DISSERTATION

SPATIAL ANALYSES OF VECTOR-BORNE DISEASE RISK FOR ALLOCATION OF DISEASE PREVENTION AND CONTROL RESOURCES

Submitted by

Anna M. Winters

Microbiology Immunology and Pathology Department

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2009

COLORADO STATE UNIVERSITY

Aug 1, 2009

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY ANNA WINTERS ENTITLED "*SPATIAL ANALYSES OF VECTOR-BORNE DISEASE RISK FOR ALLOCATION OF DISEASE PREVENTION AND CONTROL RESOURCES*" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate work

Committee Member _____
M.D. Salman, Ph.D.

Committee Member _____
Emily Zielinski-Gutierrez, Dr. P.H.

Adviser _____
Lars Eisen, Ph.D.

Co adviser _____
Chester G. Moore, Ph.D.

Department Head _____
Edward A. Hoover, Ph.D.

ABSTRACT OF DISSERTATION

SPATIAL ANALYSES OF VECTOR-BORNE DISEASE RISK FOR ALLOCATION OF DISEASE PREVENTION AND CONTROL RESOURCES

Vector-borne diseases are threats to human health in both developed and under-developed regions of the world. In 2006, for example, 247 million cases of malaria were reported causing one million deaths mostly among African children. Furthermore, approximately two-fifths of the world's population is at risk for dengue infection and more than 50 million human cases of dengue fever are estimated to occur annually in tropical and subtropical areas of the world. Vaccines are still lacking for many globally important vector-borne diseases. This places a high premium on vector control and other preventative measures, which tend to be costly and labor intensive. Particularly for public health authorities in resource-poor environments, therefore, effective response to vector-borne diseases needs to include a reliable system for cost-effective and efficient targeting of vector control and other preventative efforts in space and time.

The fundamental objective of this dissertation was to investigate spatially and temporally-explicit methods to enhance targeting of vector control and disease prevention efforts against vector-borne diseases in the developed and developing world. To achieve this objective, three vector-borne diseases were investigated: West Nile virus (WNV)

disease in Colorado, USA; plague in the West Nile region of Uganda; and dengue in Merida, Mexico.

Various methods were explored to develop and present spatial information representing risk of vector-borne diseases. Appropriate spatial scales, as well as risk-calculation methods best suited to stakeholders tasked with the role of allocating public health resources, were investigated (Chapter II). Spatial models indicating the risk of human exposure to WNV in the state of Colorado were developed using both epidemiological and entomological data (Chapters III and IV). A similar modeling approach was applied to identify risk areas for human exposure to *Yersinia pestis*, the causative agent of plague, in Uganda (Chapter V). WNV disease models for Colorado and plague models for Uganda provided robust validations indicating that the resulting spatial risk maps contained useful information for disease prevention and control efforts.

The final chapter (Chapter VI) of the dissertation expanded from spatially-explicit methods to also account for the temporal aspect of disease outbreaks or epidemics. This chapter describes the evaluation of the feasibility of an early detection system based on historically-derived thresholds of dengue incidence to identify outbreaks in a timely fashion in Merida, Mexico. The early detection system evaluated in the final chapter has the potential to provide Merida public health authorities with a resource to recognize when the current dengue burden is exceeding historical norms and may be applicable and useful in other dengue endemic areas. The chapters composing this dissertation describe the application of cross-cutting methods used to determine spatial risk in different vector-borne disease systems. Furthermore, the methods used are applicable to the current public

health situation where new vector-borne diseases are emerging and "old" diseases are resurging, further underscoring the importance of targeting limited public health resources.

Anna M. Winters
Microbiology Immunology and Pathology Department
Colorado State University
Fort Collins, CO 80523
Fall 2009

ACKNOWLEDGEMENTS

Thank you to all of my committee members, each of whom offered their time, support and counsel as I completed this dissertation. A special thanks to Dr. Lars Eisen, whose leadership and mentorship has been greatly appreciated. I am proud to be his first Ph. D. student and know that many more will benefit from his thoroughness, support and thought. Dr. Mo Salman, even with his demanding travel schedule, always set aside time to meet with me and provide statistical guidance. Dr. Chester Moore has mentored me from the very inception of my graduate career as he opened the world of vector-borne disease to me via a Fort Collins "tour of vector breeding sites". Who wouldn't want to pursue a career in vector-borne disease after such an experience? Dr. Emily Zielinski-Gutierrez brought much to my studies, including a trip to Connecticut to participate in Lyme disease focus groups – offering a uniquely political experience. Thank you also to Dr. Rebecca Eisen, who provided me with friendship and scientific direction during my fellowship at the Centers for Disease Control. Dr. Barry Beaty and Dr. Carol Blair were instrumental in providing me with funding and an open door to initiate my graduate program at Colorado State University. Finally, thank you to Benjamin Winters and the Maki family for support and encouragement through this process.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

I. INTRODUCTION

Chapter objective

This introductory chapter will establish a foundation for the dissertation research chapters found herein. The research chapters have a common objective to investigate methods to generate results allowing for improved targeting of vector control and disease prevention resources in space and time in order to efficiently and effectively limit morbidity and mortality associated with vector-borne diseases. Research chapters incorporate spatial risk modeling, early detection of incipient outbreaks, and investigation of best practices to present data to stakeholders tasked with efficient allocation of public health resources. The introduction will:

1) Outline the epidemiology of the three vector-borne diseases discussed in this dissertation: dengue, West Nile virus disease and plague;
2) Provide the rationale for applying knowledge of the specific epidemiology and risk factors associated with each of these vector-borne diseases to target disease prevention and control resources;
3) Discuss the advantages and disadvantages of different types of data used to develop spatial models and/or provide disease risk information to the various stakeholders;
4) Indicate the usefulness of spatial risk models and early detection systems for targeting public health resources.

The introduction will provide a brief synopsis for each chapter highlighting how it fits into the over-arching objective of the dissertation. Detailed topic-specific introductions are included at the beginning of each of the research chapters. The appendix provides terminology definitions for several concepts covered in the research chapters.

Importance of vector-borne diseases

Vector-borne diseases inflict major health and economic burdens upon the global community with developing countries suffering the greatest consequences.[1, 2] For example, approximately two-fifths of the world's population is at risk for dengue.[3] More than 50 million cases of dengue fever (DF) and several hundred thousand cases of dengue hemorrhagic fever (DHF) are estimated to occur annually in tropical and subtropical areas of the world.[4, 5] Malaria also inflicts a heavy burden of illness and death, especially among children and pregnant women. In 2006, there were 247 million reported cases of malaria causing one million deaths mostly occurring among African children.[6] These, and a number of other vector-borne diseases, are targeted by the World Health Organization's Special Programme for Research and Training in Tropical Diseases for reduction of disease burden in impoverished areas.[2] Vector-borne diseases are not just a concern for tropical, resource-poor regions. The 1999 recognition of West Nile virus (WNV) disease in New York, and the subsequent spread of WNV across the U.S.[7] reminded us that vector-borne diseases can also be a serious issue in developed countries.[8] WNV disease cases have now been reported from every mainland state of the U.S. Furthermore, WNV has been detected in Mexico, Canada and the Carribean.[8, 9, 10, 11] As outlined above, both developed and under-developed regions of the world are vulnerable to emerging and

resurging vector-borne diseases. This vulnerability is due, in part, to shortcomings in human resources and public health infrastructure aimed at disease prevention and control. It therefore is of great importance that efforts be made to improve disease prevention and intervention strategies.

Reducing the impact of emerging and resurging vector-borne diseases is a considerable challenge, particularly with limited tools and resources available to address the problem. For example, widespread resistance to drugs or insecticides poses a major problem to the control of many important vector-borne diseases such as dengue, malaria, and filariasis.[11] Furthermore, vaccines available for vector-borne diseases are not always widely used in endemic areas (e.g., Japanese encephalitis, tularemia, and plague) and the challenges to develop successful vaccine candidates for other vector-borne diseases (e.g. for malaria and dengue) are great.[12-14]

There are no human vaccines or antivirals available for WNV disease or dengue,[10, 15] two of the vector-borne diseases discussed in this dissertation. This makes vector control the most important current tool for interrupting the virus transmission cycles and decreasing the disease burden associated with WNV disease and dengue.[11] There is a vaccine available for plague, the third vector-borne disease discussed in this dissertation, however, the vaccine is not readily accessible nor useful against pneumonic plague, the most severe manifestation of the disease. Specifically, it is recommended only for individuals (e.g. laboratory personnel) at high risk for exposure to *Yersinia pestis*, the causative agent of plague.[16] Antibiotic therapy, when applied early in the course of infection, is very effective for treatment of plague. Even though vaccines and drug therapies are available, vector control is still important to reduce the risk for explosive

plague epidemics to occur. Severe challenges remain to provide prompt plague treatment in many areas of the world, especially in rural settings in developing countries.

Preventative measures for all three of these diseases must be implemented efficiently if they are to be beneficial and cost effective.* Additionally, resources are often limited for public health organizations, in both developed and under-developed regions of the world. The global importance of vector-borne diseases along with the limitations of public health response mechanisms place a high premium on information (e.g. environmental and social factors [risk factors]) that influence the spatial and temporal distribution of disease cases. As this dissertation will contend, a clear understanding of these risk factors and their occurrence in space and time can aid decision-making and enhance vector and disease control program performance. Furthermore, risk factors can be incorporated into spatial models indicating areas of elevated risk of exposure to specific vectors or vector-borne pathogens. Resultant risk maps can aid in decision-making of where to focus targeted interventions to reduce the transmission of the vector-borne agent versus broad, non-targeted and less cost-effective interventions. Public health resources may also be efficiently targeted by detecting an incipient outbreak early and mobilizing intervention resources in a timely fashion to disrupt the transmission cycle of the agent. Targeting public health resources is essential; especially considering that they are often severely limited. Spatial risk models and early detection of disease outbreaks provide useful tools to achieve spatially and temporally targeted resource allocation.

* For example, poorly targeted vector-control efforts (e.g., overuse of mosquito insecticides) can lead to insecticide resistance in mosquito vectors which, over time, may increase mosquito abundance – an effect opposite the intended purpose of the intervention.

Currently, many health agencies are prioritizing research agendas that focus on the identification of spatial patterns of disease risk, and the early detection of epidemics. For example, the World Health Organization convened the Dengue Scientific Working Group in 2006 to review existing knowledge of dengue and to establish priorities for future research aimed at reducing dengue morbidity and mortality through improved dengue treatment, prevention and control.[17] Their recommendations included strengthening of surveillance systems by developing and validating reliable risk indicators, developing early warning and response systems and identifying triggers of effective responses to incipient epidemics.[18] Similar research agenda priorities have been designated for WNV disease by the Centers for Disease Control and Prevention. Specifically, geographic distributions and risk factors associated with human and animal infection have been investigated to better define target areas for mosquito control in response to documented WNV activity.[19, 20]

Epidemiology of WNV disease, plague and dengue

Overview of the diseases

West Nile virus, of the genus *Flavivirus*, is maintained in transmission cycles that primarily involve birds and *Culex* mosquitoes.[10] Mosquitoes become infected when they feed on avians, including corvids (e.g., crows, magpies and jays), house sparrows, house finches and grackles.[21-23] The primary route of human infection is through the bite of an infectious mosquito. Humans are considered incidental hosts for WNV because they are unlikely to develop virus titers high enough to infect a feeding mosquito.[24] WNV occurs in Africa, Europe, the Middle East and Asia, and, since 1999, also in the Americas.[7, 25]

WNV infection can be asymptomatic or result in disease manifestations ranging from fever, headache and fatigue, to encephalitis resulting in chronic or non-resolving neurological effects or even death.[10, 25] Currently, there is no specific treatment available other than supportive care.

Table I-1 Key aspects of the three vector-borne diseases addressed in the dissertation.

	West Nile virus disease	Plague	Dengue
Infectious Agent	RNA virus; genus *Flavivirus*	*Yersinia pestis*; gram negative bacillus	RNA virus; genus *Flavivirus*; 4 virus serotypes (DEN-1, DEN-2, DEN-3, DEN-4)
Primary Vectors	*Culex* mosquitoes	Rodent fleas	*Aedes aegypti, Aedes albopictus*
Primary Reservoirs	Avians including corvids (crows, magpies and jays), house sparrows, house finches and grackles.	Rodents (e.g., rats, rock squirrels, ground squirrels, prairie dogs)	Humans; Unique dengue virus lineages also in wild primates not involved in human dengue virus cycle.
Current distribution	Africa, Europe, the Middle East, Asia. Recently emerged in the Americas.	Africa, Asia, South America, North America.	Tropical and subtropical areas of the world.
Clinical Features	West Nile fever: Fever, headache, fatigue, lymphadenopathy, eye pain. West Nile neuroinvasive disease: clinical syndromes range from febrile headache to aseptic meningitis/encephalitis, and acute flaccid paralysis.	Three clinical forms: Bubonic plague: enlarged tender lymph nodes, fever, chills. Septicemic plague: fever chills, shock, hemorrhaging. Pneumonic plague: fever chills, cough, difficulty breathing, shock.	Dengue fever: Sudden onset of fever, severe headache, myalgias, arthralgias, leucopenia, thrombocytopenia. Dengue hemorrhagic fever: Shock and hemorrhage which can lead to death.
Treatment	No specific treatment available.	Antibiotic therapy: Drugs of choice are streptomycin or gentamycin, but a number of other antibiotics are also effective.	No specific treatment available.

Plague is caused by the gram negative bacterium *Yersinia pestis*. Fleas become infected primarily by feeding on rodents including chipmunks, prairie dogs, ground squirrels, rats, mice, and other mammals that are infected with *Y. pestis*.[26, 27] *Y. pestis* may also be transmitted to humans by the bites of infected rodent fleas. Plague has a unique

and remarkable history. For centuries, plague had disastrous consequences in Asia, Africa and Europe where, at times, populations were so affected that, according to some accounts, few individuals remained to bury the dead. Epidemics included the "Black Death" which occurred in European cities in the 1300s and killed approximately one-third (20 to 30 million) of Europe's population.[27-29] Today, the World Health Organization reports 1,000 to 3,000 cases of plague every year, the majority of which occur in Africa, Asia and South America.[27] Human plague in the United States occur as scattered cases in rural areas of the west. (an average of 10 to 20 persons each year).[27] There are three clinical forms of plague. Bubonic plague is characterized by enlarged tender lymph nodes, fever, and chills. Bubonic plague may progress to septicemia characterized by fever, chills, shock and hemorrhage. The infection of the lung is termed pneumonic plague for which *Y. pestis* can be transmitted to others through the expulsion of infective respiratory droplets by coughing. Antibiotic therapy is very effective when administered promptly with streptomycin or gentamycin typically prescribed for treatment.[30]

Dengue fever (DF) and dengue hemorrhagic fever (DHF) are caused by four antigenically related virus serotypes (DEN-1, DEN-2, DEN-3, DEN-4) of the genus *Flavivirus*.[15, 31] DF and DHF occur in tropical and subtropical regions, where dengue viruses are maintained in a cycle that involves humans and *Aedes* spp. mosquitoes, principally *Aedes aegypti*. This is a domestic, day-biting mosquito that prefers to feed on humans. Dengue infection can be asymptomatic or produce a spectrum of clinical illness ranging from mild flu-like symptoms to severe and fatal hemorrhagic disease.[32] Currently, there is no specific treatment available other than supportive care.

Emerging and re-emerging vector-borne diseases

Dengue, WNV disease and plague are emerging or re-emerging vector-borne diseases, which are capable of causing disastrous consequences for health and economy of households and communities.[8, 33-35] Epidemics of dengue were first documented in the 18th century.[36] During the 1950's and 1960's, a highly successful *Ae. aegypti* eradication program to control yellow fever in the Americas resulted in dengue being eliminated from most of the region. The eradication program, however, ultimately failed to achieve its goal as countries did not sustain their eradication efforts resulting in progressive reestablishment of both *Ae. aegypti* and dengue viruses throughout the Americas.[37, 38, 39] Dengue also is a severe problem in Southeast Asia. During the aftermath of World War II, rapid urbanization and increased travel resulted in dengue hyperendemicity (co-circulation of multiple dengue virus serotypes) and the emergence of DHF in Southeast Asia.[†,11, 26, 39] DHF is now a common occurrence in Asia, tropical America and the Pacific Islands. The geographical spread of dengue viruses has been attributed to globalization with increased movement of people between continents. This, together with uncontrolled urbanization, has resulted in dramatic increases in DF and DHF worldwide.[36] In less than 20 years, the American tropics and the Pacific Islands went from having few dengue cases to a public health crisis by 2000 with insecticide resistance in the mosquito vectors, hyperendemicity, and emergence of DHF. Today, nearly two-fifths of the world's population live at risk of contracting dengue[40] making this disease the most rapidly spreading vector-borne disease.[17]

[†] DHF is a life-threatening form of dengue marked by increased vascular permeability, thrombocytopenia and hemorrhagic manifestations.

WNV has recently emerged in the Americas, causing a serious public health problem.[7, 24] WNV was first isolated from the blood of a febrile woman in the West Nile province of Uganda in 1937.[25] Later, this mosquito-borne virus was recognized as the cause of epidemics of febrile illness and sporadic encephalitis in Africa, the Mediterranean Basin, Europe, India, and Australia.[8] In recent years, WNV emerged in the Americas. After the initial outbreak in New York in 1999, WNV spread west across North America and southward into Latin America and the Caribbean.[9, 24, 41] By 2003, WNV disease cases had been reported in almost every mainland state of the U.S. The continued circulation of WNV suggests that the virus is permanently established in the Western Hemisphere. Furthermore, the Western Hemisphere epidemic is characterized by a higher rate of neuroinvasive disease compared to most previous WNV epidemics (such as those occurring in the Middle East in the 1950's). This resulted in enormous costs for vector control, testing of the blood supply and other interventions. The high rates of severe manifestations, coupled with the near certainty that WNV now is permanently established in the Western Hemisphere, emphasizes the need for vigilant surveillance followed by targeted disease prevention and control.[10, 20]

Plague is another vector-borne disease that is emerging (or increasingly being recognized) in new geographic locations. For example, in Africa, reported plague cases have increased during recent decades. More than 90% of all reported cases and deaths in the past five years have occurred in India, Madagascar, Tanzania, Mozambique, Malawi, Uganda and the Democratic Republic of Congo (DRC).[29, 42, 43] Following the re-appearance of plague in India during the 1990s, this disease was categorized as a re-emerging disease.[33, 44, 45] The examples provided by WNV disease, plague and dengue

underscore the need for development and application of new methodologies, such as spatial risk modeling, capable of producing results that help control programs to effectively target their resources to high-risk areas for pathogen transmission.

Transmission cycles

The spatial "risk" of disease occurrence, which indicates the probability of future infections or outbreaks, is influenced by the transmission dynamics of the specific vector-borne pathogen within a given geographical location.[46, 47] Major factors include: the interactions between the disease agent (*Y. pestis*, dengue viruses, WNV), the vertebrate host population (rodents, humans, birds), the pathogen vector (flea, mosquito) and the environment. Both WNV and plague are considered zoonoses as the disease agent is maintained by vertebrate animals other than humans. For both diseases, humans are considered incidental pathogen hosts becoming infected through the bite of an infected mosquito (WNV) or flea (plague). More specifically, WNV is maintained in a zoonotic cycle primarily involving birds and *Culex* mosquitoes.[25] *Y. pestis*, the causative agent of plague, is maintained primarily in zoonotic cycles by rodents and their fleas.[‡28, 29] Pneumonic plague is of special concern due to the potential for person-to-person transmission which can cause eruptive plague epidemics in settings with inadequate public health infrastructure and treatment capacity.[42, 43, 48]

Zoonotic, sylvatic dengue virus transmission cycles have been described and involve forest-dwelling *Aedes* mosquitoes and monkeys in forested habitats of West

‡ Humans are most often infected with *Y. pestis* via flea-bites during epizootic periods when rodent hosts perish from infection and infectious fleas are forced to parasitize alternative hosts, including humans.

Africa, Malaysia and Vietnam.[34, 49-52] Humans occasionally become infected with sylvatic dengue in West Africa and possibly Asia; however, humans are tangential to the maintenance cycle which involves mosquitoes and monkeys.[5, 34, 53] These sylvatic cycles involve virus strains that are phylogenetically distinct from those transmitted by domestic and anthropophilic *Ae. aegypti*. The domestic cycles of dengue virus transmission are independent of sylvatic cycles and represent the highest levels of human morbidity and mortality associated with dengue viruses.[50]

Understanding the underlying factors involved in the transmission dynamics of a specific disease agent is critical to the development of targeted prevention and control strategies. Furthermore, determination of environmental and social factors that influence risk for a given vector-borne disease is essential for spatial model development to strategically target limited public health resources.

Environmental and social determinants and heterogeneity of risk

Dengue, plague and WNV disease are similar in that both environmental and social factors contribute to pathogen transmission dynamics, although certain factors affect these disease systems very differently. An understanding of how these environmental and social factors contribute to disease risk can help us to define areas of high or low risk for specific vector-borne diseases, and aids in developing hypotheses and forming research questions of how biological mechanisms contribute to pathogen transmission and spread. In the Americas, dengue cases are mainly concentrated in urban areas where the primary vector is anthropogenic and anthropophilic and human population sizes support dengue virus transmission.[37] High human population densities,

unplanned urbanization[§] and high housing densities are all considered risk factors for dengue infection.[37] Houses that lack window screens or have screens ill-suited to prevent mosquito entrance can increase the risk of human infection.[34] Another risk factor is the lack of consistent, reliable air conditioning, a circumstance often encountered in developing regions, which contributes to an "open window syndrome" where mosquito entrance into housing premises is facilitated.[**3, 54] Lack of waste management leads to accumulation of trash and debris that collect water and serve as mosquito larval habitat. Lack of reliable sources of piped water increases water storage in and around the residence, and thus contributes to the availability of larval development sites. Residences adjacent to lots used for tire storage may experience increased risk of vector exposure as discarded tires provide larval habitat.[37] Cemeteries can also provide excellent mosquito breeding habitat as water collects in graveside flower vases and other nearby receptacles.[17, 34]

Compared to dengue, WNV disease is less concentrated in highly populated urban environments, and often occurs in semi-urban, residential areas and rural environments.[55] Main factors influencing the occurrence of WNV infection in humans include environmental variables and the bloodmeal preferences of *Culex* mosquitoes. Birds are considered the main reservoirs of WNV as they develop viral titers high enough to infect feeding mosquitoes. During the spring and summer, increases in the abundance of WNV-infected *Culex* vectors contribute to a rise in human WNV infections, especially as

---

[§] Unplanned urbanization may contribute to limited urban planning and resources, resulting in a lack of sewer systems, piped water, and trash pick-up services. These factors may contribute to the availability of larval development sites for the container-breeding dengue virus vector.

[**] Some reports indicate that women and children may be at greater risk for dengue, presumably because they spend significant amounts of time in the home environment during the daytime hours when *Ae. aegypti* preferentially feeds.

humans tend to spend more time outdoors during these seasons. High abundances of *Culex* mosquitoes is often associated with the presence of irrigated lawns or drainage systems in residential areas, irrigated agriculture, ranchlands and riparian habitats.[55-57] Human risk for severe WNV disease increases with age,[58] and time spent outdoors increases exposure to infected mosquitoes particularly during dawn and dusk hours.[59, 60] Increased human incidence of WNV disease has been associated with irrigated landscapes,[61] rural agricultural settings,[55] ranchland and riparian areas. Other studies have shown incidence of WNV disease to be associated with dead bird clusters,[22, 62, 63] early season American crow (*Corvus brachyrhynchos*) deaths,[64] increased vegetation density,[65, 66] above-average summer temperatures[67] and urban landscapes.[68]

The environmental and social determinants of plague occurrence are not as clearly understood as for WNV disease or dengue; however, the majority of recent plague outbreaks have occurred in rural areas and have been associated with availability of harborage and food resources for rodents. In some settings, human cases of plague have been correlated with food storage in the immediate home environment.[27-29] In rural areas of Africa, for example, storage of crop harvest in the peridomestic area can attract rodents increasing the risk of transmission of *Y. pestis* to humans.[29, 44, 69] Studies in New Mexico have indicated that proximity to pinon-juniper ecotones and increased precipitation are both correlated with elevated risk for plague.[70] Furthermore, failure to control fleas on dogs and cats also appear to be a possible risk factor for human exposure.[71] Plague risk has also been positively associated with habitat heterogeneity or fragmentation which may be important for inter-epizootic maintenance of *Y. pestis*.[72-74] Habitat fragmentation restricts movement of plague-susceptible small mammal populations and also separates

large populations into distinct subpopulations.[72, 73, 75-77] These subpopulations may increase the probability of local persistence of plague foci as "burn through" effects of intense epizootics will be limited because of the separate rodent populations, allowing persistent maintenance of plague transmission.

Seasonality

The transmission dynamics of the causative agents of dengue, WNV disease and plague typically exhibit seasonal trends based on environmental factors.[78] The epidemic behavior of dengue and WNV disease, for example, correlates closely with fluctuations in temperature and rainfall.[24, 53, 67, 79] Warmer temperatures shorten the extrinsic incubation period (EIP) which is defined as the time between a mosquito taking an infectious bloodmeal and being able to transmit the virus to a susceptible host.[80] A decrease in the EIP increases the probability that these mosquitoes will be able to transmit their specific pathogen to hosts within their lifetimes.[53, 81] Increased rainfall affects dengue virus transmission by providing an abundance of rain-filled containers (discarded tires, plastic containers) leading to a subsequent increase in vector abundance and human-vector contact.[37, 53] However, during dry periods residents lacking piped water may increase storage of water inside the housing structure or nearby. This provides larval development sites and may increase the abundance of dengue virus vectors during periods of reduced rainfall.[37, 53, 82] Year-to-year variation in dengue virus transmission intensity may also be observed; this can be related to social factors, changes in herd immunity, or changes in surveillance, public health staff, and/or reporting practices. For example, in the case of dengue, changes in urban development (e.g. piped water systems or enhanced or reduced

trash pickup services) may affect the number of cases from year to year. In the specific case of Merida, Mexico, the study area for Chapter VI, *Ae. aegypti* adults and dengue cases may be present throughout the year but both typically peak in the rainy season from July to October.

Culex mosquitoes, the principal vectors for WNV, exhibit a positive correlation with precipitation.[24, 83] For example, in California, summer abundance of *Cx. tarsalis* has been shown to be directly related to rainfall and snow depth, as well as river runoff. Larval habitats of *Cx. tarsalis* are closely associated with irrigated farm and ranch lands and this mosquito is associated with rural and semi-urban cases of WNV disease. Irrigation water is often the most common source of larval habitats.[57, 84] Larvae of *Cx. pipiens*, a vector that is often associated with urban outbreaks of WNV disease, are typically found in water with high organic content, such as sewer drains. *Cx. pipiens* population densities are typically highest during the dry season when water evaporates and the concentration of organics increase.[57] In Colorado, the study area for Chapters II, III and IV, WNV is transmitted primarily by *Cx. tarsalis* which is associated with irrigated and riparian areas and typically reaches peak abundance during July-August.

Plague epidemics and epizootics have also been associated with temperature and rainfall in North America and Africa.[29, 70, 85, 86] For example, increased plague activity in Africa has been associated with higher precipitation rates and cooler temperatures, while case reports have decreased during drier periods with average temperatures in excess of 27°C. Higher elevations have also been associated with increased plague occurrence.[70] The epidemic dynamics of plague, including the associations listed here, has been

researched and described; however, comprehensive explanations for these phenomena are lacking. Further studies on plague ecology therefore are warranted.

Landscape ecology and development of spatial risk models and risk maps to target resources

From Pavlovsky to present

Relationships between environmental and social determinants along with other underlying factors associated with seasonality of transmission can be exploited and modeled to identify areas and time periods of high or low disease risk.[46, 47, 87] These models may incorporate epidemiological, entomological, weather/climate and other environmental data to produce an output, typically visualized in map format, indicating the distribution of high risk versus low risk areas for vector or pathogen exposure. The basic rationale for the development of spatial risk models comes from the field of landscape ecology, where particular focus is given to the interactions among heterogeneous landscapes and the effects these interactions bear on biotic processes.[87-89] Because of the close associations between vector-borne pathogen transmission and environmental conditions, landscape features have been recognized as important determinants of pathogen transmission.[47, 87] Evgeny Pavlovsky, a Russian parasitologist, was one of the first to incorporate landscape ecology into a documented analysis of pathogen transmission. He coined the phrase "landscape epidemiology" in the 1930s based on studies in Central Asia on pathogens transmitted by arthropods.[46, 90] Pavlovsky's landscape epidemiology concept includes three fundamental principles:

(1) Diseases tend to be limited geographically; (2) Spatial variation or heterogeneity in disease occurrence arises from underlying variation in the physical and/or biological conditions that support the pathogen, its vectors, and reservoirs; (3) Mapping the abiotic and biotic conditions may identify current risk and predict future change in risk. These principles led Pavlovsky to develop the theory of natural focality of transmissive diseases (i.e., the concept of natural nidi [foci] of disease) as determined by the spatial processes operating across the landscape.[90]

Some of the earliest vector-borne diseases to be examined using the principles of landscape epidemiology were tularemia, trypanosomiasis and Crimean Congo hemorrhagic fever. Studies identified specific natural landscape features and land-use patterns by humans associated with the distribution of reservoir hosts and pathogen vectors.[91, 92] More recently, the study of environmental determinants related to vector-borne disease systems has included a number of different disease agents including mosquito- borne,[56, 61, 67, 83, 93] flea-borne,[26, 70, 72] and tick-borne agents.[94-96]

One of the most powerful applications of landscape epidemiology is the development of spatial risk models which display, in map format, distributions of areas with high or low risk of vector or pathogen exposure, theoretically allowing public health organizations to efficiently target and implement disease prevention and control measures. With the objective of improved vector-control performance, efforts to develop and apply spatial risk models have increased in recent years. Common protocols used for developing spatial risk models share several basic features including:[46, 47, 87]

1) Construction of maps displaying vector presence or abundance, reservoir presence or abundance, or disease occurrence or incidence;

2) Application of geographical information system (GIS) and remote sensing (RS) technologies (RS; continuous satellite-based collection of measurements which characterize the earth's environment)[97]
3) Identification, via statistical modeling techniques, of variables (biotic or abiotic) with the strongest association with vector presence or abundance, reservoir presence or abundance, or disease incidence or occurrence;[98]
4) Prediction of risk in areas outside the areas where the data were collected by projecting the distribution of the identified variables;
5) Use of risk map outputs from the models to guide interventions, including insecticide application, prevention practices, and other resource allocation endeavors.

This is the foundation for a number of studies conducted in recent years that share a common objective to model and map vector species distribution/abundance and disease presence or incidence. Additionally, the use of RS technologies in the field of vector-borne disease has gained momentum within the past twenty years. A common strategy for inclusion of RS data into spatial modeling development is to correlate different aspects of the environment (e.g. land use/land cover, temperature, humidity, elevation, greenness of a landscape that are remotely sensed) with vector or disease data collected by ground-based surveillance activities.[87, 99-101] Spatial modeling techniques have commonly been applied to mosquito-borne diseases. Studies have focused on identifying mosquito breeding areas (e.g. wetlands) using RS to characterize land use and land cover.[62, 65, 99, 101-107] Spatial risk models specific to a number of mosquito-borne pathogens including WNV,[108] *Plasmodium* spp. (protozoa which are the causative agents for malaria),[109]

dengue,[110-114] and Ross River virus[101] have been developed that identify areas with high risk of exposure to vectors and with potential for pathogen transmission to humans. Spatial models have also been developed for a number of tick-borne diseases, including Lyme disease, the most commonly reported vector-borne disease of humans in the U.S. and Europe.[115, 116] Models for plague have been developed, but on a more limited scale compared to mosquito-borne and tick-borne diseases. Spatial models indicating the distribution of high-risk areas for *Y. pestis* transmission have been developed for the southwestern U.S.[70, 117] Furthermore, Neerinckx et al. (2008) recently constructed a spatial model characterizing plague habitat for the African continent.[118] Many of these spatial models cover expansive areas -- even entire continents, and provide generalized maps of the distribution of risk areas for vector-borne diseases. Often, these models rely on coarse RS data that lacks the degree of resolution needed to support operational control activities at fine spatial scales.

Application of spatial risk models and risk maps

WNV disease example

Spatial models can aid in predicting levels of risk in areas where surveillance data (e.g. vector abundance, human case data) are limited. For example, WNV emerged in the U.S. in 1999 and rapidly spread across the country,[24] forcing states and counties to rapidly develop surveillance and control capability. Local funding was often very limited for WNV surveillance and many towns were unable to generate information on the abundance and distribution of local WNV mosquito vectors. In these circumstances, use of spatial modeling could have improved the use of limited control resources by

complementing existing surveillance, without compromising the integrity or usefulness of data, in the effort to define high risk areas for vector and WNV exposure.

Targeted surveillance

Spatial risk models for vector-borne diseases commonly indicate areas with high potential for pathogen transmission and can be consulted during the development of surveillance programs.[46] For example, mosquito abundance and infection rates can be used to assess the threat of pathogen transmission within an area. Developing a mosquito surveillance program based, in part, on a spatial risk model increases the probability of obtaining virus isolates by focusing mosquito trapping efforts or sentinel animals in areas with the greatest risk of pathogen transmission (i.e., areas having ecological or demographic traits conducive to the transmission of the specific vector-borne pathogen).

Targeted prevention resources

Prevention methods for vector-borne diseases vary based on the specific disease. For all vector-borne diseases, it is necessary that limited prevention resources are targeted to areas representing the highest risk of infection or greatest disease burden. Prevention resources may include education and community outreach to communicate the necessity and effectiveness of personal protective measures.[119, 120] Preventative education may also communicate the effectiveness of window screens, door screens and/or insecticide-treated materials (ITMs) to reduce the abundance of disease vectors in the living area (e.g. *Ae. aegypti*, dengue prevention).[121, 122] Furthermore, preventative campaigns may inform the public of factors contributing to plague risk. For example, in Uganda, individuals that

sleep on the floor or spend time in the food preparation huts are at increased risk for bites from rodent fleas which may be infected with *Y. pestis* (personal communication, May 2009, Emily Zielinski-Gutierrez, Rebecca Eisen). Prevention campaigns also communicate the need for repellent use in areas where WNV is present. All of these preventative efforts might be more efficiently applied by using information provided by spatial risk models and risk maps defining the location of high-risk areas.

Source reduction is another key prevention strategy that can be refined using information gained from spatial risk models. The goal of source reduction is to reduce mosquito or tick habitat (for immature and/or adult stages) and/or zoonotic host habitat (e.g. rodent harborage).[19, 25] Source reduction is arguably one of the most effective methods used to provide long-term mosquito control, although this practice is time and resource intensive.[60] Consequently, source reduction campaigns may benefit from spatial modeling and mapping by increased capacity to target initiatives to areas with high risk of pathogen transmission.[19] In dengue endemic areas, public education campaigns exist to encourage residents to dispose of trash and containers that are potential mosquito larval development sites and to encourage techniques to eliminate larvae from domestic water storage containers.[15, 34, 37, 123] Source reduction can substantially reduce mosquito abundance and the need for repeated applications of insecticides.[19] This remains a major preventative strategy, especially for reduction of dengue risk.[37, 123] In plague endemic regions, such as Uganda, where subsistence agriculture practice is very common, one strategy to reduce the number of rodents and therefore, the abundance of fleas in the peridomestic environment, is to keep crop harvest outside of the living environment (the hut).

Targeted chemical control

Chemical insecticides to control vectors are often applied when source reduction is not feasible or is insufficient given the extent of vector habitat.[37, 57] Use of chemical insecticides is often initiated when entomological surveillance indicates the presence of infected adult vectors or when human cases occur.[19, 57] For WNV disease and dengue, chemicals are used to control both the immature and adult mosquito life stages.[19, 37, 57] The objective of larviciding is to control the immature stages in their development sites before adults can emerge and disperse. When done effectively, this may maintain the adult population at a level where pathogen transmission intensity is low. Adulticiding aims to kill adults by ground or aerial application of insecticides.[57, 122] Indoor space spraying or residual spraying are especially effective against anthropophilic mosquitoes such as *Ae. aegypti*.[37] Adulticides, larvicides, and other control measures should ideally be applied in a targeted manner based on surveillance data and/or risk maps indicating areas of high risk for transmission of the vector-borne pathogen. Spatial and temporal targeting of control activities becomes especially important during epidemic periods when large numbers of human cases can occur within short periods of time.

Plague control is achieved principally by attacking the flea vectors via introduction of insecticidal dusts in commensal and/or wild rodent runways or burrows.[30] Indoor residual sprays may also be applied to reduce the flea burden and potential risk of *Y. pestis* transmission to humans.[30] Many vector control methods are expensive and labor intensive and, consequently, it is crucial that control activities be targeted with a high level of accuracy and efficiency.

Use of spatial modeling to generate improved risk maps

Traditionally, vector control initiatives have relied on simple "pin-maps" developed from previous years of surveillance data for directing their intervention campaigns. The key limitation of these pin-maps is the inability to extrapolate from a given data collection point to other areas. Spatial risk modeling addresses this by using quantitative relationships between environmental and/or demographic factors and risk of vector or pathogen exposure (for a given data collection point or area) to produce spatially continous risk maps. This allows for extrapolation to areas where surveillance data are limited or nonexistent. Based on the spatial risk model, vector control resources can be more precisely targeted to attain the greatest effectiveness in reducing disease burden.

Early detection of disease outbreaks

Static spatial risk models identify the areas at greatest risk for disease, but do not account for the temporal aspect of disease outbreaks or epidemics (e.g. the time periods at greatest risk for pathogen transmission). Early identification of an outbreak can reduce the intensity of the outbreak, and thus reduce the overall disease burden, by allowing for timely mobilization and implementation of control activities. Early detection systems alert public health practitioners to outbreak patterns early, and can also target these alerts to specific areas. Early detection systems typically monitor routine case reports for aberrations above what is historically expected. An early detection system that includes

spatial and temporal aspects can indicate not only when, but where interventions should be applied.

Early detection systems are differentiated from early warning systems which monitor risk indicators (e.g. meteorological data such as rainfall and temperature; immune status of a population) used to predict timing and location of an increase in disease prevalence.[124, 125] Early detection systems have been developed for a number of diseases.[126-128] Many of these diseases are reported by state health departments to the Centers for Disease Control and Prevention and information about epidemic alerts are made available weekly to clinicians, epidemiologists and other members of the public health field.[129, 130] Early detection of mosquito-borne disease outbreaks is crucial given their explosive nature, especially when conditions are favorable for high mosquito abundance and the level of immunity in the human population is low. Interventions that focus on reducing the number of infectious mosquitoes must be applied in a timely fashion to be most beneficial. Mosquito control interventions are only minimally effective (both in terms of cost effectiveness and reduction of pathogen transmission) if applied at the height of an outbreak or epidemic, instead of early when the number of reported cases begins to increase above the norm.

Early detection systems may aid dengue control programs, in particular, because of the volatile nature of dengue virus transmission in urban environments. As the virus is transmitted from person to person via *Ae. aegypti*, high population density contributes to eruptive dengue virus transmission during outbreak periods.[37] Mosquito control interventions take time to mobilize, and must be applied early in the outbreak or epidemic to effectively reduce the abundance of dengue virus-infected *Ae. aegypti*. Early detection

of dengue outbreaks is also essential to identify possible introduction of new virus serotypes, which can cause an increase in cases of DHF. By incorporating early detection systems into routine dengue surveillance, outbreaks may be identified in their early stages allowing appropriate interventions to be mobilized in a timely fashion.

Cost effectiveness of spatial risk models and early detection of outbreaks

Resources allocated to reduce vector-borne disease occurrence (e.g. personnel, equipment, insecticide, prevention materials, and treatment) are often very limited. Application of spatial analyses for development of risk models and risk maps to target limited resources is an investment in the front end of the intervention efforts, but once the models and maps have been developed and are used to direct intervention efforts, they have the potential to reduce vector-borne disease using fewer resources than without integration of spatial risk models into the overall management scheme. For example, use of spatial risk models is a cost-effective approach to identify areas of high risk of pathogen transmission to target larvicide or adulticides to those areas. Extrapolating these models to areas lacking surveillance data can provide helpful information of the location of high risk areas for pathogen transmission, allowing direction of intervention efforts even in areas with minimal surveillance data. Furthermore, use of spatial risk models to identify areas most at risk for pathogen transmission may allow targeted distribution of prophylaxis, prevention education, or ensure that sufficient clinical resources are available for treatment of infection. Similarly, application of an early detection system has great potential to reduce the height of an outbreak by identifying an increase in transmission early, and to direct intervention resources (e.g. mosquito control) to specific

areas in a timely fashion. Given the cost-effectiveness of spatial and temporal targeting of limited resources, significant reductions may be observed in the annual budgets of vector-borne disease control programs if early detection systems and/or spatial risk models and risk maps are integrated into intervention efforts.

An important cost-savings perspective is that GIS software are becoming more user-friendly (and thus can be used with less specialized training). There are a number of free or low-cost mapping and GIS software emerging with good capacity for basic mapping and spatial analyses. Furthermore, high quality GIS and RS data are increasingly being made available at no or minimal cost. Finally, it should be noted that mapping and GIS capacity developed for public health can also be used for a variety of purposes (e.g. city planning) which provides opportunities for cost-sharing between different branches of local government.

