardin	6
CENTRO MEDICO IMBANACO SEDE SUR	Urbanizacion Ciudad Jardin	6

APPENDIX 5: IRB APPROVAL NOTICE

Michael Desjardins <mdesjar2@uncc.edu>

IRB Notice - 18-0399

4 messages

IRB <uncc-irb@uncc.edu> Fri, Mar 22, 2019 at 7:39 AM
To: edelmel1@uncc.edu, mdesjar2@uncc.edu
Cc: uncc-irbis@uncc.edu, chammelm@uncc.edu

To: Michael Desjardins
Geography and Earth Sciences

From: Office of Research Compliance

Date: 3/22/2019
RE: Notice of Approval of Exemption with No End Date
Exemption Category: 2.Survey, interview, public observation
Study #: 18-0399

Study Title: Knowledge, attitudes, and practices regarding chikungunya, dengue, and Zika in Colombia

This submission has been reviewed by the Office of Research Compliance and was determined to meet the Exempt category cited above under 45 CFR 46.101(b). This determination has no expiration or end date and is not subject to an annual continuing review. **However, you are required to obtain IRB approval for all changes to any aspect of this study before they can be implemented.**

The Investigator Responsibilities listed below applies to this study only. Carefully review the Investigator Responsibilities.

Study Description:

Understanding the knowledge, perspectives, and behaviors of individuals susceptible to disease can improve disease surveillance, education programs, and improve the allocation of public health resources to mitigate outbreaks. It is well known that disease rates vary spatially, even at the neighborhood level. This study will administer surveys to residents in Cali and Medellin, Colombia regarding their knowledge, attitudes, and prevention practices (KAP) regarding three vector-borne diseases: dengue, chikungunya, and Zika. The survey results will shed light on the KAP variations across diseases, across different neighborhoods, and across different municipalities within the same region (i.e. Cali vs. Medellin, Colombia).

Your approved consent forms (if applicable) and other documents are available online at http://uncc.myresearchonline.org/irb/index.cfm?event=home.dashboard.irbStudyManagement&irb_id=18-0399.

Investigator's Responsibilities:

The above-cited determination has no expiration or end date and is not subject to annual continuing review.

However, the Principal Investigator needs to comply with the following responsibilities:

1. Modifications **must** be submitted for review and approval before implementing the modification. This includes changes to study procedures, study materials, personnel, etc.
2. Data security procedures must follow procedures as approved in the protocol and in accordance with ITS Guidelines for Data Handling.
3. Promptly notify the IRB (uncc-irb@uncc.edu) of any adverse events or unanticipated risks to participants or others.
4. Complete the Closure eform via IRBIS once the study is complete.
5. Be aware that this study is now included in the Office of Research Compliance (ORC) **Post-Approval Monitoring program** and may be selected for post-review monitoring at some point in the future.

APPENDIX 6: CONSENT FORM (ENGLISH)

ID#_____ Date____/_____/_____

Department of Geography and Earth Sciences

9201 University City Boulevard, Charlotte, NC 28223-0001

Consent to Participate in a Research Study

Title of the Project: Knowledge, attitudes, and practices regarding chikungunya, dengue, and Zika in Colombia

Principal Investigator: Michael R. Desjardins, Ph.D. Candidate, University of North Carolina at Charlotte and Center for Applied Geographic Information Science, USA

Co-investigators: (1) Irene Casas, PhD, Louisiana Tech University, USA;

(2) Eric Delmelle, PhD, University of North Carolina at Charlotte, USA;

(3) Colleen Hammelman, PhD, University of North Carolina at Charlotte, USA;

(4) Diana Davalos, MD, PHD, Universidad Icesi, Cali, Colombia

(5) Angela Victoria, MD, Universidad Icesi, Cali, Colombia

(6) Dayana Carbonell, MD, Universidad Libre, Cali, Colombia

Study Sponsor: None

You are invited to participate in a research study. Participation in this research study is voluntary. The information provided is to help you decide whether or not to participate. If you have any questions, please ask.

Important Information You Need to Know

- The purpose of this research is to determine if there are any geographic patterns in the way people understand and practice the recommendations of educational campaigns to control dengue fever, chikungunya, and Zika in the city of Cali.

- We are asking residents of Cali, Colombia who are age 18 and older to complete a survey about their knowledge and familiarity with dengue, chikungunya, and Zika; and how people try to protect themselves from the three previous diseases.

- You can refuse to answer any questions while completing the survey. You have the right to refuse participation at any time and the information collected from you will be deleted from the investigation.

- Please read this form and ask any questions you may have before you decide whether to participate in this research study.

Why are we doing this study?

Dengue fever, chikungunya, and Zika are diseases typical of the tropics that can have negative health impacts and also be life-threatening. Major epidemics can occur if the diseases are not controlled properly. Prevention strategies are the most effective means of control. Therefore, understanding how people learn about the strategies to prevent the disease, how they apply them to their daily lives, and how effective they consider certain preventative strategies are important to help explain the geographic patterns of the disease.

Why are you being asked to be in this research study.

You are being asked to be in this study because you are age 18 and older, and are currently living in Cali, Colombia.

What will happen if I take part in this study?

If you choose to participate you will complete a survey of around 45 questions. Questions include: demographic, neighborhood location, education, and employment characteristics; familiarity with dengue, chikungunya, and Zika; access to healthcare resources; informational sources about the diseases, how the information is used and applied; how effective individuals consider their practices based on the information learned from various sources; and if individuals participate in disease prevention techniques. The survey should take from 8-10 minutes to complete.

What benefits might I experience?

You will not benefit directly from being in this study. Others might benefit because it is important to understand the associations between where you live and your risk of getting dengue, chikungunya, and Zika. The results can improve educational resources and improve health policy to help protect you from disease.

What risks might I experience?

There are no risks if you decide to participate in this study.

How will my information be protected?

You are asked to provide your address and the name of the neighborhood where you live as part of this study. We will use your address and neighborhood to understand how your home location may influence your risk of getting dengue, chikungunya, or Zika. Your address will never be published and available to public sources (such as websites and newspapers). To further protect your privacy (identity), we'll assign a study ID code to your survey responses. While the study is active, all data will be stored in a password-protected data base that can be can be accessed by the primary researcher. Only the research team will have routine access to the study data. Other people with approval from the Investigator, may need to see the information we collect about you. Including people who work for UNC Charlotte and other agencies as required by law or allowed by federal regulations. Your individual privacy will be maintained in all written and published material resulting from the research.

How will my information be used after the study is over?

After this study is complete, study data may be shared with other researchers for use in other studies or as may be needed as part of publishing our results. The data we share will NOT include information that could identify you – such as your address that we will be collecting from you.

Will I receive an incentive for taking part in this study?

There are no incentives and you will not be paid to participate in this study.

What other choices do I have if I don't take part in this study?

You do not have to participate in this study. If you are interested in the results of this study, you can contact the principal investigator.

What are my rights if I take part in this study?

It is up to you to decide to be in this research study. Participating in this study is voluntary. Even if you decide to be part of the study now, you may change your mind and stop at any time. You do not have to answer any questions you do not want to answer.

Who can answer my questions about this study and my rights as a participant?

For questions about this research, you may contact Michael Desjardins – email: mdesjar2@uncc.edu, phone: +1(203)233-5381 and Dr. Eric Delmelle – email: Eric.Delmelle@uncc.edu.

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Office of Research Compliance at +1 (704)-687-1871 or uncc-irb@uncc.edu.

Consent to Participate

If you agree to participate in this research study, please verbally say yes to the researcher.

Be sure that any questions have been answered clearly to you and that you have a thorough understanding of the study.

Please ask any questions regarding this study's objectives and your participation.

If you have further questions that come up later, please feel free to ask a member of the research team. If you agree to participate in this study, a copy of this document will be given to you.

APPENDIX 7: CONSENT FORM (SPANISH)

Documento de identidad#_____ Fecha____/_____/_____

Department of Geography and Earth Sciences

9201 University City Boulevard, Charlotte, NC 28223-0001

Formulario de consentimiento informado

Título del proyecto: Un enfoque de métodos mixtos para la vigilancia de enfermedades transmitidas por vectores en Colombia

Investigador principal: Michael R. Desjardins, Ph.D. Candidato, Departamento de Geografía y Ciencias de la Tierra de la Universidad de Carolina del Norte en Charlotte, Estados Unidos.

Otros investigadores: (1) Irene Casas, PhD, Universidad Louisiana Tech, Estados Unidos.

(2) Eric Delmelle, PhD, Universidad de Carolina del Norte en Charlotte, Estados Unidos.