Data considerations for spatial model development

For vector-borne disease spatial risk model construction, different types of dependent data (e.g. human case count or disease incidence, entomological data, animal surveillance data) can be considered. It is important to not only understand the relative advantages and disadvantages of each type of data, but also to factor in the epidemiology of the disease in order to ensure selection of the most appropriate data.[93, 131]

Entomological data

Risk model development based on entomological data, such as presence or abundance of vectors or infected vectors, can provide information on the abiotic factors

most conducive to vector survival, proliferation and infection. Cold temperatures may cause mortality, slow arthropod development rates, or decrease the window of time for host seeking activities.[67, 79, 132-134] Warmer temperatures not only increase development rates and lengthen the window of time available for host seeking behavior (e.g. for mosquitoes between sunset and sunrise) but also can decrease the extrinsic incubation period and thus intensify the pathogen transmission cycle.[80] The distribution, abundance and composition of vector species may also be significantly affected by rainfall as it may create or increase larval habitats and provide higher humidity necessary for survival of many arthropod vectors.[78, 135-137] Landscape and vegetation type may also be a driving factor in the distribution and abundance of arthropod vectors and vertebrate hosts and can be included in a spatial model to identify high-risk areas for transmission.[47, 87]

The use of entomological data for development of spatial risk models has a strong advantage over other data types (e.g. human case data) because site-specific vector presence and abundance can be correlated with specific environmental variables in the same location that, in turn, may be used to extrapolate risk models to areas where surveillance data are lacking.[46, 93] For example, *Cx. tarsalis* abundance is associated with irrigated agriculture habitat[56] and this association has been defined based upon trapping of mosquitoes across a variety of habitats.[84] Associations of this nature between vector abundance and habitat type can be exploited for spatial risk model development and extrapolated to define the risk of transmission in areas where surveillance data are otherwise lacking or is limited.[93] Furthermore, field sampling can provide fine-scale assessments of vector abundance that may enhance the spatial resolution of risk assessments compared to other data types such as human case data.[93]

Entomological data can indicate the potential for transmission of the infectious agent before human cases of disease occur. They also supply information on potential arthropod vector species and estimates vector species abundance.[46, 57] Furthermore, entomological surveillance may provide information on infection rates in different vector species as well as information to assess vector control efforts. Many arthropod vectors transmit multiple human pathogens. Therefore, modeling the distribution and abundance of a single vector can be used to assess the potential for risk of exposure to multiple disease agents. Entomological data may also inform spatial models on the risk of transmission of specific vector-borne pathogens in natural areas or recreation areas where the human population base is inadequate to provide accurate disease incidence data.[138]

The main disadvantage for the use of entomological data for spatial model development is the requirement of many resources to obtain the data including personnel, vector trapping equipment, and identification expertise.[57] Entomological data also do not fully imply that the vectors are taking human blood meals and therefore present risk for human infection (personal communication, Rebecca Eisen, May 2009). Furthermore, entomological data do not account for the importance of human behavior such as personal protection measures or opportunities for human-vector contact.[59, 93] In the case of dengue, especially, entomological data are problematic as *Ae. aegypti* females are difficult to collect.[122] This has caused entomologists to focus on the collection of the immature larval and pupal stages of the mosquito.[122, 131] Larval and pupal indices, however, were predictive of risk of dengue virus transmission in some studies[139, 140] but showed poor predictive capability in others.[141, 142]

Vertebrate host data

Zoonotic diseases (i.e. where the disease agent is transmitted from vertebrate animals to humans, often by vectors) create the opportunity to apply data on the distribution and abundance of possible animal reservoirs in the development of spatial risk models. For example, using disease agents covered in this dissertation, spatial risk models for WNV disease have been developed using dead bird reports as indicators of viral transmission.[21, 143-145] In terms of plague risk models, rodent trapping and serosurveys can be used to identify potential plague hosts in an area, the number and kinds of infesting fleas, and whether *Y. pestis* has infected the rodent population.[146] Additionally, information on rodent die-offs related to plague epizootics[146] may be helpful additions to spatial models in order to identify the risk of transmission to humans. Clearly, there are a number of dependent datasets that may be applied to the development of spatial risk models. Creating an ideal model is often informed and/or constrained by data availability, knowledge of the epidemiology of the pathogen, as well as the intended application of the spatial risk map resulting from the model.

Epidemiological data

There are clear advantages to using epidemiological data to understand spatial patterns of human risk of exposure to vector-borne disease agents. First, human case data are mandatorily collected for notifiable diseases (any diseases required by law to be reported to government authorities)[††129] and therefore are often more readily available

[††] In the U.S. each state has its own laws concerning what diseases are reportable to the state health department. The Centers for Disease Control and Prevention is the agency charged with the responsibility

compared to entomological data. A human case of a vector-borne disease also unequivocally demonstrates human contact with the disease agent and thus represents actual rather than potential disease risk.[93, 131] Both epidemiological and entomological data provide the potential to identify socio-economic risk factors as well as environmental predictors of disease risk.

Several limitations exist for application of epidemiological data to construction of spatial risk models. First, not all vector-borne diseases are mandatorily notifiable (e.g. babeosis, Colorado tick fever, tick-borne relapsing fever).[93] Second, inherent bias may exist as case definitions and reporting practices for vector-borne diseases can vary from region to region and across time making the construction of spatial models over an extended period of time difficult. This especially holds true if the vector-borne disease is rare, which necessitates the incorporation of several years of data.[131] Additionally, epidemiologic data from humans only provides reliable information for risk in areas with an adequate population base.[93, 138] Furthermore, diagnostic bias may occur if socio-economic differences in the likelihood of seeking health care result in spatial disparities for disease detection (personal communication, Annette Bachand, May 2009). Also, the location of residence, rather than the site of probable pathogen exposure, is often the only information available for construction of spatial models. This limitation is especially challenging for diseases where exposures may occur outside of the peridomestic setting, for example, during recreational activities. Finally, asymptomatic or mild infections may not be detected. The potential for long delays in diagnostic laboratory confirmation of suspected disease cases may also contribute to under reporting, or significant delays

---

for maintaining the notifiable disease system in the U.S.. Many other countries also maintain notifiable disease systems similar to that in the U.S..

between onset date and reporting.[131] If epidemiological data are used to identify areas of risk of vector-borne diseases, consideration of the best method for presenting the spatial disease data is important. This includes either use of point locations for disease cases or aggregating disease counts or disease incidence to specific administrative boundary units. Overall, however, both epidemiological and entomological data provide useful information for the development of risk models. The incorporation of both data types into a spatial model may offset the biases of each, however, consideration of the type of biases that each data set bring to the spatial model is important to qualify the results of the spatial model. Ultimately, using entomological and epidemiological data in conjunction and with observance of possible biases in each dataset can enhance the effectiveness of spatial models to correctly identify areas of high and low risk of exposure to vector-borne pathogens.

Rationale and outline of dissertation

Vector-borne diseases inflict unacceptable health and economic burdens upon the global community. Both developed and underdeveloped regions of the world have shown vulnerability to emerging and resurging vector-borne diseases. This, in part, is due to shortcomings in human resources and public health infrastructure which reduces the effectiveness of disease prevention and control efforts to focus limited resources to areas with greatest risk of pathogen transmission. Reversing the trend of emerging and resurging vector-borne diseases is a considerable challenge especially as prospects for development of vaccines for major vector-borne diseases are not promising. Disease surveillance and application of vector control must be applied in a targeted fashion to

effectively and efficiently reduce morbidity and mortality. Targeting vector control and prevention resources is especially crucial in under developed regions of the world, where resources are severely limited. The fundamental objective of this dissertation, therefore, was to investigate spatial analyses that can be used to identify risk areas toward the goal of targeting vector control and disease prevention resources to efficiently and effectively limit morbidity and mortality associated with vector-borne diseases in the developed and developing world. To achieve this objective, three vector-borne diseases were investigated: WNV disease in Colorado, plague in the West Nile region of Uganda, and dengue in Merida, Mexico. These diseases not only exhibit unique epidemiological characteristics important to consider in achieving the objective, but also are considered public health priorities in the developed and developing world.

Within the objective to investigate spatial analyses that can be used to target vector control and prevention resources, Chapter II explores the most appropriate methods to present information for spatial risk of vector-borne disease (WNV disease) in map format. This includes the most appropriate spatial scale as well as the method (e.g. case count versus disease incidence) best suited for stakeholders tasked with the role of allocating public health resources. In Chapter III, human case data and environmental correlates are used to develop a predictive spatial model for risk of WNV exposure in eastern and western Colorado. The aim of this chapter was improved spatial targeting for costly mosquito and WNV surveillance and control schemes. This is especially important as mosquito control and personal protection against mosquito bites are the only preventative measures available to reduce human WNV disease cases. Chapter IV investigates a different approach to construction of a spatial model by assessing

entomological risk for the northern Colorado Front Range. A risk classification index was created combining data from the independently derived measures of entomological risk from Chapter IV and epidemiological risk assessed in Chapter III. The entomological model was then scaled up to cover the entire state and its performance assessed in comparison to the epidemiologic WNV disease data.

Chapter V focuses on plague, a flea-borne bacterial disease, with the purpose of describing the spatial distribution of reported clinical plague cases in the West Nile region of Uganda, identification of ecological correlates of incidence, and incorporation of these variables into a predictive model that indicates areas of plague risk. This chapter not only demonstrates that plague incidence can be modeled at a parish level scale based on environmental variables, but also identifies parishes where cases may be under reported which may serve to direct enhanced surveillance and preventative measures to decrease the burden of plague.

The final chapter of the dissertation (Chapter VI), focused on developing an early detection system for dengue outbreaks in Merida, Mexico. Historically-derived thresholds of dengue incidence were calculated and alerts identified if current incidence exceeded historical levels indicating an incipient outbreak. The thresholds developed in this chapter were specific to small regions within the city of Merida (BGSAs) that are manageable by vector control practitioners in Merida in order to apply a targeted vector control response. Despite continuous efforts by the State Health Services of Yucatan to stem the tide of dengue through vector control, there exists a disturbing trend toward increasing numbers of DHF cases in the Yucatan. Early detection of incipient dengue outbreaks can enable timely mobilization of intervention resources targeted to specific

regions within a city and increase the probability of reducing transmission prior to the height of an outbreak.

The chapters found herein have been composed in such a manner to achieve a single over-arching objective: to develop spatial models and historically-derived thresholds to target vector control and disease prevention resources to efficiently and effectively limit vector-borne disease morbidity and mortality in the developed and developing world. The methods used are cross-cutting and may be applied to numerous disease agents. This is a crucial element in the current public health environment where new vector-borne diseases are emerging and or "old" diseases are resurging. The requirement for greater efficiency will only increase as limited public health resources are pitted against the rising tide of vector-borne disease burdens in the global community.

References Cited

1. World Health Organization: The Global Burden of Disease -- 2004 Update. Geneva, 2008.

2. World Health Organization: Special Programme for Research and Training in Tropical Diseases. 2009 http://apps.who.int/tdr/svc/about.

3. WHO: Report on Dengue. Geneva: World Health Organization, Special Programme for Research and Training in Tropical Diseases (TDR), 2007; 160.

4. Gubler D: Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21rst century. Trends in Microbiology 2002; 10: 100-103.

5. Gubler DG, Kuno G: Dengue and Dengue Hemorrhagic Fever. Oxfordshire: CABI, 1997.

6. World Health Organization: Malaria. 2009. Available at: http://www.who.int/mediacentre/factsheets/fs094/en/index.html.

7. Nash D, Mostashari F, Fine A, Miller J, O'Leary D, Murray K, Huang A, Rosenberg A, Greenberg A, Sherman M, Wong S, Layton M: The outbreak of West Nile virus infection in the New York City area in 1999. N Engl J Med 2001; 344(24): 1807-14.

8. Petersen LR, Roehrig JT: West Nile virus: A reemerging global pathogen. Emerg Infect Dis 2001; 7(4): 611-614.

9. Elizondo-Quiroga D, Davis CT, Fernandez-Salas I, Escobar-Lopez R, Velasco Olmos D, Soto Gastalum LC, Aviles Acosta M, Elizondo-Quiroga A, Gonzalez-Rojas JI, Contreras Cordero JF, Guzman H, Travassos da Rosa A, Blitvich BJ, Barrett ADT, Beaty

BJ, Tesh RB: West Nile virus isolation in human and mosquitoes, Mexico. Emerg Infect Dis 2005; 11: 1449-1452.

10. Centers for Disease Control and Prevention: West Nile Virus. 2007 (Sept 21) http://www.cdc.gov/ncidod/dvbid/westnile/.

11. Gubler DJ: Resurgent vector-borne diseases as a global health problem. Emerg Infect Dis 1998; 4(3): 442-450.

12. Pierce SK, Miller LH: World Malaria Day 2009: what malaria knows about the immune system that immunologists still do not. J Immunol 2009; 182(9): 5171-7.

13. Centers for Disease Control and Prevention: Malaria Research. 2009 http://www.cdc.gov/malaria/cdcactivities/research.htm.

14. Raviprakash K, Defang G, Burgess T, Porter K: Advances in dengue vaccine development. Hum Vaccin 2009; 5(8).

15. Centers for Disease Control and Prevention: Dengue Fever. 2007 (April 21) http://www.cdc.gov/ncidod/dvbid/dengue/.

16. Gage KL, Dennis D, Tsai T: Prevention of plague: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR 1996; 45(RR-14): 1-15.

17. Special Programme for Research and Training in Tropical Diseases: World Health Organization: Report of the Scientific Working Group meeting on Dengue. Geneva: World Health Organization, 2006.

18. Farrar J, Focks D, Gubler D, Barrera R, Guzman MG, Simmons C, Kalayanarooj S, Lum L, McCall PJ, Lloyd L, Horstick O, Dayal-Drager R, Nathan MB, Kroeger A: Towards a global dengue research agenda. Trop Med Int Health 2007; 12(6): 695-9.

19. Centers for Disease Control and Prevention: Epidemic/epizootic West Nile virus in the United States: guidelines for surveillance, prevention, and control. 3rd revision. Fort Collins, CO: U.S. Department of Health and Human Services, Public Health Service, 2003.

20. United States Geological Survey and Centers for Disease Control and Prevention: ArboNET - Disease Maps 2009. Available at: http://diseasemaps.usgs.gov/.

21. Julian KG, Eidson M, Kipp AM, Weiss E, Petersen LR, Miller JR, Hinten SR, Marfin AA: Early season crow mortality as a sentinel for West Nile virus disease in humans, northeastern United States. Vector Borne Zoonotic Dis 2002; 2(3): 145-55.

22. Reisen WK, Barker CM, Carney R, Lothrop HD, Wheeler SS, Wilson JL, Madon MB, Takahashi R, Carroll B, Garcia S, Fang Y, Shafii M, Kahl N, Ashtari S, Kramer V, Glaser C, Jean C: Role of corvids in epidemiology of West Nile virus in southern California. J Med Entomol 2006; 43: 356-367.

23. Causey D, Trimble J, Hallwachs W, Brooks D, Janzen D: Migratory birds and the spread of West Nile. Science 2003; 299(5608): 821.

24. Hayes EB, Komar N, Nasci RS, Montgomery SP, O'Leary DR, Campbell GL: Epidemiology and transmission dynamics of West Nile virus disease. Emerg Infect Dis 2005; 11: 1167-1173.

25. Hayes EB, Gubler DJ: West Nile virus: epidemiology and clinical features of an emerging epidemic in the United States. Annu Rev Med 2006; 57: 181-94.

26. Tikhomirov E: Epidemiology and distribution of plague. Plague manual: epidemiology, distribution, surveillance and control. Geneva: World Health Organization, 1999; 11-37.

27. Centers for Disease Control and Prevention: Plague. 2009 http://www.cdc.gov/ncidod/dvbid/plague/.

28. Gage KL, Kosoy MY: Natural history of plague: Perspectives from more than a century of research. Ann Rev Entomol 2005; 50(1): 505-528.

29. Stenseth NC, Atshabar BB, Begon M, Belmain SR, Bertherat E, Carniel E, Gage KL, Leirs H, Rahalison L: Plague: Past, present, and future. PLoS Medicine 2008; 15: 1-5.

30. Poland J, Dennis D: Treatment of Plague. Plague Manual: Epidemiology, Distribution, Surveillance and Control. Geneva: WHO, 1999; 55-134.

31. Halstead SB: Dengue. Current Opinion in Infectious Diseases 2002; 15(5): 471-6.

32. Guzman MG, Kouri G: Dengue diagnosis, advances and challenges. Int J Infect Dis 2004; 8(2): 69-80.

33. Gubler DJ: The global emergence/resurgence of arboviral diseases as public health problems. Arch Med Res 2002; 33(4): 330-42.

34. Kyle JL, Harris E: Global spread and persistence of dengue. Ann Rev Micro 2008; 62: 71-92.

35. Suaya J, Shepard D, Beatty M: Dengue: burden of disease and costs of illness. Report on Dengue. Geneva: World Health Organization, 2006.

36. Gubler D, Clark G: Dengue/Dengue Hemorrhagic Fever: The Emergence of a Global Health Problem. Emerg Infect Dis 1995; 1(2): 55-57.

37. Pan American Health Organization: Dengue and dengue hemorrhagic fever in the Americas: guidelines for prevention and control. Washington DC: PAHO, 1994.

38. Gratz N, Knudsen A: Dengue/DHF. World Health Organization Dengue Bulletin 1997; 21.

39. Pinheiro FP, Corber SJ: Global situation of dengue and dengue haemorrhagic fever, and its emergence in the Americas. World Health Stat Q 1997; 50(3-4): 161-9.

40. World Health Organization: Dengue. 2009 (April 1) http://www.who.int/mediacentre/factsheets/fs117/en/.

41. Lorono-Pino MA, Blitvich BJ, Farfan-Ale JA, Puerto FI, Blanco JM, Marlenee NL, Rosado-Paredes EP, Garcia-Rejon JE, Gubler DJ, Calisher CH, Beaty BJ: Serologic evidence of West Nile virus infection in horses, Yucatan State, Mexico. Emerg Infect Dis 2003; 9(7): 857-859.

42. World Health Organization: Outbreak news, plague, Democratic Republic of Congo. Weekly epidemiological record 2006; 81(42): 397-408.

43. World Health Organization: Outbreak news, plague, Democratic Republic of Congo. Weekly epidemiological record 2006; 81(25): 241-248.

44. Duplantier JM, Duchemin JB, Chanteau S, Carniel E: From the recent lessons of the Malagasy foci towards a global understanding of the factors involved in plague reemergence. Vet Res 2005; 36(3): 437-53.

45. Schrag SJ, Wiener P: Emerging infectious disease - what are the relative roles of ecology and evolution. Trends Eco Evol 1995; 10(8): 319-324.

46. Ostfeld RS, Glass GE, Keesing F: Spatial epidemiology: an emerging (or re-emerging) discipline. Trends Ecol Evol 2005; 20(6): 328-36.

47. Kitron U: Risk maps: transmission and burden of vector-borne diseases. Parasitol Today 2000; 16(8): 324-5.

48. Bertherat E, Lamine KM, Formenty P, Thuier P, Mondonge V, Mitifu A, Rahalison L: Major pulmonary plague outbreak in a mining camp in the Democratic Republic of Congo: brutal awakening of an old scourge. Med Trop (Mars) 2005; 65(6): 511-4.

49. Diallo M, Ba Y, Sall AA, Diop OM, Ndione JA, Mondo M, Girault L, Mathiot C: Amplification of the sylvatic cycle of dengue virus type 2, Senegal, 1999-2000: entomologic findings and epidemiologic considerations. Emerg Infect Dis 2003; 9(3): 362-7.

50. Moncayo AC, Fernandez Z, Ortiz D, Diallo M, Sall A, Hartman S, Davis CT, Coffey L, Mathiot CC, Tesh RB, Weaver SC: Dengue emergence and adaptation to peridomestic mosquitoes. Emerg Infect Dis 2004; 10(10): 1790-6.

51. Rudnick A: Studies of ecology of dengue in Malaysia. Bull WHO 1966; 35(1): 78.

52. Rudnick A, Marchette NJ, Garcia R: Possible jungle dengue--recent studies and hypotheses. Jpn J Med Sci Biol 1967; 20 Suppl: 69-74.

53. Halstead SB: Dengue virus-mosquito interactions. Annu Rev Entomol 2008; 53: 273-91.

54. Reiter P, Lathrop S, Bunning M, Biggerstaff B, Singer D, Tiwari T, Baber L, Amador M, Thirion J, Hayes J, Seca C, Mendez J, Ramirez B, Robinson J, Rawlings J, Vorndam V, Waterman S, Gubler D, Clark G, Hayes E: Texas lifestyle limits transmission of dengue virus. Emerg Infect Dis 2003; 9(1): 86-9.

55. DeGroote JP, Sugumaran R, Brend SM, Tucker BJ, Bartholomay LC: Landscape, demographic, entomological, and climatic associations with human disease incidence of West Nile virus in the state of Iowa, USA. Int J Health Geogr 2008; 7: 19.

56. Miramontes R, Lafferty WE, Lind BK, Oberle MW: Is agricultural activity linked to the incidence of human West Nile virus? Am J Prev Med 2006; 30(2): 160-3.

57. Moore C, McLean R, Mitchell C, Nasci R, Tsai T, Calisher C, Marfin A, Moore P, Gubler D: Guidelines for Arbovirus Surveillance Programs in the United States, 1993.

58. Brown JA, Factor DL, Tkachenko N, Templeton SM, Crall ND, Pape WJ, Bauer MJ, Ambruso DR, Dickey WC, Marfin AA: West Nile viremic blood donors and risk factors for subsequent West Nile fever. Vector Borne Zoonotic Dis 2007; 7(4): 479-88.

59. Gujral IB, Zielinski-Gutierrez EC, LeBailly A, Nasci R: Behavioral risks for West Nile virus disease, northern Colorado, 2003. Emerg Infect Dis 2007; 13(3): 419-425.

60. Murphy TD, Grandpre J, Novick SL, Seys SA, Harris RW, Musgrave K: West Nile virus infection among health-fair participants, Wyoming 2003: assessment of symptoms and risk factors. Vector Borne Zoonotic Dis 2005; 5(3): 246-51.

61. Gates MC, Boston RC: Irrigation linked to a greater incidence of human and veterinary West Nile virus cases in the United States from 2004 to 2006. Prev Vet Med 2009.

62. Theophilides CN, Ahearn SC, Grady S, Merlino M: Identifying West Nile virus risk areas: The dynamic continuous-area space-time system. Am J Epidemiol 2003; 157(9): 843-854.

63. Nielsen CF, Armijos MV, Wheeler S, Carpenter TE, Boyce WM, Kelley K, Brown D, Scott TW, Reisen WK: Risk factors associated with human infection during the 2006 West Nile virus outbreak in Davis, a residential community in northern California. Am J Trop Med Hyg 2008; 78(1): 53-62.

64. Watson JT, Jones RC, Gibbs K, Paul W: Dead crow reports and location of human West Nile virus cases, Chicago, 2002. Emerg Infect Dis 2004; 10(5): 938-940.

65. Brownstein JS, Rosen H, Purdy D, Miller JR, Merlino M, Mostashari F, Fish D: Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. Vector-Borne Zoonotic Dis. 2002; 2(3): 157-164.

66. Ruiz M, Tedesco C, McTighe T, Austin C, Kitron U: Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. Int J Health Geogr 2004; 3(1): 8.

67. Reisen WK, Fang Y, Martinez VM: Effects of temperature on the transmission of West Nile virus by *Culex tarsalis* (Diptera: Culicidae). J Med Entomol 2006; 43: 309-317.

68. Ruiz MO, Walker ED, Foster ES, Haramis LD, Kitron UD: Association of West Nile virus illness and urban landscapes in Chicago and Detroit. Int J Health Geogr 2007; 6: 10.

69. Orochi-Orach: Plague outbreaks the gender and age perspective in Okoro County, Nebbi district, Uganda. Agency for Accelerated Regional Development 2002.

70. Eisen RJ, Enscore RE, Biggerstaff BJ, Reynolds PJ, Ettestad P, Brown T, Pape J, Tanda D, Levy CE, Engelthaler DM, Cheek J, Bueno R, Targhetta J, Montenieri JA, Gage KL: Human plague in the southwestern United States, 1957-2004: Spatial models of elevated risk of human exposure to *Yersinia pestis*. J Med Entomol 2007; 44(3): 530-537.

71. Mann JM, Martone WJ, Boyce JM, Kaufmann AF, Barnes AM, Weber NS: Endemic human plague in New Mexico: Risk factors associated with infection. J Infect Dis 1979; 140: 397-401.

72. Collinge SK, Johnson WC, Ray C, Matchett R, Grensten J, Cully JF, Gage KL, Kosoy MY, Loye JE, Martin AP: Landscape structure and plague occurrence in black-tailed prairie dogs on grasslands of the western USA. Landscape Eco 2005; 20(8): 941-955.

73. Eisen RJ, Gage KL: Adaptive strategies of *Yersinia pestis* to persist during inter-epizootic and epizootic periods. Vet Res 2009; 40(2): 1.

74. Winters A, Staples J, Ogen-Odoi A, Mead P, Griffith K, Owor N, Babi N, Enscore R, Eisen L, Gage K, Eisen R: Spatial risk models for human plague in the West Nile region of Uganda. Am J Trop Med Hyg 2009; 80(6): 1014-1022.

75. Davis S, Leirs H, Viljugrein H, Stenseth NC, De Bruyn L, Klassovskiy N, Ageyev V, Begon M: Empirical assessment of a threshold model for sylvatic plague. J R Soc Interface 2007; 4(15): 649-657.

76. Keeling MJ, Gilligan CA: Metapopulation dynamics of bubonic plague. Nature 2000; 407(6806): 903-6.

77. Snall T, O'Hara RB, Ray C, Collinge SK: Climate-driven spatial dynamics of plague among prairie dog colonies. Am Nat 2008; 171(2): 238-48.

78. Gage KL, Burkot TR, Eisen RJ, Hayes EB: Climate and vector-borne diseases. Am J Prev Med 2008; 35(5): 436-50.

79. Dohm DJ, O'Guinn ML, Turell MJ: Effect of environmental temperature on the ability of *Culex pipiens* (Diptera: Culicidae) to transmit West Nile virus. J Med Entomol 2002; 39(1): 221-5.

80. Higgs S, Beaty B: Natural Cycles of Vector-borne Pathogens. *In*: Marquardt WC (ed.) *Biology of Disease Vectors*. 2nd ed. Burlington: Elsevier Academic Press, 2005; 167-185.

81. Eldridge BE: Mosquitoes, the Culicidae. *In*: Marquardt WC (ed.) *Biology of Disease Vectors*. 2nd ed. Burlington: Elsevier Academic Press, 2005; 95-112.

82. Pant CP, Yasuno M: Field studies on gonotropic cycle of *Aedes aegypti* in Bangkok, Thailand. J Med Entomol 1973; 10(2): 219-223.

83. Wegbreit J, Reisen WK: Relationships among weather, mosquito abundance, and encephalitis virus activity in California: Kern County 1990-98. J Am Mosq Control Assoc 2000; 16(1): 22-27.

84. Reisen WK, Milby MM, Meyer RP: Population dynamics of adult *Culex* mosquitoes (Diptera: Culicidae) along the Kern River, Kern County, California, in 1990. J Med Entomol 1992; 29(3): 531-43.

85. Enscore RE, Biggerstaff BJ, Brown TL, Fulgham RE, Reynolds PJ, Engelthaler DM, Levy CE, Parmenter RR, Montenieri JA, Cheek JE, Grinnell RK, Ettestad PJ, Gage KL: Modeling relationships between climate and the frequency of human plague cases in the southwestern United States, 1960-1997. Am J Trop Med Hyg 2002; 66(2): 186-96.

86. Parmenter RR, Yadav EP, Parmenter CA, Ettestad P, Gage KL: Incidence of plague associated with increased winter-spring precipitation in New Mexico. Am J Trop Med Hyg 1999; 61(5): 814-21.

87. Kitron U: Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. J Med Entomol 1998; 35(4): 435-45.

88. Turner M: Landscape ecology: the effect of pattern on process. Annu. Rev. Ecol. Syst. 1989; 20: 171-197.

89. Turner M, Gardner R: Quantitative Methods in Landscape Ecology. New York: Springer-Verlag, 1990.

90. Pavlovsky E: Natural Nidality of Transmissible Diseases. Urbana: University of Illinois Press, 1966.

91. Ford J: The role of the trypanosomiasis in African ecology Oxford: Clarendon, 1971.

92. Hoogstraal H: The epidemiology of tick-borne Crimean-Congo hemorrhagic fever in Asia, Europe, and Africa. J Med Entomol 1979; 15: 307-417.

93. Eisen RJ, Eisen L: Spatial modeling of human risk of exposure to vector-borne pathogens based on epidemiological versus arthropod vector data. J Med Entomol 2008; 45(2): 181-92.

94. Haemig PD, Waldenstrom J, Olsen B: Roadside ecology and epidemiology of tick-borne diseases. Scand J Infect Dis 2008; 40(11-12): 853-8.

95. Maupin GO, Fish D, Zultowsky J, Campos EG, Piesman J: Landscape ecology of Lyme disease in a residential area of Westchester County, New York. Am J Epidemiol 1991; 133(11): 1105-13.

96. Killilea ME, Swei A, Lane RS, Briggs CJ, Ostfeld RS: Spatial dynamics of Lyme disease: a review. Ecohealth 2008; 5(2): 167-95.

97. Kalluri S, Gilruth P, Rogers DJ, Szczur M: Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: a review. PLoS Pathog 2007; 26(3): 1361-1371.

98. Rogers DJ: Models for vectors and vector-borne disease. Adv Parasitol 2006; 62: 1-35.

99. Rogers DJ, Randolph SE: Studying the global distribution of infectious diseases using GIS and RS. Nat Rev Microbiol 2003; 1(3): 231-7.

100. Beck LR, Lobitz BM, Wood BL: Remote sensing and human health: New sensors and new opportunities. Emerg Infect Dis 2000; 6(3): 217-227.

101. Dale P, Ritchie SA, Territo BM, Morris CD, Muhar A, Kay BH: An overview of remote sensing and GIS for surveillance of mosquito vector habitats and risk assessment. J Vector Ecol 1998; 23(1): 54-61.

102. Moncayo AC, Edman JD, Finn JT: Application of geographic information technology in determining risk of eastern equine encephalomyelitis virus transmission. J Am Mosq Control Assoc 2000; 16(1): 28-35.

103. Lothrop HD, Reisen WK: Landscape affects the host-seeking patterns of *Culex tarsalis* (Diptera: Culicidae) in the Coachella Valley of California. J Med Entomol 2001; 38(2): 325-332.

104. Reisen WK, Lothrop HD, Presser SB, Hardy JL, Gordon EW: Landscape ecology of arboviruses in southeastern California - temporal and spatial patterns of enzootic activity in Imperial Valley, 1991-1994. J Med Entomol 1997; 34(2): 179-188.

105. Wood BL, Beck LR, Washino RK, Hibbard KA, Salute JS: Estimating high mosquito-producing rice fields using spectral and spatial data. Int. J. Remote Sensing 1992; 13: 2813-2826.

106. Diuk-Wasser MA, Bagayoko M, Sogoba N, Dolo G, Toure MB, Traore SF, Taylor CE: Mapping rice field anopheline breeding habitats in Mali, West Africa, using Landsat ETM+ sensor data. Int J Remote Sense 2004; 25: 359-376.

107. Kleinschmidt I, Omumbo J, Briet O, van de Giesen N, Sogoba N, Mensah NK, Windmeijer P, Moussa M, Teuscher T: An empirical malaria distribution map for West Africa. Trop Med Int Health 2001; 6(10): 779-786.

108. Rogers DJ, Myers M, Tucker CJ, Smith P, White DJ: Predicting the distribution of West Nile fever in North America using satellite sensor data. Photogramm Eng Rem S 2002; 68: 181-220.

109. Hay SI, Omumbo J, Kraig M, Snow R: Earth observation geographic information systems and Plasmodium falciparum malaria in Sub-Saharan Africa. Adv Parasitol 2000; 47(174-206).

110. Endy TP, Nisalak A, Chunsuttiwat S, Libraty DH, Green S, Rothman AL, Vaughn DW, Ennis FA: Spatial and temporal circulation of dengue virus serotypes: a prospective study of primary school children in Kamphaeng Phet, Thailand. Am J Epidemiol 2002; 156(1): 52-9.

111. Getis A, Morrison AC, Gray K, Scott TW: Characteristics of the spatial pattern of the dengue vector, *Aedes aegypti*, in Iquitos, Peru. Am J Trop Med Hyg 2003; 69(5): 494-505.

112. Morrison AC, Getis A, Santiago M, Rigau-Perez JG, Reiter P: Exploratory space-time analysis of reported dengue cases during an outbreak in Florida, Puerto Rico, 1991-1992. Am J Trop Med Hyg 1998; 58(3): 287-298.

113. Rotela C, Fouque F, Lamfri M, Sabatier P, Introini V, Zaidenberg M, Scavuzzo C: Space-time analysis of the dengue spreading dynamics in the 2004 Tartagal outbreak, Northern Argentina. Acta Tropica 2007; 103(1): 1-13.

114. Rogers DJ, WIlson A, Hay SI, Graham K: The global distribution of yellow fever and dengue. Adv Parasitol 2006; 62: 181-220.

115. Barbour A, Fish D: The biological and social phenomenon of Lyme disease. Science 1993; 260: 1610-1616.

116. Eisen RJ, Lane RS, Fritz CL, Eisen L: Spatial patterns of Lyme disease risk in California based on disease incidence data and modeling of vector-tick exposure. Am J Trop Med Hyg 2006; 75(4): 669-676.

117. Eisen RJ, Glass GE, Eisen L, Cheek J, Enscore RE, Ettestad P, Gage KL: A spatial model of shared risk for plague and hantavirus pulmonary syndrome in the southwestern United States. Am J Trop Med Hyg 2007; 77(6): 999-1004.

118. Neerinckx SB, Peterson AT, Gulinck H, Deckers J, Leirs H: Geographic distribution and ecological niche of plague in sub-Saharan Africa. Int J Health Geogr 2008; 7: 54.

119. Kabasele K: When covering up is a good thing. Family physicians' role in educating the public about West Nile virus. Can Fam Physician 2005; 51: 797-9, 802-4.

120. Aquino M, Fyfe M, MacDougall L, Remple V: Protective behavior survey, West Nile virus, British Columbia. Emerg Infect Dis 2004; 10(8): 1499-501.

121. Kroeger A, Lenhart A, Ochoa M, Villegas E, Levy M, Alexander N, McCall PJ: Effective control of dengue vectors with curtains and water container covers treated with insecticide in Mexico and Venezuela: cluster randomised trials. BMJ 2006; 332(7552): 1247-52.

122. Focks D: A review of entomological sampling methods and indicators for dengue vectors. Geneva: WHO, 2003.

123. Manrique-Saide P, Davies CR, Coleman PG, Rebollar-Tellez E, Che-Medoza A, Dzul-Manzanilla F, Zapata-Peniche A: Pupal surveys for *Aedes aegypti* surveillance and potential targeted control in residential areas of Merida, Mexico. J Am Mosq Control Assoc 2008; 24(2): 289-98.

124. Abeku TA, Hay SI, Ochola S, Langi P, Beard B, de Vlas SJ, Cox J: Malaria epidemic early warning and detection in African highlands. Trends Parasitol 2004; 20(9): 400-5.

125. Kuhn K, Campbell-Lendrum DH, Haines A, Cox J: Using climate to predict infectious disease epidemics: World Health Organization, 2005.

126. Moore M, Chan E, Lurie N, Schaefer A, Varda D, Zambrano J: Strategies to improve global influenza surveillance: a decision tool for policymakers. BMC Public Health 2008; 28(8).

127. Yang W, Xing H, Wang H, Lan Y, Sun Q, Hu S, Lu W, Yuan Z, Chen Y, Dong B: A study on early detection for seven infectious diseases. Zhonghua Liu Xing Bing Xue Za Zhi 2004; 25(12): 1039-1041.

128. Lewis R, Nathan N, Diarra L, Belanger F, Paquet C: Timely detection of meningococcal meningitis epidemics in Africa. Lancet 2001; 358(9278): 287-293.

129. MMWR: Summary of notifiable diseases -- United States, 2006. MMWR 2008; 55(53): 1-100.

130. Thacker S, Choi K, Brachman P: The surveillance of infectious diseases. JAMA 1983; 249: 1191-1185.

131. Eisen L, Lozano-Fuentes S: Use of mapping and spatial and space-time modeling approaches in operational control of *Aedes aegypti* and dengue. PLoS Negl Trop Dis 2009; 3(4): e411.

132. Mail GA, McHugh RA: Relation of temperature and humidity to winter survival of *Culex pipiens* and *Culex tarsalis*. Mosquito News 1961; 21: 252-254.

133. Reeves WC, Hardy JL, Reisen WK, Milby MM: Potential effect of global warming on mosquito-borne arboviruses. J Med Entomol 1994; 31(3): 323-332.

134. Reisen WK, Hardy JL, Presser SB, Chiles RE: Seasonal variation in the vector competence of *Culex tarsalis* (Diptera: Culicidae) from the Coachella Valley of California for Western equine encephalomyelitis and St. Louis encephalitis viruses. J Med Entomol 1996; 33(3): 433-437.

135. Singh N, Sharma VP: Patterns of rainfall and malaria in Madhya Pradesh, central India. Ann Trop Med Parasitol 2002; 96(4): 349-59.

136. Thammapalo S, Chongsuwiwatwong V, McNeil D, Geater A: The climatic factors influencing the occurrence of dengue hemorrhagic fever in Thailand. Southeast Asian J Trop Med Public Health 2005; 36(1): 191-6.

137. Wu PC, Guo HR, Lung SC, Lin CY, Su HJ: Weather as an effective predictor for occurrence of dengue fever in Taiwan. Acta Trop 2007; 103(1): 50-7.

138. Eisen L, Eisen RJ: Need for improved methods to collect and present spatial epidemiologic data for vector borne diseases. Emerg Infect Dis 2007; 13(12): 1816-20.

139. Sanchez L, Vanlerberghe V, Alfonso L, Marquetti Mdel C, Guzman MG, Bisset J, van der Stuyft P: *Aedes aegypti* larval indices and risk for dengue epidemics. Emerg Infect Dis 2006; 12(5): 800-6.

140. Chadee DD, Williams FL, Kitron UD: Impact of vector control on a dengue fever outbreak in Trinidad, West Indies, in 1998. Trop Med Int Health 2005; 10(8): 748-54.

141. Sulaiman S, Pawanchee ZA, Arifin Z, Wahab A: Relationship between Breteau and House indices and cases of dengue/dengue hemorrhagic fever in Kuala Lumpur, Malaysia. J Am Mosq Control Assoc 1996; 12(3 Pt 1): 494-6.

142. Barrera R, Delgado N, Jimenez M, Valero S: Eco-epidemiological factors associated with hyperendemic dengue haemorrhagic fever in Maracay City, Venezuela. Dengue Bull 2002; 26: 84-95.

143. Beroll H, Berke O, Wilson J, Barker IK: Investigating the spatial risk distribution of West Nile virus disease in birds and humans in southern Ontario from 2002 to 2005. Popul Health Metr 2007; 5: 3.

144. Cooke WH, Grala K, Wallis RC: Avian GIS models signal human risk for West Nile virus in Mississippi. Int J Health Geogr 2006; 5: 36.

145. Roberts RS, Foppa IM: Prediction of equine risk of West Nile virus infection based on dead bird surveillance. Vector Borne Zoonotic Dis 2006; 6(1): 1-6.

146. Gage KL: Plague Surveillance. Plague Manual: Epidemiology, Distribution, Surveillance and Control. Geneva: WHO, 1999; 135-165.

II. Determination and Presentation of Spatial Patterns of Vector-borne Disease Occurrence Based on Epidemiologic Data: A Case Study for West Nile Virus Disease in Colorado

Anna M. Winters[1,2], Mark J. Delorey[1], Rebecca J. Eisen[1], Chester G. Moore[2], Roger S. Nasci[1], Emily Zielinski-Gutierrez[1], W. John Pape[3], and Lars Eisen[2]*

[1]Division of Vector-Borne Infectious Diseases, National Center for Vector-Borne, Zoonotic and Enteric Diseases, Centers for Disease Control and Prevention, Fort Collins, Colorado 80522; [2]Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado 80523; [3]Communicable Disease Program, Colorado Department of Public Health and Environment, Denver, Colorado 80246.

Abstract

We used epidemiologic data for West Nile virus (WNV) disease in Colorado from 2003 and 2007 to determine: (1) the degree to which estimates of vector-borne disease occurrence is influenced by spatial scale (i.e., county vs. census tract); and (2) the extent of concordance between spatial risk patterns based on case counts versus incidence. Analyses revealed that county-scale presentation accounted for only ~50% of the overall variance in incidence compared to census tract level presentation suggesting that sub-county scales provide beneficial information helpful for stakeholder communities. There was high concordance between spatial patterns of WNV disease incidence and case counts for census tract (83%) but not for county (50%) or zip code (31%). We discuss the relevance of these findings for practices to develop spatial epidemiologic data for vector-borne diseases and to present these data to stakeholder communities.