(3) Colleen Hammelman, PhD, Universidad de Carolina del Norte en Charlotte, Estados Unidos.

(4) Diana Davalos, MD, PHD, Universidad Icesi, Cali, Colombia.

(5) Angela Victoria, MD, Universidad Icesi, Cali, Colombia.

(6) Dayana Carbonell, MD, Universidad Libre, Cali, Colombia

Patrocinador del estudio: Ninguno

Te invitamos a participar en un estudio de investigación. Su participación en este proyecto de investigación es completamente voluntaria. La información proporcionada es para ayudarlo a decidir si desea o no participar. Si tiene alguna pregunta, por favor pregunte.

Información importante que necesita saber

- El objetivo de esta investigación es determinar si existen patrones relacionados con la geografía o los espacios y la forma en que las personas entienden y practican recomendaciones de campañas educativas para controlar el dengue, el chikungunya y el Zika en las ciudad de Cali, Colombia.

- Estamos pidiendo a los residentes de Cali, Colombia que tengan 18 años o más que completen una encuesta sobre su conocimiento y familiaridad con el dengue, el chikungunya y el Zika; y cómo las personas tratan de protegerse de las tres enfermedades anteriores.

- Puede negarse a responder ciertas preguntas mientras completa la encuesta. Tiene el derecho de rechazar la participación en cualquier momento y su información recolectada será eliminada del estudio.

- Por favor, lea este formulario y haga cualquier pregunta que pueda tener antes de decidir si desea participar en este estudio de investigación.

¿Por qué estamos haciendo este estudio?

El dengue, el chikungunya y el Zika son enfermedades típicas de las regiones tropicales que pueden tener efectos negativos para la salud y también pueden poner en peligro la vida. Pueden ocurrir grandes epidemias si las enfermedades no se controlan adecuadamente. Las estrategias de prevención son los medios más efectivos de control, por esto es importante conocer cómo las personas aprenden las estrategias para prevenir la enfermedad, cómo las aplican a su vida diaria y qué tan efectivas consideran estas estrategias, para así lograr explicar los patrones geográficos de la enfermedad.

¿Por qué te piden que participes en este estudio de investigación?

Se le pide que participe en este estudio porque tiene 18 años o más y actualmente vive en Cali, Colombia.

¿Qué pasará si participo en este estudio?

En esta investigación, se pide a las personas que completen una encuesta de 45 preguntas. Las preguntas incluyen: características demográficas, ubicación del barrio, educación, empleo; familiaridad con el dengue, chikungunya y Zika; acceso a recursos sanitarios; fuentes de información sobre las enfermedades, cómo se utiliza y aplica la información; cómo los individuos consideran sus prácticas en relación a la información aprendida de varias fuentes; y si los

individuos participan en técnicas de prevención de enfermedades. La encuesta dura entre 8-10 minutos para completarse.

¿Qué beneficios podría experimentar?

No hay beneficios si participa en este estudio. Otros pueden beneficiarse porque es importante comprender las asociaciones entre el lugar donde vive y su riesgo de desarrollar dengue, chikungunya y Zika. Los resultados pueden mejorar los recursos educativos y la política de salud para ayudarlo a protegerse de las enfermedades.

¿Qué riesgos podría experimentar?

No hay riesgos si participa en este estudio.

¿Cómo se protegerá mi información?

Se le solicita que proporcione su dirección y el nombre del vecindario donde vive como parte de este estudio. Usaremos su dirección y vecindario para comprender cómo la ubicación de su hogar puede influir en su riesgo de desarrollar dengue, chikungunya o Zika. Su dirección nunca se publicará y estará disponible para fuentes públicas (como sitios web y periódicos). Para proteger aún más su privacidad (identidad), asignaremos un código de identificación del estudio a las respuestas de su encuesta. Mientras el estudio esté activo, todos los datos se almacenarán en una base de datos protegida por contraseña a la que podrá acceder el investigador principal. Solo el equipo de investigación tendrá acceso de rutina a los datos del estudio. Es posible que otras personas con la aprobación del investigador necesiten ver la información que recopilamos sobre usted. Incluyendo a las personas que trabajan para UNC Charlotte y otras agencias según lo exige la ley o lo permiten las regulaciones federales. Su privacidad individual se mantendrá en todo el material escrito y publicado que resulte de la investigación.

¿Cómo se utilizará mi información después de que termine el estudio?

Después de completar este estudio, los datos del estudio se pueden compartir con otros investigadores para su uso en otros estudios o según sea necesario como parte de la publicación de nuestros resultados. Los datos que compartimos NO incluirán información que pueda identificarlo, como su dirección que recopilaremos de usted.

¿Recibiré un incentivo por participar en este estudio?

No hay incentivos y no se le pagará por participar en este estudio.

¿Qué otras opciones tengo si no participo en este estudio?

No tienes que participar en este estudio. Si está interesado en los resultados de este estudio, puede comunicarse con el investigador principal.

¿Cuáles son mis derechos si participo en este estudio?

Su participación en este proyecto de investigación es completamente voluntaria. Puede negarse a responder ciertas preguntas mientras completa la encuesta. Tiene el derecho de rechazar la participación en cualquier momento y su información recolectada será eliminada del estudio. En caso de que se disguste o algo relacionado con la encuesta le genere molestia, el investigador que le esté ayudando con la misma puede retirar su participación del estudio y referirlo a donde necesite para recibir la ayuda apropiada.

¿Quién puede responder mis preguntas sobre este estudio y mis derechos como participante?

Si tiene alguna pregunta sobre esta investigación, puede comunicarse con Michael Desjardins al teléfono: (203) 233-5381 (Estados Unidos), o escribir al correo: mdesjar2@uncc.edu.

Si tiene alguna pregunta sobre sus derechos como individuo en el proyecto de investigación, debe comunicarse (de forma anónima, si lo desea) al Comité de Uso Humano de la Universidad de Carolina del Norte en Charlotte, al correo: uncc-irb@uncc.edu

Consentimiento para participar

Si acepta participar en este estudio de investigación, diga verbalmente que sí al investigador.

Asegúrese de que todas las dudas que tenga acerca de la investigación sean resueltas de forma clara por alguno de los investigadores que le esté ayudando con la encuesta.

Puede realizar cualquier pregunta con respecto a los objetivos de este estudio y su participación en la investigación.

Si tiene preguntas que surgen más adelante, no dude en preguntar a un miembro del equipo de investigación. Si acepta participar en este estudio, se le entregará una copia de este documento.

APPENDIX 8: REQUEST TO SURVEY IN HEALTHCARE FACILITY (SPANISH)

Department of Geography and Earth Sciences
9201 University City Boulevard, Charlotte, NC 28223-0001

Doctor(a)
<<nombre del pesonaje>>
Gerente / Director
<<nombre de la IPS>>
L.C.

Asunto: solicitud de autorización para realizar encuestas en la salas de espera de <<nombre de la IPS>>

Reciba un cordial saludo,

Le escribimos como parte de un grupo de investigación, que tiene por objetivo determinar si existen patrones y agrupaciones geográficas en la forma cómo las personas entienden y practican recomendaciones de las campañas educativas para controlar el dengue, el chikungunya y el Zika en la ciudad de Cali, Colombia. Nuestro grupo de investigación cuenta con profesionales de las Universidades ICESI, Libre, Louisiana Tech, Carolina del Norte en Charlotte y DIME Clínica Neurocardiovascular.

Estamos solicitando autorización para administrar encuestas en las áreas de la sala de espera de su <<nombre de la IPS>>. El proyecto fue aprobado por el comité de ética de investigación de la Universidad de Carolina del Norte en Charlotte (caso # 18-0399)

Les anexamos un resumen del proyecto y estaremos muy atentos si requieren información adicional. Para nosotros es clave obtener una respuesta a esta solicitud lo antes posible, tener los datos de la persona de contacto y ampliar la información que su organización requiera.

Agradeciéndole de antemano la atención prestada y muy atentos a resolver cualquier inquietud que se pueda presentar.

Atentamente,

Michael Desjardins
Universidad de Carolina del Norte

Alejandro Varela Villegas
DIME Clínica Neurocardiovascular

Charlotte, Estados Unidos Cali, Colombia

Department of Geography and Earth Sciences
9201 University City Boulevard, Charlotte, NC 28223-0001

Solicitud de autorización para administrar encuestas

Título del proyecto: Un enfoque de métodos mixtos para la vigilancia de enfermedades transmitidas por vector en Colombia

Investigador principal: Michael R. Desjardins, Ph.D. Candidato, Departamento de Geografía y Ciencias de la Tierra de la Universidad de Carolina del Norte en Charlotte, Estados Unidos.