KEY WORDS: Information delivery, mapping, spatial scale, vector-borne disease, West Nile virus disease

Introduction

In the last two decades, technological capacity to map and model spatial patterns of risk for exposure to arthropod vectors and vector-borne pathogens has progressed rapidly.[1-4] Geographical Information System (GIS) and Remote Sensing (RS) software have become more user-friendly and are now complemented by easy-to-use tools to assess spatial and space-time clustering, such as SaTScan™ and the DYCAST system.[5, 6] New mapping software, such as Google Earth™ and MS Virtual Earth™,[7, 8] provides basic and easy-to-use capacity to generate not only spatial data overlaid on pre-existing satellite imagery or map representations but also dynamic illustrations of space-time patterns that can be played as "movie clips".[9-11] These developments provide novel capacity to determine and present spatial patterns of disease incidence, or of occurrence of vectors and vector-borne pathogens. Indeed, GIS and other mapping technologies are now routinely used both in academic institutions and by public health agencies at national, state, county and city levels in the U.S.. This is accompanied by explosive development in the field of web-based information delivery which now provides an effective medium to distribute maps to a wide range of stakeholders including the medical community, vector control practitioners, policy makers and the public at large.[12, 13] Using West Nile virus (WNV) disease as an example, maps showing spatial distributions of WNV disease cases or WNV disease incidence are readily available from the Centers for Disease Control and Prevention website (http://www.cdc.gov/ncidod/dvbid/westnile/index.htm), the U.S. Geological Survey Disease Maps web site (http://diseasemaps.usgs.gov/index.html), and from many state or

local health department websites in WNV disease endemic areas. Such maps can be used as tools to target limited prevention, surveillance and control resources to high-risk areas for WNV exposure, and to inform the public about local risk levels.

However, with this new technological capacity to determine and present spatial risk patterns comes a series of questions regarding how it should be used constructively in public health. Benefits and drawbacks of using entomological versus epidemiological data in spatial risk assessments have been discussed previously for important vector-borne diseases in North America.[9, 14] Here, we focus on epidemiologic data and use WNV disease in Colorado during the outbreak years of 2003 (total of 2,947 WNV disease cases reported from the state) and 2007 (578 reported WNV disease cases) as a case study, and apply quantitative statistical methods to explore two questions which have not previously been addressed for vector-borne diseases in the U.S.: (1) How are estimates of vector-borne disease occurrence influenced by spatial scale (i.e., county vs. census tract?); and (2) What is the extent of concordance among spatial risk patterns based on disease case counts versus disease incidence for commonly used spatial boundary units such as county, census tract and zip code?

Materials and Methods

Epidemiologic and GIS-based data.

The study was based on WNV disease cases reported to the Colorado Department of Public Health and Environment (CDPHE) during 2003 and 2007 which represent two outbreak years in Colorado with 2,947 WNV disease cases reported in 2003 and 578 cases reported in 2007. The ratios of WNV fever to WNV neuroinvasive disease were

~3.9:1 in 2003 and ~ 5.0:1 in 2007 suggesting that detection and reporting of cases was relatively consistent during these years. Our spatial analyses included all reported WNV disease cases, including WNV fever and WNV neuroinvasive disease, because our main interest was in risk of exposure to the WNV rather than disease manifestation. The use of all WNV disease cases, rather than only WNV neuroinvasive disease cases, was further justified by that analysis of data from 2003 and 2007 showed significant correlations between numbers of WNV fever and WNV neuroinvasive disease cases for both the county scale (excluding counties reporting no WNV disease cases; 2003, Spearman's rank correlation: $\rho_s = 0.851$, n = 47, $P < 0.001$; 2007,: $\rho_s = 0.684$, n = 34, $P < 0.001$) and the census tract scale (excluding census tracts reporting no WNV disease cases; 2003: $\rho_s = 0.128$, n = 737, $P < 0.001$; 2007: $\rho_s = 0.161$, n = 312, $P < 0.005$). Data for "true" infection rates for WNV in Colorado are not available because many infections are inapparent or cause only mild symptoms that are unlikely to result in visits to physicians, laboratory confirmation of WNV exposure and case reporting. Furthermore, inclusion of WNV fever cases provided a more suitable sample size for development of a robust spatial model.

The epidemiologic database provided by CDPHE included, for each case, information for county, zip code and census tract of residence, and date of onset of symptoms. No personal identifiers were included in the database. The epidemiologic database was complemented with GIS-derived data for geographical boundaries (county, zip code and census tract; Environmental Systems Research Institute [ESRI], Redlands, CA) and 2004 human population (data from the U.S. Census Bureau provided by ESRI).

WNV disease cases were aggregated to county, zip code and census tract units and cumulative disease incidence (hereinafter referred to as incidence per 100,000 person-years) was calculated for 2003 and 2007 combined. Combining cases from 2003 and 2007 was justified as the spatial patterns of WNV disease incidence in Colorado were similar in these outbreak years for the county scale (Spearman rank correlation; ρ_s = 0.737, n = 64, $P < 0.001$), the census tract scale ($\rho_s = 0.434$, n = 1,075, $P < 0.001$) and the zip code scale ($\rho_s = 0.459$, n = 443, $P < 0.001$). Aggregating cases to county, zip code and census tract units was based on the assumption that the likely WNV exposure site was located within the specific boundary unit where the residence was located. This undoubtedly introduces some degree of error due to occupational, recreational or travel exposures where individuals are exposed outside of their resident zip code, census tract or county. However, in the absence of reliable information for specific exposure sites, use of residence as the assumed exposure location was the best available solution.

Spatial model to partition variance of WNV disease incidence at county and census tract scales.

WNV disease incidence data for the census tract and county spatial boundary units in Colorado were used to explore the degree to which representations of spatial variability in risk are influenced when data are aggregated to county versus census tract. Zip codes were not used in this analysis because, unlike census tracts, they do not always nest within counties. To identify how variance of WNV disease incidence was partitioned across counties and within counties, a generalized linear mixed-effects (GLME) model was fitted to the data with the response variable being cumulative incidence per 100,000

person-years (2003 and 2007 combined) reported to the specific spatial unit (census tract or county). The population of each spatial unit was assumed to be fixed and the model assumed a binomial distribution for the responses. The GLME model specification is

$$\eta_{ij} = \beta_0 + u_i + v_{ij}$$

where η_{ij} is the linear predictor for the jth census tract in county i, β_0 is the intercept, u_i and v_{ij} are random effects for county i and census tract j in county i. Case counts were modeled according to the logistic model[16] and u_i and v_{ij} were distributed as a conditionally autoregressive (CAR) model[17, 18] as exploratory data analysis of the residuals revealed spatial dependence among the data (based on Moran's I statistic and neighbor and distance weights matrices). The model was fitted using only the "distance" spatial weights matrix (assuming that the strength of the correlation between spatial units is inversely proportional to the distances between their centroids) because previous studies[19] indicated that unintended correlation results can occur when a "neighbor" spatial weights matrix (equal correlation among adjacent spatial units) is used in a CAR model.

A hot-spot analysis based on the Getis-Ord Gi* statistic (ArcGIS 9.2, Spatial Analyst, ESRI, Redlands, CA) was conducted to determine presence of local clustering of census tracts with either high or low WNV disease incidence based on Z-score values.[20] A high and positive Z score value indicates that a census tract is surrounded by other census tracts reporting high WNV disease incidence ("hot-spot"). A high but negative Z-score value indicates that a census tract is surrounded by census tracts reporting low WNV disease incidence ("cool-spot"). Examples of "hot-spots" within counties reporting

overall low incidence and "cool-spots" within counties reporting overall high WNV disease incidence were displayed in map format.

Disease case count versus incidence.

Zip code, census tract and county were used to determine the extent of concordance for spatial patterns of areas characterized by high risk of exposure to WNV based on WNV disease case counts versus WNV disease incidence. As before, this analysis was conducted using combined cumulative WNV disease data for 2003 and 2007. For each spatial unit (zip code, census tract, county) and disease risk estimate (case count, case incidence) we systematically categorized risk by quartiles. WNV disease case counts and WNV disease incidences falling within the 4^{th} quartile were considered "high risk" and used to determine the degree of concordance for spatial patterns of high-risk areas for case count versus case incidence for each of the three spatial boundary units examined. This was achieved by contingency table analysis. In addition, we determined the overall degree of correlation between case counts and disease incidence for the three spatial scales using Spearman's rank correlation.

Statistical analyses and map development.

Statistical analyses were conducted using the S-PLUS® v. 8.0 (TIBCO Software Inc., Palo Alto, CA) and JMP® 7.0.1 (SAS Institute Inc., Cary, NC) statistical packages. Maps for WNV disease cases and incidence were developed using ArcGIS 9.2.

Results

Basic description of spatial patterns for WNV disease case counts and WNV disease incidence.

During 2003 and 2007, a total of 3,525 human WNV disease cases were reported to the Colorado Department of Public Health and Environment (CDPHE). Larimer, Boulder and Weld counties in north-central Colorado accounted for the largest numbers of WNV disease cases during these two years with totals of 650, 537 and 507, respectively. In contrast, the highest WNV disease incidences occurred in the northeastern part of the state, with Logan, Sedgwick and Phillips counties reporting 293, 273 and 228 cases, respectively, per 100,000 person-years compared to 92--116 cases per 100,000 person-years for Boulder, Larimer, and Weld counties (Fig. 1). In the western, mountainous part of Colorado, Mesa and Delta counties reported the highest incidence rates (16--27 cases per 100,000 person-years) (Fig. 1). Spatial patterns for WNV disease case counts and WNV disease incidence are displayed visually by county, census tract and zip code in Fig. 2.

Spatial model to partition variance of WNV disease incidence at county and census tract scales.

Significant spatial autocorrelation was detected from the residuals of the generalized linear mixed-effects model for both county (Neighbor weights matrix, Moran's $I = 0.63$, $P < 0.01$; Distance weights matrix, Moran's $I = 0.48$, $P < 0.01$) and census tract (Neighbor weights matrix, Moran's $I = 0.22$, $P < 0.01$; Distance weights matrix, Moran's $I = 0.08$, $P = 0.03$). To model the observed spatial dependence, a conditional autoregressive model was fit to the data.[17] Positive spatial correlation among

census tracts occurred for multiple counties in north-central Colorado (northern Front Range area), two counties in the southeastern part of the state, and for two counties to the southwest (Fig. 3). Negative spatial correlation among census tracts occurred in counties dispersed throughout the state (Fig. 3). The scale factors, σ_u^2 and σ_v^2 had a ratio of:

$$\frac{\ddot{\Phi}_u^2}{\ddot{\Phi}_v^2} \approx \frac{4.5}{3.2} \approx 1.41$$

with a 95% confidence interval of 0.91 -- 2.26, indicating no statistical difference from a ratio of 1 and suggesting that the spatial model partitions the total variability between counties and census tracts similarly. That is, after accounting for spatial dependence, the results indicate that: (1) variability in WNV disease incidence within counties is approximately the same as the variability between counties; and (2) county-scale presentations of spatial WNV disease incidence patterns account for only ~ 50% of the variance in WNV disease incidence that can be shown by presenting census tract level data.

The Getis-Ord Gi* statistic identified "hot-spot" census tracts that were surrounded by other census tracts reporting high WNV disease incidence and "cool-spot" census tracts surrounded by other low WNV disease incidence census tracts. We found numerous instances where "hot-spots" or "cool-spots" within counties are obscured when WNV disease incidence is displayed at the county scale. Denver County provided an example of census tract "hot-spots" occurring within a county reporting overall low incidence (17.09 cases per 100,000 person-years). As shown in Fig. 4a, eight statistically significant ($P < 0.05$) census tract "hot-spots" were identified within this county. Conversely, "cool-spots" occurred in several counties in north-central Colorado that reported high overall WNV disease incidence (Larimer, Weld, and Morgan counties;

range = 109.03 – 134.48 cases per 100,000 person-years). Within these counties, 14 census tracts were statistically significant "cool-spots" (Fig. 4b). These examples illustrate the finding that aggregating WNV disease incidence to the county scale can obscure "hot-spots" or "cool-spots" discernable at the finer census tract scale.

Concordance between spatial patterns based on disease case count versus incidence.

Spatial patterns for WNV disease case counts and WNV disease incidence are shown by county, census tract and zip code in Fig. 2. Positive correlations between case counts and disease incidence occurred for all three spatial units but was much stronger for census tract ($\rho_s = 0.877$, n = 1,075, $P < 0.001$) than for zip code ($\rho_s = 0.238$, n = 443, $P < 0.001$) or county ($\rho_s = 0.558$, n = 64, $P < 0.001$). A similar pattern among the spatial units was detected for high-risk areas falling in the 4th quartile for each disease measure. High-risk counties based on WNV disease case counts were distributed throughout the Front Range in the central part of the state, whereas high-risk counties based on WNV disease incidence more commonly were located in far eastern Colorado (Fig. 2). Of the 12 counties classified as high risk based on WNV disease incidence, 6 were also classified as high risk based on WNV disease case counts (50% concordance between spatial patterns for high-risk counties based on WNV disease case count versus WNV disease incidence; Table 1).

High-risk zip codes based on WNV disease case counts occurred in three distinct clusters in the north-central, northeastern and south-central parts of Colorado, whereas high-risk zip codes based on WNV disease incidence were shifted to the far eastern parts of the state (Fig. 2). Of the 74 zip codes classified as high risk based on WNV disease

incidence, 23 were also classified as high risk based on WNV disease case counts (31% concordance; Table 1). In contrast, we found far higher concordance (83%) between spatial patterns for high risk based on WNV disease case count versus WNV disease incidence for the census tract scale (Table 1). High-risk census tracts were ubiquitous in northeastern Colorado, occurred commonly to the southeast, and were found only sporadically in the western, mountainous part of the state (Fig. 2).

Discussion

We used West Nile virus (WNV) disease in Colorado as a case study to quantitatively examine: (1) the degree to which estimates of vector-borne disease incidence is influenced by spatial scale (i.e., county vs. census tract); and (2) the extent of concordance between spatial risk patterns based on disease case counts versus disease incidence for commonly used spatial boundary units. The analyses revealed that variability in WNV disease incidence within counties is approximately the same as the variability between counties, and that county-scale presentations of spatial WNV disease incidence patterns therefore account for only ~50% of the variance in WNV disease incidence that can be shown by presenting census tract level data. Use of the county scale also was found to mask "hot-spots" evident at finer scale (census tract or zip code) in counties with low overall WNV disease incidence. Further, there was high concordance between spatial patterns of areas with high risk for exposure to WNV based on WNV disease incidence and WNV disease case counts for the census tract scale but not for the county or zip code scales. Below we discuss the relevance of these findings for practices

to develop spatial epidemiologic data for vector-borne diseases in the U.S. and to present these data to stakeholder communities.

There are stakeholder communities with an interest in spatial patterns of risk for contracting diseases caused by vector-borne pathogens. In the specific case of WNV disease, stakeholders include federal, state and local public health agencies, mosquito control programs, health care providers, purveyors of disease prevention products, and the general public. These stakeholders have needs for spatial information that differ not only in terms of scale but also in type of information. For example, a mosquito control program aiming to implement control activities to suppress vector mosquitoes and reduce the burden of WNV disease likely will be most interested in finding out where high numbers of WNV disease cases occur at sub-county scales in order to focus expensive prevention efforts. On the other hand, a member of the public seeking information to help determine his/her personal risk of exposure to WNV, and the need for use of personal protective measures such as repellents, may be more interested in spatial risk estimates based on WNV disease incidence (which account for population size) in the area of interest. The challenge presented to public health map-makers is to present stakeholders with a package of suitable and easy-to-understand information for spatial risk patterns in electronic map formats while at the same time protecting patient privacy and carefully considering benefits and drawbacks to determination and presentation of risk assessments at different spatial scales.[21]

Basic options to present information for spatial risk of vector-borne diseases in map formats include point locations for disease cases or aggregation of disease case counts or disease incidence to administrative boundary units (summarized in Table 2). A map

showing individual case point locations is obviously the most precise way to present spatial disease data. However, this has distinct disadvantages including: (1) the possibility that the address of residence is not the site of pathogen exposure; (2) a lack of accounting for population size; and (3) in some countries, including the U.S., strict regulations to guide the use of patient health information.[22-24] The latter issue can be addressed by random offsets from the actual location of the patient's residence but this essentially means that an inaccurate map is presented.

A more commonly used approach to avoid privacy issues is to aggregate disease case counts or disease incidence to administrative boundaries. This in turn raises the issue of the modifiable areal unit problem,[25] which occurs when numerical results vary when the same set of data is grouped at different levels of spatial resolution, and raises the question of which boundary unit best captures the variability of spatial vector-borne disease data without compromising data quality.[21] In the U.S., the Centers for Disease Control and Prevention and nearly all individual state health agencies provide spatial WNV disease information to the public at the county scale. The lone exception is the Colorado Department of Public Health and Environment which, in addition to county-based information, also provides maps for WNV disease incidence by census tract.

Benefits and drawbacks of county versus sub-county scales.

Although our results provide a compelling argument for display of risk patterns for exposure to vector-borne pathogens at sub-county scales, there are several problems that need to be considered before sub-county information is presented to end-users. There is no question that sub-county variability exists for risk of exposure to mosquito and tick

vectors of human pathogens such as WNV and the Lyme disease spirochete, *Borrelia burgdorferi*, in the U.S..[26-30] The basic problem when working with sub-county spatial risk patterns developed based on epidemiologic data is to determine which of the resulting patterns are real and which are likely to be analysis artifacts. Such artifacts may occur for several reasons including: (1) case files for common vector-borne diseases, such as WNV disease and Lyme disease, often lack information for likely site of vector and pathogen exposure and thus the address of residence, in actuality, is not the exposure location; (2) information that a case has occurred may result in other nearby cases being detected through increased risk perception and health care seeking; and (3) lack of access to health care among lower income zip codes or census tracts may prevent reporting and thus mask the presence of disease in those areas. This is especially problematic in terms of milder disease manifestations for which detection and treatment is "optional" as is the case with WNV fever. These problems, of course, occur also at the county scale but can be assumed to have greater impact at sub-county scales. One way to evaluate the accuracy of sub-county scale risk patterns that are based on epidemiologic data is to develop complementary spatial models based on entomological risk measures such as abundance of vectors or pathogen-infected vectors and compare the spatial patterns based on epidemiologic versus entomological data.[14, 30, 31] Concordance between epidemiologic and entomological risk measures can validate sub-county scale risk patterns, whereas discordance indicates the need for additional investigations. For example, ground-based entomological surveillance in areas with high projected epidemiologic risk but low projected entomological risk can be used to assess whether the observed epidemiologic pattern represents "real" risk or more likely is a data artifact.

When choosing the most appropriate spatial scale to use for presentation of epidemiologic data for vector-borne diseases to stakeholder communities, we are faced with a situation where use of the county scale obscures variability in spatial risk patterns evident at sub-county scales. However, use of sub-county scales introduces more potential error in terms of actual pathogen exposure location not falling within the spatial boundary unit containing the case's residence. Further studies are needed to determine the extent of this error for county versus sub-county scales for various vector-borne diseases. Use of sub-county units with small population sizes may also present the problem of unstable incidence rates, however this is also reflective of a temporally dynamic emerging infectious disease such as WNV disease.[21] Numerous spatial statistic smoothing methods exist to deal with the problem of rate instability including local-area averaging or geostatistical smoothing such as kriging.[32,33] Our findings also highlight the need to present maps of vector-borne disease incidence at either county or sub-county scales together with information on the limitations for the scale at which the data are presented.

Presentation of disease case counts versus disease incidence.

Figure 2 provides a powerful visual example of the value of side-by-side presentations of spatial disease patterns based on case counts versus case incidence. At the county scale, there was low overall correlation between disease incidence and case counts as well as poor concordance (50%) for counties categorized as high risk for WNV exposure based on case counts versus case incidence. Because some stakeholders are better served knowing disease case counts (e.g., mosquito control programs) whereas

other stakeholders need information based on disease incidence (e.g., policy makers), our findings argue for presentations of WNV disease data at the county scale that include maps showing WNV disease case counts as well as WNV disease incidence. Concordance between high-risk areas determined by case counts versus case incidence was also poor for the zip code scale (31%) but much higher for the census tract scale (83%). This pattern of higher concordance for census tracts than for either zip codes or counties in Colorado likely results, in part, from census tracts having a more uniform population size (mean population of 4,427, s.d. = 2,321) than either zip codes (mean population of 10,742, s.d. = 13,584) or counties (mean population of 74,355, s.d. = 148,158).

Use of WNV fever cases.

The spatial analyses presented here included all reported WNV disease cases (2003 and 2007) including WNV fever and WNV neuroinvasive disease as our main interest was not disease manifestation, but rather how to best display disease occurrence in map format. Inclusion of all reported WNV disease cases also increased the sample size contributing to a more robust spatial model. Several limitations do accompany the inclusion of WNV fever cases into the analyses, however. These limitations include: (1) WNV disease cases manifesting in the less severe WNV fever form often go unreported compared to neuroinvasive disease cases, in which the severity typically requires hospitalization, thereby resulting in laboratory confirmation and case reporting; (2) reporting of WNV fever cases may be differential (e.g., individuals exhibiting WNV fever symptoms who have sufficient access to health care have a higher likelihood of

seeking health services compared to individuals in rural areas, or those who lack resources to obtain health care). County scale analyses may mask the effects of differential reporting of WNV fever as the unit typically encompasses many different social and economic conditions. However, zip code and especially census tracts tend to contain a more uniform population in terms of social structure and economic conditions, which may tend to add bias to displays of WNV fever at these finer scales. Future research could repeat the analyses conducted here and weight WNV fever cases to account for possible reporting bias. To account for possible bias, prospectively collected WNV neuroinvasive disease cases could also be added to the current epidemiological database and the analyses repeated using only neuroinvasive disease cases.

Applicability to other vector-borne diseases in North America

The analytical methods employed in our study on WNVD in Colorado are broadly applicable to vector-borne diseases in North America where humans are incidental pathogen hosts. This includes a wide range of diseases caused by pathogens transmitted by fleas (e.g., plague), mosquitoes (e.g., eastern equine encephalitis, La Crosse encephalitis, St. Louis encephalitis, and WNV disease) and ticks (e.g., babesiosis, Colorado tick fever, human granulocytic anaplasmosis, human monocytic ehrlichiosis, Lyme disease, Rocky Mountain spotted fever, tick-borne relapsing fever, and tularemia). The same methods also may be applicable to mosquito-borne diseases where humans serve as important or primary pathogen hosts (e.g., dengue and malaria), but this needs to be corroborated in future studies.

Future directions.

Our study demonstrates the potential value of using sub-county scales to determine and present spatial assessments of risk for vector-borne pathogens based on epidemiologic data. However, further studies are needed to determine: (1) the potential drawbacks of moving from county to sub-county scale, especially with regard to potential increases in error due to pathogen exposure occurring outside of the census tract or zip code of residence; (2) logistical concerns related to changes in data collection practices; and (3) potential bias due to inclusion of WNV fever cases in the analyses. Another important research need is development of spatial risk models based on entomological risk measures to complement risk assessments based on epidemiologic data.

There also is need for extensive research on delivery mechanisms for spatial risk maps and other risk assessment information to stakeholder communities, especially through web-based information delivery mechanisms. This includes: (1) gaining a better understanding of what type of information different types of stakeholders feel that they require; and (2) determining optimal map and text formats to ensure that the message we aim to transmit in fact is clear to the user. Evaluating the effect of various data presentations on risk (e.g. maps of WNV diseases case counts versus disease incidence) also merits future research as the role of perceived threat is well documented in relation to the public's use of preventative actions.

Acknowledgments

We thank Saul Lozano-Fuentes of Colorado State University for helpful discussions.

Financial Support: The study was funded, in part, by a grant from the Centers for Disease Control and Prevention (T01/CCT822307) and a contract from the National Institutes of Allergy and Infectious Diseases (N01-AI-25489).

Authors' addresses: Anna M. Winters, Mark J. Delorey, Rebecca J. Eisen, Roger S. Nasci, and Emily Zielinski-Gutierrez, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, 3150 Rampart Rd. Fort Collins, CO 80522; Lars Eisen and Chester G. Moore, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO 80523; W. John Pape, Communicable Disease Program, Colorado Department of Public Health and Environment, 4300 Cherry Creek Drive South, Denver, CO 80246.

Table II-1 Concordance between spatial patterns for high-risk areas.
Concordance between spatial patterns for high-risk areas based on WNV disease incidence versus WNV disease case counts for the county, zip code and census tract scales in Colorado.

	WNV disease case count		
WNV disease incidence	High risk[a]	Other[b]	% concordance
	County scale		
High risk[a]	6	6	50%
Other[b]	6	46	88%
	Zip code scale		
High risk	23	51	31%
Other	51	318	86%
	Census tract scale		
High risk	158	33	83%
Other	33	851	96%

[a]Spatial units with WNV disease case counts or WNV disease incidences falling within the 4th quartile for each spatial scale were considered high risk; [b]Spatial units either lacking cases and with incidence of 0, or falling within the 1st to 3rd quartile.

Table II-2 Options for presentation of spatial patterns of vector-borne diseases.

Method	Advantages	Disadvantages
Disease case point locations	Fine-scale information allowing for precise communication of disease risk. Facilitates development of predictive spatial risk models from knowledge of case point locations (based on kriging inter-polation, associations of case locations with environmental factors, etc). Not subject to modifiable areal unit problem.	Privacy issues associated with dissemination of information for disease case locations (can be addressed by random offsets from actual case locations). Address of residence may not be the site of pathogen exposure. May simply reflect population density.
Aggregate of disease cases	Useful for targeting vector control resources to the areas with the highest case loads. Improved likelihood that the probable site of pathogen exposure is included within the boundary unit associated with the address of residence.	Less precise compared to disease case point locations. Subject to modifiable areal unit problem. May simply reflect population density.
Aggregate of disease incidence	Disease risk measure that accounts for population density and thus reflects level of personal risk. Improved likelihood that the probable site of pathogen exposure is included within the boundary unit associated with the address of residence.	Less precise compared to disease case point locations. Under-represent total variance. Subject to modifiable areal unit problem. Problematic for boundary units with low population bases.

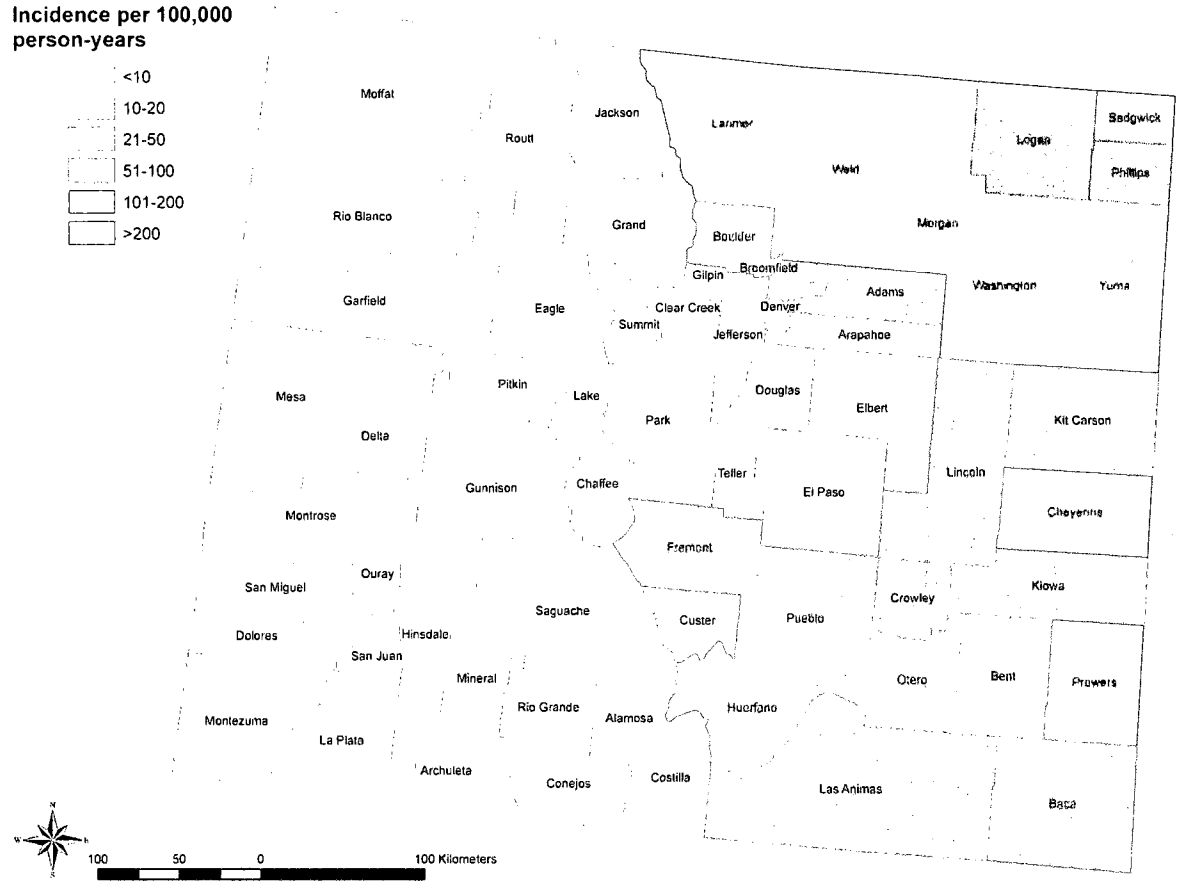

Figure II-1 West Nile virus disease incidence per 100,000 person-years.
These data for Colorado is displayed by county based on combined data (cumulative incidence) for the 2003 and 2007 outbreak years.

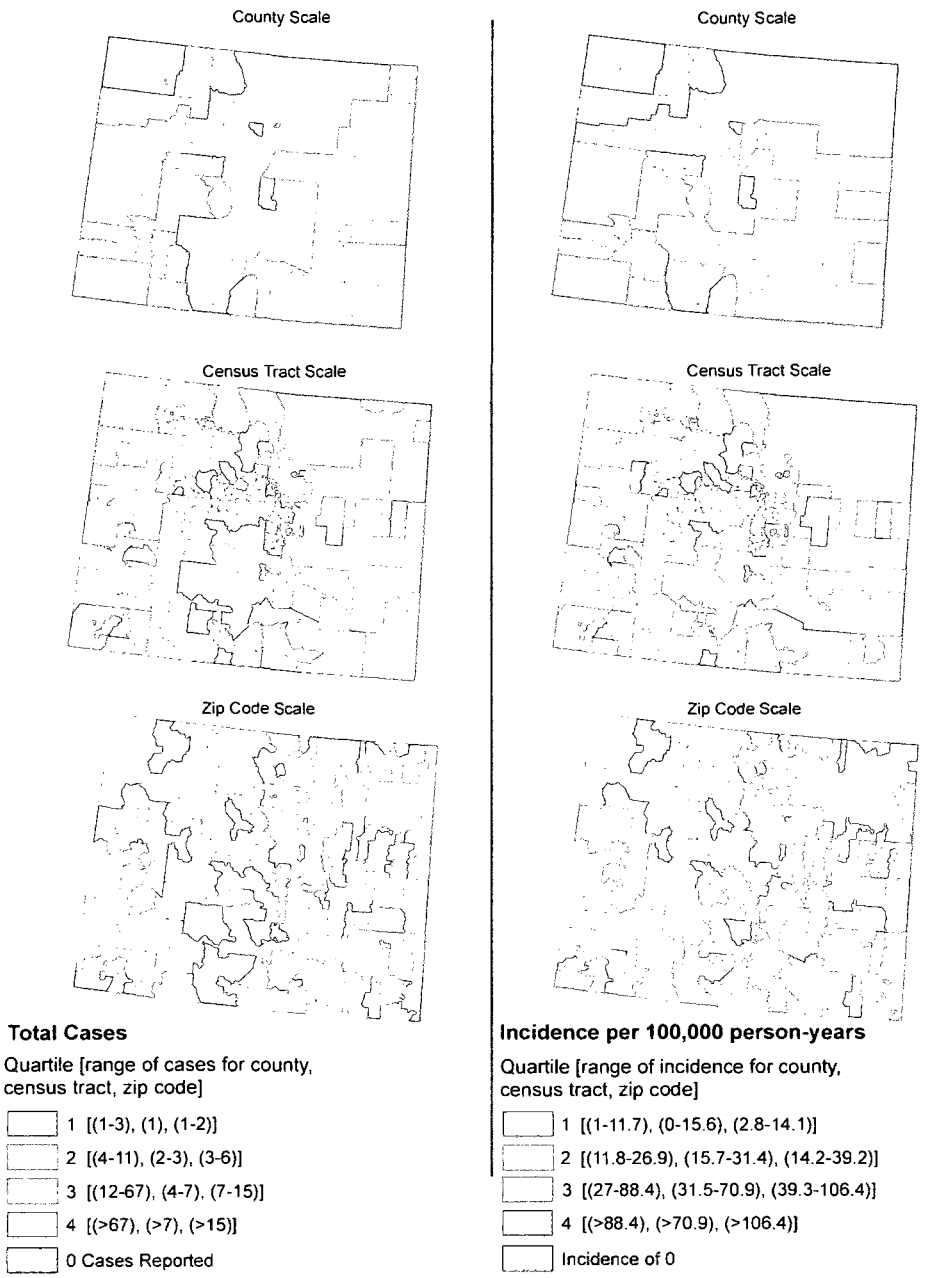

Figure II-2 West Nile virus disease case counts and incidence per 100,000 person-years.

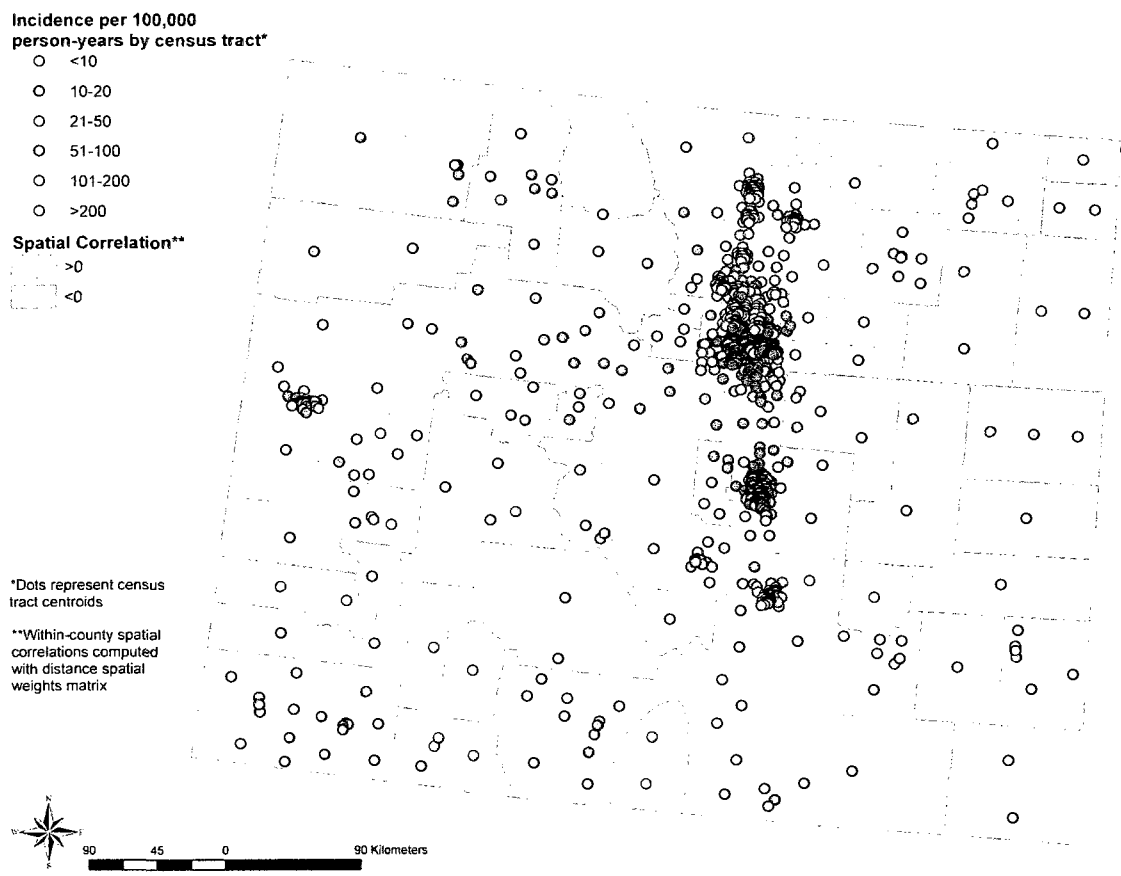

Figure II-3 Spatial correlation of WNV diseae incidence among census tracts.
Spatial correlation of West Nile virus disease incidence among census tracts for counties in Colorado based on combined West Nile virus disease data for 2003 and 2007. Counties with positive spatial correlation among census tracts are shown in purple, counties with negative spatial correlation among census tracts in light blue, and counties containing too few census tracts to calculate spatial correlation in white. Points indicate locations of census tract centroids and are color coded by West Nile virus disease incidence.

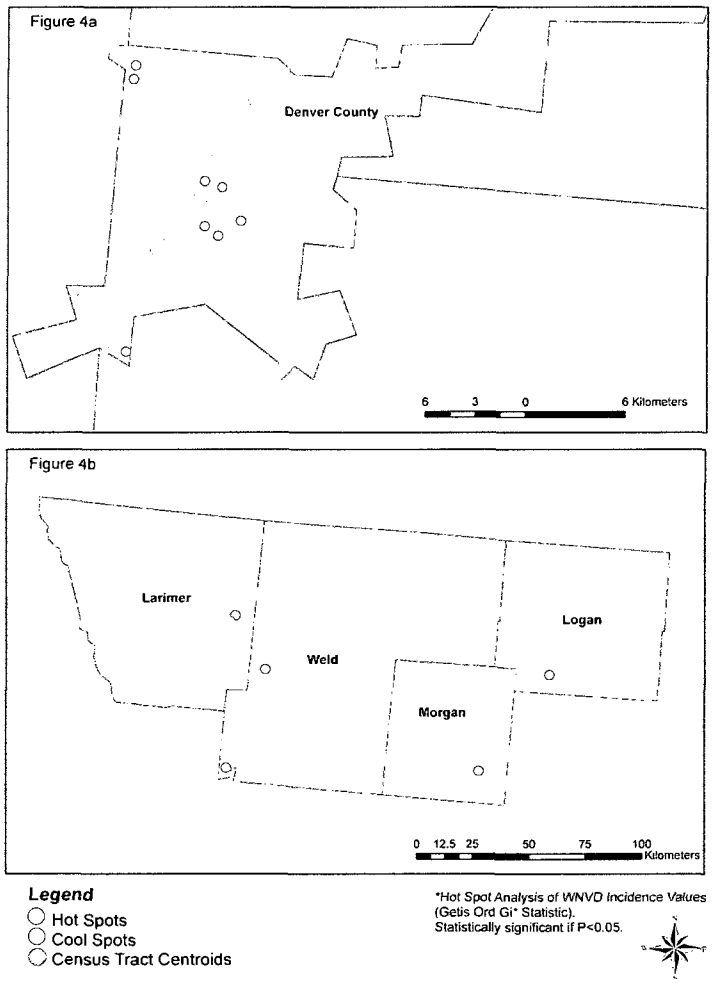

Figure II-4 Hot spots and cools spots of WNV disease incidence.
Hot spots and cool spots of West Nile virus disease incidence based on combined data for 2003 and 2007 for census tracts located within selected counties in Colorado: (A) Denver County with a low overall incidence (17.09 cases per 100,000 person-years), (B) Larimer, Weld, and Morgan counties with high overall incidences; range = (109.03 -- 134.48 cases per 100,000 person-years). Points indicate locations of census tract centroids and are color coded to indicate presence of statistically significant ($P < 0.05$) hot or cool spots of West Nile virus disease incidence

References Cited

1. Beck LR, Lobitz BM, Wood BL, 2000. Remote sensing and human health: New sensors and new opportunities. *Emerg Infect Dis 6:* 217-227.

2. Kalluri S, Gilruth P, Rogers D, Szczur M, 2007. Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: A review. *PLoS Pathog 3:* e116.

3. Kitron U, 1998. Landscape ecology and epidemiology of vector-borne diseases: Tools for spatial analysis. *J Med Entomol 35:* 435-445.

4. Rogers DJ, Randolph SE, 2003. Studying the global distribution of infectious diseases using GIS and RS. *Nature Rev Microbiol 1:* 231-237.

5. SaTScan™. (http://www.satscan.org/). 2009.

6. Theophilides CN, Ahearn SC, Grady S, Merlino M, 2003. Identifying West Nile virus risk areas: The dynamic continuous-area space-time system. *Am J Epidemiol 157:* 843-854.

7. GoogleEarth. (http://earth.google.com/). 2009.

8. MSVirtualEarth. (http://www.microsoft.com/virtualearth/). 2009.

9. 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 Negl Trop Dis 3: In Press.*

10. Kamadjeu R, 2009. Tracking the polio virus down the Congo River: A case study on the use of Google Earth™ in public health planning and mapping. *Int J Health Geogr 8:* 4.

11. Lozano-Fuentes S, Elizondo-Quiroga D, Farfan-Ale JA, Loroño-Pino MA, Garcia-Rejon J, Gomez-Carro S, Lira-Zumbardo V, Najera-Vazquez R, Fernandez-Salas I, Calderon-Martinez J, Dominguez-Galera M, Mis-Avila P, Coleman M, Morris N, Moore CG, Beaty BJ, Eisen L, 2008. Use of Google Earth to strengthen public health capacity and facilitate management of vector-borne diseases in resource-poor environments. *Bull WHO 86:* 718-725.

12. Gao S, Mioc D, Yi X, Anton F, Oldfield E, Coleman D, 2009. Towards Web-based representation and processing of health information. *Int J Health Geogr 8:* 3.

13. Bell BS, Hoskins RE, Pickle LW, Wartenberg D, 2006. Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. *Int J Health Geogr 5:* 49.