Otros investigadores: (1) Irene Casas, PhD, Universidad Louisiana Tech, Estados Unidos.
(2) Eric Delmelle, PhD, Universidad de Carolina del Norte en Charlotte, Estados Unidos.
(3) Colleen Hammelman, PhD, Universidad de Carolina del Norte en Charlotte, Estados Unidos.
(4) Diana Davalos, MD, PHD, Universidad Icesi, Cali, Colombia.
(5) Angela Victoria, MD, Universidad Icesi, Cali, Colombia.
(6) Dayana Carbonell, MD, Universidad Libre, Cali, Colombia
(7) Alejandro Varela, MD, Gerente de DIME DIME - Clínica Neurocardiovascular de Cali.

Fechas del Proyecto: 1 de junio - 30 de junio (Cali, Colombia).

¿Qué estamos solicitando?

Quisiéramos obtener autorización de su institución de salud (centro de salud/puesto de salud/clínica/hospital) para administrar encuestas a las personas que se encuentran en el área de la sala de espera. La descripción de nuestro proyecto y el impacto que puede tener en el conocimiento de salud pública en Cali se detallan a continuación.

Propósito del proyecto

- El objetivo de esta investigación es determinar si existen patrones entre la forma en que las personas entienden y practican recomendaciones de campañas educativas para controlar el dengue, el chikungunya y el Zika en la ciudad de Cali, Colombia y el espacio geográfico.

- Estamos solicitando a los residentes de Cali, Colombia mayores de 18 años, que completen una encuesta sobre su conocimiento y familiaridad con el dengue, el chikungunya y el Zika; y cómo se protegen de estas tres enfermedades.

- El proyecto fue aprobado por el comité de ética de investigación de la Universidad de Carolina del Norte en Charlotte (caso # 18-0399).

¿Por qué estamos haciendo este estudio?

El dengue, el chikungunya y el Zika son enfermedades típicas de las regiones tropicales que pueden tener efectos negativos para la salud y también pueden poner en peligro la vida. Pueden ocurrir grandes epidemias si las enfermedades no se controlan adecuadamente. Las estrategias de prevención son los medios más efectivos de control, por esto es importante conocer cómo las personas aprenden las estrategias para prevenir la enfermedad, cómo las aplican a su vida diaria y qué tan efectivas consideran estas estrategias, para así lograr explicar los patrones geográficos de la enfermedad.

Procedimiento de encuesta

Invitaremos a las personas en la sala de espera a participar en el estudio. Si aceptan, se les proporcionará un formulario de consentimiento y procederán a completar una encuesta de alrededor de 40 preguntas. Las preguntas incluyen: características demográficas, ubicación del barrio, educación, empleo; familiaridad con el dengue, chikungunya y Zika; acceso a recursos sanitarios; fuentes de información sobre las enfermedades, cómo se utiliza y aplica la información; cómo los individuos consideran sus prácticas en relación a la información aprendida de varias fuentes; y si los individuos participan en técnicas de prevención de enfermedades. La encuesta toma entre 8-10 minutos para completar.

Si nuestra solicitud es aceptada

Necesitaremos:

(1) El nombre de la persona de contacto en las instalaciones.

(2) Una carta de autorización que indica que nuestro equipo puede administrar la encuesta en las instalaciones.

Información de contacto principal

Michael R. Desjardins

- correo electrónico – mdesjar2@uncc.edu
- teléfono - +1(203)233-5381 (estados unidos)

Dr. Alejandro Varela

- correo electrónico – gerenciageneral@dime.com.co

APPENDIX 9: SEMI-STRUCTURED INTERVIEW (ENGLISH)

Survey ID_____ Survey #_____ Date_____

Questions for medical directors or health officials in Cali

1. What is your profession and what are your main responsibilities?

2. How long have you been a health professional in Cali?

3. During an endemic year of dengue, what percentage of your patients were diagnosed with dengue, zika and / or chikungunya?

4. What are the most effective preventive measures against *Aedes* mosquitoes?

5. In the health care center where you work, do you have educational campaigns in your facilities or within the community?

6. Since the first appearance of chikungunya and Zika in Colombia, has the awareness of the two diseases been the same as dengue?

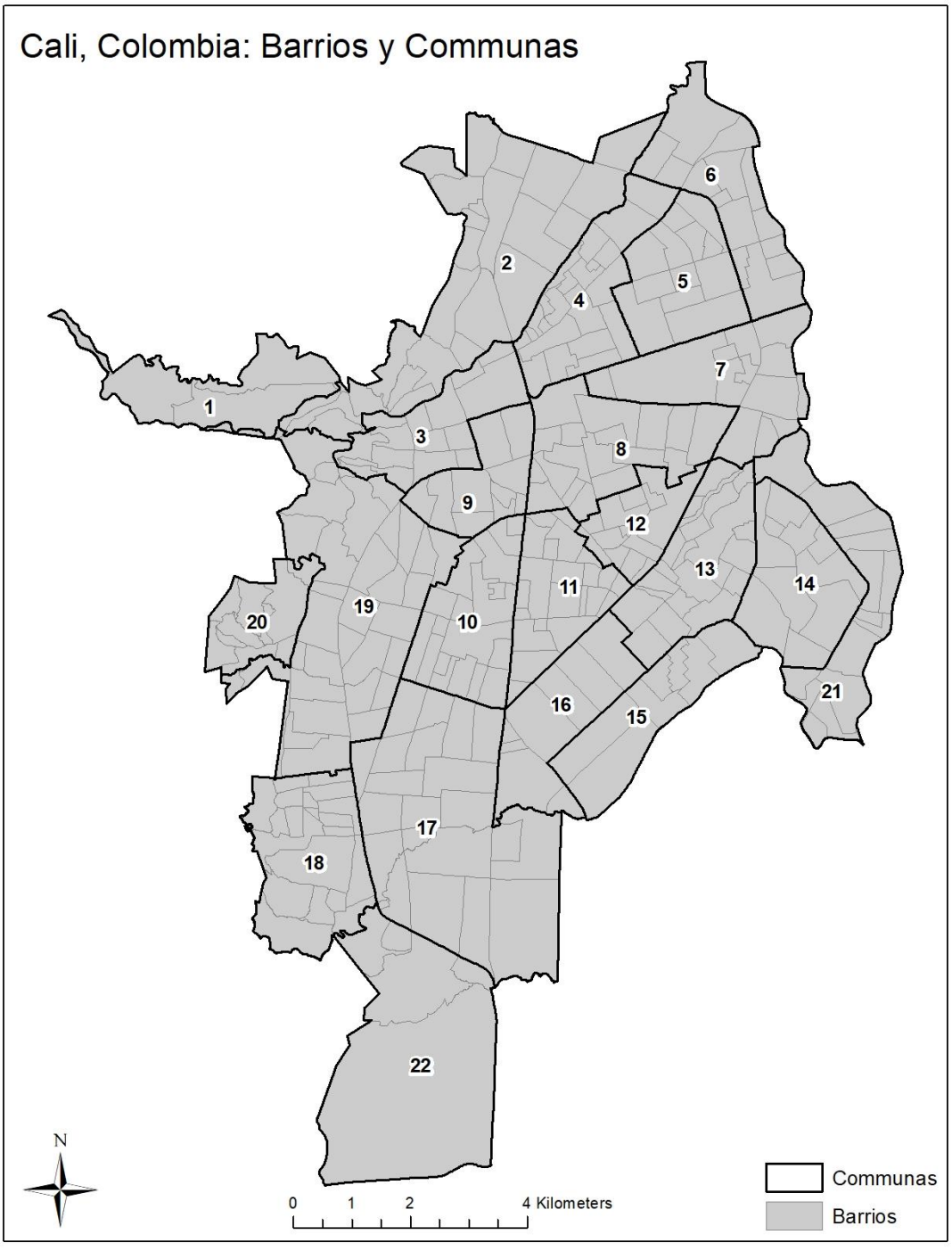
7. Are there data and research needs that can improve decision making to reduce the transmission of dengue, chikungunya and Zika?

8. Do you think that Valle del Cauca should carry out more campaigns for control and education in the eradication of *Aedes*?

9. How can surveys, in this case, the KAP administered in health centers in Cali, improve education and prevention strategies related to dengue, chikungunya and Zika?

10. On the map below, indicate what neighborhoods are at the highest risk of dengue, chikungunya and Zika (circle the approximate locations).

ID de entrevista_____ Entrevista #_____ Fecha_____



APPENDIX 10: SEMI-STRUCTURED INTERVIEW (SPANISH)

ID de entrevista_____ Entrevista #_____ Fecha_____

Preguntas para directores médicos o funcionarios de salud en Cali

1. Cuál es su profesión y cuáles son sus principales responsabilidades?

2. Cuánto tiempo lleva como profesional de la salud en Cali?

3. Durante un año endémico del dengue, ¿qué porcentaje de sus pacientes fueron diagnosticados con dengue, zika y / o chikungunya?

4. Cuáles son las medidas preventivas más efectivas contra los mosquitos *Aedes*?

5. En el centro de atención médica donde labora, ¿ tiene campañas educativas en sus instalaciones o dentro de la comunidad?

6. Desde la primera aparición de chikungunya y zika en Colombia, ¿se ha definido como enfermedades iguales que el dengue?

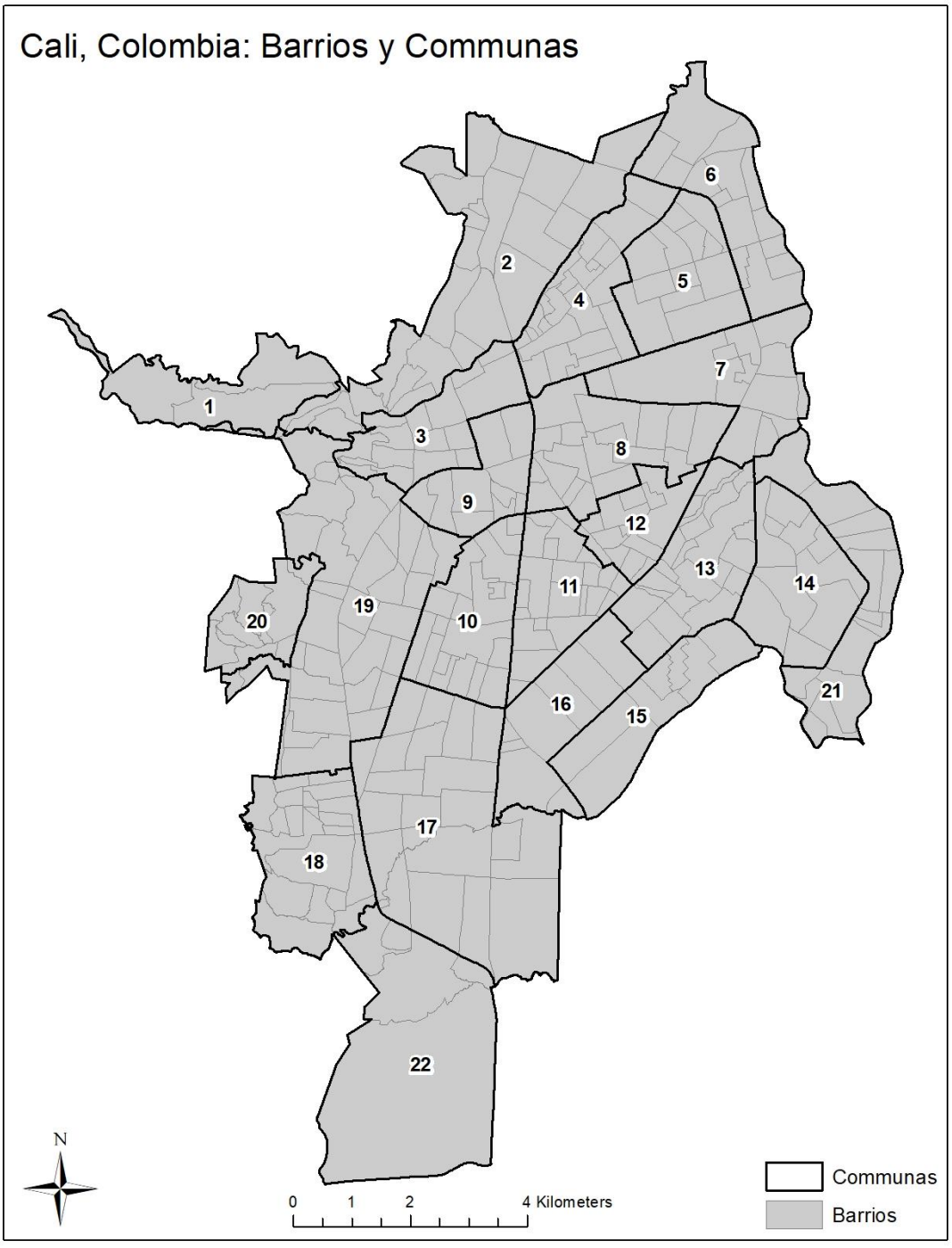
7. Existen datos y necesidades de investigación que puedan mejorar la toma de decisiones para reducir la transmisión del dengue, chikungunya y Zika?

8. Consideras que el Valle del Cauca debería hacer más campañas para el control y educación en la erradicación del *Aedes*?

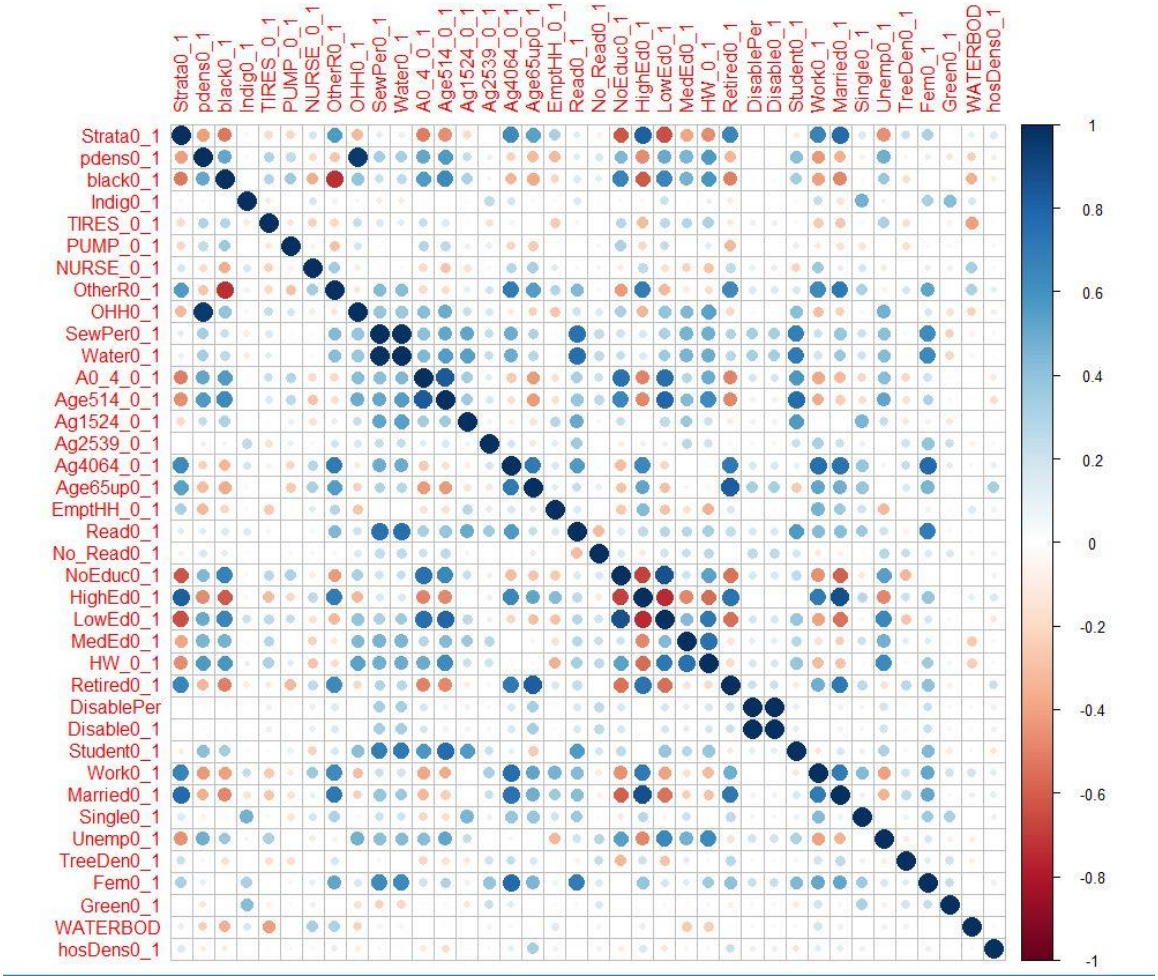
9. Cómo pueden las encuestas, en este caso el CAP administrado en los centros de salud en Cali, mejorar las estrategias de educación y prevención relacionadas con el dengue, el chikungunya y el Zika?