14. Eisen RJ, Eisen L, 2008. Spatial modeling of human risk of exposure to vector-borne pathogens based on epidemiological versus arthropod vector data. *J Med Entomol 45:* 181-192.

15. CDPHE. Colorado Department of Public Health and Environment (http://www.cdphe.state.co.us/dc/zoonosis/wnv/index.html). 2009.

16. Hardin J, Hilbe J: Generalized Linear Models and Extensions, 2nd Edition ed. College Station: Stata Press, 2007.

17. Fotheringham A, Brunsdon C, Charlton M: Geographically Weighted Regression. West Sussex: John Wiley and Sons Ltd, 2002.

18. Cressie N: Statistics for Spatial Data. New York: John Wiley and Sons, 1993.

19. Wall M, 2004. A close look at the spatial structure implied by the CAR and SAR models. *J Stat Planning and Inference 121:* 311-324.

20. Getis A, Ord J, 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis 24:* 189-206.

21. Rytkonen MJ, 2004. Not all maps are equal: GIS and spatial analysis in epidemiology. *Int J Circumpolar Health 63:* 9-24.

22. Curtis A, Mills JW, Leitner M, 2006. Keeping an eye on privacy issues with geospatial data. *Nature 441:* 150.

23. HIPAA. Health Insurance Portability and Accountability Act of 1996 (http://www.cms.hhs.gov/HIPAAGenInfo/Downloads/HIPAALaw.pdf.). 1996.

24. VanWey LK, Rindfuss RR, Gutmann MP, Entwisle B, Balk DL, 2005. Confidentiality and spatially explicit data: Concerns and challenges. *Proc Natl Acad Sci USA 102:* 15337-15342.

25. Openshaw S: The Modifiable Areal Unit Problem. Norwich, UK: Geo Books, 1984.

26. Diuk-Wasser MA, Brown HE, Andreadis TG, Fish D, 2006. Modeling the spatial distribution of mosquito vectors for West Nile virus in Connecticut, USA. *Vector-Borne Zoonotic Dis 6:* 283-295.

27. Eisen RJ, Eisen L, Lane RS, 2006. Predicting density of *Ixodes pacificus* nymphs in dense woodlands in Mendocino County, California, based on geographical information systems and remote sensing versus field-derived data *Am J Trop Med Hyg 74* 632-640.

28. Nielsen CF, Armijos MV, Wheeler S, Carpenter TE, Boyce WM, Kelley K, Brown D, Scott TW, Reisen WK, 2008. Risk factors associated with human infection during the 2006 West Nile virus outbreak in Davis, a residential community in northern California. *Am J Trop Med Hyg 78:* 53-62.

29. Rodgers SE, Mather TN, 2006. Evaluating satellite sensor-derived indices for Lyme disease risk prediction. *J Med Entomol 43:* 337-343.

30. Winters AM, Bolling BG, Beaty BJ, Blair CD, Eisen RJ, Meyer AM, Pape WJ, Moore CG, Eisen L, 2008. Combining mosquito vector and human disease data for improved assessment of spatial West Nile virus disease risk. *Am J Trop Med Hyg 78:* 654-665.

31. Eisen RJ, Lane RS, Fritz CL, Eisen L, 2006. Spatial patterns of Lyme disease risk in California based on disease incidence data and modeling of vector-tick exposure. *Am J Trop Med Hyg 75:* 669-676.

32. Elliot P, Wakefield J, Best N, Briggs D: Spatial Epidemiology Methods and Applications. New York: Oxford University Press, 2000.

33. Waller LA, Gotway CA: Applied Spatial Statistics for Public Health. Hoboken: Wiley, 2004.

III. Predictive Spatial Models for Risk of West Nile Virus Exposure in Eastern and Western Colorado

This chapter has been published in the American Journal of Tropical Medicine and Hygiene and is cited as Winters AM, Eisen RJ, Lozano-Fuentes S, Moore CG, Pape WJ, Eisen L. 2008. Predictive Spatial Models for risk of West Nile virus exposure in eastern and western Colorado. AJTMH 79 (4):581-590.

Anna M. Winters[1], Rebecca J. Eisen[2], Saul Lozano-Fuentes[1],

Chester G. Moore[1], W. John Pape[3], and Lars Eisen[1]

[1]Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado 80523; [2]Division of Vector-Borne Infectious Diseases, National Center for Vector-Borne, Zoonotic and Enteric Diseases, Centers for Disease Control and Prevention, Fort Collins, Colorado 80521; [3]Communicable Disease Program, Colorado Department of Public Health and Environment, Denver, Colorado 80546.

Abstract

Mosquito control and personal protection against mosquito bites are the only preventative measures available to reduce human West Nile virus (WNV) disease. Improved spatial targeting is desirable for costly mosquito and WNV surveillance and control schemes. We used multivariate regression modeling to develop spatial models for risk of exposure to WNV in western and eastern Colorado based on associations between Geographic Information System-derived environmental data and zip code of residence for 3,659 human WNV disease cases from 2002--2006. Models were robust with user accuracies for correct classification of risk areas of 70--87% for presence of WNV disease and 67--80% for high incidence of WNV disease. The importance of selecting a suitable model development area in an ecologically and climatically diverse environment was demonstrated by models based on data from the eastern plains landscape performing poorly in the mountainous western part of Colorado and vice versa.

KEY WORDS: West Nile virus disease, Colorado, Geographic Information Systems (GIS), spatial modeling

Introduction

West Nile virus (WNV) disease, which is caused by an arthropod-borne virus in the family Flaviviridae, has emerged as a serious public health problem in the United States.[1, 2] After spreading west from the initial focus in New York, the virus reached Colorado in 2002[2] and caused an epidemic in 2003 with 2,947 human cases and 63 fatalities reported to the Colorado Department of Public Health and Environment[3]. In the western United States, WNV is maintained in a natural transmission cycle primarily involving birds and *Culex* mosquitoes and is considered to most commonly be transmitted to humans by *Cx. tarsals.*[4-6] In the absence of a vaccine against WNV for use in humans,[7-10] mosquito control and personal protection against mosquito bites are the only preventative measures available to reduce risk of WNV exposure.[11-13] Effective implementation of mosquito and WNV control schemes requires knowledge of local spatial and temporal patterns of *Culex* mosquitoes and WNV. Spatial factors influencing risk of WNV exposure in humans include location of larval habitats, microclimate, and local abundance of hosts for female mosquitoes, especially birds serving as WNV reservoirs.[2, 6] Temporal factors affecting *Culex* vector abundance include the extent of annual flooding caused by snowmelt runoff and seasonal rainfall and temperature patterns.[14-19] High local abundances of *Culex* vectors enhance the enzootic transmission potential of WNV and may lead to intense viral transmission within the enzootic bird-mosquito cycle followed by incidental infections of humans and domestic animals.[2, 12]

Although it is well understood that spatially targeted and timely implementation of mosquito control measures is critical to reduce mosquito abundance and risk of WNV

transmission, deciding where and when to implement costly vector control measures is not a trivial matter. Detection of human WNV disease cases does not provide sufficient lead time for effective mosquito control because the peak in abundance of *Cx. tarsalis* in Colorado precedes the peak in human cases by several weeks to a month.[20, 21] Human WNV cases typically peak in late July and August coinciding with a rapid decline of populations of host-seeking *Cx. tarsalis* females. Emergency mosquito control initiated in response to peaks in human WNV cases is therefore of limited use. Surveillance for WNV seropositive birds, which typically is achieved through monitoring of sentinel chicken flocks or by testing of dead birds,[22-27] or WNV-infected mosquitoes, likewise may fail to provide sufficient lead time to implement effective mosquito control. To be most effective, mosquito control measures need to be implemented proactively based on factors discernible before the risk of a WNV outbreak can be determined unequivocally (e.g., weather patterns, *Culex* mosquito abundance). This is a conundrum facing local decision-makers across the United States charged with making decisions regarding resource allocation for mosquito control activities.

Knowledge of historical spatial and temporal patterns of abundance of *Culex* vectors and human WNV disease cases provide crucial information to facilitate implementation of proactive and evidence-based mosquito surveillance and control schemes.[15, 28-34] For large parts of Colorado, risk patterns can only be discerned from epidemiological data because mosquito control programs and, consequently, entomological data are lacking. To improve the knowledge of spatial risk of WNV exposure in Colorado, we used data for human WNV disease cases from 2002--2006 to first determine fine-scale (zip code-based) WNV disease incidence and thereafter identify associations between

environmental factors and WNV disease incidence through multivariate regression modeling. A secondary aim was to determine whether models based on data from the plains landscape of eastern Colorado were applicable to the mountainous western part of the state and vice versa.

Materials and Methods

Study area

The topographically diverse state of Colorado encompasses short-grass prairie in the eastern plains, high mountains of the Continental Divide running north to south in the central part of the state, and canyons and high plateaus to the west. The high elevation and mid-latitude interior continent location of Colorado results in a cool, dry climate with considerable daily and seasonal temperature fluctuations.[35] Precipitation in the eastern Colorado plains is seasonal with 70--80% of the annual total precipitation occurring from April through September. Along the foothills at the eastern edge of the Rocky Mountains (the Colorado Front Range), daily and seasonal temperature fluctuations are less severe and precipitation increases as one moves from the foothills into montane areas. Eastern Colorado is primarily rural agricultural land to the east with increasing focal human populations to the west along the Front Range (e.g., Colorado Springs, Denver, Boulder-Longmont, Loveland-Fort Collins). Two major rivers flow through eastern Colorado, the South Platte River to the northeast and the Arkansas River to the southeast. Western Colorado topography includes canyons and plateaus with winter weather colder, but less variable, than eastern Colorado. The valleys of west-central and southwestern Colorado

have a relatively mild climate and precipitation exhibits less seasonality compared to eastern Colorado. The Colorado River flows through west central Colorado.

Epidemiological data

The study included 3,659 WNV disease cases reported to the Colorado Department of Public Health and Environment (CDPHE) by zip code during 2002--2006. All included disease cases were from zip codes reporting a mean cooling degree day (CDD) value that exceeded a cut-off of 20.7 during the month of July (1961--1990, data from Climate Source LLC, Corvallis, OR). This CDD cutoff was chosen because entomological data collected along elevation-climate gradients in Larimer County in north-central Colorado in 2006 demonstrated that risk of exposure to *Culex* WNV vectors was minimal in areas with a CDD value below 20.7 (L. Eisen, unpublished data). Based on this CDD cutoff, the high mountain portion of central Colorado was excluded from the study as an area with minimal risk of exposure to *Culex* WNV vectors (Fig. 1). Few WNV disease cases (n = 42; 1.1% of the total for the state) were reported from zip codes in the excluded area during 2002--2006 and we consider it probable that the vast majority of these cases resulted from out-of-zip code exposures to infected *Culex* mosquitoes at lower elevations. Data from 301 zip codes to the east and 59 zip codes to the west of the high mountains were used to calculate zip code-based cumulative WNV disease incidences per 100,000 person-years during 2002--2006. Cumulative cases were aggregated to zip code rather than census tract because the variability of census tract area (<1-8,947 km^2) was more pronounced than zip code area (<1-5,131 km^2), especially in high-density urban and sparsely populated rural regions of the state. Quality of data from

Colorado for probable sites of exposure (point locations) for the relatively mild and common WNV disease is highly variable (reflecting differences in information gathering practice and interest of individual physicians) and precludes reliable use of point locations as the basis for determining environmental correlates of WNV disease risk. We do, however, consider it a valid working assumption that most exposures occurred within the zip code of residence because human habitation in the semi-arid Colorado landscape tends to cluster in the vicinity of mosquito larval habitat (rivers, lakes, ponds, irrigated lands). Major recreational areas in western Colorado and the Front Range area cluster in montane habitats where *Culex* vectors are rare or lacking and, thus, WNV exposure is unlikely to commonly be associated with out-of-zip code recreational activities.

Environmental data

Geographic Information System (GIS)-based data included:

1) administrative boundaries including state, county and zip code (Environmental Systems Research Institute; ESRI, Redlands, CA);

2) long-term (1961--1990) climate data (mean monthly and annual temperature, cooling degree, heating degree and growing degree days; mean monthly and annual precipitation, snowfall, and relative humidity; mean annual length of freeze-free period; Julian date of first and last snowfall; at 2 x 2 km spatial resolution; Climate Source LLC);

3) U.S. Geological Survey 30-meter national elevation data-set;

4) land cover classification based on the 30-meter 2001 National Land Cover Dataset (NLCD);

5) annual average Normalized Difference Vegetation Index (NDVI) data from 2005 (1x1 km spatial resolution; derived from NOAA Advanced Very High Resolution Radiometer images).

GIS-based data were extracted by zip code using ArcGIS 9.2 (ESRI). All layers were projected to North American Albers Equal Area Conic projection and North American 1983 datum.

Model construction

Multivariate logistic regression models, based on the above-mentioned environmental data, were developed to predict spatial patterns of: 1) zip codes with WNV disease cases present; and 2) zip codes with high incidence of WNV disease. Separate models were developed for eastern versus western Colorado due to the climatic and topographic variability between the two regions which may result in different environmental determinants for WNV disease incidence. Logistic regression modeling, rather than linear regression modeling, was used because WNV disease incidence data were not normally distributed and transformation could not make them so (Shapiro-Wilk test; $P < 0.05$ for both eastern and western Colorado). High WNV disease incidence was classified using the fourth quartile cut-off value for zip code-based incidence in eastern Colorado ($\geq$ 38.39 cases per 100,000 person-years) or western Colorado ($\geq$ 25.75 cases per 100,000 person-years). Because the range of disease incidence by zip code was greater in eastern Colorado (0--754.72 cases per 100,000 person-years) versus western Colorado (0--512.82 cases per 100,000 person-years), separate cut-off values for high WNV disease incidence were used to ensure that model development would be representative of the specific region.

Model development for eastern or western Colorado was based on random selection of 75% of the zip codes for each region, with the remaining 25% of zip codes reserved for model validation (Fig. 1). Forward stepwise regression was used to identify three or more candidate models each for presence of WNV disease (eastern and western Colorado) and high incidence of WNV disease (eastern and western Colorado). Covariates included in the models (probability to enter of 0.25) were restricted to variables significantly associated with WNV disease incidence in univariate tests of association (Wilcoxon's test; $P < 0.05$) but not strongly correlated with each other (Spearman's rank correlation; $\rho_s < 0.8$). A goodness of fit test was applied to determine whether the model covariates adequately explained the distribution of WNV disease incidence. Receiver Operating Curves (ROCs) assessed the overall discrimination of the model based on the area under the curve (AUC). The AUC provides a threshold-independent measure of the overall accuracy of the model and was also used to determine the optimal probability cut-off which characterized each grid cell as presence versus absence of WNV disease or high versus lower WNV disease incidence. Models with the lowest Aikaike Information Criterion (AIC) were considered the most parsimonious models, but models within two AIC units of the minimum AIC value were considered competing.[36] The best models for presence of WNV disease and high incidence of WNV disease in western and eastern Colorado had the lowest or competing AIC values and provided the most robust validations.

The models are described with the following equation:

$$Logit(P) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k \qquad [1]$$

where P is the probability of a grid cell being classified as having WNV disease cases present or high WNV disease incidence, β_0 is the intercept and $\beta_1 \ldots \beta_k$ represent the coefficients associated with each independent variable $x_1 \ldots x_k$.

To create GIS-data layers predictive of presence or high incidence of WNV disease, the following equation (derived from equation 1) was used to express the probability that a particular cell is classified as predicted presence or high incidence of WNV disease cases:

$$P = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}{[(1 + \exp(\beta_0 + \beta_1 x + \beta_2 x_2 + \ldots + \beta_k x_k)]} \quad [2]$$

Data layers predictive of presence or high incidence of WNV disease were first constructed by applying the eastern and western models at the zip code scale. More fine-scale predictive data layers (2 x 2 km pixel size) were generated by applying specific model equations to the raster data using the raster calculator of ArcGIS 9.2 (ESRI Redland, CA). For each zip code, the maximum probability value (P) for having WNV disease cases present or high WNV disease incidence among raster cells was extracted using zonal statistics (ArcGIS Spatial Analyst, ESRI). User and producer accuracies were maximized simultaneously by optimizing the probability cut-off value. Cells with a probability greater than the cut-off value were classified as 'presence' or 'high incidence' for WNV disease, depending upon the model. All other cells were considered 'absence' or 'lower incidence'. We also evaluated the predictive capability of the eastern Colorado models for western Colorado and the western models for eastern Colorado to investigate the potential for applying regional models statewide. This included extrapolation of the model developed for eastern Colorado to western Colorado and evaluation against

cumulative WNV disease incidence data from western Colorado. The same procedure was repeated for the other half of the state. Results of statistical tests were considered significant when $P < 0.05$.

Results

Summary of West Nile virus disease data

Mean cumulative incidence (2002--2006) of WNV disease by zip code for the eastern and western parts of Colorado was similar with 29.63 cases per 100,000 person-years in the east (range, 0--754.72) and 29.13 cases per 100,000 person-years in the west (range, 0--512.82). Larimer, Boulder and Weld counties in the Front Range area reported the highest case numbers across the state (totals of 618, 513 and 489 cases for 2002--2006, respectively). On the other hand, the highest cumulative incidences of WNV disease occurred in the eastern part of the state, with the northeastern counties of Logan, Sedgwick and Phillips reporting 100, 154 and 122 cases per 100,000 person-years, respectively, during 2002--2006, compared to 35--44 cases per 100,000 person-years for Boulder, Larimer, and Weld. Mesa and Delta counties (located in western Colorado), and Fremont County (located in central Colroado), reported relatively high incidence rates (33--49 cases per 100,000 person-years).

Predictive spatial models for risk of exposure to West Nile virus

Multivariate logistic regression models based on GIS-derived environmental data were developed to predict areas where WNV disease was present, or where WNV disease incidence was high, for western and eastern Colorado (Table 1). Lack of fit tests indicated that all models included sufficient numbers of covariates ($P > 0.05$ in all cases)

and whole model tests denoted good overall fit for all models ($P < 0.05$ for western Colorado models and $P < 0.001$ for eastern Colorado models) (Table 1). Parameter estimates for covariates of selected models are shown in Table 2. Models for presence of WNV disease and high incidence of WNV disease in eastern and western Colorado based upon zip code and 2 x 2 km raster scales are shown in Figures 2-5. The selected western and eastern Colorado models for presence of WNV disease were not the most parsimonious models, but were considered competing based on AIC values and ultimately were chosen because they provided the most sensitive and specific validations. Selected models included NDVI in May as a covariate in the eastern model and precipitation in May in the western model for presence of WNV disease. Eastern Colorado had a higher percentage of area with predicted presence and high incidence of WNV disease (67 and 27%, respectively) compared to the western part of the state (40 and 12%, respectively) (Fig. 2-5). The models predicted high incidence of WNV disease along river basins including the Platte, Arkansas and Colorado rivers, and in the northern Colorado Front Range (Fig. 4-5).

Evaluation of models

Models were validated internally (against the zip code data set used to construct the models) as well as externally (against the zip code data set not included in model development). Validations were conducted at zip code scale and also at a finer 2 x 2 km scale. Only the external validation results are reported in the text but data for both internal and external validation results are shown in Tables 3 and 4. User and producer accuracies were calculated for evaluation of the models. User accuracy (errors of commission) calculates the probability that the modeled data represents reality and

provides a measurement of how well the models represent disease risk. Producer accuracy (errors of omission) is a measure of the accuracy of the particular modeling scheme and measures what percentage of the area was modeled correctly.[37]

External zip code scale validation – Presence of WNV disease. Producer accuracies for categorizing areas with WNV disease present as likely to contain cases, or areas with WNV disease absent as likely to lack cases, were 77.78% and 50.00%, respectively, for western Colorado and 89.83% and 50.00%, respectively, for eastern Colorado (Table 3). User accuracy for presence of WNV disease cases in zip codes predicted by the model to contain cases was 70.00% in western Colorado and 86.89% in eastern Colorado. Zip codes classified as lacking WNV disease did not yield disease cases 60.00% of the time in western Colorado and 57.14% of the time in eastern Colorado (Table 3). Overall, the models provided an accurate fit to the actual spatial pattern of zip codes with presence of WNV disease in western Colorado and most areas in eastern Colorado except the east-central region of the state where disease presence was under-predicted (Figure 2).

External zip code scale validation – High incidence of WNV disease. Producer accuracies for categorizing areas with high WNV disease incidence as likely to have high incidence, or areas with lower incidence as likely to have lower incidence were 66.67% and 88.89%, respectively, in western Colorado and 58.82% and 91.38%, respectively, for eastern Colorado (Table 3). User accuracy for high WNV disease incidence in zip codes predicted by the model to have high incidence was 80.00% in western Colorado and 66.67% in eastern Colorado. Zip codes predicted to have lower incidence also reported lower disease incidence 80.00% of the time in western Colorado and 88.33% of the time

in eastern Colorado (Table 3). Regarding the spatial distribution of errors of omission, the high incidence model under-predicted disease risk along a portion of the Arkansas River in the south-central region of the state (Figure 4).

External fine 2 x 2 km scale validation – Presence of WNV disease. The 2 x 2 km scale model for presence of WNV disease had producer accuracies for categorizing areas with WNV disease present as likely to contain cases, or areas with WNV disease absent as likely to lack cases, of 88.89% and 66.67%, respectively, in western Colorado and 88.14% and 35.17%, respectively, in eastern Colorado. User accuracies were 80.00% in western Colorado and 85.25% in eastern Colorado for occurrence of WNV disease cases in areas predicted by the model to have WNV disease cases present. Areas predicted to lack WNV disease cases contained no cases 80.00% of the time in western Colorado and 41.67% of the time in eastern Colorado (Table 4). Similar to the zip code-based predicted presence models, the raster-based models fit the actual spatial pattern of areas with presence of WNV disease in the western region and the majority of the eastern region but under-predicted disease presence in the east-central region of the state (Fig. 3). The raster-based model for eastern Colorado did, however, predict presence of raster cells with risk within zip codes where WNV disease was reported but not predicted at the zip-code scale.

External fine 2 x 2 km scale validation – High incidence of WNV disease. The 2 x 2 km scale model for high incidence of WNV disease had producer accuracies for categorizing areas with high WNV disease incidence as likely to have high incidence, or areas with lower incidence as likely to have lower incidence, of 66.67% and 88.89%, respectively, in western Colorado and 58.82% and 80.00%, respectively, in eastern

Colorado. User accuracies were 80.00% in western Colorado and 62.50% in eastern Colorado for occurrence of high incidence of WNV disease in areas predicted by the model to have high incidence. Areas classified as lower disease incidence also had lower incidence 80.00% of the time in western Colorado and 77.42% of the time in eastern Colorado (Table 4). The raster-based high incidence models showed a spatial pattern of disease incidence similar to reported WNV disease. Similar to the zip code-based high incidence model, the raster-based high incidence model under-reported disease risk along a portion of the Arkansas River at the western edge of the eastern plains. The raster-based high incidence model did, however, show small patches of risk within zip codes reporting high incidence of WNV disease in this area (Fig. 5). To investigate how these small patches of predicted high incidence of WNV disease coincided with population density, the population per square kilometer was mapped by census tract and compared to the raster-based high incidence model (Fig. 6). This revealed that the small patches of model-predicted high incidence of WNV disease occurring along the aforementioned portion of the Arkansas River commonly coincided with population centers (e.g., Pueblo).

Extrapolation of models to other regions in Colorado

We also investigated whether zip code-based models developed for eastern or western Colorado provide accurate predictions for presence or high incidence of WNV disease in the other part of the state. Model predictions were validated at the zip code scale against actual presence or high incidence of WNV disease for all zip codes in the validation area (i.e., 301 eastern zip codes or 59 western zip codes). When applied to western Colorado, the selected eastern Colorado model for presence of WNV disease gave producer accuracies for categorizing areas with WNV disease present as likely to

contain cases, or areas with WNV disease absent as likely to lack cases, of 65.71% and 50.00%, respectively (Table 5). User accuracy in western Colorado for occurrence of WNV disease cases in areas predicted by the eastern model as WNV disease present was 65.71%. Zip codes in western Colorado classified by the eastern model as lacking WNV disease cases contained no cases 50.00% of the time (Table 5). The eastern model for presence of WNV disease predicted cases to occur in areas of western Colorado that did not report WNV disease cases, especially to the far northwest (Fig. 7).

When applied to western Colorado, the selected model for high WNV disease incidence in eastern Colorado gave producer accuracies for categorizing areas with high WNV disease incidence as likely to have high incidence, or areas with lower incidence as likely to have lower incidence, of 41.18 and 76.19%, respectively, and user accuracies of 41.18 and 76.19%, respectively (Table 5). This model fit the actual spatial pattern for areas with high WNV disease incidence in the west-central region along the Colorado River basin, but over-predicted high WNV disease incidence to the northwest (Fig. 7).

When applied to eastern Colorado, the selected western model for presence of WNV disease cases gave producer accuracies for categorizing areas with WNV disease present as likely to contain cases, or areas with WNV disease absent as likely to lack cases, of 84.23% and 36.67%, respectively (Table 5). User accuracy for occurrence of WNV disease in areas predicted by the western model to contain cases was 84.23%. Zip codes in eastern Colorado classified by the western model as lacking WNV disease contained no cases 36.67% of the time (Table 5). As shown in Fig. 8, the western model for presence of WNV disease over-predicted presence of WNV disease in eastern Colorado.

When applied to eastern Colorado, the selected western model for high WNV disease incidence gave producer accuracies for categorizing areas with high incidence as likely to have high incidence, or areas with lower incidence as likely to have lower incidence, of 25.68% and 75.77%, respectively (Table 5). User accuracy for occurrence of high incidence of WNV disease in areas of eastern Colorado considered by the western model as high incidence was 25.68%. Zip codes in eastern Colorado classified by the western model as having lower WNV disease incidence had lower incidence 75.77% of the time (Table 5). The western model predicted only the southeastern corner of the eastern part of Colorado to include areas with high WNV disease incidence thus overlooking a large number of zip codes with actual high WNV disease incidence to the northeast (Fig. 8).

Overall, the zip code-based models performed better when applied to the model development area (Table 3) than the other part of the state (Table 5). This was most pronounced for the models of high WNV incidence, with user accuracies for categorizing areas with high incidence as likely to have high incidence of 66.67 -- 80.00% within the respective model development areas but 25.68 -- 41.18% when the models were applied to the other part of the state.

Discussion

West Nile virus is the leading cause of arboviral disease in the United States with >15,000 human cases and >600 fatalities since 1999.[38, 39] Despite the fact that Colorado has emerged as a major focal point of WNV disease, with 3,703 human cases and 76 fatalities reported from 2002--2006, mosquito control is still inconsistent or lacking in

large parts of the state including areas severely afflicted by WNV disease in the eastern plains region. Because entomological data on *Culex* vector abundance is scarce from areas in Colorado lacking mosquito control programs, we have developed models based on associations between environmental and epidemiological data to assess statewide spatial risk of WNV exposure at multiple scales. The resulting models identify specific environmental factors associated with WNV disease occurrence or high incidence of WNV disease, and can be used to assess statewide mosquito control needs.

As expected, the models predicted high incidence of WNV disease along river basins including the Platte, Arkansas and Colorado rivers, and in the northern Colorado Front Range. Eastern Colorado had greater coverage of areas with predicted presence and high incidence of WNV disease compared to the western part of the state. This may be due to several factors. First, human population density in eastern Colorado (572 per km^2) is 12-fold greater than in western Colorado (46 per km^2); the larger human population in the east likely results in a greater likelihood of WNV cases occurring in this region. This was reflected in the model outcomes which showed large areas of predicted high incidence of WNV disease occurring in populated regions especially within the northern Front Range. In more rural areas with focal population centers (e.g. southern Front Range, eastern plains region, and the central western region) small patches of predicted high incidence of WNV disease overlapped with or occurred in close proximity to population centers (Fig. 6). Second, eastern Colorado has a higher coverage of irrigated agricultural land (18%) compared to the western part (10%), and ponds, small lakes, and riparian areas are common within and adjacent to population centers in the northern Front Range severely impacted by WNV disease (e.g., Boulder, Fort Collins, Greeley,

Loveland). Irrigated agricultural land and ponds, lakes, and riparian areas provide larval habitat for *Cx. tarsalis*[18, 19, 40] and thus may support enzootic WNV amplification cycles. Third, agricultural areas occur commonly on the edges of population centers throughout eastern Colorado. These may provide "source habitats" for *Cx. tarsalis*, with movement of mosquitoes into nearby urban areas during the summer months when *Cx. tarsalis* is known to shift from feeding on birds to mammals, including humans.[41-43] Fourth, there could be systematic differences in the level of detection and reporting of WNV cases between counties in the eastern and western parts of Colorado. Further research on the spatiotemporal adaptation of mosquito populations and WNV transmission in and around population centers in Colorado is urgently needed to improve mosquito control practices.

The selected eastern and western Colorado models included temperature variables as important predictors of WNV disease (Tables 1--2). This was not surprising because warm temperatures result in more rapid development rates among the immature life stages and a shorter gonotrophic cycle in *Cx. tarsalis* as well as a shorter extrinsic incubation period of WNV in *Culex* vectors.[16, 17, 44, 45] Interestingly, in eastern Colorado we found a positive association between WNV disease and temperature in July (i.e., a negative association with heating degree days which reflects demand for heating energy and therefore has a negative relationship with temperature), but a negative association with temperature in August (i.e., a positive association with heating degree days). This may be explained by: (1) high summer temperatures ultimately leading to reduction in availability of natural larval habitat in the semi-arid climate of eastern Colorado and (2) possible negative effects of high temperatures on WNV replication or transmission which previously was demonstrated for western equine encephalitis virus.[44, 46] The selected

western Colorado models included positive associations with cooling degree days in May (cooling degree days measure the demand for cooling energy and has a positive association with temperature) and temperature in March. This association indicates that early season warm stretches are critical for mosquito population build-up and sets the stage for WNV disease outbreaks in the mountainous, high elevation landscape of western Colorado (mean zip code elevation was 2,080 m in the western region as compared to 1,672 m in the eastern region).

Precipitation was predictive of WNV disease in both eastern and western Colorado (Tables 1--2). The association with WNV disease was positive for spring precipitation (April-May) in both regions but negative for summer (July) precipitation in the east. This is not surprising because high spring precipitation likely combines with snowmelt run-off to create natural larval habitat, whereas heavy summer rains can flush out eggs and immatures from larval habitats, especially for *Cx. pipiens* in urban or semi-urban environments[47]. Further research is needed to determine if the effect of summer precipitation on *Cx. tarsalis* populations is more detrimental in urban or semi-urban environments where man-made waste water systems concentrate water flow than in rural areas where such systems are lacking.

Normalized Difference Vegetation Index (NDVI) in May was positively associated with WNV disease in the eastern models. A high NDVI value is associated with abundance of green vegetation, which often is found along riparian areas and around lakes and ponds providing larval habitat[32, 48]. NDVI was not significantly associated with WNV disease in the western model, perhaps because the distinct elevation and temperature gradients occurring in the mountainous landscape of western Colorado,

relative to the eastern plains, result in climate factors being more basic determinants of *Culex* abundance than abundance of green vegetation.

The user accuracy of the selected models in predicting areas with WNV disease present or with high WNV disease incidence generally was high. Predictive models for presence of WNV disease reliably identified areas that reported WNV disease cases; user accuracies at the zip code scale (external validation) ranged from 70% for western Colorado to 87% for eastern Colorado. The predictive models for high incidence of WNV disease most reliably identified areas that did not report high incidence of WNV disease; user accuracies at the zip code scale (external validation) were 80% for western Colorado and 88% for eastern Colorado. The models for high incidence of WNV disease, which are unlikely to over-predict risk areas, are well suited for determination of critical statewide mosquito control needs.

Several areas in eastern and western Colorado not reporting WNV disease cases were classified by the models as having predicted presence or high incidence of WNV disease. This error of commission may be attributed, in part, to persons without or with only mild symptoms of WNV disease not seeking medical attention. Based upon calculations by the Centers for Disease Control and Prevention, only 1 in 150 (0.7%) infections of WNV result in neuroinvasive disease, which typically requires hospitalization.[49] Applying this ratio to the 757 West Nile virus disease cases from Colorado that resulted in neuroinvasive disease (encephalitis or meningitis) from 2002--2006, the actual number of WNV infections in Colorado during this time period likely exceeded 110,000. Calculations by the Centers for Disease Control and Prevention also estimate that approximately 20% of likely WNV infections result in clinical disease (and

thus may be recognized and reported). Therefore, of the 110,000 likely infections in Colorado from 2002--2006, approximately 22,000 cases of clinical disease likely occurred of which 3,703 WNV disease cases (16.8%) were actually reported. Behavioral practices, such as use of repellents or people spending minimal time outdoors during the peak of the WNV transmission season,[13, 50, 51] as well as mosquito control practices may have contributed to the errors of commission as neither were considered in our model.

Errors of omission also occurred, where WNV disease was reported within areas predicted by our models to have no or low risk of WNV disease. Comparison of the zip code-based versus the raster-based high incidence models did, however, demonstrate that the more fine-scale model predicted presence of patches of risk within zip codes with reported high disease incidence but not containing enough coverage of high risk pixels to classify the entire zip code as predicted high risk. These patches of risk identified by the raster models, but overlooked by the zip code models, illustrate the inherent problems of using a coarse-scale unit to display disease risk and provide merit towards the use of fine-scale, raster-based models. Further, small patches of high risk identified by the raster-based models were most likely accurate as they often coincided with small (densely populated) census tracts with high reported WNV disease incidence (Fig. 6).

Errors of omission may have also resulted from travel or recreationally related WNV exposure outside the zip code of residence. Unfortunately, reliable information on point locations for probable site of WNV exposure was not available for the development of our models. Rather, the models were based on human cases of WNV disease aggregated to the zip code level with a basic assumption that WNV exposure most likely occurred within the zip code of residence. Lack of reliable information for probable sites

of pathogen exposure for common and relatively mild vector-borne diseases, such as Lyme disease and WNV disease, is emerging as a major impediment to the development of spatial epidemiology models for vector-borne diseases in the United States.[52] Comprehensive case investigations including determination of the probable pathogen exposure site for vector-borne diseases are typically only conducted for plague and the costs and time associated with such investigations are prohibitive for more common and less severe diseases. This underscores the need for a discussion in the public health community of the importance of determining the probable site of WNV exposure and, possibly, for physicians routinely taking standardized travel histories from WNV disease patients.

The multivariate logistic regression models were based on cumulative disease incidence aggregated to zip codes (dependent variable) and raster-based environmental covariates (independent variables). All predictive models were first displayed and validated at the zip code scale. In order to construct more fine-scale continuous spatial risk surfaces, the same predictive models were applied to 2 x 2 km raster data. External validation of the zip code versus fine-scale 2 x 2 km models showed the zip code models to most accurately identify areas where WNV disease was present or of high incidence in eastern Colorado. In western Colorado, the zip code and fine-scale 2 x 2 km high incidence models had identical user and producer accuracies, and the fine-scale 2 x 2 km model for presence of WNV disease in western Colorado validated more robustly when compared to the zip code model. The validation results thus did not provide conclusive evidence for which scale was more appropriate. The lack of point locations of probable WNV exposure for model development likely reduced the usefulness of the fine-scale

approach, especially in eastern Colorado. On the other hand, a key advantage of fine-scale spatial risk models is an enhanced ability to target small but important high risk areas occurring within large areas of low risk and obscured at the zip code scale.

Understanding the appropriate scale for model development and the feasibility of scaling-up a model for risk of WNV exposure to a larger area is critical. We therefore investigated how risk models developed for western or eastern Colorado performed when applied to the other area. The results showed that the models (especially the high incidence models) performed far better in the part of Colorado for which they were developed and underscored the importance of selecting a proper scale for model development based on an understanding of the physical environment (the plains landscape of eastern Colorado versus the mountainous landscape of western Colorado) rather than an administrative boundary (the state of Colorado). For example, the model for high WNV disease incidence developed for western Colorado failed to detect large areas with high risk for WNV exposure in the northeastern part of the state (Fig. 8), and the model developed for eastern Colorado predicted high WNV disease incidence in parts of northwestern Colorado where very few cases actually occurred (Fig. 7). Although Colorado may be an extreme example in terms of topographic, ecological, and climatic variability, it underscores the importance of careful selection of model development areas for epidemiological models of vector-borne diseases. Based on similarities in the physical environment, the model from eastern Colorado may perform well throughout the Central Plains and the model from the western part of the state may be applicable to much of the intermountain west.

Finally, our study highlights one important problem related to mosquito control in Colorado, namely the lack of control activities in rural areas of the eastern plains at high risk for WNV exposure. From a statewide perspective, mosquito control activities accurately target high risk urban areas along the Colorado River in the west and in the northern Front Range in east-central Colorado, but smaller communities along the Arkansas and Platte rivers in the eastern plains severely impacted by WNV disease and lacking means to pay for mosquito control have been left by the wayside. WNV disease will remain a public health concern in Colorado for the foreseeable future and needs to be addressed in a more comprehensive and evidence-based fashion than the current piecemeal approach which is largely based on control by commercial outfits on year-to-year contracts.

Acknowledgements

The study was funded by grants from the Centers for Disease Control and Prevention (T01/CCT822307), Colorado State University College of Veterinary Medicine and Biomedical Sciences, and by National Institutes of Health contract AI-25489.

Corresponding author: Anna M. Winters, Department of Microbiology, Immunology and Pathology, Colorado State University, Infectious Disease Annex - 1690 Campus Delivery, Fort Collins, CO 80523-1690; Telephone: 970-491-8745; FAX: 970-491-8707; e-mail: anna.winters@colostate.edu

Table III-1. Multivariate logistic regression models
Multivariate logistic regression models for presence of West Nile virus disease cases and high West Nile virus disease incidence for western and eastern Colorado.

Area and model type	Model No.	k[a]	Negative log-likelihood	AIC Values[b] AIC	ΔAIC	Whole model P-value	Goodness of fit P -value	AUC[c]	Independent model covariates[d]
Western Colorado									
Presence of WNV disease	1	2	25.56	55.12	0.61	0.004	0.16	0.768	CDD in May
	2	3	2.26	54.51	0	0.004	0.20	0.781	CDD in May, RH in August
	3*	4	23.92	55.84	1.33	0.009	0.18	0.780	CDD in May, RH in August, Precip in May
High incidence of WNV disease	1*	2	21.07	43.15	0	0.007	0.47	0.763	Temp in March
	2	3	20.96	47.93	1.78	0.023	0.43	0.748	Temp in March, Annual RH
	3	3	20.74	47.48	1.33	0.018	0.45	0.791	Temp in March, NDVI in June
	4	4	20.65	49.30	3.15	0.037	0.43	0.777	Temp in March, Annual RH, NDVI in June
Eastern Colorado									
Presence of WNV disease	1	2	102.00	208.00	6.84	<0.001	0.80	0.690	Precip in July
	2	3	101.59	209.18	8.02	<0.001	0.83	0.704	Precip in July, HDD in July
	3	4	96.58	201.16	0	<0.001	0.92	0.750	Precip in July, HDD in July, Precip in April
	4*	5	96.35	202.69	1.53	<0.001	0.92	0.754	Precip in July, HDD in July, Precip in April, NDVI in May
High incidence of WNV disease	1	2	101.34	206.68	25.96	<0.001	0.74	0.819	Elev
	2	3	96.46	198.91	18.19	<0.001	0.93	0.831	Elev, Precip in July
	3	4	91.16	190.33	9.61	<0.001	0.98	0.853	Elev, Precip in July, HDD in August
	4	5	85.82	181.64	0.93	<0.001	0.99	0.866	Elev, Precip in July, HDD in August, Snow in September
	5*	6	84.36	180.72	0	<0.001	0.99	0.873	Elev, Precip in July, HDD in August, Snow in September, NDVI in May

*Selected models.
[a] k, the number of estimated parameters included in the model; [b] AIC, Akaike Information Criterion: ΔAIC = AIC for model – AIC for most parsimonious model; [c] AUC, area under the Receiver Operator Curve; [d] CDD, cooling degree days; Elev, elevation; HDD, heating degree days; Precip, precipitation; RH, relative humidity; Snow, snowfall; Temp, temperature.

Table III-2. Parameter estimates for multivariate logistic regression models
Parameter estimates for selected multivariate logistic regression models predicting areas with West Nile virus disease present, or with high incidence of West Nile virus disease for easter and western Colorado.

Area, model type and model covariates	Parameter estimate		Likelihood ratio test		
	Estimate	S.E.	χ^2	df	*P*
Western Colorado					
Presence of WNV disease					
Intercept	-10.66	6.51	2.68	1	0.10
Cooling degree days in May (CDD)	0.23	0.09	5.89	1	0.01
Relative humidity in August (%)	0.16	0.11	2.03	1	0.15
Precipitation in May (mm)	0.02	0.03	0.65	1	0.42
High incidence of WNV disease					
Intercept	-2.86	0.92	9.55	1	0.00
Temperature in March (°C)	0.05	0.02	5.63	1	0.01
Eastern Colorado					
Presence of WNV disease					
Intercept	-1.57	3.91	0.02	1	0.69
Precipitation in April (mm)	0.06	0.06	5.62	1	0.02
Precipitation in July (mm)	-0.04	-0.04	7.06	1	0.01
Heating degree days in July (HDD)	-0.02	-0.02	9.90	1	0.00
NDVI in May	0.02	0.02	0.62	1	0.43
High incidence of WNV disease					
Intercept	3.65	6.32	0.033	1	0.56
Elevation (m)	-0.01	0.01	10.10	1	0.00
Precipitation in July (mm)	-0.06	0.02	8.74	1	0.00
Heating degree days in August (HDD)	0.03	0.06	3.12	1	0.08
Snowfall in September (mm)	-0.07	0.02	8.28	1	0.00
NDVI in May	0.05	0.03	2.79	1	0.10

Table III-3. Validation of zip code scale models

Zip code scale-based validation of models for predicted presence or high incidence of West Nile virus disease for western and eastern Colorado based upon the internal build data set and external validation data sets.