10. En el siguiente mapa, indique dónde se encuentran los vecindarios de mayor riesgo de dengue, chikungunya y Zika (circule los lugares aproximados).

ID de entrevista_____ Entrevista #_____ Fecha_____



APPENDIX 11: CORRELATION MATRIX FOR VARIABLES IN TABLE 5



APPENDIX 12: R CODES FOR MODELS 1 & 2 IN CHAPTER 4

```
setwd("C:/temp ")

#install packages - (install them first in case they have never been
installed before)
install.packages("sp")
install.packages("spdep")
install.packages("rgdal")
install.packages("classInt")
install.packages("spData")
install.packages("RColorBrewer")
install.packages("dplyr")
install.packages("CARBayesST")
install.packages("olsrr")
install.packages("htmltools")
install.packages("scales")
install.packages("Rtools")
install.packages("stats")
install.packages("corrplot")
install.packages("caret")
install.packages("FactoMineR")
install.packages("factoextra")
install.packages("CARBayes")
install.packages("ggpubr")
if(!require(devtools)) install.packages("devtools")
devtools::install_github("kassambara/ggpubr")


#load packages
library(sp) #spatial analysis package
library(spData) #spatial data package
library(spdep) #spatial modeling package
library(rgdal)  # useful to work with esri shapefiles
library(RColorBrewer)  #useful to choose color palette (see Appendix A)
library(classInt)   #useful to make choropleth maps
library(dplyr)
library(CARBayesST)
library(CARBayes)
library(scales)
library(olsrr)
library(htmltools)
library(stats)
library(corrplot)
library(caret)
library(gridExtra)
library(grid)
library(FactoMineR)
library(factoextra)
library(Rtools)
library(devtools)


DENFGIS <- readOGR(dsn="H:/", layer="Cali_Final")
```

```
W <- poly2nb(DENFGIS, row.names = NULL, snap=sqrt(.Machine$double.eps),
+queen=TRUE, useC=TRUE, foundInBox=NULL)
W.list <- nb2listw(W.nb, style = "B")
W <- nb2mat(W.nb, style = "B")

DENFData = read.csv("H:/DENF1516_Final6.csv")
DENFData$logSMR <- log(DENFData$SMR)

SMR.av <- summarise(group_by(DENFData,ID2), SMR.mean = mean(SMR))
DENFGIS@data$SMR <- SMR.av$SMR.mean

#spatial weight matrix
W.nb <- poly2nb(DENFGIS, row.names = SMR.av$ID2)
W.list <- nb2listw(W.nb, style = "B")
W <- nb2mat(W.nb, style = "B")


### PCA Analysis - long way... ###
DENFDataCorr = read.csv("J:/Fall_2019/GEOMED/Data/DENF1516_corr1.csv")

apply(DENFDataCorr, 2, var) #compute variance

DF.cov <- cov(DENFDataCorr) # Calculate eigenvalues & eigenvectors
DF.eigen <- eigen(DF.cov) # Calculate eigenvalues & eigenvectors
str(DF.eigen) # Calculate eigenvalues & eigenvectors

(phi <- DF.eigen$vectors[,1:49])
phi <- -phi
phi

PC1 <- as.matrix(DENFDataCorr) %*% phi[,1]
PC2 <- as.matrix(DENFDataCorr) %*% phi[,2]
PC3 <- as.matrix(DENFDataCorr) %*% phi[,3]
PC4 <- as.matrix(DENFDataCorr) %*% phi[,4]
PC5 <- as.matrix(DENFDataCorr) %*% phi[,5]
PC6 <- as.matrix(DENFDataCorr) %*% phi[,6]

PVE <- DF.eigen$values / sum(DF.eigen$values)
round(PVE, 6)

## PCA REAL
pca_result <- prcomp(DENFDataCorr, scale = TRUE)
names(pca_result)
pca_result$center
pca_result$scale
pca_result$rotation
pca_result$rotation <- -pca_result$rotation
pca_result$rotation
pca_result$x <- - pca_result$x
head(pca_result$x)
pca_result$sdev
(VE <- pca_result$sdev^2)
PVE <- VE / sum(VE)
round(PVE, 3)

pr<-prcomp(DENFDataCorr, scale = FALSE)
summary(pr) # two PCs for cumulative proportion of >80%
```

```
newdat<-pr$x[,1:2]



trans = preProcess(DENFDataCorr[,1:49], method=c("BoxCox", "center",
"scale", "pca")) #transform
PC = predict(trans, DENFDataCorr[,1:49])
head(PC,3)
trans$rotation

##Final PCA - USE THIS CODE
DENFDataCorr = read.csv("H:/Fall_2019/GEOMED/Data/DENF1516_corr2.csv")
#after removing vars from 1st PCA
res.pca <- PCA(DENFDataCorr, graph = FALSE)
print(res.pca)

eigenvalues <- res.pca$eig
head(eigenvalues[, 1:2])


library("factoextra")
fviz_screeplot(res.pca, ncp=10)

head(res.pca$var$coord)
fviz_pca_var(res.pca)

head(res.pca$var$cos2)
fviz_pca_var(res.pca, col.var="cos2") +
scale_color_gradient2(low="white", mid="blue",
+high="red", midpoint=0.5) + theme_minimal()

head(res.pca$var$contrib)
# Contributions of variables on PC1
fviz_pca_contrib(res.pca, choice = "var", axes = 1)
# Contributions of variables on PC2
fviz_pca_contrib(res.pca, choice = "var", axes = 2)
# Contributions of variables on PC3
fviz_pca_contrib(res.pca, choice = "var", axes = 3)

# Total contribution on PC1-PC3
fviz_pca_contrib(res.pca, choice = "var", axes = 1:3)

# Coordinates of variables
head(res.pca$var$coord,n=49)
head(res.pca$var$cos2,n=49)
####################################

mydata.cor = cor(DENFDataCorr)
corrplot(mydata.cor)
palette = colorRampPalette(c("green", "white", "red")) (38)
heatmap(x = mydata.cor, col = palette, symm = TRUE)

DENFDataVIF = read.csv("J:/Fall_2019/GEOMED/Data/DENF1516_vif1.csv")

formulaVIF <- lm(SMR ~ + NURSE_0_1 + TIRES_0_1 + TreeDen0_1 + Water0_1
+ pdens0_1 , data=DENFDataVIF)
ols_vif_tol(formulaVIF)
```

```
ols_coll_diag(formulaVIF)

formulaDENF <- Observed ~ offset(log(Expect)) + FAC1 + FAC2 + NURSE_0_1
+ TIRES_0_1 + pdens0_1 + WATERBOD +
TreeDen0_1 + TavgL5 + DTRMaxL4 + RelHRngL3 + RainTL3 + RainDL5 +
CoolDL2 + WarmDL5
# Age514_0_1 + LowEd0_1 + Ag4064_0_1 + WarmD0_1 + Married0_1 + Fem0_1 +
SewPer0_1 + DTRAvg0_1 + Unemp0_1
# lag test
install.packages("Hmisc")
install.packages("knitr")
library("Hmisc")
library("knitr")
DENFAvgTCor =
read.csv("J:/Dissertation/Data/Cali/chp2/cross_correlations/Avg_temp_la
gs_30_weeks.csv")
res <- cor(DENFAvgTCor)
round(res,2)

res2 <- rcorr(as.matrix(DENFAvgTCor))
res2


##

formulaDENF <- Observed ~ offset(log(Expect)) + FAC1 + FAC2 + NURSE_0_1
+ TIRES_0_1 + pdens0_1 + WATERBOD +
TreeDen0_1 + TavgL5 + DTRMaxL4 + RelHRngL3 + RainTL3 + RainDL5 +
CoolDL2 + WarmDL5
formulaDENF <- Observed ~ offset(log(Expect)) + FAC1 + FAC2 + NURSE_0_1
+ TIRES_0_1 + pdens0_1 + WATERBOD +
TreeDen0_1 + DTRMaxL4 + RelHRngL3 + RainTL3 + RainDL5 + WarmDL5
modelDENF <- glm(formula = formulaDENF, family = "poisson", data =
DENFData)

resid.glm <- residuals(modelDENF)
summary(modelDENF)$coefficients
summary(modelDENF)$dispersion

moran.mc(x = resid.glm[1:340], listw = W.list, nsim = 10000) # week 1
moran.mc(x = resid.glm[341:680], listw = W.list, nsim = 10000) # week 2


#ST-CAR model
modelCARDENF <- ST.CARlinear(formula = formulaDENF, family = "poisson",
data = DENFData,
+W = W, burnin = 20000, n.sample = 220000, thin = 10)

model2 <-  ST.CARar(formula = formulaDENF, family = "poisson", data =
DENFData, W = W,
+burnin = 20000, n.sample = 220000, thin = 10)
print(model2)
model3 <-  ST.CARar(formula = formulaDENF, family = "poisson", data =
DENFData, W = W, burnin = 30000,
+n.sample = 220000, thin = 10)
print(model3)
```

```
parameter.summary <- summarise.samples(exp(model2$samples$beta[ , -1]),
quantiles = c(0.5, 0.025, 0.975))
round(parameter.summary$quantiles, 3)

DENFData$Rate <- (DENFData$Observed/DENFData$Expected)*1000
modelDENFN <-glm.nb(SMR ~ FAC1 + FAC2 + NURSE_0_1 + TIRES_0_1 +
pdens0_1 + WATERBOD + TreeDen0_1 + TavgL5 +
+DTRMaxL4 + RelHRngL3 + RainTL3 + RainDL5 + CoolDL2 + WarmDL5, data =
DENFData)

summary(modelDENFN)


#posterior distributions for the covariate effects
colnames(model2$samples$beta) <- c("Intercept", "FAC1NEW2", "FAC2NEW2",
"NURSE_0_1","TIRES_0_1","pdens0_1", "WATERBOD",
+"TreeDen0_1","Tavg0_1","DTRMax0_1","RelHRng0_1","RainT0_1","RainD0_1",
"CoolD0_1", "WarmD0_1")
plot(exp(model2$samples$beta[ , -1]))

rate.est <- matrix(model2$fitted.values / DENFData$TOTAL,nrow =
nrow(W), byrow = FALSE)
rate.est <- as.data.frame(rate.est)

rate.est <- matrix(model2$fitted.values, nrow = nrow(W), byrow = FALSE)
write.table(rate.est,file="H:\\DENF_Post.csv")

colnames(rate.est) <- c("r30", "r31", "r32", "r33", "r34", "r35",
"r36", "r37", "r38", "r39", "r40")
DENFGIS@data <- data.frame(DENFGIS@data, rate.est)
breakpoints <- c(0, quantile(SMR.av$SMR.mean, seq(0.001, 0.005,
0.001)), 0.001)
spplot(DENFGIS, c("r31", "r33", "r35", "r37", "r39", "r40"), names.attr
= c("Rate 31", "Rate 33", "Rate 35", "Rate 37",
+"Rate 39", "Rate 40"), xlab = "Easting", ylab = "Northing", scales =
list(draw = TRUE), at = breakpoints,
+col.regions = terrain.colors(n = length(breakpoints - 1)),
par.settings=list(fontsize=list(text=15)))

# disease correlations
DCZCorr1516 = read.csv("J:/Fall_2019/GEOMED/Data/DCZ_Corr15-16.csv")
DCCorr15 = read.csv("J:/Fall_2019/GEOMED/Data/DC_Corr15.csv")
DCZCorr16 = read.csv("J:/Fall_2019/GEOMED/Data/DCZ_Corr16.csv")

mydata.cor = cor(DCCorr15)
res3 <- cor.test(mydata.cor$DENF, mydata.cor$CHIK, method = "spearman")
```