Model classification	Actual classification for presence of WNV disease cases or high WNV disease incidence					
	Yes	No	% correct[a]	Yes	No	% correct[a]
Internal validation	**Western Colorado**			**Eastern Colorado**		
Predicted presence of WNV disease cases[b]						
Yes	21	5	80.77%	158	24	86.81%
No	5	13	72.22%	24	20	45.45%
% correct[c]	80.77%	72.22%		86.81%	45.45%	
Predicted high incidence of WNV disease cases[d]						
Yes	5	6	45.45%	33	24	57.89%
No	6	27	81.82%	24	145	85.80%
% correct[c]	45.45%	81.82%		57.89%	85.80%	
External validation						
Predicted presence of WNV disease cases[b]						
Yes	7	3	70.00%	53	8	86.89%
No	2	3	60.00%	6	8	57.14%
% correct[c]	77.78%	50.00%		89.83%	50.00%	
Predicted high incidence of WNV disease cases[d]						
Yes	4	1	80.00%	10	5	66.67%
No	2	8	80.00%	7	53	88.33%
% correct[c]	66.67%	88.89%		58.82%	91.38%	

[a] User accuracy (commission error).

[b] Probability cut-off value of $P \geq 0.50$ for western Colorado and $P \geq 0.70$ for eastern Colorado.

[c] Producer accuracy (omission error).

[d] Probability cut-off value of $P \geq 0.42$ for western and eastern Colorado.

Table III-4. Fine-scale validation of models

Fine-scale (2 x 2 km) validation of models for predicted presence or high incidence of WNV disease for western and eastern Colorado based upon the internal build data set and external validation data sets.

Model classification	Actual classification for presence of WNV disease cases or high WNV disease incidence					
	Yes	No	% correct[a]	Yes	No	% correct[a]
Internal validation	**Western Colorado**			**Eastern Colorado**		
Predicted presence of WNV disease cases[b]						
Yes	21	5	80.77%	158	24	86.81%
No	5	13	72.22%	24	20	45.45%
% correct[c]	80.77%	72.22%		86.81%	45.45%	
Predicted high incidence of WNV disease cases[d]						
Yes	5	6	45.45%	38	19	66.67%
No	6	27	81.82%	19	83	81.37%
% correct[c]	45.45%	81.82%		66.67%	81.37%	
External validation						
Predicted presence of WNV disease cases[b]						
Yes	8	2	80.00%	52	9	85.25%
No	1	4	80.00%	7	5	41.67%
% correct[c]	88.89%	66.67%		88.14%	35.71%	
Predicted high incidence of WNV disease cases[d]						
Yes	4	1	80.00%	10	6	62.50%
No	2	8	80.00%	7	24	77.42%
% correct[c]	66.67%	88.89%		58.82%	80.00%	

[a] User accuracy (commission error).

[b] Probability cut-off value of $P \geq 0.50$ for western Colorado and $P \geq 0.70$ for eastern Colorado.

[c] Producer accuracy (omission error).

[d] Probability cut-off value of $P \geq 0.42$ for western and eastern Colorado.

Table III-5. Validation of zip code scale models in other regions of Colorado
Validation of zip code scale models for predicted presence or high incidence of WNV disease developed for eastern or western Colorado when applied to the other part of the state.

	Actual classification for presence of WNV disease cases or high WNV disease incidence					
Model classification	**Eastern models applied to western Colorado**			**Western models applied to eastern Colorado**		
	Yes	No	% correct[a]	Yes	No	% correct[a]
Predicted presence of WNV disease cases[b]						
Yes	23	12	65.71%	203	38	84.23%
No	12	12	50.00%	38	22	36.67%
% correct[c]	65.71%	50.00%		84.23%	36.67%	
Predicted high incidence of WNV disease cases[d]						
Yes	7	10	41.18%	19	55	25.68%
No	10	32	76.19%	55	172	75.77%
% correct[c]	41.18%	76.19%		25.68%	75.77%	

Models developed in eastern Colorado were validated against all 59 western zip codes and models developed in western Colorado against all 301 eastern zip codes.

[a] User accuracy (commission error).

[b] Probability cut-off value of $P \geq 0.78$ for western Colorado and $P \geq 0.71$ for eastern Colorado.

[c] Producer accuracy (omission error).

[d] Probability cut-off value of $P \geq 0.460$ for western Colorado and $P \geq 0.34$ for eastern Colorado.

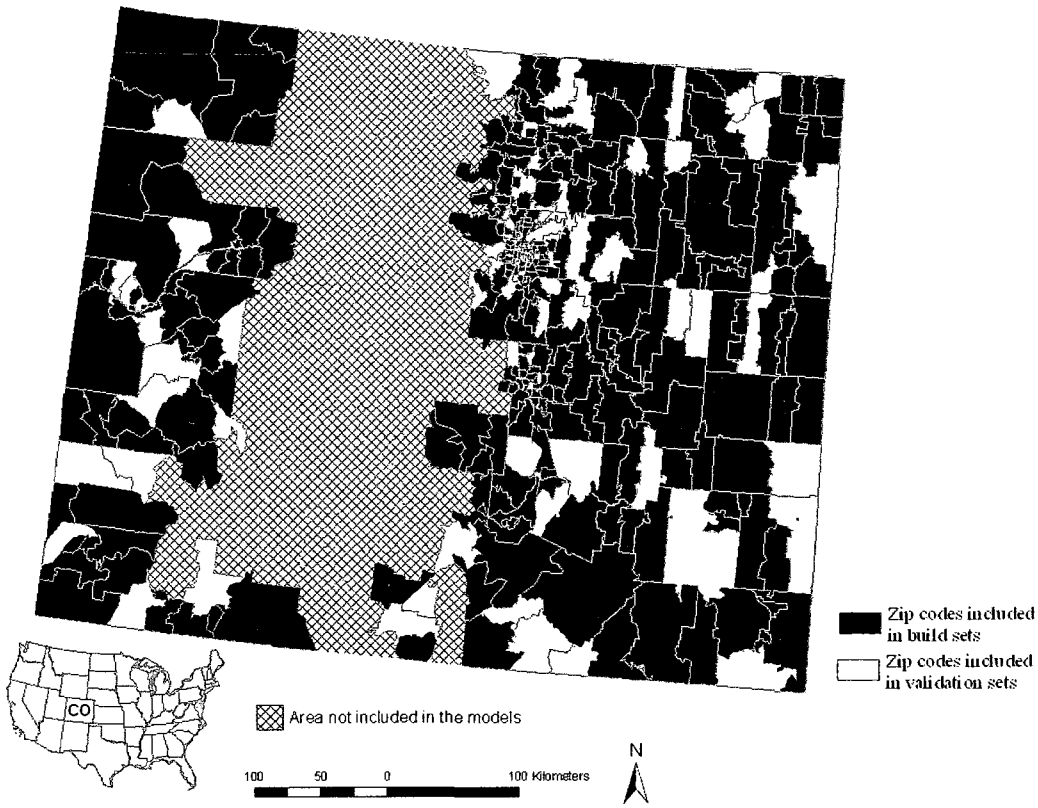

Figure III-1. Zip codes used to build and validate models
Distribution of zip codes in western and eastern Colorado used to build and validate models for predicted presence and high incidence of West Nile virus disease. The central part of the state, which is dominated by high mountains where Culex vectors for West Nile virus are rare or lacking, was excluded from the models.

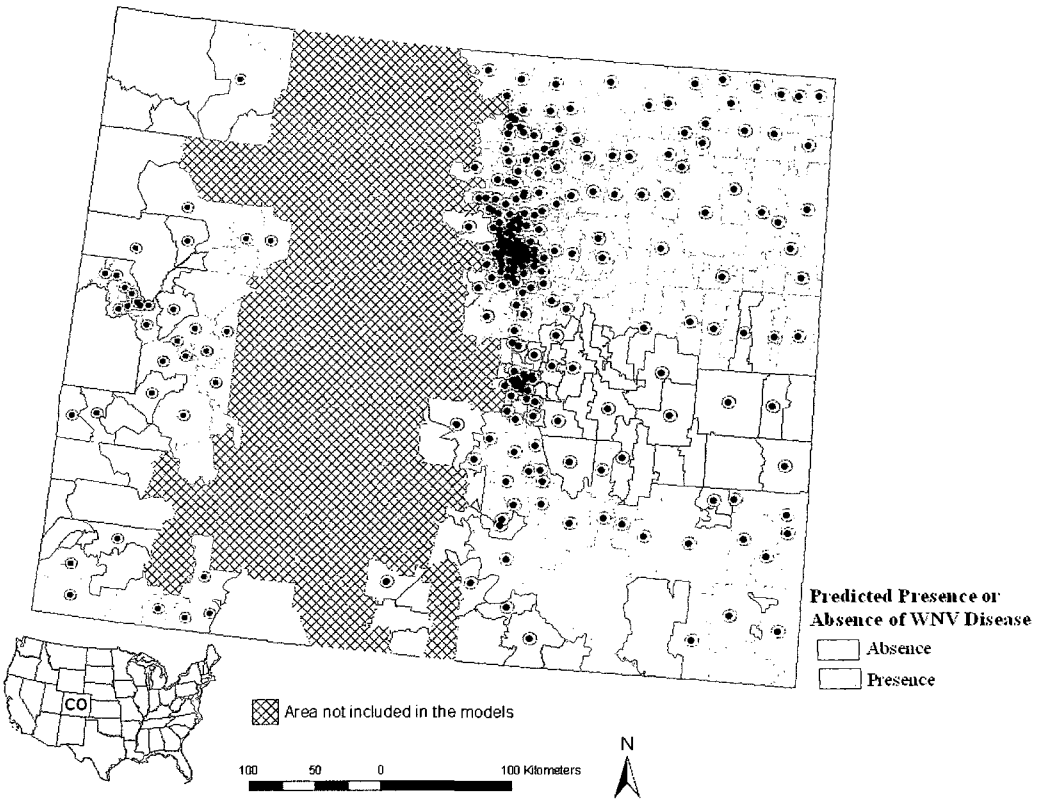

Figure III-2. Predicted presence of West Nile virus disease by zip code
Predicted presence of WNV disease in western and eastern Colorado for the zip code scale-based model. Zip codes with WNV disease cases reported during 2002--2006 are indicated by black circles.

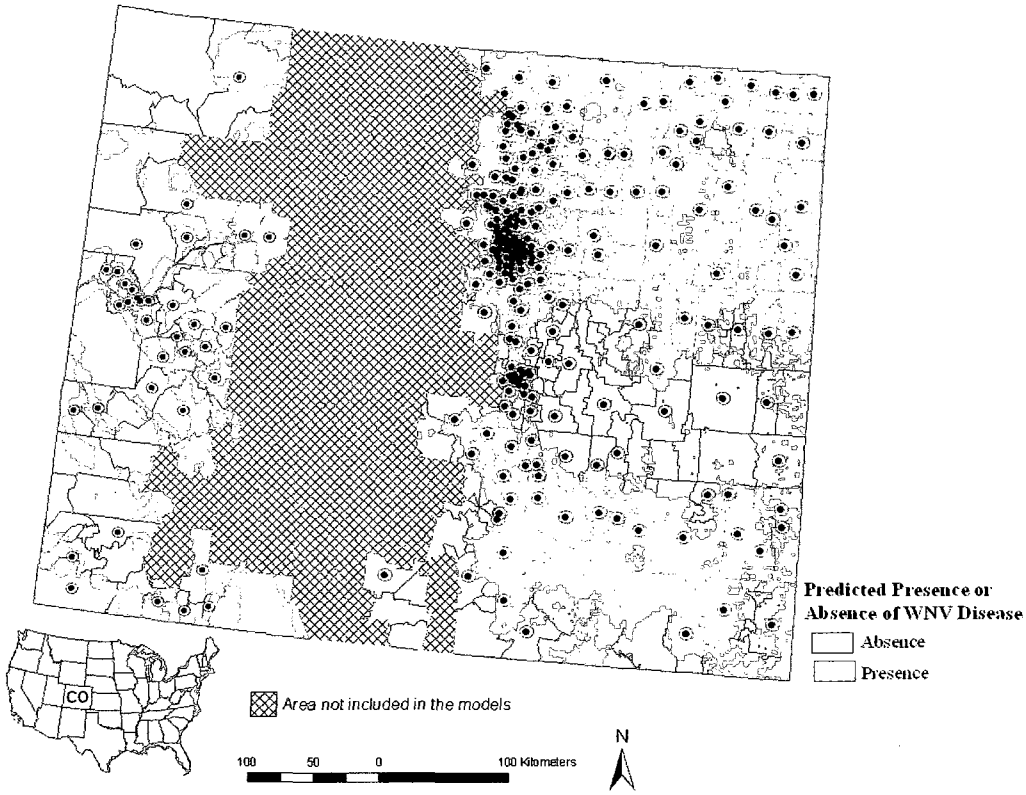

Figure III-3. Predicted presence of West Nile virus disease for the fine-scale model
Predicted presence of WNV disease in western and eastern Colorado for the fine-scale (2 x 2 km) model. Zip codes with WNV disease cases reported during 2002--2006 are indicated by black circles.

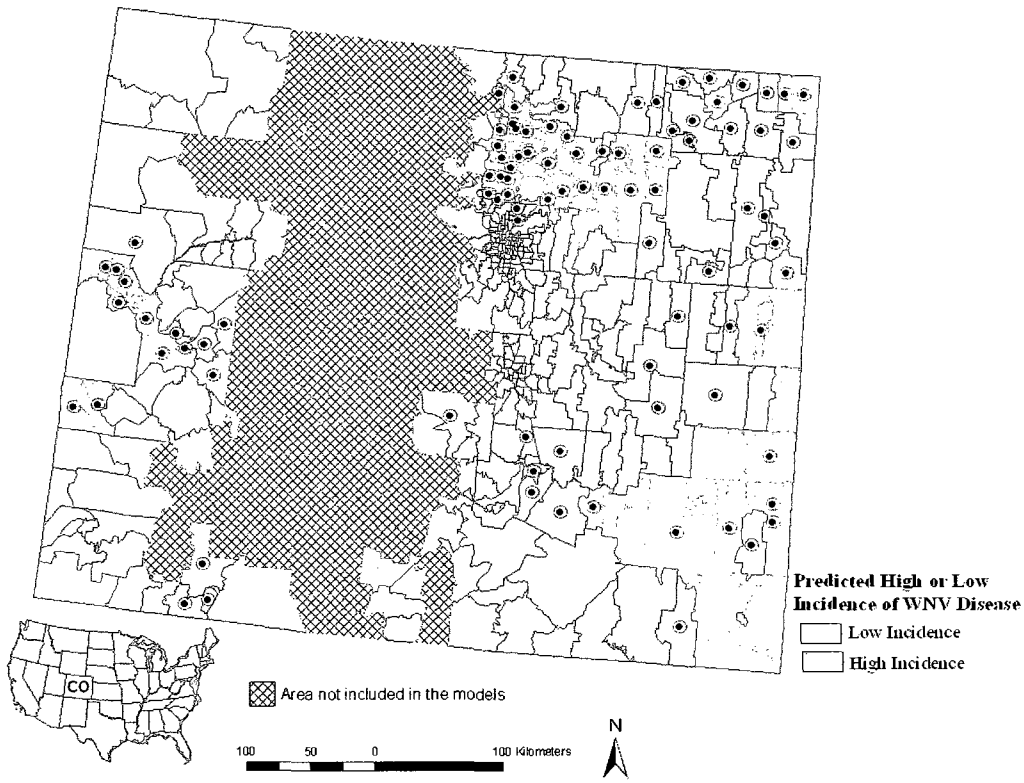

Figure III-4. Predicted high incidence of West Nile virus disease in Colorado – zip code
Predicted presence of high cumulative incidence of WNV disease in western Colorado (> 25.75 cases per 100,000 person-years) and eastern Colorado (> 38.39 cases per 100,000 person-years) from 2002--2006 for the zip code scale-based model. The high incidence cut-offs were chosen based upon the fourth quartile cut-off value for zip code-based incidence in either region. Zip codes reporting high incidences of WNV disease during 2002--2006 are indicated by black circles.

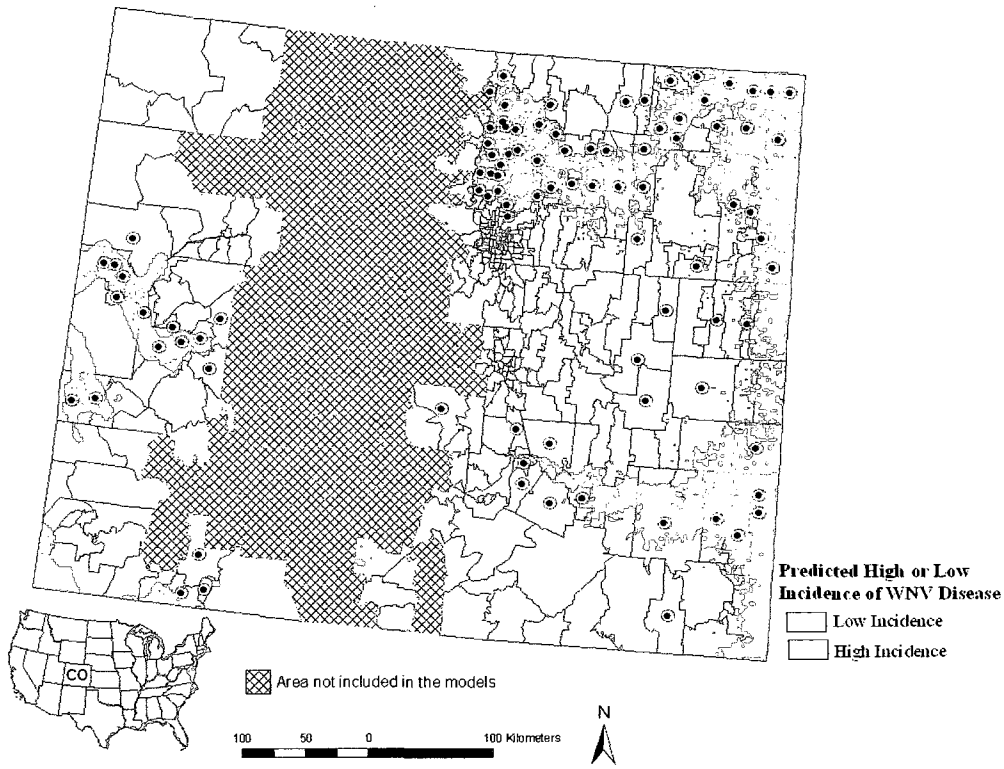

Figure III-5. Predicted high incidence of West Nile virus disease in Colorado – fine scale
Predicted presence of high cumulative incidence of WNV disease in western Colorado (> 25.75 cases per 100,000 person-years) and eastern Colorado (> 38.39 cases per 100,000 person-years) during 2002--2006 for the fine-scale (2 x 2 km) model. The high incidence cut-offs were chosen based upon the fourth quartile cut-off value for zip code–based incidence in either region. Zip codes reporting high incidences of WNV disease during 2002--2006 are indicated by black circles.

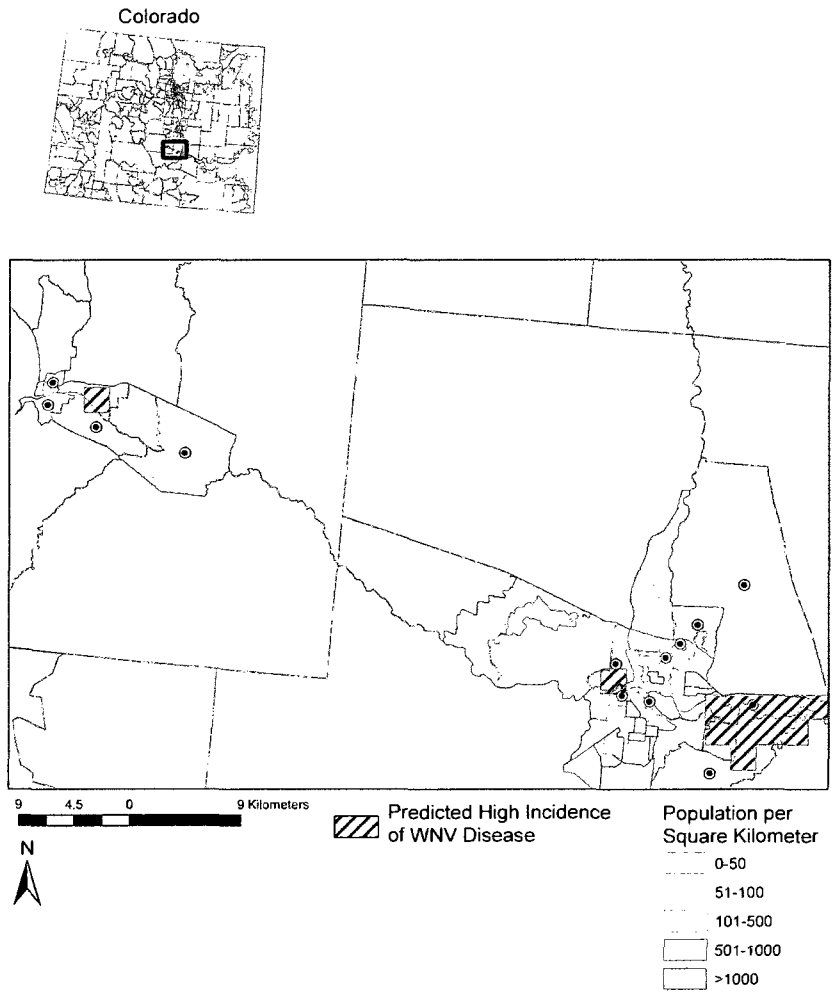

Figure III-6. Predicted high incidence and population centers
Along the portion of the Arkansas River where the raster-based model under-predicted disease risk (see inset map), we found that small patches of model-predicted high incidence of WNV disease commonly coincided with population centers (e.g., Pueblo, Colorado). Population density per square kilometer is displayed at the census tract level and census tracts with high WNV disease incidence during 2002--2006 are indicated by black circles.

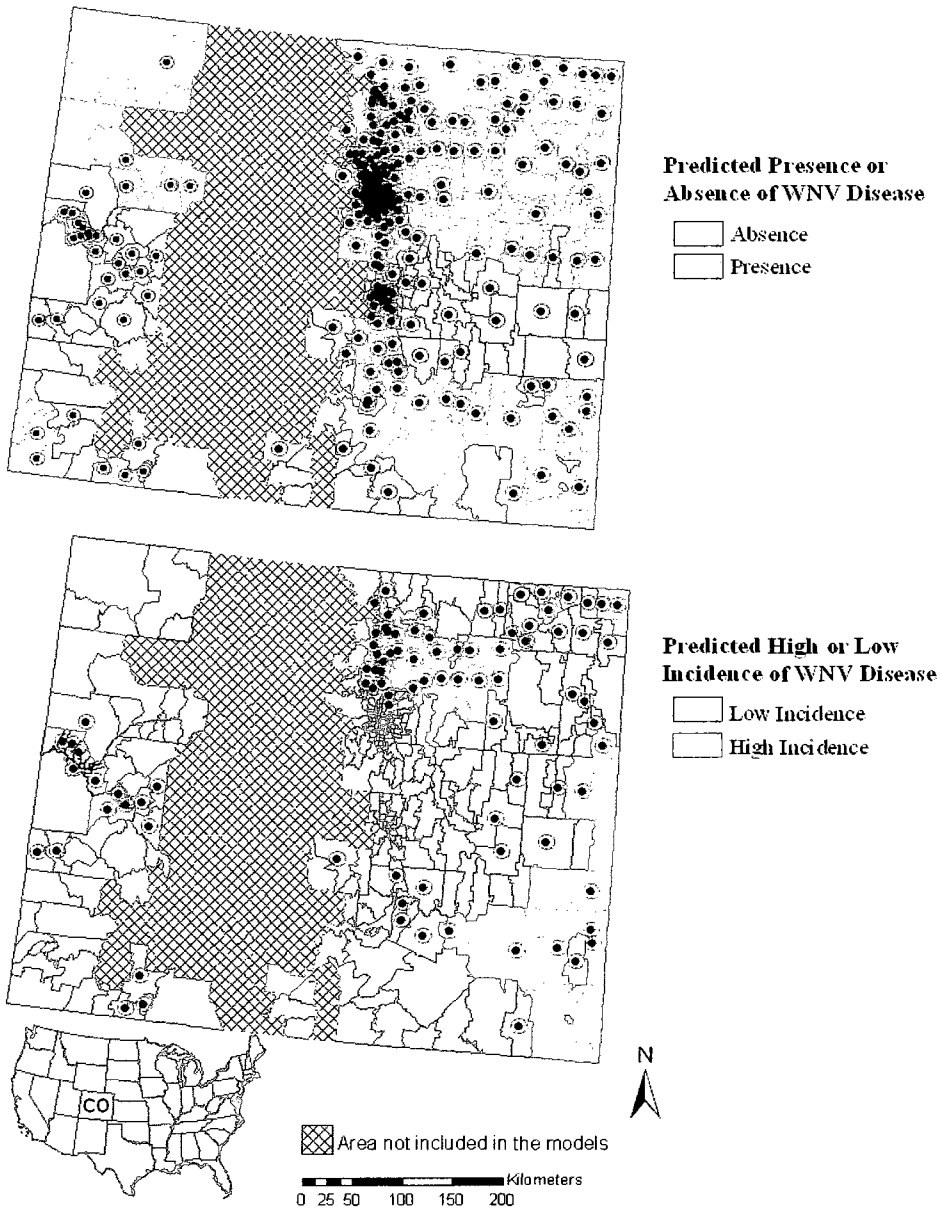

Figure III-7. Statewide application of the eastern zip code model
Outcome for the eastern zip code-based model for predicted presence and predicted high incidence of WNV disease when applied statewide. Zip codes with WNV disease cases reported (top map) or with high WNV disease incidence (bottom map) during 2002--2006 are indicated by black circles.

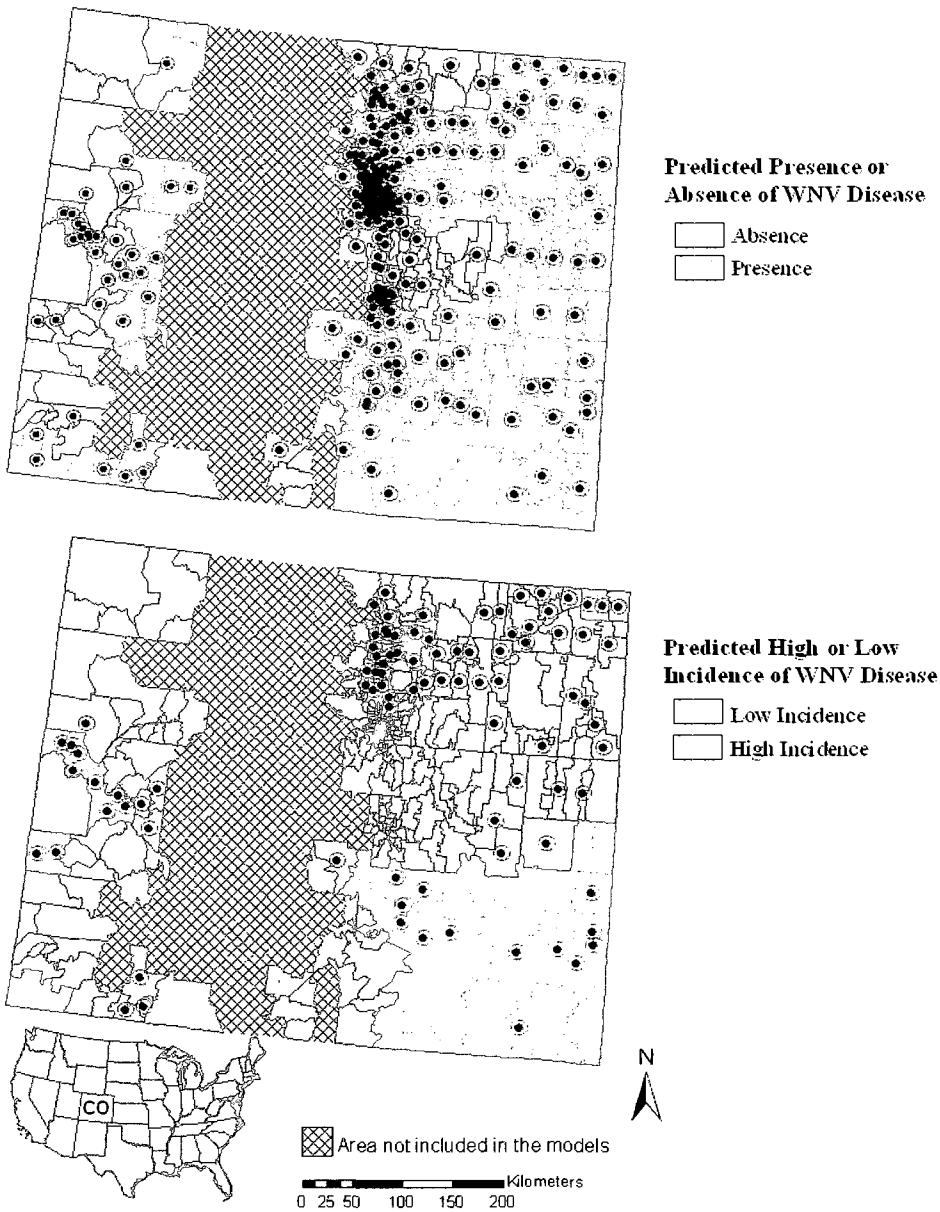

Figure III-8. Statewide application of the western zip code model
Outcome for the western zip code-based model for predicted presence and predicted high incidence of WNV disease when applied statewide. Zip codes with WNV disease cases reported (top map) or with high WNV disease incidence (bottom map) during 2002--2006 are indicated by black circles.

References Cited

1. Nash D, Mostashari F, Fine A, Miller J, O'Leary D, Murray K, Huang A, Rosenberg A, Greenberg A, Sherman M, Wong S, Layton M: The outbreak of West Nile virus infection in the New York City area in 1999. N Engl J Med 2001; 344(24): 1807-14.

2. Hayes EB, Komar N, Nasci RS, Montgomery SP, O'Leary DR, Campbell GL: Epidemiology and transmission dynamics of West Nile virus disease. Emerg Infect Dis 2005; 11: 1167-1173.

3. West Nile Virus. 2007 (August 8) http://www.cdphe.state.co.us/dc/zoonosis/wnv/.

4. Turell MJ, Dohm DJ, Sardelis MR, Oguinn ML, Andreadis TG, Blow JA: An update on the potential of north American mosquitoes (Diptera: Culicidae) to transmit West Nile Virus. J Med Entomol 2005; 42(1): 57-62.

5. Goddard LB, Roth AE, Reisen WK, Scott TW: Vector competence of California mosquitoes for West Nile Virus. Emerg Infect Dis 2002; 8(12): 1385-1391.

6. Reisen W, Lothrop H, Chiles R, Madon M, Cossen C, Woods L, Husted S, Kramer V, Edman J: West Nile virus in California. Emerg Infect Dis 2004; 10(8): 1369-1378.

7. Monath TP: Prospects for development of a vaccine against the West Nile virus. Ann N Y Acad Sci 2001; 951: 1-12.

8. Gould LH, Fikrig E: West Nile virus: a growing concern? J Clin Invest 2004; 113(8): 1102-7.

9. Mason PW, Shustov AV, Frolov I: Production and characterization of vaccines based on flaviviruses defective in replication. Virology 2006; 351(2): 432-43.

10. Granwehr BP, Lillibridge KM, Higgs S, Mason PW, Aronson JF, Campbell GA, Barrett AD: West Nile virus: where are we now? Lancet Infect Dis 2004; 4(9): 547-56.

11. Petersen LR, Marfin AA: West Nile virus: a primer for the clinician. Ann Intern Med 2002; 137(3): 173-9.

12. Sampathkumar P: West Nile virus: Epidemiology, clinical presentation, diagnosis, and prevention. Mayo Clinic Proceedings 2003; 78(9): 1137-1144.

13. Herrington J: Pre-West Nile virus outbreak: perceptions and practices to prevent mosquito bites and viral encephalitis in the United States. Vector Borne Zoonotic Dis 2003; 3(4): 157-73.

14. Wegbreit J, Reisen WK: Relationships among weather, mosquito abundance, and encephalitis virus activity in California: Kern County 1990-98. J Am Mosq Control Assoc 2000; 16(1): 22-27.

15. Reisen WK, Lothrop HD, Hardy JL: Bionomics of *Culex tarsalis* (Diptera: Culicidae) in relation to arbovirus transmission in southeastern California. J Med Entomol 1995; 32(3): 316-27.

16. Reisen WK: Effect of temperature on *Culex tarsalis* (Diptera: Culicidae) from the Coachella and San Joaquin Valleys of California. J Med Entomol 1995; 32(5): 636-645.

17. Reisen WK, Fang Y, Martinez VM: Effects of temperature on the transmission of West Nile virus by *Culex tarsalis* (Diptera: Culicidae). J Med Entomol 2006; 43: 309-317.

18. Reisen WK, Lothrop HD, Presser SB, Hardy JL, Gordon EW: Landscape ecology of arboviruses in southeastern California - temporal and spatial patterns of enzootic activity in Imperial Valley, 1991-1994. J Med Entomol 1997; 34(2): 179-188.

19. Reisen WK, Lothrop HD, Presser SB, Milby MM, Hardy JL, Wargo MJ, Emmons RW: Landscape ecology of arboviruses in southern California - temporal and spatial

patterns of vector and virus activity in Coachella Valley, 1990-1992. J Med Entomol 1995; 32(3): 255-266.

20. Bolling BG, Moore CG, Anderson SL, Blair CD, Beaty BJ: Entomological studies along the Colorado Front Range during a period of intense West Nile virus activity. J Am Mosq Control Assoc 2007; 23(1): 37-46.

21. Smith GC, Moore CG, Davis T, Savage HM, Thapa AB, Shrestha SL, Karabatsos N: Arbovirus surveillance in northern Colorado, 1987 and 1991. J Med Entomol 1993; 30: 257-261.

22. Eidson M, Kramer L, Stone W, Hagiwara Y, Schmit K: Dead bird surveillance as an early warning system for West Nile virus. Emerging Infectious Diseases 2001; 7(4): 631-635.

23. Eidson M, Komar N, Sorhage F, Nelson R, Talbot T, Mostashari F, McLean R: Crow deaths as a sentinel surveillance system for West Nile virus in the northeastern United States, 1999. Emerging Infectious Diseases 2001; 7(4): 615-620.

24. Mostashari F, Kulldorff M, Hartman JJ, Miller JR, Kulasekera V: Dead bird clusters as an early warning system for West Nile virus activity. Emerging Infectious Diseases 2003; 9(6): 641-646.

25. Julian KG, Eidson M, Kipp AM, Weiss E, Petersen LR, Miller JR, Hinten SR, Marfin AA: Early season crow mortality as a sentinel for West Nile virus disease in humans, northeastern United States. Vector Borne Zoonotic Dis. 2002; 2(3): 145-155.

26. Guptill SC, Julian KG, Campbell GL, Price SD, Marfin AA: Early-season avian deaths from West Nile virus as warnings of human infection. Emerging Infectious Diseases 2003; 9(4): 483-484.

27. Nemeth NM, Beckett S, Edwards E, Klenk K, Komar N: Avian mortality surveillance for West Nile virus in Colorado. Am J Trop Med Hyg 2007; 76(3): 431-7.

28. Brownstein JS, Rosen H, Purdy D, Miller JR, Merlino M, Mostashari F, Fish D: Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. Vector Borne Zoonotic Dis 2002; 2(3): 157-64.

29. DiMenna MA, Bueno R, Jr.,, Parmenter RR, Norris DE, Sheyka JM, Molina JL, LaBeau EM, Hatton ES, Glass GE: Emergence of West Nile virus in mosquito (Diptera: Culicidae) communities of the New Mexico Rio Grande Valley. J. Med. Entomol. 2006; 43: 594 -599

30. DiMenna MA, Bueno R, Jr., Parmenter RR, Norris DE, Sheyka JM, Molina JL, Labeau EM, Hatton ES, Roberts CM, Glass GE: Urban habitat evaluation for West Nile virus surveillance in mosquitoes in Albuquerque, New Mexico. . J. Am. Mosq. Contr. Assoc. 2007; 23(2): 153-160.

31. Ruiz M, Tedesco C, McTighe T, Austin C, Kitron U: Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. Int. J. Health Geogr. 2004; 3(1): 8.

32. Diuk-Wasser MA, Brown HE, Andreadis TG, Fish D: Modeling the spatial distribution of mosquito vectors for West Nile virus in Connecticut, USA. Vector Borne Zoonotic Dis 2006; 6(3): 283-295.

33. Godsey MS, Jr., Blackmore MS, Panella NA, Burkhalter K, Gottfried K, Halsey LA, Rutledge R, Langevin SA, Gates R, Lamonte KM, Lambert A, Lanciotti RS, Blackmore CG, Loyless T, Stark L, Oliveri R, Conti L, Komar N: West Nile virus epizootiology in the southeastern United States, 2001. Vector Borne Zoonotic Dis 2005; 5(1): 82-9.

34. White DJ: Vector surveillance for West Nile virus. Ann NY Acad Sci 2001; 951: 74-83.

35. Climate of Colorado. 2007 (August) http://www.wrcc.dri.edu/.

36. Burnham K, Anderson D: Model selection and multimodel inference: A practical information-theoretic approach. New York: Springer, 2002.

37. Congalton R, Green K: Assessing the accuracy of remotely sensed data: principles and practices. Boca Raton: Lewis Publishers, 1999.

38. Higgs S, Schneider BS, Vanlandingham DL, Klingler KA, Gould EA: Nonviremic transmission of West Nile virus. Proceedings of the National Academy of Sciences of the United States of America 2005; 102(25): 8871-4.

39. Prevention CfDCa: West Nile Virus. 2007 (Sept 21) http://www.cdc.gov/ncidod/dvbid/westnile/.

40. Miramontes R, Lafferty WE, Lind BK, Oberle MW: Is agricultural activity linked to the incidence of human West Nile virus? Am J Prev Med 2006; 30(2): 160-3.

41. Kilpatrick AM, Kramer LD, Campbell SR, Alleyne EO, Dobson AP, Daszak P: West Nile virus epidemics in North America are driven by shifts in mosquito feeding behavior. PLoS Biology 2006; 4: 606-610.

42. Tempelis CH, Francy DB, Hayes RO, Lofy MF: Variations in feeding patterns of seven culicine mosquitoes on vertebrate hosts in Weld and Larimer counties, Colorado. Am. J. Trop. Med. Hyg. 1967; 16(1): 111-119.

43. Tempelis CH, Reeves WC, Bellamy RE, Lofy MF: A three-year study of the feeding habits of *Culex tarsalis* in Kern County, California. American Journal of Tropical Medicine and Hygiene 1965; 14: 170-7.

44. Reisen WK, Meyer RP, Presser SB, Hardy JL: Effect of temperature on the transmission of western equine encephalomyelitis and St. Louis encephalitis viruses by *Culex tarsalis* (Diptera: Culicidae). J Med Entomol 1993; 30(1): 151-60.

45. Hagstrum DW, Workman EB: Interaction of temperature and feeding rate in determining the rate of development of larval *Culex tarsalis* (Diptera: Culicidae). Annals of the Entomological Society of America 1971; 64: 668-671.

46. Kramer LD, Hardy JL, Presser SB: Effect of temperature of extrinsic incubation on the vector competence of Culex tarsalis for Western Equine encephalomyelitis virus. American Journal of Tropical Medicine & Hygiene 1983; 32(5): 1130-9.

47. Beehler JW, Mulla MS: Effects of organic enrichment on temporal distribution and abundance of culicine egg rafts. J Am Mosq Control Assoc 1995; 11(2 Pt 1): 167-71.

48. Ward MP, Ramsay BH, Gallo K: Rural cases of equine West Nile virus encephalomyelitis and the normalized difference vegetation index. Vector Borne Zoonotic Dis 2005; 5(2): 181-8.

49. Mostashari F, Bunning ML, Kitsutani PT, Singer DA, Nash D, Cooper MJ, Katz N, Liljebjelke KA, Biggerstaff BJ, Fine AD, Layton MC, Mullin SM, Johnson AJ, Martin DA, Hayes EB, Campbell GL: Epidemic West Nile encephalitis, New York, 1999: results of a household-based seroepidemiological survey. Lancet 2001; 358(9278): 261-4.

50. Gujral IB, Zielinski-Gutierrez EC, LeBailly A, Nasci R: Behavioral risks for West Nile virus disease, northern Colorado, 2003. Emerg Infect Dis 2007; 13(3): 419-425.

51. O'Leary DR, Marfin AA, Montgomery SP, Kipp AM, Lehman JA, Biggerstaff BJ, Elko VL, Collins PD, Jones JE, Campbell GL, Petersen LR: The epidemic of West Nile virus in the United States, 2002. Vector Borne Zoonotic Dis 2004; 4(1): 61-70.