APPENDIX 13: R CODES FOR MODELS 3 & 4 IN CHAPTER 4

```
setwd("C:/temp ")

#install packages - (install them first in case they have never been
installed before)
install.packages("sp")
install.packages("spdep")
install.packages("rgdal")
install.packages("classInt")
install.packages("spData")
install.packages("RColorBrewer")
install.packages("dplyr")
install.packages("CARBayesST")
install.packages("olsrr")
install.packages("htmltools")
install.packages("scales")
install.packages("Rtools")
install.packages("stats")
install.packages("corrplot")
install.packages("caret")
install.packages("FactoMineR")
install.packages("factoextra")
install.packages("CARBayes")


#load packages
library(sp) #spatial analysis package
library(spData) #spatial data package
library(spdep) #spatial modeling package
library(rgdal)  # useful to work with esri shapefiles
library(RColorBrewer)  #useful to choose color palette (see Appendix A)
library(classInt)   #useful to make choropleth maps
library(dplyr)
library(CARBayesST)
library(CARBayes)
library(scales)
library(olsrr)
library(htmltools)
library(stats)
library(corrplot)
library(caret)
library(gridExtra)
library(grid)
library(FactoMineR)
library(factoextra)

CHIKGIS <- readOGR(dsn="H:/", layer="Cali_Final")

W <- poly2nb(CHIKGIS, row.names = NULL, snap=sqrt(.Machine$double.eps),
queen=TRUE, useC=TRUE, foundInBox=NULL)
W.list <- nb2listw(W.nb, style = "B")
W <- nb2mat(W.nb, style = "B")

CHIKData = read.csv("H:/CHIK1516_Final3.csv")
CHIKData$logSMR <- log(CHIKData$SMR)
```

```
SMR.av <- summarise(group_by(CHIKData,ID2), SMR.mean = mean(SMR))
CHIKGIS@data$SMR <- SMR.av$SMR.mean

#spatial weight matrix
W.nb <- poly2nb(CHIKGIS, row.names = SMR.av$ID2)
W.list <- nb2listw(W.nb, style = "B")
W <- nb2mat(W.nb, style = "B")


formulaCHIK <-  Observed ~ offset(log(Exp)) + FAC1NEW2 + FAC2NEW2 +
NURSE_0_1 + TIRES_0_1 + pdens0_1 + WATERBOD + TreeDen0_1 + TAvgL6 +
DTRMaxL8 + RHRngL5 + TrainL5 + RainDL6 + CoolDL1 + WarmDL6

modelCHIK <- glm(formula = formulaCHIK, family = "poisson", data =
CHIKData)
resid.glm <- residuals(modelCHIK)
summary(modelCHIK)$coefficients
summary(modelCHIK)$dispersion

moran.mc(x = resid.glm[4761:5100], listw = W.list, nsim = 10000) # week
14

#ST-CAR linear model
modelCARCHIK <- ST.CARar(formula = formulaCHIK, family = "poisson",
data = CHIKData, W = W, burnin = 20000, n.sample = 220000, thin = 10)

print(modelCARCHIK)
parameter.summary <- summarise.samples(exp(modelCARCHIK$samples$beta[ ,
-1]), quantiles = c(0.5, 0.025, 0.975))
round(parameter.summary$quantiles, 3)

rate.est <- matrix(modelCARCHIK$fitted.values, nrow = nrow(W), byrow =
FALSE)
write.table(rate.est,file="H:\\CHIK_Post.csv")



#posterior distributions for the covariate effects
colnames(modelCARCHIK$samples$beta) <- c("Intercept", "FAC1NEW2",
"FAC2NEW2", "NURSE_0_1","TIRES_0_1","pdens0_1", "WATERBOD",
"TreeDen0_1","Tavg0_1","DTRMax0_1","RelHRng0_1","RainT0_1","RainD0_1","
CoolD0_1", "WarmD0_1")
plot(exp(modelCARCHIK$samples$beta[ , -1]))
```