52. Eisen L, Eisen RJ: Need for improved methods to collect and present spatial epidemiologic data for vectorborne diseases. Emerg Infect Dis; 13: 1816-1820.

IV. Combining Mosquito Vector and Human Disease Data for Improved Assessment of Spatial West Nile Virus Disease Risk.

This chapter has been published in the American Journal of Tropical Medicine and Hygiene and is cited as Winters AM, Bolling BG, Beaty BJ, Blair CD, Eisen RJ, Meyer AM, Pape WJ, Moore CG, Eisen L. 2008. Combining mosquito vector and human disease data for improved assessment of spatial West Nile virus disease risk. AJTMH 78(4):654-665.

Anna M. Winters[1], Bethany G. Bolling[1], Barry J. Beaty[1], Carol D. Blair[1], Rebecca J. Eisen[2], Andrew M. Meyer[1], W. John Pape[3], Chester G. Moore[1], and Lars Eisen[1]

[1]Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523; [2]Division of Vector-Borne Infectious Diseases, National Center for Vector-Borne, Zoonotic and Enteric Diseases, Centers for Disease Control and Prevention, Fort Collins, CO 80521; [3]Colorado Department of Public Health and Environment, Denver, CO 80246

Abstract

Assessments of spatial risk of exposure to vector-borne pathogens that combine vector and human disease data are needed for areas encompassing large tracts of public land with low population bases. We addressed this need for West Nile virus (WNV) disease in the northern Colorado Front Range by developing not only a spatial model for entomological risk of exposure to *Culex tarsalis* WNV vectors and an epidemiological risk map for WNV disease but also a novel risk classification index combining data for these independently derived measures of entomological and epidemiological risk. Risk of vector exposure was high in the densely populated eastern plains portion of the Front Range but low in cooler montane areas to the west that are sparsely populated but used heavily for recreation in the summer. The entomological risk model performed well when applied to the western, mountainous part of Colorado and validated against epidemiologic data.

Key words: *Culex tarsalis*, Climate, Colorado, GIS, Risk, West Nile virus disease

Introduction

West Nile virus (WNV) disease has emerged as a serious problem in the Central Plains and Rocky Mountains with epidemics causing hundreds to thousands of reported human cases in Colorado, Montana, Nebraska, Wyoming, and the Dakotas in 2003 and in Idaho and Utah in 2006 (data from the Centers for Disease Control and Prevention; http://www.cdc.gov/ncidod/dvbid/westnile). In 2003, Colorado alone reported 2,947 cases of human disease with 63 fatalities (data from the Colorado Department of Public Health and Environment; http://www.cdphe.state.co.us/dc/zoonosis/wnv/). The northern part of the Colorado Front Range, which spans the transition zone from the Central Plains to the Rocky Mountains, was one of the focal points of this epidemic with three counties (Larimer, Boulder, Jefferson) accounting for 1,124 reported WNV cases in 2003 and 187 additional cases during 2004-2006. The case load for this 3-county area notably exceeds that of most eastern states.

Spatial patterns of risk of human exposure along the northern Colorado Front Range to locally occurring WNV vectors such as *Culex pipiens* and, especially, *Cx. tarsalis*[1, 2, 3, 4, 5, 6, 7] are intriguing because this area: 1) extends from the western edge of the Central Plains into the eastern edge of the Rocky Mountains and encompasses habitats ranging from prairie through foothills shrub, montane and subalpine forest, and to alpine areas; 2) includes dramatic elevation-climate gradients potentially affecting the ability of *Culex* mosquitoes to establish breeding populations; and 3) exhibits strong spatial clustering of the human population to the plains habitat in the east.

Culex tarsalis is, based on its high vector efficiency[1, 2], common occurrence[4, 5, 6, 8, 9, 10, 11], and willingness to bite mammals[7, 12, 13] considered a primary vector of WNV to humans in Colorado and indeed the western United States. Along the northern Colorado Front Range, this mosquito occurs commonly in heavily populated plains areas at low elevations (e.g., the cities of Boulder, Loveland, and Fort Collins)[4, 5, 6]. There is, however, a lack of detailed knowledge regarding the distribution and abundance of *Cx. tarsalis* in montane areas to the west. These montane areas include large tracts of public land (e.g., Rocky Mountain National Park, Roosevelt National Forest) used heavily in the summer for recreational activities by the Front Range population as well as tourists. Information regarding *Cx. tarsalis* in montane areas of the western United States typically have been qualitative and restricted to records of collection sites[14, 15, 16]. Because montane areas in the northern Colorado Front Range and, indeed, many other parts of the western United States are dominated by public access lands and have very low population bases, epidemiologic data are of limited use to assess risk of WNV exposure and need to be complemented with models for entomological risk of exposure to *Cx. tarsalis* and other potential WNV vectors.

The primary goal of this study was to create a comprehensive spatial risk model, combining entomological risk of exposure to *Cx. tarsalis* and epidemiological risk of WNV exposure, for a 3-county area in the northern Colorado Front Range extending from prairie landscapes at the western edge of the Central Plains into montane and subalpine/alpine habitats at the eastern edge of the Rocky Mountains. A secondary goal was to explore how the entomological risk model performed, compared to epidemiologic data, when scaled-up from the ecologically and climatically diverse Larimer County

model development area to four different regions of Colorado: eastern plains, northern Colorado Front Range, southern Colorado Front Range, western mountains and high plateau.

Materials and Methods

Development of entomological risk model: Field sampling of Culex tarsalis

Fieldwork targeted two habitat-climate-elevation gradients in the central and southern parts of Larimer County: the Poudre River and Big Thompson River corridors (Fig. 1). The general climate in this area is characterized by cold winters and hot summers with low humidity, averaging ≈ 400 mm of precipitation per year at lower elevations (Mountain States Weather Services, Fort Collins, CO). Mosquito sampling was conducted on 5-7 occasions (roughly every two weeks in each site) during 20 June - 13 September 2006. This time period includes the peak seasonal activity by *Cx. tarsalis* in Larimer County[4, 5]. Sampling included 10 sites located along each river corridor (total of 20 sites) at elevations ranging from ≈ 1,500-2,350 m (Fig. 1). Mosquitoes were collected using CO_2-baited CDC miniature light traps (John W. Hock Company, Gainesville, FL) suspended ≈ 1.5 m above the ground and operated from afternoon (1500-1700 hours) until morning (0800-1000 hours). Each sampling site held two traps baited with ≈ 1 kg of dry ice and located directly along the aforementioned rivers or their tributaries. Trees lining the sampling sites included cottonwood (*Populus* spp.), willow (*Salix* spp.), and, at higher elevations, aspen (*Populus tremuloides*). For sites located along the Poudre River, we also determined temperature and relative humidity, at ≈ 1 m

above the ground on the north-facing side of a tree, every 30 min throughout the study period using HOBO H8 Pro series loggers (Onset Computer Corporation, Pocasset, MA).

Collected mosquitoes were examined with a dissecting microscope and identified to species using published keys[15, 17]. Site-specific mean abundances of *Cx. tarsalis* per trap night were calculated for the period 20 June - 13 September 2006; this included data from 10-14 trap nights in each site. Mean abundance over a prolonged time period was used to minimize the effects of: (1) variable weather conditions on different trapping nights on mosquito activity and (2) differences in the seasonal timing of peak abundance of *Cx. tarsalis* along the elevation-climate gradients examined.

Development of entomological risk model: Geographic Information System database

Site locations were mapped with a GPS receiver (Trimble Geo XT; Trimble Corp., Sunnyvale, CA). Site-specific 30-yr mean climate data (mean monthly and annual minimum, mean, and maximum temperature; mean monthly and annual cooling, heating, and growing degree days; mean monthly and annual precipitation, snowfall, and relative humidity; mean and median annual length of freeze-free period; median Julian date of first and last snowfall) were derived from Geographic Information System (GIS)-based data for 1961-1990 (2x2 km spatial resolution; Climate Source LLC, Corvallis, OR) using ArcGIS9.2 (ESRI, Redlands, CA). Climate data layers were generated by Climate Source LLC using the PRISM modeling system and Gaussian filter resolution enhancement. Site-specific data on elevation were derived from the U.S. Geological Survey 30-meter digital national elevation data-set. Selected environmental site characteristics are provided in Table 1.

Development of entomological risk model: Model construction

Associations between site-specific long-term climate data (annual and monthly data for 1961-1990) and mean abundance of *Cx. tarsalis* per trap night during 20 June - 13 September 2006 (hereinafter referred to as abundance of *Cx. tarsalis*) were explored by multivariate regression models based on data from 13 sites yielding at least 0.2 mosquitoes per trap night. This excluded seven sites located at elevations above 1,950 m and found to yield no or very few mosquitoes (0-0.1 per trap night). Long-term climate data are useful for recognizing relative differences in climatic conditions between spatial locations and thus can serve as the basis for spatial modeling. Our use of climate data from 1961-1990 and mosquito data from 2006 was based on the assumption that differences in climatic conditions between sampling sites in 2006 were consistent with the differences for the long-term averages. This indeed was the case for one key climate factor available both as temperature logger-derived data from 2006 and long-term 1961-1990 GIS-derived data; site-specific mean June-August temperatures in 2006 were strongly correlated with 1961-1990 averages for these months in the same sites (pairwise correlation coefficient = 0.891, n = 8, $P = 0.003$). Use of long-term climate data available as spatially continuous GIS data layers also enabled development of spatially continuous predictive models for mosquito abundance based on a far wider range of potentially important climate factors than those available for the mosquito sampling year of 2006.

Elevation, climate factors, and mosquito abundance (natural log transformed data) were all normally distributed (Shapiro-Wilk test; $P > 0.05$ in all cases). Results for selected environmental variables in univariate tests are shown in Table 2. A model to predict abundance of *Cx. tarsalis* (dependent variable) was constructed using a forward

stepwise multiple regression approach including five biologically meaningful covariates (Table 2; mean temperature for April-September; cumulative cooling degree days for June-August; mean relative humidity for March-May; mean precipitation for January-February; snowfall in April) and using a probability to enter covariates of 0.25.

The equation for the resulting model (which was based on the relationship between cooling degree days and abundance of *Cx. tarsalis*) was applied, at a 2x2 km spatial resolution, to the cooling degree day data layer using the raster calculator of ArcGIS9.2 in order to create a continuous data layer predictive of abundance of *Cx. tarsalis*. The inclusion of Boulder and Jefferson counties in this model is justified because ranges for the single covariate included in the model (cooling degree days) are similar in these counties to the range occurring in the Larimer County model development area; the minimum cooling degree day value was identical for all three counties and the maximum values for Boulder and Jefferson counties exceeded that for Larimer County by no more than 20% (Table 3). Further, the areas of Boulder and Jefferson counties with cooling degree day values exceeding the maximum for Larimer County were only 17% and 14% of the total areas of these counties, respectively.

The model for projected entomological risk of exposure to *Cx. tarsalis* was categorized, at a spatial resolution of 2x2 km, as follows: low (< 1 mosquito per trap night), moderate (1-5 mosquitoes per trap night), or high (> 5 mosquitoes per trap night). This cooling degree day-based model for abundance of *Cx. tarsalis* does not, however, account for presence of open, dry habitats with low suitability for *Cx. tarsalis* within areas classified as moderate or high risk. We therefore created an analysis mask based on GIS data layers for presence of surface water (streams, rivers, irrigation canals, lakes,

ponds) obtained from the U.S. Geological Survey National Hydrography Dataset and landcover from the Colorado Gap analysis project. The mask was designed to uniformly include non-irrigated areas (natural habitats or non-irrigated cattle grazing land) at distances exceeding a given cut-off value (hereinafter referred to as buffer distance) from perceived larval habitat into the lowest category of entomological risk (< 1 mosquito per trap night). *Culex tarsalis* can disperse over large distances in irrigated agricultural settings in California[10, 18, 19, 20] but empirical data for dispersal of adults from larval habitats in the semi-arid landscapes of Colorado have been lacking. We initially considered four different potential buffer distances: 100, 250, 500 and 1,000 m. However, the 100 m and 250 m buffer distances proved unrealistic as preliminary data collected from three separate sites in southeastern Larimer County (Butterfly Woods Natural Area, Dixon Reservoir, Douglass Reservoir) during 28 June – 12 July 2007 indicate that abundance of *Cx. tarsalis* can remain high (> 5 mosquitoes per trap night) at distances up to 250 m from larval habitat detectable from GIS-based data layers (data not shown). Spatial models predicting moderate or high entomological risk of exposure to *Cx. tarsalis* were therefore developed using a conservative buffer distance of 500 m around perceived larval habitat (streams, rivers, irrigation canals, lakes, ponds, irrigated agricultural land).

Development of epidemiological risk maps

The epidemiological risk map for the Larimer-Boulder-Jefferson area was based on data for WNV disease cases reported to the Colorado Department of Public Health and Environment from 2002-2006 and with a known census tract of residence; this included 618 cases for Larimer County, 514 cases for Boulder County, and 181 cases for Jefferson County. Census tract population data used were from 2004 (data provided by ESRI and

based on information from the U.S. Census Bureau). Cumulative incidence (hereinafter referred to as incidence) of WNV disease per 100,000 person-years, 2002-2006, was calculated and displayed by census tract during using ArcGIS9.2. Epidemiological risk similarly was mapped for the entire state of Colorado; this included data for 3,749 WNV disease cases occurring during 2002-2006 and with a known census tract of residence.

Comparison of entomological and epidemiological risk

The Spatial Analyst in ArcGIS9.2 was used to calculate the percentage of each census tract in Larimer-Boulder-Jefferson covered by area representing moderate (> 1 *Cx. tarsalis* per trap night) or high (> 5 *Cx. tarsalis* per trap night) projected entomological risk of exposure to *Cx. tarsalis*. We also created a 100x100 m grid and calculated average straight line (Euclidean) distances from the center of a census tract to the nearest area with moderate or high projected entomological risk (distance = 0 m if such areas were contained within the census tract). Thereafter, associations were explored between percentage of a census tract classified as area with moderate or high projected entomological risk or average distance to the nearest area with moderate or high projected entomological risk and WNV disease incidence. Because results for moderate and high entomological risk were similar, only data for comparisons between high entomological risk and WNV disease incidence are presented here.

The potential for applying the entomological risk model beyond the Larimer-Boulder-Jefferson area was evaluated by determination of the strength of the association between percentage of census tract classified as area with high projected entomological risk and WNV disease incidence for four different regions of Colorado: the eastern plains, the northern Colorado Front Range, the southern Colorado Front Range, and the

western mountains and high plateau (see Fig. 6 for a map showing the area included in each region). Statistical analyses were carried out using the JMP® statistical package[21] and results were considered significant when $P < 0.05$.

Results

Entomological risk model for the Larimer-Boulder-Jefferson area

We used a forward stepwise regression modeling approach to develop a model capable of explaining variability in abundance of *Cx. tarsalis* (Table 1) from climate data with relevance to mosquito biology and available as GIS data layers at a 2x2 km spatial resolution (Table 2; mean temperature for April-September; cumulative cooling degree days for June-August; mean relative humidity for March-May; mean precipitation for January-February; snowfall in April). The resulting model was based on cumulative cooling degree days for June-August (x) (ln *Cx. tarsalis* per trap night = -3.2485 + 0.0122x; ANOVA: $F_{1,11} = 28.22$, adjusted $r^2 = 0.758$, $P < 0.001$). Lack of fit test indicated that a sufficient number of independent variables (n = 1) were included in the model ($P = 0.41$). There was no spatial autocorrelation for either abundance of *Cx. tarsalis* in the sites included in the modeling effort (Moran's I = 0.01, Z[I] = 1.1) or for the residuals of the linear regression of abundance of *Cx. tarsalis* on cooling degree days (I = -0.08, Z[I] = 0).

The equation for the model predicting abundance of *Cx. tarsalis* was applied to GIS-based cooling degree day data clipped to the Larimer-Boulder-Jefferson area and projected entomological risk of exposure to *Cx. tarsalis* was categorized as low (< 1 mosquito per trap night), moderate (1-5 mosquitoes per trap night), or high (> 5

mosquitoes per trap night). Buffers around perceived larval habitat were applied to uniformly include non-irrigated areas located more than 500 m from larval habitat in the lowest risk category. The spatial pattern of areas with high projected entomological risk of exposure to *Cx. tarsalis* in the Larimer-Boulder-Jefferson area is shown in Fig. 2. High risk areas include southeastern Larimer County, eastern Boulder County, and northeastern Jefferson County.

Epidemiological risk map for the Larimer-Boulder-Jefferson area

The epidemiological risk map for the Larimer-Boulder-Jefferson area is shown in Fig. 3 with five risk categories: 0-10, 11-20, 21-40, 41-60, and > 60 WNV disease cases per 100,000 person-years from 2002-2006. Census tracts with WNV disease incidences > 60 cases per 100,000 person-years occurred exclusively in the eastern part of the targeted counties, especially from southeastern Larimer County through eastern Boulder County and into the northeastern edge of Jefferson County. Overall, WNV disease incidence decreased from the plains in eastern Larimer and Boulder counties (typically > 40 WNV cases per 100,000 person-years) to montane areas to the west in these counties or to Jefferson County in the south (most commonly < 20 WNV cases per 100,000 person-years).

A combined entomological and epidemiological risk classification index

We also created a census tract-based risk classification index for combined entomological and epidemiological risk of exposure to vectors and WNV in the Larimer-Boulder-Jefferson area. This novel risk classification index is based on percentage coverage by area with high projected risk of exposure to *Cx. tarsalis* and WNV disease incidence and includes five risk classes: Very Low, Low, Moderate, High, and Very High

(see Table 4 for further description of these risk classes). The risk classification index map, which is shown in Fig. 4, indicates that areas with high to very high risk classification indices cluster from southeastern Larimer County through Boulder County and into northeastern Jefferson County. The western, montane parts of Larimer and Boulder counties and most of Jefferson County were classified as low to very low risk.

Comparison of spatial patterns of entomological and epidemiological risk

We found a strong association between presence of areas with high projected entomological risk of exposure to *Cx. tarsalis* and incidence of WNV disease at the census tract scale in the Larimer-Boulder-Jefferson area. WNV disease incidence was positively associated with percentage coverage (0, 1-33, 34-67, 68-100%) by areas with high projected entomological risk (Fig. 5A; WNV disease incidences based on combined case and population data for all census tracts falling within a given category of entomological risk). Statistical analysis was based on data from individual census tracts and revealed a significant positive correlation between coverage of area by high projected entomological risk and WNV disease incidence (Spearman's rank correlation; $\rho_s = 0.409$, $n = 253$, $P < 0.001$). Further, census tracts containing 68-100% area with high projected entomological risk had higher WNV disease incidence among residents (median of 26.9 cases per 100,000 person-years; n = 132 census tracts) than census tracts either lacking such areas (median of 4.7 cases per 100,000 person-years; n = 41 census tracts) or with 1-33% coverage (median of 11.4 cases per 100,000 person-years; n = 38 census tracts) or 34-67% coverage (median of 10.1 cases per 100,000 person-years; n = 42 census tracts) (Wilcoxon Rank Sum test with chi-square approximation: $\chi^2 \geq 10.68$, df = 1, $P < 0.002$ in all three cases). Finally, WNV disease incidence was higher for census tracts with 34-

67% or 1-33% coverage of areas with high projected entomological risk than for census tract lacking such areas ($\chi^2 \geq 6.73$, df = 1, $P < 0.01$ in both cases).

We also found that WNV disease incidence was negatively associated with average distance to nearest area with high projected entomological risk (Fig. 5B; WNV disease incidences based on combined case and population data for all census tracts falling within a given distance category). Statistical analysis was again based on data from individual census tracts and revealed that WNV disease incidence for census tracts containing areas with high projected entomological risk (median of 17.7 cases per 100,000 person-years; n = 212 census tracts) was four-fold higher than for census tracts lacking such areas (median of 4.7 cases per 100,000 person-years; n = 41 census tracts) ($\chi^2 = 28.60$, df = 1, $P < 0.001$).

Performance of model when applied to different regions of Colorado

We also determined how the entomological risk model performed, compared to epidemiologic data, when scaled-up from the ecologically and climatically diverse Larimer County model development area to four different regions of Colorado: eastern plains, northern Front Range, southern Front Range, western mountains and high plateau (Fig. 6). Overall, the entomological risk model performed well in the northern Front Range and western mountains and high plateau but very poorly in the eastern plains (Fig. 6; WNV disease incidences based on combined case and population data for all census tracts within a region of the state falling within a given category of entomological risk). The percentage of an individual census tract covered by area classified as high projected entomological risk was significantly positively correlated with WNV disease incidence not only for the Larimer-Boulder-Jefferson area ($\rho_s = 0.551$, n = 253, $P < 0.001$) but also

for an expanded portion of the northern Front Range ($\rho_s = 0.356$, n = 427, $P < 0.001$), western Colorado ($\rho_s = 0.740$, n = 151, $P < 0.001$), and the southern Front Range ($\rho_s = 0.292$, n = 75, $P = 0.01$). In contrast, we found a negative correlation between coverage with high projected entomological risk and WNV disease incidence in the eastern plains ($\rho_s = -0.304$, n = 422, $P < 0.001$).

Census tracts in the northern Front Range containing 68-100% area with high projected entomological risk had higher WNV disease incidence among residents (median of 10.5 cases per 100,000 person-years; n = 305 census tracts) than census tracts lacking such areas (median of 0 cases per 100,000 person-years; n = 77 census tracts) or with 1-33% coverage of area with high entomological risk (median of 4.7 cases per 100,000 person-years; n = 25 census tracts) ($\chi^2 = 50.34$, df = 1, $P < 0.001$ and $\chi^2 = 5.69$, df = 1, $P = 0.02$, respectively). A similar pattern was seen for western Colorado where census tracts containing 68-100% area with high projected entomological risk had higher WNV disease incidence among residents (median of 25.8 cases per 100,000 person-years; n = 32 census tracts) than census tracts lacking such areas (median of 0 cases per 100,000 person-years; n = 93 census tracts) or with 1-33% coverage of area with high entomological risk (median of 11.8 cases per 100,000 person-years; n = 17 census tracts) ($\chi^2 = 76.95$, df = 1, $P < 0.001$ and $\chi^2 = 8.16$, df = 1, $P = 0.004$, respectively). In the southern Front Range, census tracts containing 68-100% area with high projected entomological risk had higher WNV disease incidence among residents (median of 25.7 cases per 100,000 person-years; n = 63 census tracts) than census tracts with 0-33% coverage (median of 9.3 cases per 100,000 person-years; n = 7 census tracts) ($\chi^2 = 6.63$, df = 1, $P = 0.01$). In striking contrast, census tracts in eastern Colorado containing 100%

area with high projected entomological risk had six-fold lower WNV disease incidence among residents (median of 6.3 cases per 100,000 person-years; n = 391 census tracts) than census tracts with 0-67% coverage (median of 39.9 cases per 100,000 person-years; n = 20 census tracts) ($\chi^2 = 24.08$, df = 1, $P < 0.001$).

Figure IV-7 shows the scaled-up spatial model for high projected entomological risk of exposure to *Cx. tarsalis*, in relation to census tract-based WNV disease incidence, for the northern Front Range and western Colorado.

Potential impact of climate warming on mosquito abundance

We used data from HOBO loggers operated in eight sites along the Poudre River during 2006 to estimate the relationship between elevation (x) and daily mean temperature during the peak mosquito breeding season in June-August (Fig. 8). Based on the linear relationship between daily mean June-August temperature in 2006 (y) and elevation (x) (y = 28.64793 - 0.00538x; $F_{1,6} = 22.76$, $r^2 = 0.791$, $P = 0.003$), incremental elevation increases of 100 m along the Poudre River corresponded to an incremental decrease in mean daily June-August temperature in 2006 of 0.5 °C per 100 m elevation change unit. It therefore follows that a realistic future climate warming scenario resulting in mean daily summer temperature increases of 1-2 °C [22] would cause a shift toward current temperature conditions in the future occurring at elevations 200-400 m higher than today. This realistic scenario would most likely result in both expansion in the range of *Cx. tarsalis* toward higher elevations in the Colorado Front Range and increased mosquito abundance near the current upper altitudinal (climate) limit where *Cx. tarsalis* now is present but scarce.

Discussion

Spatial assessments of risk of exposure to WNV have tended to focus on either entomological risk of vector exposure[23, 24, 25, 26, 27, 28] or disease risk based on avian or equine WNV surveillance[29, 30, 31, 32] or human case data.[33, 34, 35, 36, 37] However, each of these risk measures has inherent weaknesses. Entomological risk does not account for the importance of human behavior such as use of mosquito repellents[6]. Avian WNV surveillance either requires costly serosurveys or is subject to variability in levels of detection of dead birds[38, 39]. Epidemiologic data from humans only provide reliable information for risk in areas with an adequate population base and typically are based on address of residence rather than likely exposure site[40, 41]. Following earlier efforts for other mosquito-borne arboviruses,[42] several studies have attempted to address these issues by combining entomological risk measures (vector larval habitat, vector abundance, abundance of infected mosquito pools) with either avian WNV surveillance data[43] or human WNV disease data (case locations, disease incidence).[44, 45, 46, 47, 48, 49] However, these studies typically either were restricted to risk assessments at a crude county spatial scale[44, 47, 48] or failed to generate continuous spatial risk surfaces.[43, 46, 49] Perhaps the most complete previous approach comes from a Mississippi study combining spatial continuous environmental data with zip-code based incidence of WNV disease in humans.[45] This approach was, however, to some extent impeded by a low case load of human WNV disease (276 cases were reported from Mississippi during the 2002-2003 study period).

To account for the inherent weakness of using either entomological or epidemiological risk measures separately, we developed not only a spatial model for

entomological risk of exposure to *Cx. tarsalis* (Fig. 2) and an epidemiological risk map for WNV disease (Fig. 3) but also a novel risk classification index combining data for the independently derived measures of entomological and epidemiological risk (Fig. 4). This included a 3-county area (Larimer-Boulder-Jefferson) in the northern Colorado Front Range severely afflicted by WNV disease; the three counties reported 1,313 human cases from 2002-2006. Risk of exposure to *Cx. tarsalis* and WNV was found to be high in the densely populated eastern plains portion of the northern Front Range but low in cooler montane areas to the west that are sparsely populated but used heavily for recreation.

Development of spatial risk models for vector-borne diseases in the western United States that combine entomological and epidemiological risk measures, as previously accomplished for Lyme disease in California,[41, 50] is crucial because this part of the country encompasses large tracts of public land where risk cannot be assessed by epidemiologic data alone due to low population bases. Public lands can, however, be heavily used for recreational purposes and thus represent significant risk of human exposure to vector-borne pathogens. In this study, we were able to use an entomological risk model to demonstrate that Rocky Mountain National Park in far southwestern Larimer County, which receives > 3 million annual visitors with a summer peak in visitor numbers coinciding with vector mosquito and WNV activity in Colorado,[4, 5] currently presents minimal risk of exposure to *Cx. tarsalis* vectors. This was corroborated by the fact that mosquito sampling during the summer of 2006 in six locations in a heavily used portion of Rocky Mountain National Park yielded a single *Cx. tarsalis* (L. Eisen, unpublished data).

Our epidemiological risk assessment highlights the importance of using an appropriate spatial scale, such as census tract or zip-code, for presentation of incidence of vector-borne diseases in the west, where counties tend to cover large areas and often include considerable ecological and climatic variability[40]. This is important both to detect areas of low risk within counties with overall high risk (e.g., WNV disease in the Larimer-Boulder-Jefferson area [Fig. 3] or Lyme disease in north coastal California[50]) and to detect small isolated areas of high risk within counties with overall low risk (e.g., West Nile virus disease in parts of western Colorado [Fig. 7] or Lyme disease in southern California[50]). Additional research efforts relating to spatial eco-epidemiology of WNV disease in Colorado are underway.

Access to census tract-based WNV disease incidence data from 2002-2006 for the state of Colorado provided an opportunity to examine how the entomological risk model for the Larimer-Boulder-Jefferson area performed when scaled-up to a larger area of the northern Front Range (Larimer-Boulder-Jefferson-Douglas-El Paso), the southern Front Range, the western Colorado high plateau and mountains, and the eastern Colorado plains. The robust performance, relative to epidemiologic data, of the scaled-up entomological risk model in western Colorado can be attributed to the Larimer County model development area including a climate gradient similar to that for western Colorado. This demonstrates the value of developing spatial models for entomological risk of vector exposure within small but topographically, climatically, and ecologically diverse geographical areas. On the other hand, the Larimer County-derived entomological risk model performed very poorly in the eastern plains. This is likely because Larimer County includes only a limited portion of the climate variability existing in eastern

Colorado. Efforts are underway to generate data for abundance of *Cx. tarsalis* along a climate gradient extending from the Front Range into the eastern plains along the Big Thompson and South Platte rivers. These data will be used to develop a separate model for entomological risk of exposure to *Cx. tarsalis* in the eastern plains.

Finally, our data on abundance of *Cx. tarsalis* along climate-elevation gradients in the Colorado Front Range and the relationship between elevation and temperature in this area suggest that the spatial distribution and abundance patterns of *Cx. tarsalis* are sensitive to climate warming. We expect projected climate warming over the next 50 yr in Colorado, which includes summer temperature increases of 1-2 °C[22], to shift the cool end of the distribution of *Cx. tarsalis* several hundred meters upward in elevation in the northern Front Range. This, in conjunction with potential changes in availability of mosquito larval habitat following expected decreases in mountain snow-pack and subsequent river and stream flooding activity, will impact future patterns of risk of exposure to *Cx. tarsalis* in the Rocky Mountain region. The Colorado Front Range is exceptionally well suited for long-term empirical studies on the effect of climate warming on spatial patterns of distribution and abundance of *Cx. tarsalis* and spatial patterns of presence of WNV in local mosquito populations.

Ackowledgements

We gratefully acknowledge the U.S. Forest Service and the cities of Fort Collins and Loveland for permission to collect mosquitoes, and S.L. Anderson and J. Montgomery for technical assistance. The study was funded, in part, by a grant from the Centers for

Disease Control and Prevention (T01/CCT822307) and a contract with the National Institutes of Health (AI-25489).

Table IV-1. *Culex tarsalis* abundance and selected environmental characteristics
Abundance of *Cx. tarsalis* from late June to mid-September 2006 and selected environmental characteristics for 20 sites in Larimer County, Colorado.

Site	Mean no. of *Cx. tarsalis* per trap night	Elevation (m)	Mean annual climate data[a]				
			Temp. (°C)	Cooling degree days[b]	Relative humidity (%)	Precipitation (mm)	Snowfall (mm)
Lee Martinez[c]	7.6±11.5	1,510	9.4	495	52	392	1,396
Namaqua[c]	7.3±4.3	1,520	9.3	479	52	366	1,155
Glade Park[c]	4.3±3.7	1,540	9.2	459	52	389	1,147
Bellvue[c]	30.3±58.9	1,560	8.7	450	52	414	1,445
Picnic Rock[c]	2.5±3.3	1,610	7.6	381	53	459	1,843
Gateway[c]	0.4±0.6	1,640	7.7	340	53	443	1,727
Narrows[c]	4.5±7.2	1,690	8.5	388	53	436	1,303
Viestenz Smith[c]	4.6±9.2	1,730	7.8	357	53	456	1,551
Ouzel[c]	0.2±0.4	1,750	6.5	221	54	492	2,244
Idylwilde[c]	0.4±0.8	1,840	7.3	280	54	462	1,679
Stove Prairie[c]	0.3±0.8	1,860	6.0	174	54	496	2,347

Table IV-1, continued

Site	Mean no. of *Cx. tarsalis* per trap night	Elevation (m)	Mean annual climate data[a]				
			Temp. (°C)	Cooling degree days[b]	Relative humidity (%)	Precipitation (mm)	Snowfall (mm)
Drake[c]	0.2±0.4	1,870	6.8	280	54	467	1,778
Waltonia	0	1,970	6.7	253	54	460	1,762
Dutch George[c]	0.2±0.6	2,000	5.7	153	54	451	2,188
Big Thompson	0	2,090	6.4	208	54	457	1,764
Eggers	0.1±0.3	2,110	5.2	96	53	381	1,860
Sleepy Hollow	0	2,120	6.2	155	54	456	1,743
Dadd Gulch	0.1±0.3	2,130	4.8	92	53	330	1,853
Glen Comfort	0	2,220	5.9	74	54	393	1,199
Bliss	0.1±0.3	2,360	1.9	0	54	465	3,418

[a] Mean values for 1961-1990 based on GIS-derived data (2x2 km spatial resolution). Site locations determined with a GPS receiver.

[b] Cooling degree days are calculated as the number of degree days exceeding a baseline of 65 °F (18.3 °C).

[c] Sites included in climate-based model for abundance of *Cx. tarsalis*.

Table IV-2. Linear relationships between climate data and abundance of *Culex tarsalis*
Adjusted r^2-values and statistical significances for linear relationships between monthly or annual long-term (1961-1990) average climate data and abundance of *Cx. tarsalis* for 13 sites in Larimer County, Colorado, from late June to mid-September 2006.

	Adjusted r^2-value for linear relationship between climate variable and abundance of *Cx. tarsalis* (natural log transformed)[a]				
Time period	Mean temperature	Cooling degree days[b]	Relative humidity	Precipitation	Snowfall
January	0.499**	0^{NS}	0.047^{NS}	**0.731***	0.541**
February	0.710***	0^{NS}	0.520**	**0.715***	0.532**
March	0.708***	0^{NS}	**0.687***	0.420**	0.491**
April	**0.733***	0^{NS}	**0.615***	0.333*	**0.665***
May	**0.719***	0.300*	**0.670***	0.051^{NS}	0.543**
June	**0.715***	**0.756***	-0.046^{NS}	-0.083^{NS}	0^{NS}
July	**0.717***	**0.754***	0.470**	0.201^{NS}	0^{NS}
August	**0.712***	**0.759***	0.529**	0.496**	0^{NS}
September	**0.715***	0.726***	0.516**	0.278*	0.373*
October	0.683***	0.153^{NS}	-0.091^{NS}	0.369*	0.553**
November	0.679***	0^{NS}	0.111^{NS}	0.687***	0.461**
December	0.590**	0^{NS}	0.010^{NS}	0.612**	0.541**
Annual	0.715***	0.755***	0.796***	0.498**	0.579***

Climate variable-specific data shown in bold were used (as mean or cumulative data for indicated months) in a multivariate forward stepwise regression modeling approach to determine the association between climate conditions and abundance of *Cx. tarsalis*.

[a] ANOVA: NS, No significant association ($P > 0.05$); * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

[b] Cooling degree days are calculated as the number of degree days exceeding a baseline of 65 °F (18.3 °C).

Table IV-3. Climate ranges for Larimer, Boulder and Jefferson counties.

	Annual climate ranges (2x2 km spatial resolution)[a]				
County	Mean temperature (°C)	Cooling degree days[b]	Relative humidity (%)	Precipi-tation (mm)	Snow-fall (mm)
Larimer	-0.1–9.7	0–548	51–57	321–1,265	854–17,382
Boulder	-0.1–10.3	0–661	51–60	355–1,255	993–17,721
Jefferson	3.7–10.0	0–639	49–55	364–605	1336–4,281

[a] Mean values for 1961-1990 based on GIS-derived data.

[b] Cooling degree days are calculated as the number of degree days exceeding a baseline of 65 °F (18.3 °C).

Table IV-4. Risk classification index scheme for West Nile virus exposure
Risk classification index scheme for West Nile virus exposure in Larimer-Boulder-Jefferson based on percentage coverage of census tract by area with high projected risk of exposure to *Cx. tarsalis* and West Nile virus disease incidence.

Risk classification index	Number of census tracts	Coverage of census tract by area with high projected risk of exposure to *Cx. tarsalis*	Census tract-based WNV disease incidence per 100,000 person-years, 2002-2006
Very Low	29	0-10%	< 5
Low	27	0-10%	> 5
Low	15	11-50%	< 5
Moderate	28	11-50%	> 5
Moderate	71	> 50%	< 20
High	41	> 50%	> 20
Very High	42	> 75%	> 40

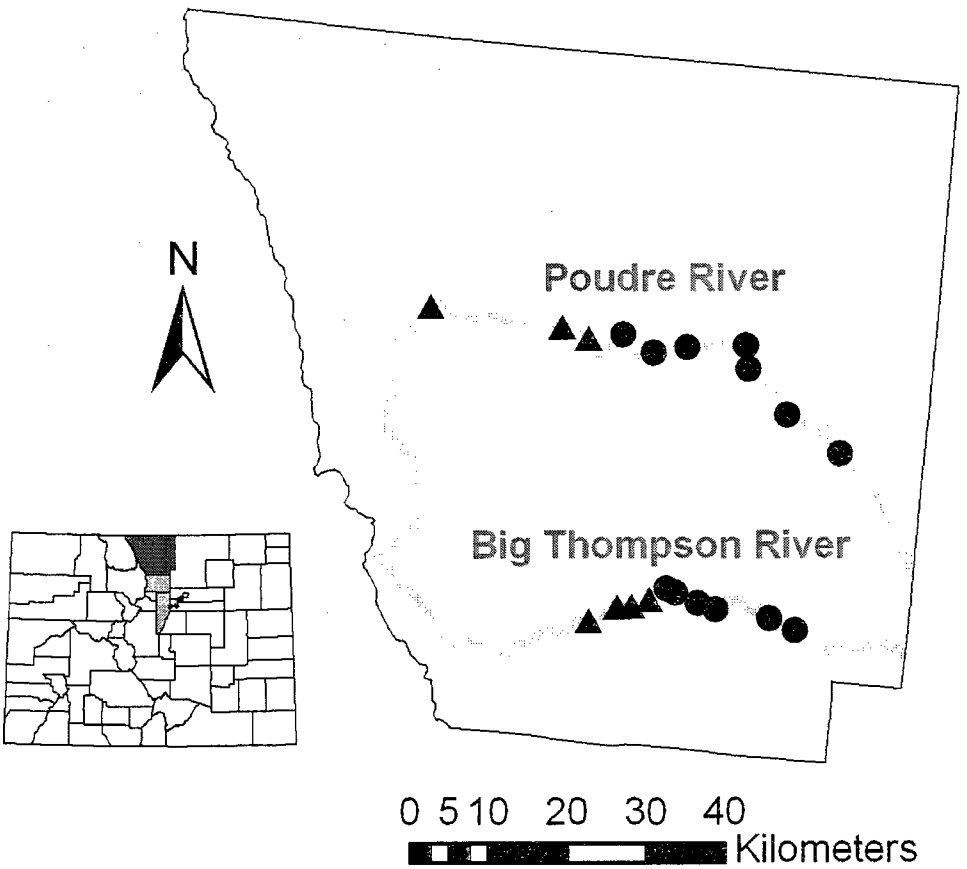

Figure IV-1. Location of Larimer County sites sampled for *Culex tarsalis*.
Location of Larimer County sites (black circles and triangles) sampled for Cx. tarsalis in 2006 and of the Poudre River and Big Thompson River. Sites included in modeling of abundance of Cx. tarsalis based on climate data are shown as black circles. The inset map shows the location of Larimer County (dark gray) and Boulder and Jefferson counties (light gray) in Colorado.

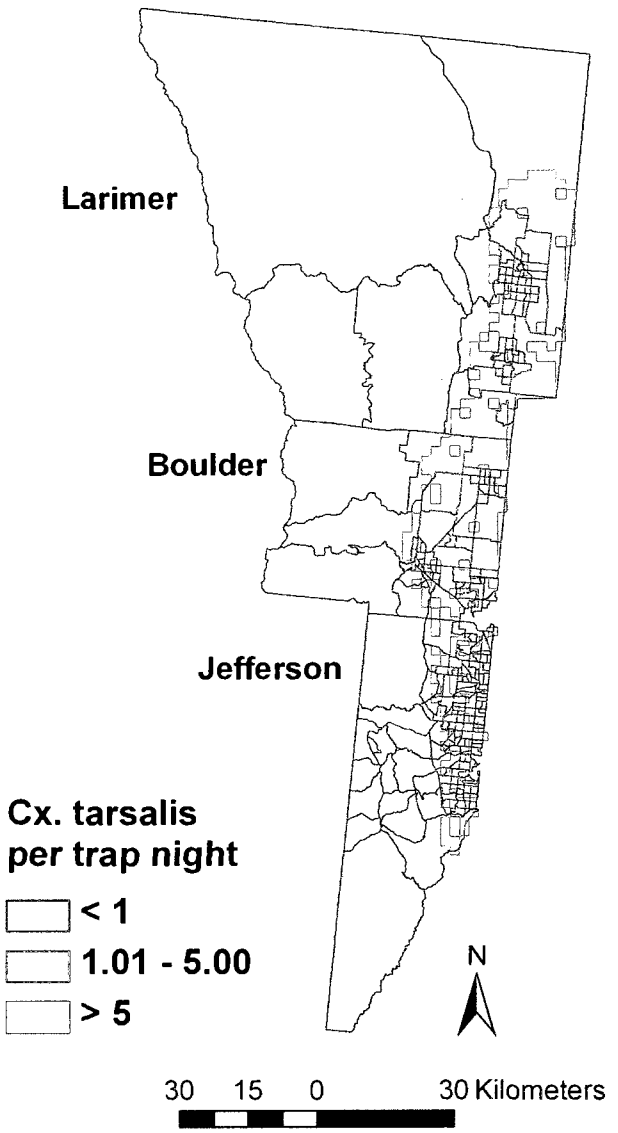

Figure IV-2. Projected entomological risk of exposure to *Culex tarsalis*
Projected entomological risk of exposure to Cx. tarsalis in the Larimer-Boulder-Jefferson area of the northern Colorado Front Range based on a cooling degree day model and where non-irrigated areas located more than 500 m from perceived larval habitat are included in the lowest abundance category. This figure appears in color at www.ajtmh.org.