APPENDIX 14: R CODES FOR MODELS 5 & 6 IN CHAPTER 4

```
setwd("C:/temp ")

#install packages - (install them first in case they have never been
installed before)
install.packages("sp")
install.packages("spdep")
install.packages("rgdal")
install.packages("classInt")
install.packages("spData")
install.packages("RColorBrewer")
install.packages("dplyr")
install.packages("CARBayesST")
install.packages("olsrr")
install.packages("htmltools")
install.packages("scales")
install.packages("Rtools")
install.packages("stats")
install.packages("corrplot")
install.packages("caret")
install.packages("FactoMineR")
install.packages("factoextra")
install.packages("CARBayes")
install.packages("AER")


#load packages
library(sp) #spatial analysis package
library(spData) #spatial data package
library(spdep) #spatial modeling package
library(rgdal)  # useful to work with esri shapefiles
library(RColorBrewer)  #useful to choose color palette (see Appendix A)
library(classInt)   #useful to make choropleth maps
library(dplyr)
library(CARBayesST)
library(CARBayes)
library(scales)
library(olsrr)
library(htmltools)
library(stats)
library(corrplot)
library(caret)
library(gridExtra)
library(grid)
library(FactoMineR)
library(factoextra)
library(AER)

ZikaGIS <- readOGR(dsn="H:/", layer="Cali_Final")

W <- poly2nb(DENFGIS, row.names = NULL, snap=sqrt(.Machine$double.eps),
queen=TRUE, useC=TRUE, foundInBox=NULL)
W.list <- nb2listw(W.nb, style = "B")
W <- nb2mat(W.nb, style = "B")

ZikaData = read.csv("H:/Zika16_Final3.csv")
```

```
ZikaData$logSMR <- log(ZikaData$SMR)
ZikaData$Rate <- (ZikaData$Observed/ZikaData$Expected)*1000


SMR.av <- summarise(group_by(ZikaData,ID2), SMR.mean = mean(SMR))
ZikaGIS@data$SMR <- SMR.av$SMR.mean

#spatial weight matrix
W.nb <- poly2nb(ZikaGIS, row.names = SMR.av$ID2)
W.list <- nb2listw(W.nb, style = "B")
W <- nb2mat(W.nb, style = "B")


formulaZika <- Observed ~ offset(log(Expected)) + FAC1NEW2 + FAC2NEW2 +
NURSE_0_1 + TIRES_0_1 + pdens0_1 + WATERBOD + TreeDen0_1 + TavgL8 +
RHRngL7 + TrainL5 + RainDL5 + CoolDL3 + WarmDL3

formulaZika1 <- Observed ~ offset(log(Expected))

# Age514_0_1 + LowEd0_1 + Ag4064_0_1 + WarmD0_1 + Married0_1 + Fem0_1 +
SewPer0_1 + DTRAvg0_1 + Unemp0_1

modelZika <- glm(formula = formulaZika, family = "poisson", data =
ZikaData)
modelZika2 <- glm(formula = formulaZika1, family = "poisson", data =
ZikaData)

resid.glm <- residuals(modelZika)
summary(modelZika)$coefficients
summary(modelZika)$dispersion


resid.glm <- residuals(modelZika2)
summary(modelZika2)$coefficients
summary(modelZika2)$dispersion

moran.mc(x = resid.glm[681:1020], listw = W.list, nsim = 10000) # week
3

dispersiontest(modelZika,trafo=NULL)

#negative binomial
modelZikaN <-glm.nb(Rate ~ FAC1NEW2^2 + FAC2NEW2^2 + NURSE_0_1^2 +
TIRES_0_1^2 + pdens0_1^2 + WATERBOD^2 + TreeDen0_1^2 + TavgL8^2 +
RHRngL7^2 + TrainL5^2 + RainDL5^2 + CoolDL3^2 + WarmDL3^2, data =
ZikaData)
summary (modelZikaN)
resid.glm <- residuals(modelZika)
summary(modelZika)$coefficients
summary(modelZika)$dispersion


#ST-CAR ar model
modelCARZika <- ST.CARar(formula = formulaZika, "poisson", data =
ZikaData, W = W, burnin = 20000, n.sample = 220000, thin = 10)
```

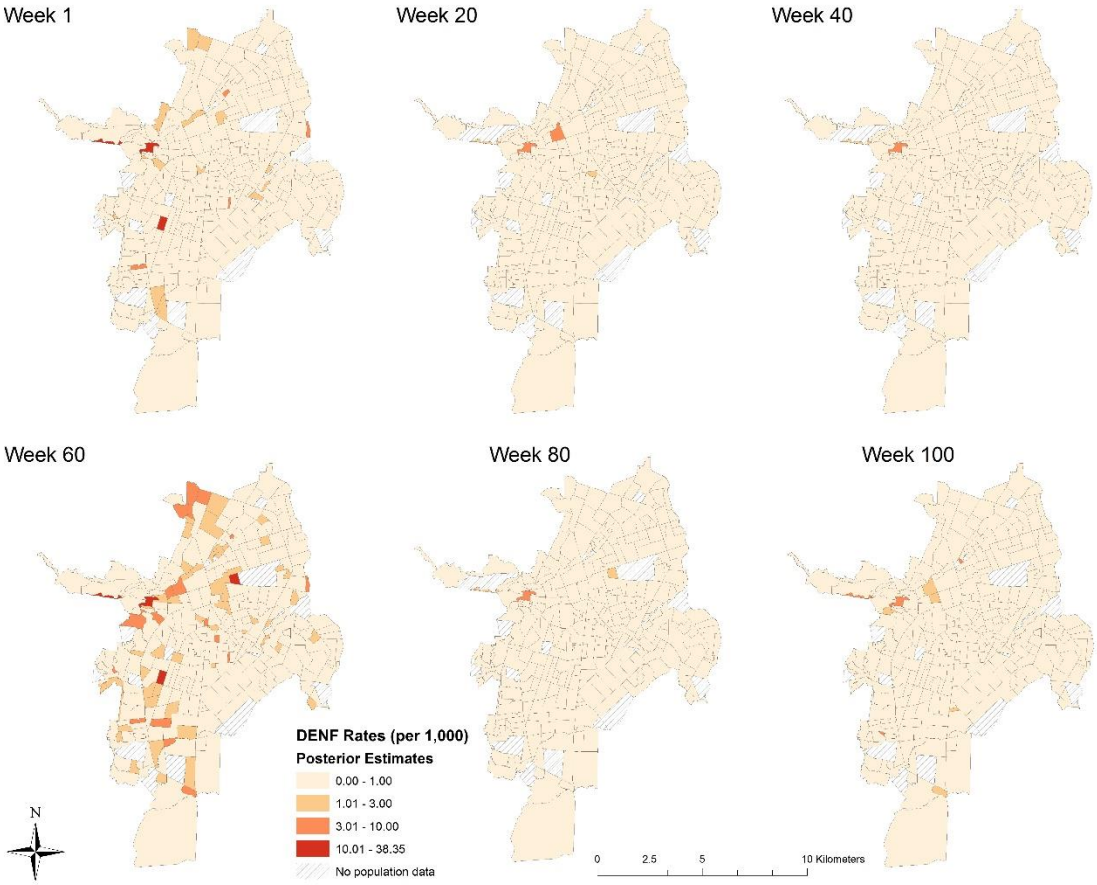
```
modelCARZika2 <- ST.CARar(formula = formulaZika1, "poisson", data =
ZikaData, W = W, burnin = 20000, n.sample = 220000, thin = 10)
print(modelCARZika2)

parameter.summary <- summarise.samples(exp(modelCARZika2$samples$beta[
, -1]), quantiles = c(0.5, 0.025, 0.975))
round(parameter.summary$quantiles, 3)

print(modelCARZika)
parameter.summary <- summarise.samples(exp(modelCARZika$samples$beta[ ,
-1]), quantiles = c(0.5, 0.025, 0.975))
round(parameter.summary$quantiles, 3)

#posterior distributions for the covariate effects
colnames(modelCARZika$samples$beta) <- c("Intercept", "FAC1NEW2",
"FAC2NEW2", "NURSE_0_1","TIRES_0_1","pdens0_1", "WATERBOD",
"TreeDen0_1","Tavg0_1","DTRMax0_1","RelHRng0_1","RainT0_1","RainD0_1","
CoolD0_1", "WarmD0_1")
plot(exp(modelCARZika$samples$beta[ , -1]))

#ST-CAR ar model - binomial
modelCARZika <- ST.CARar(formula = formulaZika, "poisson", data =
ZikaData, W = W, burnin = 20000, n.sample = 220000, thin = 10)


print(modelCARZika)
parameter.summary <- summarise.samples(exp(modelCARZika$samples$beta[ ,
-1]), quantiles = c(0.5, 0.025, 0.975))
round(parameter.summary$quantiles, 3)

rate.est <- matrix(modelCARZika$fitted.values, nrow = nrow(W), byrow =
FALSE)
write.table(rate.est,file="H:\\Zika_Post.csv")
```

APPENDIX 15: R CODES FOR DATA ANALYSIS IN CHAPTER 5

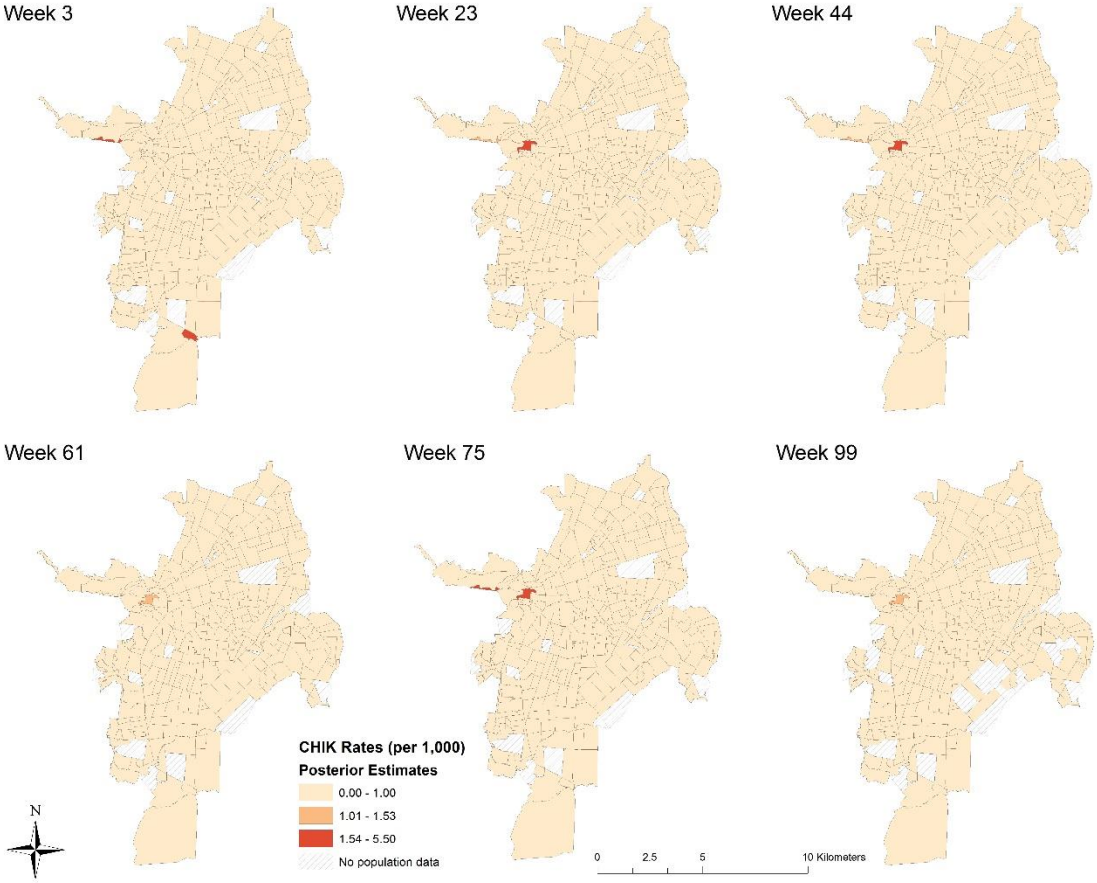
```
setwd("C:/temp ")

#load packages
library(sp) #spatial analysis package
library(spData) #spatial data package
library(spdep) #spatial modeling package
library(rgdal)  # useful to work with esri shapefiles
library(RColorBrewer)  #useful to choose color palette (see Appendix A)
library(classInt)    #useful to make choropleth maps
library(dplyr)
library(CARBayesST)
library(CARBayes)
library(scales)
library(olsrr)
library(htmltools)
library(stats)
library(corrplot)
library(caret)
library(gridExtra)
library(grid)
library(FactoMineR)
library(factoextra)
library(corrplot)
library(MASS)
options(tibble.print_max = Inf)


KAPData = read.csv("J:/Dissertation/Surveys/read_surveys.csv")

#chi2 tests
tbl = table(KAPData$K5D, KAPData$EstrataNEW)
tbl
chisq.test(tbl)
ctbl = cbind(tbl[,"Middle"], tbl[,"Low"])
ctbl
chisq.test(ctbl)

t.test(ctbl)



tbl = table(KAPData$PG3, KAPData$EstrataNEW)
tbl
chisq.test(tbl)

##

formulaVIF <- lm(SUM_K ~ + EstrataNEW + Educ + Sex + Age + Occupation +
CivilStatus + Children + PersonHH + Race + PG4 + PG1 + PG2 + PG3 +
PG5.1 + PG5.2 ,data=KAPData)
ols_vif_tol(formulaVIF)
ols_coll_diag(formulaVIF)

#Knowledge models
```

```
formulaKD <- SUM_K ~ EstrataNEW + Sex + factor(CivilStatus) +
factor(Race) + factor(Occupation)

modelKD <- glm(formula = formulaKD, family = poisson(link = "log"),
data = KAPData)

resid.glm <- residuals(modelKD)
summary(modelKD)$coefficients


#Attitude Models
formulaA <- SumA ~ Educ + factor(Race) + factor(Occupation) + PG1

modelA <- glm(formula = formulaA, family = poisson(link = "log"), data
= KAPData)

resid.glm <- residuals(modelA)
summary(modelA)$coefficients

#Practice Models
formulaPrac <- SumPrac ~ Sex + PG1 + PG2 + PG3 + PG4 + PG5.2

modelPrac <- glm(formula = formulaPrac, family = poisson(link = "log"),
data = KAPData)

resid.glm <- residuals(modelPrac)
summary(modelPrac)$coefficients
```

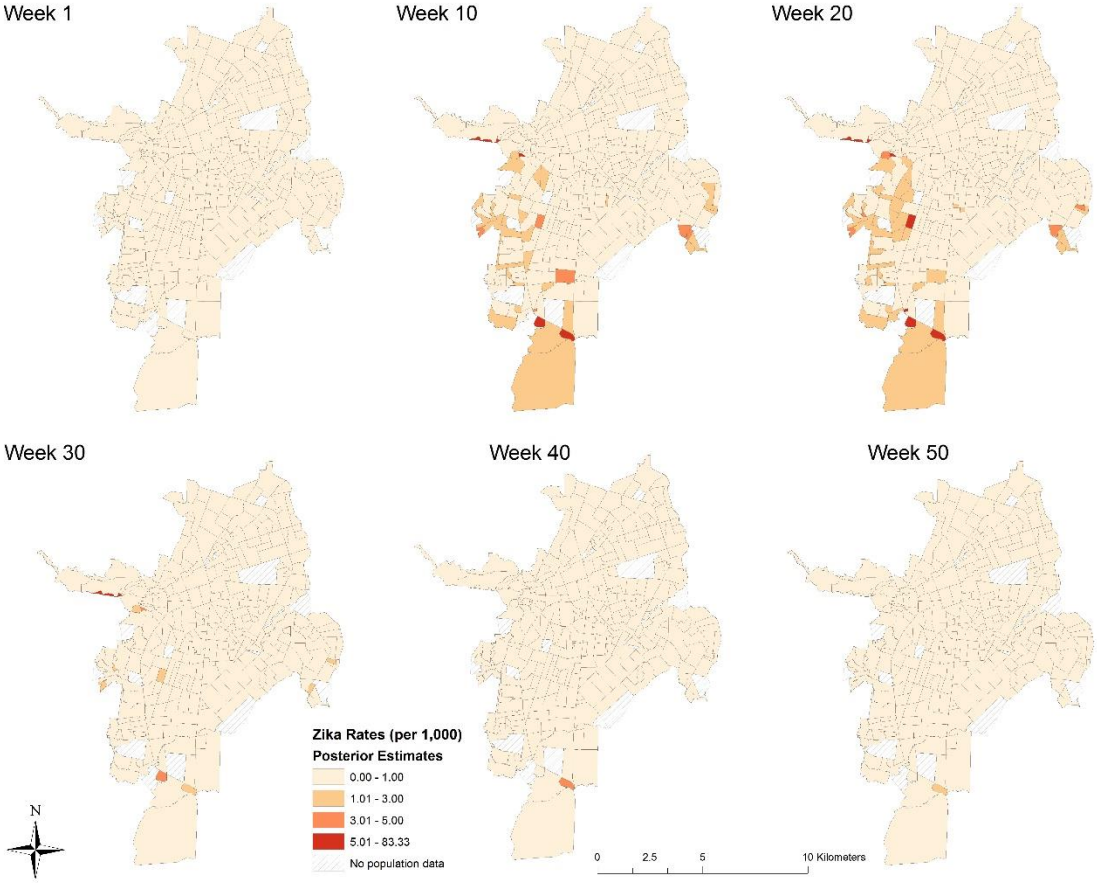
APPENDIX 16: TEMPORAL CROSS-SECTIONS OF DENF RATES (POSTERIOR ESTIMATES) BETWEEN 2015 AND 2016 IN CALI



APPENDIX 17: TEMPORAL CROSS-SECTIONS OF CHIK RATES (POSTERIOR ESTIMATES) BETWEEN 2015 AND 2016 IN CALI



APPENDIX 18: TEMPORAL CROSS-SECTIONS OF ZIKA RATES (POSTERIOR ESTIMATES) IN 2016



APPENDIX 19: REFERENCE MAP OF COLOMBIA