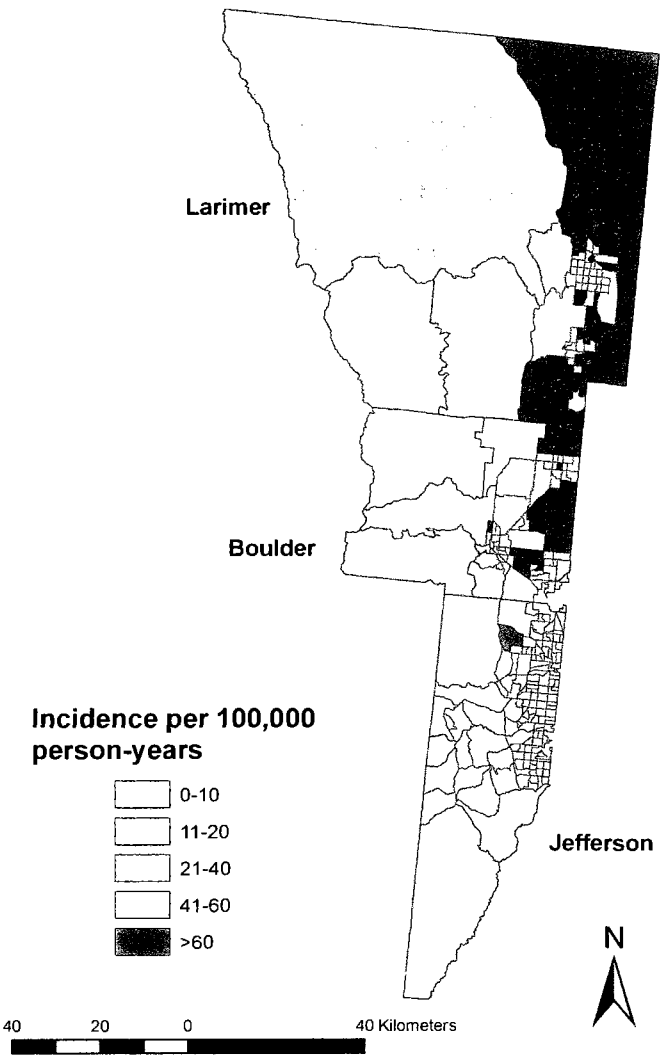

Figure IV-3. West Nile virus disease incidence by census tract, 2002-2006
West Nile virus disease incidence per 100,000 person-years, 2002-2006, by census tract in the Larimer-Boulder-Jefferson area of the northern Colorado Front Range. This figure appears in color at www.ajtmh.org.

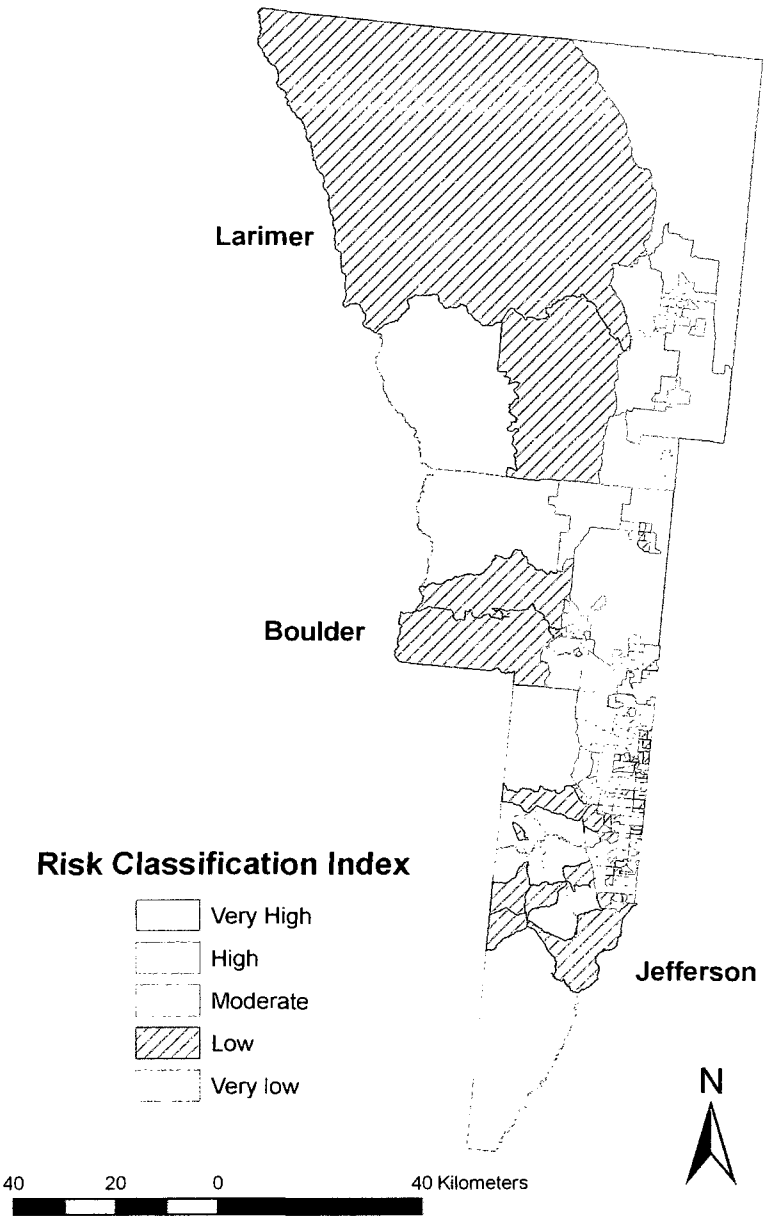

Figure IV-4. Risk classification index
Census tract-based risk classification index for combined entomological and epidemiological risk of exposure to Cx. tarsalis and West Nile virus in the Larimer-Boulder-Jefferson area of the northern Colorado Front Range. This figure appears in color at www.ajtmh.org.

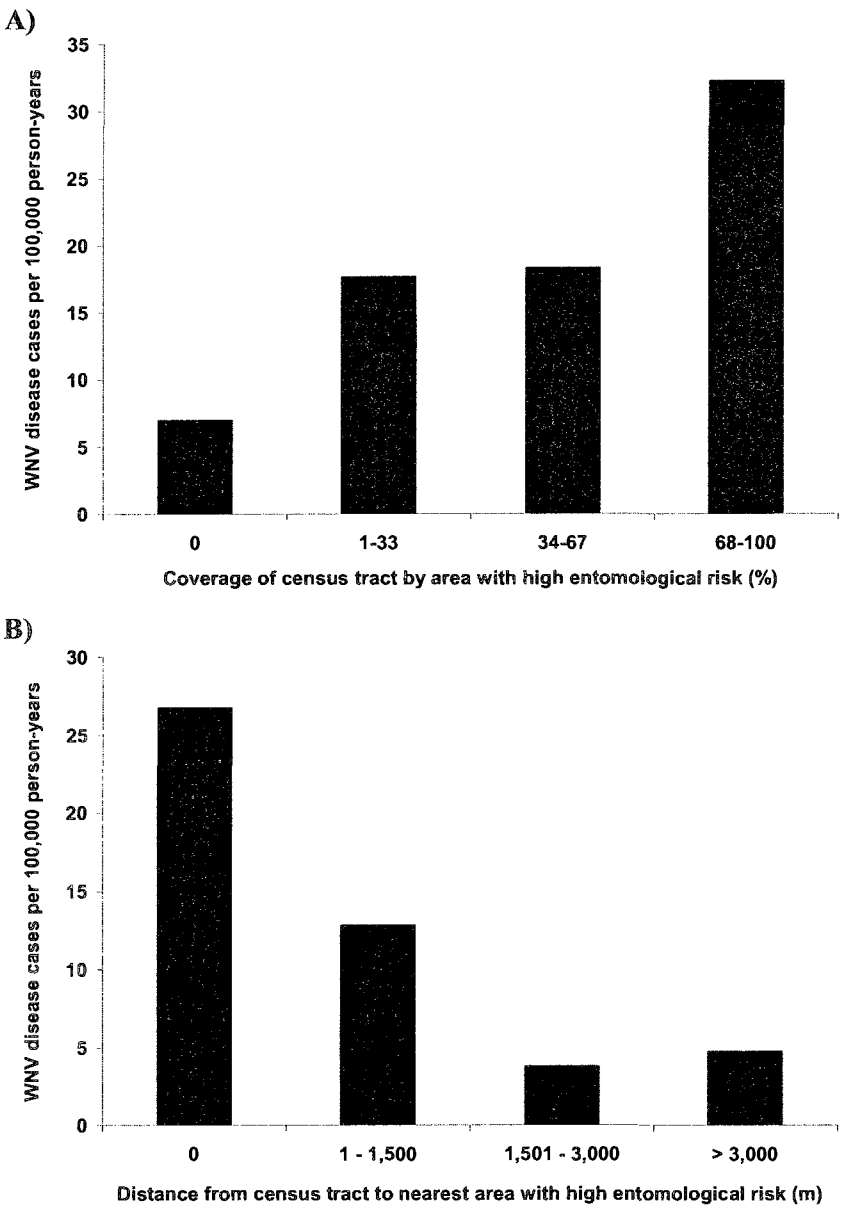

Figure IV-5. West Nile virus disease incidence compared to entomological risk
West Nile virus disease incidence per 100,000 person-years, 2002-2006, in the Larimer-Boulder-Jefferson area of the northern Colorado Front Range in relation to A) percentage of census tract covered by area with high projected entomological risk of exposure to Cx. tarsalis and B) distance to an area with high projected entomological risk (distance = 0 m if such an area is contained within a specific census tract). WNV disease incidences are based on combined case and population data for all census tracts falling within a given category for coverage of or distance to entomological risk area.

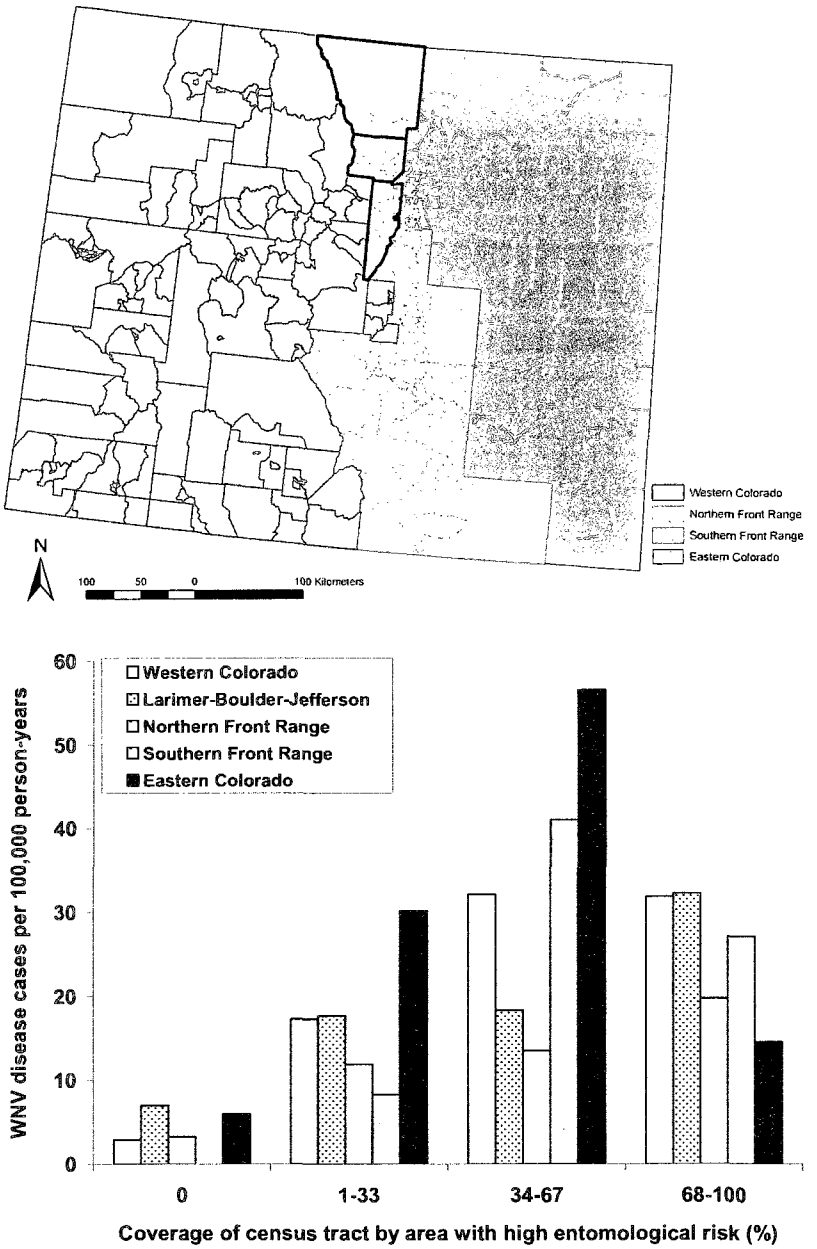

Figure IV-6. Coverage of census tract by area with high entomological risk
Relationship between coverage of census tract by area with high projected entomological risk of exposure to Cx. tarsalis and West Nile virus disease incidence per 100,000 person-years, 2002-2006, in five different regions of Colorado: the western Colorado high plateau and mountains, the Larimer-Boulder-Jefferson 3-county portion of the northern Front Range (indicated by bold county outlines), the northern Front Range, the southern Front Range, and the eastern Colorado plains. WNV disease incidences are based on combined case and population data for all census tracts within each region falling within a given category for coverage of area with high entomological risk.

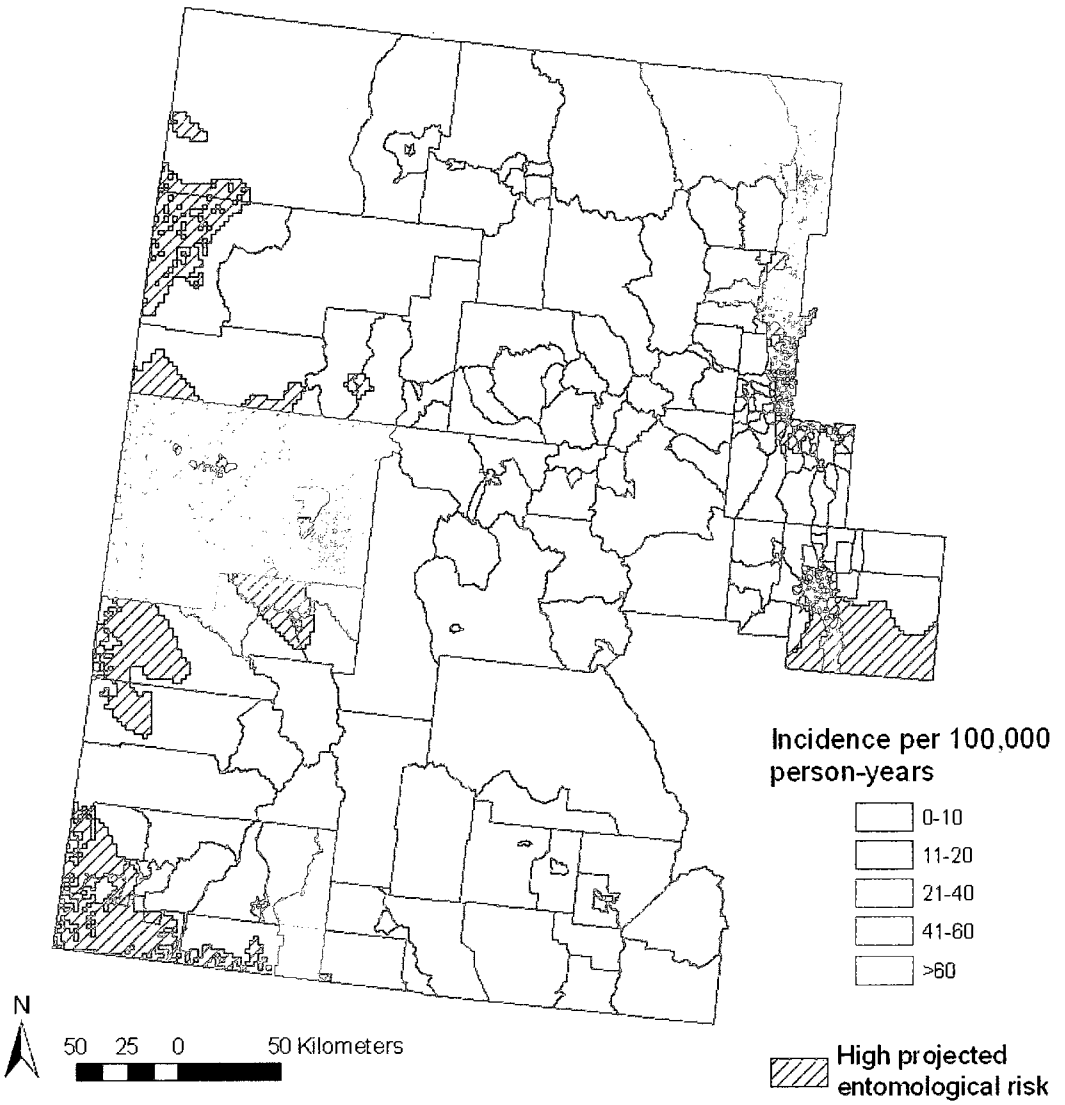

Figure IV-7. High entomological risk and disease incidence by census tract
Spatial distribution of areas with high projected entomological risk of exposure to *Cx. tarsalis* in the northern Front Range and western Colorado in relation to census tract-based West Nile virus disease incidence per 100,000 person-years, 2002-2006. This figure appears in color at www.ajtmh.org.

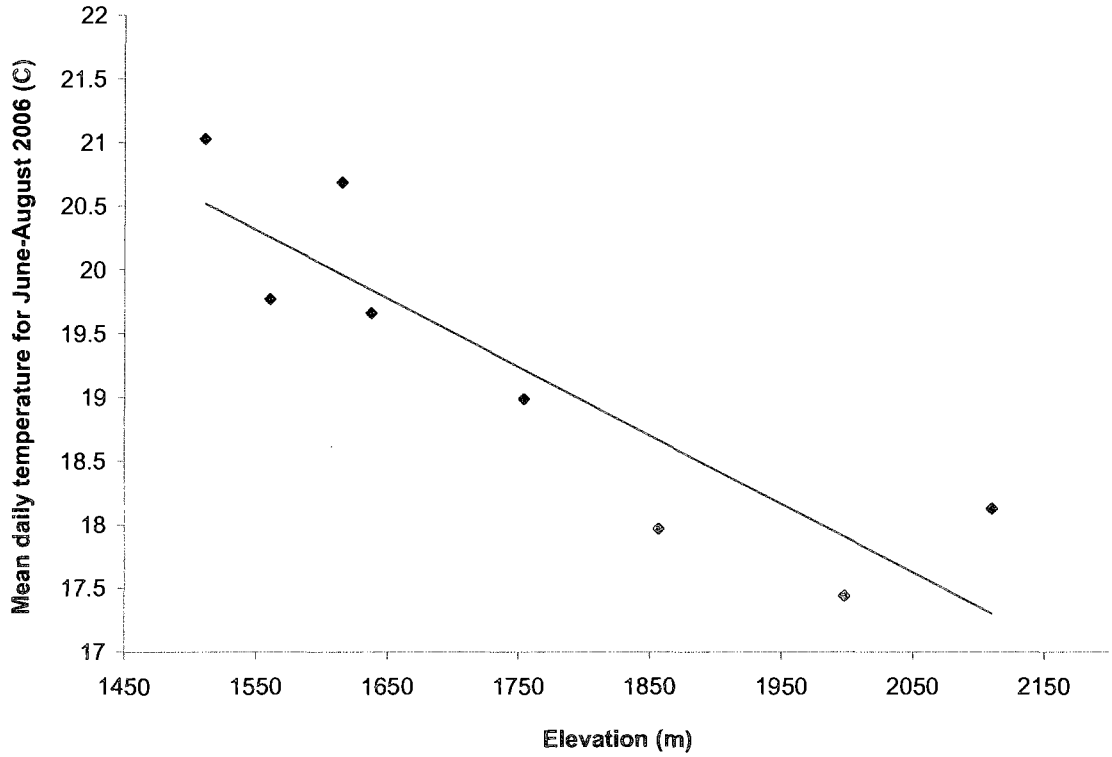

Figure IV-8. Relationship between elevation and mean daily temperature
Relationship between elevation (x) and mean daily temperature during June-August 2006 (y) along the Poudre River corridor in Larimer County (y = 28.64793 - 0.00538x; ANOVA; F1,6 = 22.76, r2 = 0.791, P = 0.003).

References Cited

1. Goddard LB, Roth AE, Reisen WK, Scott TW, 2002. Vector competence of California mosquitoes for West Nile Virus. Emerg. Infect. Dis. 8: 1385-1391.
2. Turell MJ, Dohm DJ, Sardelis MR, O'Guinn ML, Andreadis TG, Blow JA, 2005. An update on the potential of North American mosquitoes (Diptera: Culicidae) to transmit West Nile virus. J. Med. Entomol. 42: 57-62.
3. Turell MJ, Sardelis MR, O'Guinn ML, Dohm DJ, 2002. Potential vectors of West Nile virus in North America. Curr. Topics Microbiol. Immunol. 267: 241-252.
4. Bolling BG, Moore CG, Anderson SL, Blair CD, Beaty BJ, 2007. Entomological studies along the Colorado Front Range during a period of intense West Nile virus activity. J. Am. Mosq. Contr. Assoc. 23: 37-46.
5. Smith GC, Moore CG, Davis T, Savage HM, Thapa AB, Shrestha SL, Karabatsos N, 1993. Arbovirus surveillance in northern Colorado, 1987 and 1991. J. Med. Entomol. 30: 257-261.
6. Gujral IB, Zielinski-Gutierrez EC, LeBailly A, Nasci R, 2007. Behavioral risks for West Nile virus disease, northern Colorado, 2003. Emerg. Infect. Dis 13: 419-425.
7. Tempelis CH, Francy DB, Hayes RO, Lofy MF, 1967. Variations in feeding patterns of seven culicine mosquitoes on vertebrate hosts in Weld and Larimer counties, Colorado. Am. J. Trop. Med. Hyg. 16: 111-119.
8. Tsai TF, Smith GC, Happ CM, Kirk LJ, Jakob WL, Bolin RA, Francy DB, Lampert KJ, 1989. Surveillance of St. Louis encephalitis virus vectors in Grand Junction, Colorado, in 1987. J. Am. Mosq. Contr. Assoc. 5: 161-165.

9. Tsai TF, Smith GC, Ndukwu M, Jakob WL, Happ CM, Kirk LJ, Francy DB, Lampert KJ, 1988. Entomologic studies after a St. Louis encephalitis epidemic in Grand Junction, Colorado. Am. J. Epidemiol. 128: 285-297.

10. Reisen WK, Lothrop HD, 1995. Population ecology and dispersal of *Culex tarsalis* (Diptera: Culicidae) in the Coachella Valley of California. J. Med. Entomol. 32: 490-502.

11. Wegbreit J, Reisen WK, 2000. Relationships among weather, mosquito abundance, and encephalitis virus activity in California: Kern County 1990-98. J. Am. Mosq. Contr. Assoc. 16: 22-27.

12. Hess AD, Hayes RO, 1970. Relative potentials of domestic animals for zooprophylaxis against mosquito vectors of encephalitis. Am. J. Trop. Med. Hyg. 19: 327-334.

13. Lothrop HD, Reisen WK, 2001. Landscape affects the host-seeking patterns of *Culex tarsalis* (Diptera: Culicidae) in the Coachella Valley of California. J. Med. Entomol. 38: 325-332.

14. Baker M, 1961. The altitudinal distribution of mosquito larvae in the Colorado Front Range. Trans. Am. Entomol. Soc. 87: 231-246.

15. Harmston FC, Lawson FA, 1967. Mosquitoes of Colorado. Atlanta, GA: U.S. Department of Health, Education and Welfare, Public Health Service.

16. Carpenter SJ, 1969. Observations on the distribution and ecology of mountain *Aedes* mosquitoes in California. XII. Other species found in the mountains. Calif. Vector Views 16: 27-34.

17. Darsie RF, Jr., Ward RA, 2005. Identification and geographical distribution of the mosquitoes of North America, north of Mexico. Gainesville, FL: University Press of Florida.

18. Bailey SM, Eliason DA, Hoffman BL, 1965. Flight and dispersal of the mosquito *Culex tarsalis* Coquillett in the Sacramento Valley of California. Hilgardia 37: 73-113.

19. Dow RP, Reeves WC, Bellamy RE, 1965. Dispersal of female *Culex tarsalis* into a larvicided area. Am. J. Trop. Med. Hyg. 14: 656-670.

20. Reisen WK, Milby MM, Meyer RP, Pfuntner AR, Spoehel J, Hazelrigg JE, Webb JP, Jr., 1991. Mark-release-recapture studies with *Culex* mosquitoes (Diptera: Culicidae) in southern California. J. Med. Entomol. 28: 357-371.

21. Sall J, Creighton L, Lehman A, 2005. JMP Start Statistics. Third Edition. Belmont, CA: Brooks/Cole.

22. Leung LR, Qian Y, Bian X, Washington WM, Han J, Roads JO, 2004. Mid-century ensemble regional climate change scenarios for the western United States. Climatic Change 62: 75-113.

23. Diuk-Wasser MA, Brown HE, Andreadis TG, Fish D, 2006. Modeling the spatial distribution of mosquito vectors for West Nile virus in Connecticut, USA. Vector-Borne & Zoonotic Dis. 6: 283-295.

24. Zou L, Miller SN, Schmidtmann ET, 2006. Mosquito larval habitat mapping using Remote Sensing and GIS: Implications of coalbed methane development and West Nile virus. J. Med. Entomol. 43: 1034-1041.

25. Shone SM, Ferrao PN, Lesser CR, Norris DE, Glass GE, 2001. Analysis of mosquito vector species abundances in Maryland using geographic information systems. Ann. N.Y. Acad. Sci. 951: 364-368.

26. Shaman J, Stieglitz M, Stark C, Le Blancq S, Cane M, 2002. Using a dynamic hydrology model to predict mosquito abundances in flood and swamp water. Emerg. Infect. Dis. 8: 6-13.

27. Rios J, Hacker CS, Hailey CA, Parsons RE, 2006. Demographic and spatial analysis of West Nile virus and St. Louis encephalitis in Houston, Texas. J. Am. Mosq. Contr. Assoc. 22: 254-263.

28. Kronenwetter-Koepel TA, Meece JK, Miller CA, Reed KD, 2005. Surveillance of above- and below-ground mosquito breeding habitats in a rural Midwestern community: Baseline data for larvicidal control measures against West Nile virus vectors. Clin. Med. Res. 3: 3-12.

29. Gibbs SEJ, Wimberly MC, Madden M, Masour J, Yabsley MJ, Stallknecht DE, 2006. Factors affecting the geographic distribution of West Nile virus in Georgia, USA:2002-2004. Vector-Borne Zoonotic Dis. 6: 73-82.

30. Theophilides CN, Ahearn SC, Grady S, Merlino M, 2003. Identifying West Nile virus risk areas: The dynamic continuous-area space-time system. Am. J. Epidemiol. 157: 843-854.

31. Ward MP, 2006. Spread of equine West Nile virus encephalomyelitis during the 2002 Texas epidemic. Am J Trop Med Hyg 74: 1090-1095.

32. Ward MP, Ramsay BH, Gallo K, 2005. Rural cases of equine West Nile virus encephalomyelitis and the Normalized Difference Vegetation Index. Vector-Borne & Zoonotic Dis. 5: 181-188.

33. Brownstein JS, Rosen H, Purdy D, Miller JR, Merlino M, Mostashari F, Fish D, 2002. Spatial analysis of West Nile virus: Rapid risk assessment of an introduced vector-borne zoonosis. Vector-Borne & Zoonotic Dis. 2: 157-164.

34. Corrigan RLA, Waldner C, Epp T, Wright J, Whitehead SM, Bangura H, Young E, Townsend HGG, 2006. Prediction of human cases of West Nile virus by equine cases, Saskatchewan, Canada, 2003. Prev. Vet. Med. 76: 263-272.

35. Yiannakoulias NW, Schopflocher DP, Svenson LW, 2006. Modelling geographic variations in West Nile virus. Can. J. Publ. Health 97: 374-378.

36. Ruiz MO, Tedesco C, McTighe T, Austin C, Kitron U, 2004. Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. Int. J. Health Geogr. 3: 8-19.

37. Ruiz MO, Walker ED, Foster ES, Haramis LD, Kitron U, 2007. Association of West Nile virus illness and urban landscapes in Chicago and Detroit. Int. J. Health Geogr. 6: 10-20.

38. Ward MR, Stallknecht DE, Willis J, Conroy MJ, Davidson WR, 2006. Wild bird mortality and West Nile virus surveillance: Biases associated with detection, reporting, and carcass persistence. J. Wildl. Dis. 42: 92-106.

39. Patnaik JL, Juliusson L, Vogt RL, 2007. Environmental predictors of human West Nile virus infections, Colorado. Emerg. Infect. Dis. 13: 1788-1790.

40. Eisen L, Eisen RJ, 2007. Need for improved methods to collect and present spatial epidemiologic data for vectorborne diseases. Emerg. Infect. Dis. 13: 1816-1820.

41. Eisen RJ, Eisen L, 2008. Spatial modeling of human risk of exposure to vector-borne pathogens based on epidemiological versus arthropod vector data. J. Med. Entomol. 45: In Press.

42. Lothrop HD, Reisen WK, 1999. A geographical information system to manage mosquito and arbovirus surveillance and control data in the Coachella Valley of California. J. Am. Mosq. Contr. Assoc. 15: 299-307.

43. Takeda T, Whitehouse CA, Brewer M, Gettman AD, Mather TN, 2003. Arbovirus surveillance in Rhode Island: Assessing potential ecologic and climatic correlates. J. Am. Mosq. Contr. Assoc. 19: 179-189.

44. Brownstein JS, Holford TR, Fish D, 2004. Enhancing West Nile virus surveillance, United States. Emerg. Infect. Dis. 10: 1129-1133.

45. Cooke W, III, Grala K, Wallis R, 2006. Avian GIS models signal human risk for West Nile virus in Mississippi. Int. J. Health Geogr. 5: 36-56.

46. Kutz FW, Wade TG, Pagac BB, 2003. A geospatial study of the potential of two exotic species of mosquitoes to impact the epidemiology of West Nile virus in Maryland. J. Am. Mosq. Contr. Assoc. 19: 190-198.

47. Lukacik G, Anand M, Shusas EJ, Howard JJ, Oliver J, Chen H, Backenson PB, Kauffman EB, Bernard KA, Kramer LD, White DJ, 2006. West Nile virus surveillance in mosquitoes in New York State, 2000-2004. J. Am. Mosq. Contr. Assoc. 22: 264-271.

48. Mongoh MN, Kaitsa ML, Dyer NW, 2006. Environmental and ecological determinants of West Nile virus occurrence in horses in North Dakota, 2002. Epidemiol. Infect. 135: 57-66.

49. Orme-Zavaleta J, Jorgensen J, D'Ambrosio B, Altendorf E, Rossignol PA, 2006. Discovering spatio-temporal models of the spread of West Nile virus. Risk Analysis 26: 413-422.

50. Eisen RJ, Lane RS, Fritz CL, Eisen L, 2006. Spatial patterns of Lyme disease risk in California based on disease incidence data and modeling of vector-tick exposure. Am. J. Trop. Med. Hyg. 75: 669-676.

V. Spatial Risk Models for Human Plague in the West Nile Region of Uganda

This chapter has been published in the American Journal of Tropical Medicine and Hygiene and is cited as Winters AM, Staples JE, Ogen-Odoi A, Mead PS, Griffith K, Owor N, Babi N, Enscore RE, Eisen L, Gage KL, Eisen RJ. 2009. **Spatial Risk Models for Human Plague in the West Nile Region of Uganda.** ***AJTMH 80 (6):1014-1022.***

Anna M. Winters[1, 2]*, J. Erin Staples[1], Asaph Ogen-Odoi[3], Paul S. Mead[1], Kevin Griffith[1], Nicholas Owor[3], Nackson Babi[3], Russell E. Enscore[1], Lars Eisen[2], Kenneth L. Gage[1], and Rebecca J. Eisen[1]

[1]Division of Vector-Borne Infectious Diseases, National Center for Vector-Borne, Zoonotic and Enteric Diseases, Centers for Disease Control and Prevention, Fort Collins, Colorado; [2]Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado; [3]Uganda Virus Research Institute, Entebbe, Uganda.

Abstract

The West Nile region of Uganda represents an epidemiologic focus for human plague in east Africa. However, limited capacity for diagnostic laboratory testing means few clinically diagnosed cases are confirmed and the true burden of disease is undetermined. The aims of the study were (1) describe the spatial distribution of clinical plague cases in the region, (2) identify ecological correlates of incidence, and (3) incorporate these variables into predictive models that define areas of plague risk. The model explained 74% of the incidence variation and revealed that cases were more common above 1,300 m than below. Remotely-sensed variables associated with differences in soil or vegetation were also identified as incidence predictors. The study demonstrated that plague incidence can be modeled at parish-level scale based on environmental variables and identified parishes where cases may be under-reported and enhanced surveillance and preventative measures may be implemented to decrease the burden of plague.

KEY WORDS: human plague, Uganda, West Nile region, Geographic Information Systems (GIS), spatial model

Introduction

Plague, a severe bacterial disease caused by *Yersinia pestis*, has a world-wide distribution.[1] In recent decades, the majority of human cases have been reported from eastern or southern Africa and Madagascar.[2-4] The West Nile region of Uganda, located in the northwest of the country, represents the current primary epidemiological focus in that country.[5, 6] However, due to limited capacity for diagnostic laboratory testing in that region, few clinically diagnosed cases are confirmed and the true burden of disease is yet to be determined.

Yersinia pestis is maintained primarily in zoonotic cycles by rodents and their fleas. Although the bacterium can be transmitted through direct contact or inhalation of infectious respiratory droplets, humans are most often infected with *Y. pestis* via flea-bites during epizootic periods when rodent hosts perish from infection and infectious fleas are forced to parasitize alternative hosts, including humans.[1] Plague infections have a short incubation period (typically 2--6 days for the bubonic form of the disease) followed by acute onset of illness. If antibiotic treatment is delayed or inadequate, the disease is often fatal.[2] Outcome of infection is improved by early diagnosis followed by appropriate antibiotic therapy.[1, 7-14]

Raising awareness among health care providers, environmental health specialists, and the public in areas where transmission to humans is most likely to occur could aid in prevention and control of plague in Africa. In the United States, fine-scale spatial risk models have been constructed to identify areas posing the greatest threat to humans within key plague foci such as the four corners area.[15, 16] Although a plague model has

been constructed for the African continent[17], its resolution is very coarse and public health utility limited. Fine-scale models developed in African plague foci are needed to define high risk areas and are in line with the World Health Organization's recommendations for implementation of public health interventions including predictive surveillance and preventative control for plague.[18, 19]

In the study, we sought to (1) describe the spatial distribution of suspect and probable human plague cases in the West Nile region, (2) identify remotely sensed ecological correlates of parish-level disease incidence, and (3) incorporate these variables into a predictive model that defines areas of plague risk in the region. This is useful for defining existing epidemiological foci, and for identifying areas where enhanced surveillance and preventative measures may be implemented to decrease the burden of plague.

Materials and Methods

Study area

Two districts, Arua and Nebbi, located in the West Nile region of Uganda were included in the study (Figure 1). These districts, each comprised of three counties, encompass an elevation gradient from 626--1,668 m with low-lying areas in the northeast rising to higher elevations across the Rift Valley escarpment which runs from north to south roughly bisecting the districts. The eastern parts of Arua and Nebbi districts, which include portions of Padyere and Madi-Okollo counties, are characterized by sandy soil and low rainfall while the western highlands of Okoro and Vurra counties contain lush vegetation, fertile soil, and numerous rivers and tributaries. The annual rainfall average in

this region is 1,250 mm with March-June experiencing less reliable rainfall and heavier and more reliable precipitation occurring from late August through November (http://www.nebbi.go.ug/; http://www.arua.go.ug/).[20, 21] The 2002 Uganda Population and Housing Census reported a population of 833,928 in Arua district and 435,360 in Nebbi district (Uganda Bureau of Statistics, 2002).

Epidemiological Data

An epidemiological database of reported human plague cases within Nebbi and Arua districts was developed based on review of health records from a total of 31 health facilities (27 district clinics and 4 hospitals) (Figure 1). The year 1999 was selected as the starting year because a uniform recording system was initiated that year and required clinics to record all consults and the presumptive diagnoses. All cases for which plague was listed as the primary clinical diagnosis were extracted from clinic and hospital log books. Information was available for age, sex, place of residence, date of clinic visit, concurrent diagnoses, and treatment. Additionally, Health Management Information System (HMIS) forms from the district Ministry of Health (MOH) offices were reviewed and information pertaining to plague cases such as onset of illness and outcome, including date of death, was extracted and cross-referenced with the information obtained in the health clinics. Finally, beginning in 2000 in Okoro County, a separate standardized reporting form was used that included more detailed information about the cases including environmental data (e.g., reported rodent die-offs) for the site of residence.

Standard criteria for diagnosis of plague in Uganda are sudden onset of fever, chills, malaise, headache or prostration accompanied by either painful regional lymphadenopathy (bubonic), hematemesis or hematochezia (septicemic), or cough with

hemoptysis (pneumonic). For the purpose of this study, a suspect plague case was defined as a patient who was seen in a health facility where he/she was diagnosed and treated for plague. A probable plague case was a patient seen at the referral hospital in the region where he/she was diagnosed and treated for plague and there was additional environmental data suggesting ongoing plague activity, such as a rat die-off in the area where the patient likely contracted the infection. The ratio of suspect cases by parish relative to total suspect cases was calculated; the same calculation was completed for probable cases by parish. The proportion probable cases was subtracted from the proportion suspect cases and a Wilcoxon's signed-rank test applied to the difference setting the mean equal to zero to identify whether the proportional representation of suspect cases versus probable cases was similar.

Cases were excluded from the analysis if (1) the village of residence was located in the Democratic Republic of Congo (DRC); or (2) the parish of residence was not listed. Although village of residence was known for the majority of cases, many village locations have not been georeferenced and, therefore, could not be used to create spatial risk models. Parishes, which represent the next smallest spatial unit and have been georeferenced and linked with population data, therefore serve as the spatial unit for analyses. Cumulative plague incidence rates for 1999--2007 by parish were calculated based on population data acquired from the 2002 Uganda Population and Housing Census (Uganda Bureau of Statistics, 2002).

Environmental Data

Geographic Information System (GIS)-based data used in the spatial modeling included: (1) administrative boundaries within Uganda including district, county, and

parish (International Livestock Research Institute [ILRI], 2006); (2) 90-m digital elevation model (DEM) (Shuttle Radar Topography Mission [SRTM] Elevation Data Set, 2008; accessed August 2008 at http://seamless.usgs.gov/); (3) Multipurpose Landcover database, 1 km resolution (Africover; Geographic Information Support Team [GIST], 2003); and (4) Landsat 7 Enhanced Thematic Mapper Plus [ETM+] images dated January 1, 2007 (Row/Path: 58/172) and January 10, 2007 (Row/Path: 58/173). The Landsat imagery is remotely sensed data and was collected by the Landsat 7 satellite. This satellite scans the entire earth's surface and collects 8 bands or channels of reflected energy that may be used to discriminate between earth surface materials as these materials (e.g., soil versus vegetation) have unique amounts of emitted and reflected radiation which vary by wavelength and may be captured by satellite sensors.[22] The Landsat images were acquired through a cooperative agreement with the National Geospatial-Intelligence Agency [NGA]. Images were captured during clear atmospheric conditions and radiometric and geometrical distortions of the imagery had been corrected. Landsat images from 1 and 10 January 2007 were mosaiced together based on spatial reference to cover the entire study area which included two different paths that are covered on different dates. To test whether raster values from each separate image were similar within the overlapping region, 30 points were randomly selected from each of the individual Landsat images within the overlap area. There was no difference for values from the two images at the randomly selected points (Wilcoxon's signed-rank test, df = 29, $P = 1.0$) which confirmed that the mosaicing process had been successful.

Landsat ETM+ band 6 (low gain and high gain bands) was converted from digital number (DN) to absolute radiance and then to surface temperature (°C) using the raster

calculator in ArcGIS 9.3 (ESRI, Redlands, CA).[23-27] Additionally, several dimensionless measures indicating landscape characteristics were calculated based on manipulation of Landsat bands using the bandmath function in ENVI 4.5.[22, 28, 29] These included the Normalized Difference Vegetation Index (NDVI), wetness, brightness and greenness indexes. NDVI (based on Landsat band 3 and band 4) is directly related to the photosynthetic capacity of plant canopies and therefore may be used to detect coverage of green vegetation. A higher NDVI pixel value indicates abundance of green biomass. Wetness, greenness and brightness indices are compositive values calculated through tasseled cap transformation which weights the sums of separate Landsat bands resulting in relative measures of soil brightness, greenness of the vegetation, and wetness of the land cover.[22, 29] For example, in the greenness image, a higher pixel value indicates greater biomass in that area; a higher pixel value in the wetness image indicates greater moisture status; a high value for brightness is indicative of bare soil.

Because of the paucity of classified vegetation and soil layers with fine spatial resolution for this area, it was necessary to use pure band values as proxies for vegetation and soil variability. Individual Landsat bands measure distinct wavelengths within the electromagnetic spectrum that may be used to discriminate between different landscape variables, however it is difficult to make conclusions about specific landscape characteristics based on individual band values compared to indices[22]. Minimum, maximum, and average values for all indices, individual Landsat bands (1-5 and 7, 30 x 30 m spatial resolution; band 6, thermal infrared band, 60 x 60 m resolution; band 8, panchromatic band, 15 x 15 m), and DEM data were extracted for each parish using the

zonal statistics function of the ArcGIS 9.3 Spatial Analyst; all of these variables were continuous.

In addition, a dichotomous variable for mean elevation by parish was created based on an elevation cutoff of 1,300 m. This cut-off value was selected because qualitative observations indicated plague to be present in parishes with mean elevations above the cutoff, but absent in parishes at lower elevations. Finally, a variety field was calculated for the Multipurpose Landcover database (Africover) indicating the number of separate land classes within each parish. All layers were projected to World Geodetic System (WGS) 1984 projection.

Multivariate logistic regression model

We applied a multivariate logistic regression model to determine the probability of plague case occurrence by parish. Covariates included in the logistic regression model were chosen via forward stepwise regression (probability to enter of 0.25) and were restricted to variables significantly associated with plague incidence in univariate tests of association (Wilcoxon's test, $P < 0.05$) but not strongly correlated with each other (Spearman's rank correlation, $\rho_s < 0.8$). A goodness of fit test was applied to determine whether the model covariates adequately explained the distribution of plague occurrence. Receiver Operating Curves (ROCs) assessed the overall discrimination of the model based on the area under the curve (AUC), which was also used to determine the optimal probability cut-off to characterize each parish as having elevated or low risk for plague thereby optimizing sensitivity and specificity simultaneously.

The logistic model is described by the following equation:

$$Logit(P) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k \qquad [1]$$

where P is the probability of a parish being classified as having elevated risk, β_0 is the intercept and $\beta_1 \ldots \beta_k$ represent the coefficients associated with each independent variable $x_1 \ldots x_k$. The probability that a particular parish is classified as having elevated or low risk for plague cases to occur can be derived from equation 1 using the following expression:

$$P = \exp(\beta_0 + \beta_1 x_1 + \cdots + + \beta_k x_k) / [1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + + \beta_k x_k)] \quad [2]$$

Each parish was assigned a single probability value. The optimal probability cut-off value was chosen by maximizing sensitivity and specificity simultaneously using receiver operating characteristic (ROC) curves. Parishes with values above the probability threshold value were classified as having elevated risk, while all others were considered low risk.

Clinic buffer

Minimal access to public transportation makes access to health care facilities difficult in parts of the study area and may result in underreporting of plague cases from some parishes. We therefore only included parishes with centroids within 15.4 km of the nearest designated health clinic in the logistic model. The 15.4 km "clinic buffer" distance was chosen as it represented the longest travel distance from a reported plague case to a health clinic within the 1999--2007 study period. The clinic buffer excluded Ayivu county in Arua district and Jonam county in Nebbi district thereby decreasing the study area to include parishes within 2 counties (Vurra and Madi-Okollo) in Arua district and 2 counties (Padyere and Okoro) in Nebbi District.

Multivariate linear regression model

To predict parish-scale incidence of human plague from environmental GIS/RS-based data, multivariate linear regression models were constructed using data from the parishes classified by the logistic regression model as areas with elevated plague risk.[31] Candidate models were identified using forward stepwise regression (probability to enter of 0.25). Predictive variables were restricted to those significantly associated with plague incidence in univariate tests of association (Wilcoxon's test; $P < 0.05$) but not strongly correlated with each other (Spearman's rank correlation; $\rho_s < 0.8$). Models with the lowest Akaike Information Criterion (AIC) were considered the most parsimonious models, but models within two AIC units of the minimum AIC value were considered competing.[32] The best model had the lowest AIC and provided the most robust validation. A Moran's I statistic was calculated using ArcGIS 9.3 to ensure that the residuals of the selected final model were not spatially autocorrelated. The predictive capability of the linear regression model was evaluated by regressing actual reported incidence on predicted incidence. The predictive equation was extrapolated to all parishes included in the study area and not excluded by the clinic buffer. If a parish was predicted to have an incidence less than zero, it was assigned an incidence equal to zero. Statistical analyses were carried out using the JMP statistical package[33], and results were considered significant when $P < 0.05$.

Results

Summary of epidemiological data

A total of 2,011 human plague cases reported from January 1999 to December 2007 were included in the epidemiological database. Of these, 152 cases were excluded

from the analysis because either the village of residence was located in the Democratic Republic of Congo (DRC) or parish was not listed. The remaining 1,859 cases were comprised of 76 probable cases and 1,783 suspect cases. From 1999--2007, the annual mean number of reported cases was 199 (range 11--445) with the greatest number of cases reported in 2001. Onset of plague symptoms followed a seasonal pattern with cases increasing in September, peaking in November and decreasing in December and January; few cases were reported in the remaining months. A further description of epidemiologic features and clinical presentation of plague cases will be described in more detail in a separate publication.

Qualitative analysis of the data revealed that suspect and probable cases had similar spatial trends such that case numbers were higher in the western reaches of Nebbi and Arua districts, especially at elevations above 1,300 m. Calculation of the mean difference between suspect and probable cases across all parishes indicated that the proportional representation of suspect cases versus probable cases was similar (Wilcoxon's signed rank test: df = 25, $P = 1.0$) and justified the use of both probable and suspect cases in the analysis.

Logistic model for elevated or low risk of plague

A multivariate logistic regression model based on GIS/RS-derived environmental data was developed to predict parishes with elevated versus low plague risk (Table 1). The model revealed that elevated parish-level plague risk was positively associated with mean parish elevation being >1,300 m and with maximum brightness and average wetness, but negatively associated with average greenness. A lack of fit test indicated that the most parsimonious model included sufficient numbers of covariates with appropriate

functional relationships ($\chi^2 = 56.27$, df = 111, $P > 1.0$) and a whole model test denoted good overall fit ($\chi^2 = 96.69$, df = 4, $P < 0.001$). Accuracy of the best model, based on the area under the ROC curve, was 0.96 indicating that randomly selected presence and absence pairs would be correctly ordered by their probability scores 96% of the time. Probability of plague case occurrence was calculated based on the model described, and dichotomized into "elevated risk" or "low risk" categories based on a cut-off probability value ($P = 0.55$) which was derived from the ROC curve and simultaneously optimized the sensitivity and specificity of the model.

The model predicted 46 (40%) of included parishes to report plague cases. These parishes were located in the western region of the study area, along the DRC border. Nearly all parishes predicted to have elevated plague risk were located above the 1,300 m elevation cutoff. Evaluation of the model against actual case reports by parish revealed a sensitivity of 93% (i.e., 93% of the parishes with plague reported were correctly classified by the model as having elevated plague risk). Model specificity was 92% (i.e., 92% of parishes where plague was not reported were correctly identified as having low plague risk). The model predicted plague cases in 6 parishes where no cases were actually reported, thereby producing a positive predictive value of 87%. Three parishes below the 1,300 m elevation cutoff actually reported plague cases, but were misclassified by the logistic model, and were predicted to pose a low risk decreasing the negative predictive value to 96% (Table 2).

Linear regression model for plague incidence

The linear regression model was developed based on 49 parishes either predicted by the logistic regression model to have plague cases present (n = 46), or misclassified by

that model as low risk, but where cases actually had been reported (n = 3). The best linear regression model indicated that within these parishes, plague incidence continued to be positively associated with elevation above 1,300 m as seen in the logistic model. Positive associations were also observed between plague incidence and the parish level average value of band 3, minimum elevation values in parishes that were above the 1,300 m (e.g. low areas within a high plateau), surface temperature, and the variety of land-cover classes (Africover) (Table 3). Plague incidence was negatively associated with average values of band 7 and minimum brightness (Table 3). Combined, these variables explained 68% of the total variation in parish-level incidence ($F_{7,41} = 12.58$, $r^2 = 0.68$, $P < 0.001$). Parameter estimates for covariates of the selected model are shown in Table 3. The model predicted cases to occur within parishes throughout the model build area, including parishes in northern Vurra County and eastern Okoro County where elevations were lower than 1,300 m.

Two parishes in northern Vurra County with very high plague incidences (Ayavu parish, incidence = 64.7 cases per 1,000 population; Ozoo parish, incidence = 54.1 cases per 1,000 population) prevented the model residuals from achieving a normal distribution. Once these parishes were removed from the analysis, the residuals of the model were normally distributed (Shapiro-Wilks test, W = 0.98, $P = 0.66$). The residuals from the entire model development area (including Ayavu and Ozoo parishes) were not spatially autocorrelated (Moran $I = -0.08$, $Z[I] = -0.92$). Regressing actual incidence on predicted incidence showed that 74% of the variation in actual incidence was explained by the predictive model ($F_{1,47} = 130.42$, $r^2 = 0.74$, $P < 0.001$) (Figure 3).

Extrapolation of the linear regression model

The linear regression model explained approximately 50% of the total variation in parish-level plague incidence when extrapolated east to all parishes within the 4 county study area and in the clinic buffer ($F_{1,114} = 114.05$, $r^2 = 0.50$, $P < 0.001$). The most noteworthy areas of discordance between the predictive model and actual reported incidence were (1) those in which very high incidence was expected but no cases were reported or (2) areas where actual incidence was much higher than predicted. Specifically, three parishes in northern Vurra County did not report any cases from 1999--2007 whereas the model predicted incidences ranging from 26.0--48.1 cases per 1,000 population (Figure 3). Two of these parishes were noted as outliers in the analysis of residuals. Conversely, three parishes located in central and southern Vurra County were predicted to report cumulative incidences of 29.0--42.5 cases per 1,000 population (1999--2007), but case reports were higher than predicted (42.6--64.7 cases per 1,000 population) (Figure 3).

Discussion

Previous studies have identified the West Nile region of Uganda as an epidemiological focus for plague.[5, 6] However, within this region the spatial distribution of case occurrence and the ecological factors associated with elevated plague risk are poorly defined. We have addressed this by (1) mapping parish-level incidence of plague in Arua and Nebbi districts from 1999--2007 and (2) developing statistical models to predict plague incidence based on elevation and remotely-sensed environmental variables. The study demonstrated that plague incidence can be modeled at the parish-

level scale based on environmental variables and identified several parishes where plague may be under-reported and enhanced plague surveillance is called for.

Similar to previously published models of plague risk for the United States[15, 34, 35] and descriptive historical observations from Africa[20, 21, 36, 37], elevation was an important predictor of risk within the study area in Uganda's West Nile region. Parishes located in western counties and situated above the Rift Valley escarpment (average parish elevation >1,300 m) reported a higher incidence of plague than counties located below the escarpment. Consistently recorded meteorological data are lacking from this area but the higher elevation sites are perceived to receive some of the highest rainfall in the West Nile region and experience lower temperatures than neighboring low-lying areas which report average monthly temperatures of 22--35°C.[21] This combination of elevated rainfall and moderate temperature is consistent with conditions observed in other plague-endemic regions of Africa such as Kenya, Tanzania, Democratic Republic of Congo, Madagascar and southern Africa as observed by Davis upon evaluation of the relationship between seasonal occurrence of plague and meteorological variables in localities throughout Africa.[37, 38] Initiation of plague activity in Africa was associated with increased precipitation and temperatures above 15°C while case reports decreased during drier periods and when average temperatures exceeded 27°C, an observation similar to ones made earlier in Vietnam.[39-41] Likewise, in North America occurrence of plague epidemics and epizootics have previously been associated with temperature and rainfall.[35, 37, 42-44] The reasons for why occurrence of plague cases is related to temperature in Africa are unknown and merit further studies into both the ecology of plague in rodent reservoirs, flea vectors, and *Y. pestis* transmission dynamics and potential alterations in human

behavior which might increase the risk of *Y.pestis* transmission. For instance, it has been proposed that lower temperatures in this region cause individuals to sleep in the cooking hut near the fire to stay warm which will increase their risk for plague by bringing them into contact with rodent fleas responsible for plague transmission.[21] Alternatively, temperature and relative humidity could influence host-seeking behavior of fleas.

Accuracy of the models to predict case occurrence or incidence of reported plague cases was improved by incorporating remotely-sensed environmental indicators (e.g., temperature, brightness, land cover) together with elevation. However, because these variables were aggregated by parish in order to match the finest spatial resolution that was georeferenced for plague cases, the biological meaning of these indicators needs to be interpreted with care. Several variables included in the models are indicative of differences in soil or vegetation types (e.g., Landsat bands 3 and 7, brightness, greenness, and wetness).[22, 29] It is difficult to interpret the significance of individual Landsat bands that are included in the models as in what environmental characteristics a high or low band value indicates. Although these Landsat bands may not provide specific information on the environmental characteristics correlated with plague incidence, the Landsat satellites provide global coverage and provide opportunities to extrapolate the model to areas where surveillance information is lacking (e.g., the DRC located west of the study area). Further classification of vegetation and soil types in the West Nile region is required to identify precise environmental correlates of risk and develop hypotheses regarding biological linkages between the remotely-sensed indicators and risk of *Y. pestis* transmission from rodent-flea cycles to humans.

Plague incidence also was positively associated with the number of land cover types occurring within parishes. Others have suggested that habitat heterogeneity or fragmentation may be important for inter-epizootic maintenance of *Y. pestis*.[1, 45-49] For example, sufficiently large plague-susceptible small mammal populations or communities may be separated into distinct subpopulations or metapopulations by landscape features that restrict movement and thus either allow some subpopulations to be unaffected by plague epizootics or at least slow epizootic spread enough for some populations to recover prior to being affected by the next epizootic. Thus, habitat heterogeneity could increase the probability of local persistence of plague foci. This underscores the need for studies on rodent and fleas faunas, rodent movement patterns, and the effect of agricultural practices within clearly defined habitat types in plague-endemic areas in Africa.

Although 74% of the variation in parish level incidence was explained by the model, there were some notable outliers that affected model performance. We believe the most critical areas of discordance between actual and predicted plague incidence are those parishes that did not report cases, but where the model expects cases to occur. This includes Eruba, Kuluva and Ezuku parishes in northern Vurra county which were expected to report incidences of 25--50 cases per 1,000 population but no cases were observed between 1999--2007 (Figure 3). Possible reasons for the overprediction include underreporting of cases during the study period due to unstable political conditions along the DRC border. Also, model development was based on nine years of epidemiological surveillance. However, plague is characterized by long periods of quiescence followed by rapidly spreading epizootics.[1, 50] Therefore, it is possible that some parishes that did not

report cases but were identified as ecologically conducive for plague activity could have been sampled during a locally quiescent period. For example, the model predicted an incidence of 18 cases per 1,000 population for Central Parish in southeastern Okoro County, yet no cases were reported from 1999--2007. However, a retrospective record review from 1986-1998 indicated hundreds of plague cases were reported to Paidha Clinic located within Central Parish. Similarly, during 2008 a patient with plague was reported from Ezuku parish in northern Vurra County where no cases were reported from 1999-2007. Together, these observations suggest that the areas identified by the model as high risk are ecologically conducive for plague activity but outbreaks may occur more sporadically. Additionally, ecological similarities between other parishes where plague was reported and the outliers in northern Vurra County lead to the speculation that the overprediction of the model in these parishes may be due to reporting anomalies. These parishes in northern Vurra County are therefore prime candidates for prospective studies and enhanced plague surveillance.

There were several parishes in southern and central Vurra County where the model expected fewer cases than actually occurred. Our model development was based on clinically diagnosed plague cases because standard laboratory confirmation of cases was not a common practice in this area.[51, 52] It is possible that during outbreaks the index of suspicion for plague was elevated and this may have resulted in over-reporting of plague cases in some parishes. Additionally, the underprediction of the model may be attributed to overreporting of plague cases from health clinics bordering the Democratic Republic of Congo (DRC) due to patients crossing the border to seek care in Uganda. While we attempted to account for this by excluding patients who reported that their

village of residence was in DRC, village of residence was self-reported and anecdotal evidence suggests that some patients from DRC may inaccurately state that their village of residence is in Uganda.

Plague prevention and control recommendations typically include implementation of vector control, reduction of rodent abundance through elimination of harborage or food sources in and around homes[12, 53-55], or rodent control measures when appropriate[54], maintenance of surveillance activities that can aid in predicting where future cases are likely to occur[19], and education campaigns aimed at raising awareness of the disease among health care workers and the public.[56] The spatial resolution of the model is useful for identifying parishes at elevated risk for plague, which may aid in identifying clinics that are most likely to see plague patients and suggests areas where surveillance should be enhanced. This information can be used to target educational campaigns and increase awareness at the parish level. However, because risk within individual parishes is considered homogeneous by the model, it is not sufficient for identifying villages that are most at risk and therefore is of limited use for spatially targeting vector control activities. Future development of spatial risk models based on exposure site locations for plague cases (and negative control locations) would provide fine resolution outputs that may be used to target vector control activities and could serve as a tool for further targeting enhanced surveillance activities. Such information could also explain parish-level model errors. For example, if villages at highest risk within a parish are a long distance to a health care clinic, cases may be under-reported. Identifying these villages could aid in mobilizing resources to these under-served areas at elevated risk. In addition, fine-scale spatial risk modeling could aid in refining the understanding of the ecological factors that

are predictive of risk. This information would be useful for extrapolating the risk assessment model to areas where consistent enhanced epidemiological surveillance activities are lacking and where the burden of disease remains unclear. For example, models from the West Nile region in Uganda very likely are applicable to areas with similar elevations and ecological characteristics in the northeastern part of the DRC directly across the border from the study area.

Acknowledgments

We thank Chris Sexton for assisting in the extraction and compilation of epidemiological data for the area; Patrick Kariuki and Wm. Steve Helterbrand for geographic information system data for the study area; Emily Zielinski-Gutierrez, Heidi E. Brown and Ingrid Weber for helpful discussions.

Authors' addresses: Anna M. Winters, J. Erin Staples, Paul S. Mead, Kevin Griffith, Russell E. Enscore, Kenneth L. Gage, and Rebecca J. Eisen, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, 3150 Rampart Rd. Fort Collins, CO 80522; Asaph Ogen-Odoi, deceased, formerly of the Uganda Virus Research Institute, Entebbe, Uganda, Nackson Babi, Nicholas Owor, also Uganda Virus Research Institute; Anna M. Winters, Lars Eisen, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO 80523; E-mail: AWinters1@cdc.gov (Anna M. Winters).

Table V-1. Multivariate logistic regression model parameter estimates
Parameter estimates for the multivariate logistic regression model predicting elevated versus low risk of human plague in portions of Arua and Nebbi districts, Uganda.

	Parameter estimate		Likelihood ratio test		
Model covariates	Estimate	S.E.	χ^2	df	P
Intercept	2.90	2.78	1.09	1	0.30
1,300 m elevation cut-off	6.48	1.62	16.06	1	0.00
Wetness (mean)	0.29	0.08	11.98	1	0.00
Greenness (mean	-0.13	0.07	3.20	1	0.07
Brightness (maximum)	0.004	0.00	0.83	1	0.36

Table V-2. Evaluation of the logistic regression model
Evaluation of the logistic regression model for parishes classified as having elevated or low plague risk compared to reported plague cases.

	Actual classification for presence of plague cases		
Predicted risk of plague[a]	Yes	No	% correct
Elevated	40	6	87%[d]
Low	3	67	96%[e]
% correct	93%[b]	92%[c]	

[a] Probability cut-off value of $P \geq 0.55$; [b] Sensitivity; [c] Specificity
[d] Positive predictive value; [e] Negative predictive value

Table V-3. Multivariate linear regression models
Multivariate linear regression models for incidence of human plague including model parameters.

							AIC Values			Model Parameters						
Model no.	k	SSE	df	MSE	r^2	r^2 adjusted	AIC	ΔAIC	Intercept	Elev Cutoff	B3 Mean	B7 Mean	Elev Min	Surface Temp	Land cover Variety	Bright Min
1	2	8334.78	47	177.34	0.07	0.05	118.86	39.48	0.09	10.10						
2	3	7175.58	46	155.99	0.20	0.16	108.62	29.25	-60.06	16.54	1.41					
3	4	4752.27	45	105.61	0.47	0.43	101.59	22.22	-74.71	30.17	4.95	-3.78				
4	5	4066.52	44	92.42	0.55	0.50	95.87	16.50	-140.59	25.32	7.04	-5.47	0.04			
5	6	3354.07	43	78.00	0.63	0.58	91.29	11.92	-207.24	20.65	8.48	-6.63	0.05	5.29		
6	7	3036.92	42	72.31	0.66	0.61	84.75	5.37	-226.00	21.59	8.91	-6.85	0.06	5.08	2.67	
7*	8	2845.86	41	69.41	0.68	0.63	82.24	2.87	-222.39	22.34	8.42	-6.22	0.05	6.78	2.51	-0.17

*Selected model; k, the number of estimated parameters included in the model; SSE, sum of squared errors; df, degrees of freedom; MSE, mean squared error; AIC, Aikaike Information Criterion; ΔAIC = AIC for model – AIC for most parsimonious model; Elev Cutoff, a binary variable indicating parishes with mean elevation above or below 1,300 m; B3 and B7 Mean, mean values for Landsat band 3 and band 7; Elev Min, minimum elevation within a parish; Surface temp, surface temperature calculated based on Landsat band 6; Landcover Variety, heterogeneity (number) of land cover types within a parish; Bright Min, minimum soil brightness within a parish.

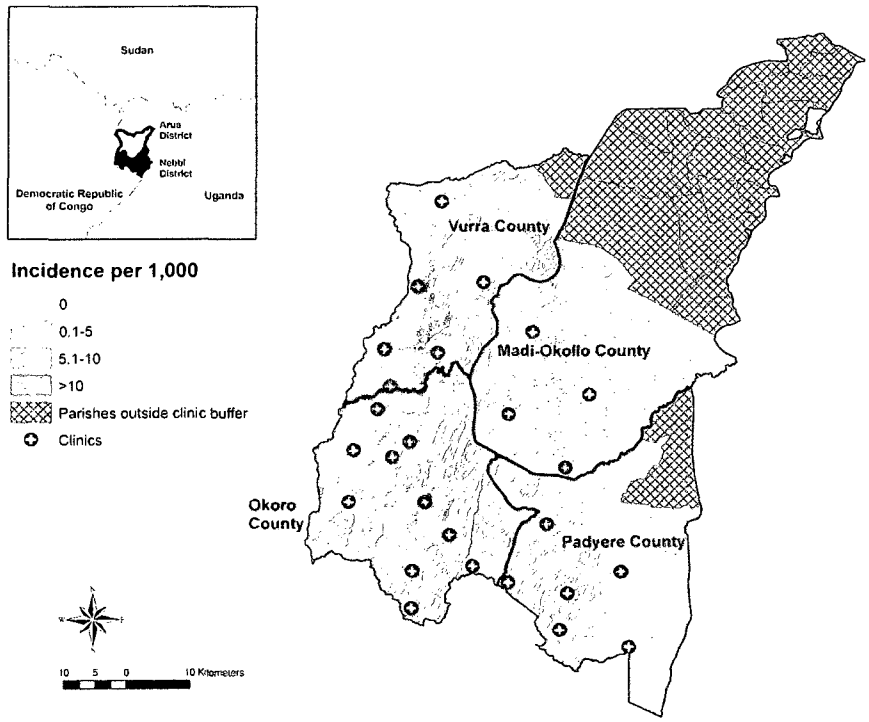

Figure V-1. Incidence of human plague by parish
Reported cumulative incidence per 1,000 population of human plague (1999-2007) by parish in Arua and Nebbi districts.

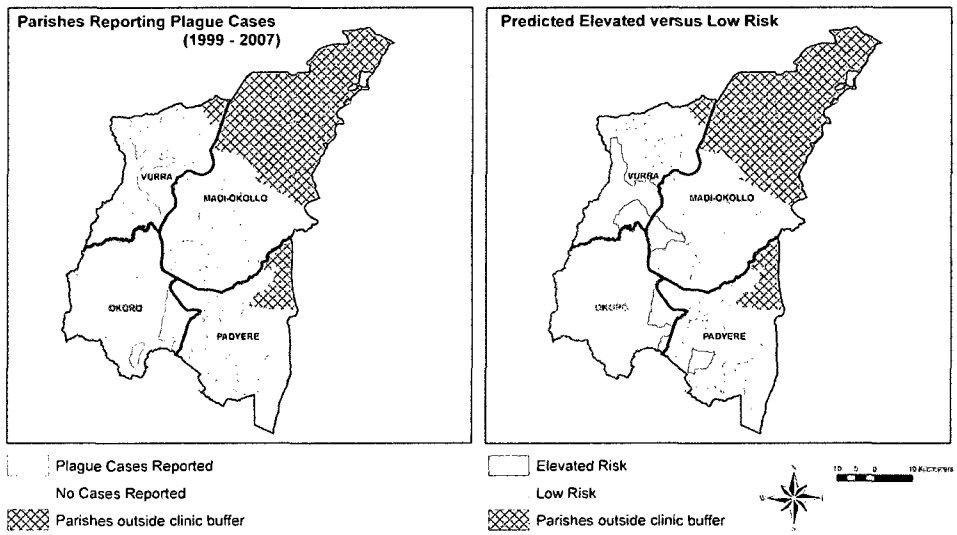

Figure V-2. Plague case reporting compared to model predictions
Parishes reporting plague cases compared to the prediction from the multivariate logistic regression model for elevated versus low risk of plague.

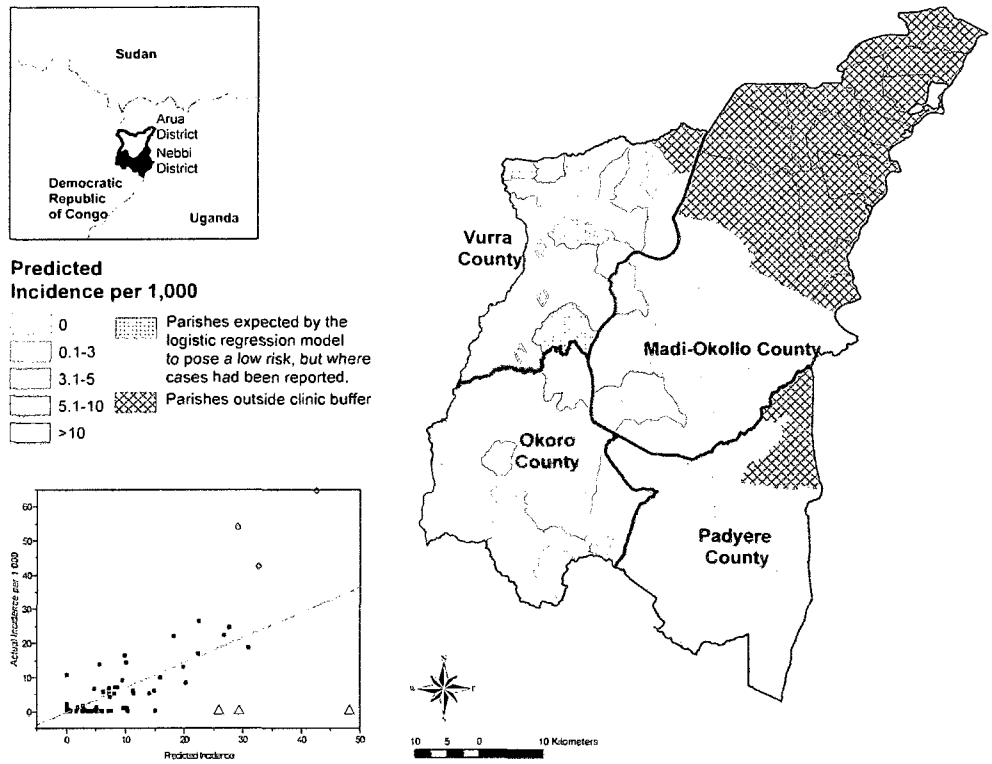

Figure V-3. Multivariate linear regression model predicting plague incidence
Multivariate linear regression model predicting plague incidence per 1,000 population. Inset graph identifies parishes in Vurra County where the model over (diamond) or under (asterisk) predicts incidence.

References Cited

1. Gage KL, Kosoy MY, 2005. Natural history of plague: Perspectives from more than a century of research. *Ann Rev Entomol 50:* 505-528.

2. Dennis DT, Gage KL: Plague. In: Cohen J, Powderly WG, eds. Infectious Diseases, vol 2. London: Mosby, 2003; 1641-1648.

3. WHO, 2004. Human plague in 2002 and 2003. *Wkly Epidemiol Rec 79:* 301-308.

4. WHO, 2005. Outbreak news index 2005. *Wkly Epidemiol Rec 80:* 433-440.

5. Kilonzo BS, 1999. Plague epidemiology and control in eastern and southern Africa during the period 1978 to 1997. *Central African Journal of Medicine 45:* 70-76.

6. WHO, 1983. Weekly Epidemiological Record. *Wkly Epidemiol Rec 58:* 265-272.

7. Aikimbajev A, Meka-Mechenko T, Temiralieva G, Bekenov J, Sagiyev Z, Kaljan K, Mukhambetova AK, 2003. Plague in Kazakhstan at the present time. *Przegl Epidemiol 57:* 593-598.

8. Hull HF, Montes JM, Mann JM, 1987. Septicemic plague in New Mexico. *J Infect Dis 155:* 113-118.

9. Mann JM, Schmid GP, Stoesz PA, Skinner MD, Kaufmann AF, 1982. Peripatetic plague. *JAMA 247:* 47-48.

10. MMWR, 1994. Human plague - United States, 1993-1994. *MMWR 43:* 242-246.

11. MMWR, 2002. Imported plague: New York City, 2002. *MMWR 52:* 725-728.

12. MMWR, 2006. Human plague - four states, 2006. *MMWR 55:* 940-943.

13. Crook LD, Tempest B, 1992. Plague. A clinical review of 27 cases. *Arch Intern Med 152:* 1253-1256.

14. Pollitzer R: Plague. World Health Organization Monograph Series No. 22. Geneva, Switzerland: World Health Organization, 1954.

15. Eisen RJ, Reynolds PJ, Ettestad P, Brown T, Enscore RE, Biggerstaff BJ, Cheek J, Bueno R, Targhetta J, Montenieri JA, Gage KL, 2007. Residence-Linked Human Plague in New Mexico: A Habitat-Suitability Model. *Am J Trop Med Hyg 77:* 121-125.

16. Nakazawa Y, Williams R, Peterson AT, Mead P, Staples E, Gage KL, 2007. Climate change effects on plague and tularemia in the United States. *Vector Borne Zoonotic Dis 7:* 529-540.

17. Neerinckx SB, Peterson AT, Gulinck H, Deckers J, Leirs H, 2008. Geographic distribution and ecological niche of plague in sub-Saharan Africa. *Int J Health Geogr 7:* 54.

18. Barnes AM, 1980. Plague surveillance and control. *WHO Chronicle 34:* 139-143.

19. Gage KL: Plague Surveillance. Plague Manual: Epidemiology, Distribution, Surveillance and Control. Geneva: WHO, 1999; 135-165.

20. Hopkins GHE. Report on rats, fleas and plague in Uganda. East African Standard, Ltd., 1949; 52.

21. Orochi-Orach, 2002. Plague outbreaks the gender and age perspective in Okoro County, Nebbi district, Uganda. *Agency for Accelerated Regional Development.*

22. Jensen J: Remote Sensing of the Environment - An Earth Resource Perspective, 2nd ed. Upper Saddle River: Prentice Hall, 2007.

23. Chander G, Markham B, 2003. Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. *IEEE Transactions on Geoscience and Remote Sensing 41:* 2674-2677.

24. Chen Y, Guo D, Zhou T, 1999. Noise removal of thermal image and the determination of the temperatures. *Remote Sensing Information 11:* 7-9.

25. Han-qiu X, Ben-qing C, 2004. Remote sensing of the urban heat island and its changes in Xiamen City of SE China. *J Environ Sci 16:* 276-281.

26. Landsat 7 Science Data Users Handbook. *Available at: http://landsathandbook.gsfc.nasa.gov/handbook/handbook_toc.html. Accessed Dec 3, 2008, 2008.*

27. USGS - EROS Data Center, 2006. MRLC 2001 Image Processing Procedure.

28. Getting Started with ENVI - Version 4.5. *Available at: http://www.ittvis.com/ProductServices/ENVI/ProductDocumentation.aspx. Accessed Dec 3, 2008,*

29. Jensen J: Introductory Digital Image Processing: A Remote Sensing Perspective, 2nd ed. Upper Saddle River: Prentice Hall, 1995.

30. Eisen RJ, Borchert JN, Holmes JL, Amatre G, Van Wyk K, Enscore RE, Babi N, Atiku LA, Wilder AP, Vetter SM, Bearden SW, Montenieri JA, Gage KL, 2008. Early-phase transmission of *Yersinia pestis* by cat fleas (*Ctenocephalides felis*) and their potential role as vectors in a plague-endemic region of Uganda. *Am J Trop Med Hyg 78:* 949-956.

31. Gelman A, Hill J: Data Analysis Using Regression and Multilevel/Hierarchical Models. New York: Cambridge University Press, 2007.

32. Burnham K, Anderson D: Model selection and multimodel inference: A practical information-theoretic approach. New York: Springer, 2002.

33. Sall J, Creighton L, Lehman A. JMP Start Statistics. 3rd ed. Belmont: Brooks/Cole, 2005; 584.

34. Eisen RJ, Enscore RE, Biggerstaff BJ, Reynolds PJ, Ettestad P, Brown T, Pape J, Tanda D, Levy CE, Engelthaler DM, Cheek J, Bueno R, Targhetta J, Montenieri JA, Gage KL, 2007. Human plague in the southwestern United States 1957-2004: spatial models of elevated risk of human exposure to *Yersinia pestis*. *J Med Entomol 44:* 530-537.

35. Eisen RJ, Glass GE, Eisen L, Cheek J, Enscore RE, Ettestad P, Gage KL, 2007. A spatial model of shared risk for plague and hantavirus pulmonary syndrome in the southwestern United States. *Am J Trop Med Hyg 77:* 999-1004.

36. Velimirovic B, Zikmund V, Herman J, 1968. Plague in the Lake Edwards focus: the Democratic Republic of Congo, 1960-1966. *Z Tropenmed Parsit 20:* 373-387.

37. Davis DH, 1949. Current methods of controlling rodents and fleas in the campaign against bubonic plague and murine typhus. *J R Sanit Inst 69:* 170-175.

38. Davis DH, 1953. Plague in Africa from 1935 to 1949; a survey of wild rodents in African territories. *Bull WHO 9:* 665-700.

39. Cavanaugh DC, Marshall JD, Jr., 1972. The influence of climate on the seasonal prevalence of plague in the Republic of Vietnam. *J Wildl Dis 8:* 85-94.

40. Cavanaugh DC, Dangerfield HG, Hunter DH, Joy RJ, Marshall JD, Jr., Quy DV, Vivona S, Winter PE, 1968. Some observations on the current plague outbreak in the Republic of Vietnam. *Am J Public Health Nations Health 58:* 742-752.

41. Marshall JD, Ouy DV, Gibson FL, Dung TC, Cavanaugh DC, 1967. Ecology of plague in Vietnam: commensal rodents and their fleas. *Mil Med 132:* 896-903.

42. Enscore RE, Biggerstaff BJ, Brown TL, Fulgham RE, Reynolds PJ, Engelthaler DM, Levy CE, Parmenter RR, Montenieri JA, Cheek JE, Grinnell RK, Ettestad PJ, Gage KL, 2002. Modeling relationships between climate and the frequency of human plague cases in the southwestern United States, 1960-1997. *Am J Trop Med Hyg 66:* 186-196.

43. Parmenter RR, Yadav EP, Parmenter CA, Ettestad P, Gage KL, 1999. Incidence of plague associated with increased winter-spring precipitation in New Mexico. *Am J Trop Med Hyg 61:* 814-821.

44. Gage KL, 2008. Climate and vectorborne diseases. *Am J Prev Med 35:* 436-450.

45. Eisen RJ, Gage KL, 2009. Adaptive strategies of *Yersinia pestis* to persist during inter-epizootic and epizootic periods. *Vet Res 40:* 1.

46. Keeling MJ, Gilligan CA, 2000. Metapopulation dynamics of bubonic plague. *Nature 407:* 903-906.

47. Snall T, O'Hara RB, Ray C, Collinge SK, 2008. Climate-driven spatial dynamics of plague among prairie dog colonies. *Am Nat 171:* 238-248.

48. Collinge SK, Johnson WC, Ray C, Matchett R, Grensten J, Cully JF, Gage KL, Kosoy MY, Loye JE, Martin AP, 2005. Landscape structure and plague occurrence in black-tailed prairie dogs on grasslands of the western USA. *Landscape Ecology 20:* 941-955.

49. Davis S, Leirs H, Viljugrein H, Stenseth NC, De Bruyn L, Klassovskiy N, Ageyev V, Begon M, 2007. Empirical assessment of a threshold model for sylvatic plague. *J R Soc Interface 4:* 649-657.

50. Barnes AM, Maupin GO, 1982. Observations on the biting of humans by *Euhoplopsyllus-Glacialis-Affinis* (Siphonaptera, Pulicidae) and a review of is plague-transmission potential. *J Med Entomol 19:* 748-749.

51. Poland JD: Diagnosis and Clinic Manifestations. Plague Manual: Epidemiology, Distribution, Surveillance and Control. Geneva: World Health Organization, 1999.

52. Chu MC. Laboratory manual of plague diagnostics. Geneva: Centers for Disease Control and Prevention and World Health Organization, 2000; 129.

53. Akiev AK, 1982. Epidemiology and incidence of plague in the world, 1958-79. *Bull WHO 60:* 165-169.

54. Gratz NG: Control of Plague Transmission. Plague Manual: Epidemiology, Distribution, Surveillance and Control. Geneva: World Health Organization, 1999; 97-134.

55. Poland JD, Barnes AM: Plague. In: Steele JH, ed. CRC Handbook Series in Zoonoses. Section A: Bacterial, Rickettsial and Mycotic diseases. Volume I. Boca Raton: CRC Press Inc., 1979; 515-559.

56. Gage KL: National Health Services in Prevention and Control. Plague Manual: Epidemiology, Distribution, Surveillance and Control. Geneva: WHO, 1999; 167-171.

VI. Evaluation of an Early Detection System for Dengue in Merida, Mexico

Anna M. Winters

Abstract

Early detection of disease outbreaks can aid public health practitioners in the timely application of interventions to reduce disease case numbers. I used dengue case data from Merida, Mexico to evaluate the feasibility of an early detection system based on comparison with historical data to identify dengue outbreaks in their early stages. A "window of vigilance" was defined from epidemiological weeks 24—36 during which outbreaks may be detected early by monitoring the number of dengue cases and comparing against historical data from non-outbreak years. Furthermore, individual thresholds were developed for basic geostatistical areas (BGSAs) within Merida to provide alerts when current incidence exceeded historic levels indicating that interventions should be targeted to these areas. The number of alerts varied by threshold algorithm and the southeast region of Merida reported the greatest number of alerts per BGSA. Use of the early detection system may aid the local vector control program to target their response activity during outbreak times, when large numbers of dengue cases are reported, and response to all cases is not be feasible due to time and resource limitations. In conclusion, the early detection system has potential to provide Merida public health authorities with a practical resource to recognize when the current dengue burden is exceeding historical norms. The same approach may also be applicable and useful in other dengue endemic areas.

KEY WORDS: dengue, early detection, strategic planning, vector control, Google Earth, Mexico

Introduction

In public health practice, aberrations of reported disease case numbers can be monitored by early detection systems to alert of building outbreak or epidemic events that may necessitate intervention. Early detection systems are differentiated from early warning systems which monitor indicators (e.g., meteorological data such as rainfall and temperature, or the immune status of a population) used to predict timing and location of an increase in disease prevalence.[1, 2] Early detection systems have been developed for a number of diseases.[3-5] In the U.S., many of these diseases are reported by state health departments to the Centers for Disease Control and Prevention and epidemic alert information is made available weekly to clinicians, epidemiologists and other members of the public health field.[6, 7] Early detection of mosquito-borne disease outbreaks is crucial given their often explosive nature. Interventions that focus on reducing the number of infected mosquitoes must be applied in a timely fashion to be most effective (mosquito control interventions are only minimally effective if applied at the height of an outbreak/epidemic, instead of early when the number of reported disease cases begins to increase above the norm). Early detection systems have been developed for several mosquito-borne diseases including malaria,[1, 8-12] dengue,[13-15] and Ross River virus disease.[16, 17] Multiple methods have been applied and results vary; however the general theme is that early detection for vector-borne disease is cost effective[8] and although it offers only limited lead time for preparation and implementation of preventative measures, it can aid in a timely and effective response to reduce morbidity and mortality.

Early detection systems may aid dengue control programs, in particular, because of the volatile nature of dengue virus transmission in urban environments. Dengue

viruses are transmitted primarily by *Aedes aegypti,* a highly anthropogenic and anthropophilic mosquito vector which thrives in domestic environments.[18, 19, 20, 21] High population density, housing which lacks air conditioning or adequate screening and allows entrance of mosquito vectors, lack of piped water which necessitates water storage (and creates larval development sites) and presence of water-collecting debris (e.g. small plastic containers, tires) all contribute, directly or indirectly, to dengue virus transmission.[22-25] If these risk factors are combined with the introduction of a new dengue virus serotype, the intensity of dengue virus transmission can become very high, because herd immunity to the new serotype may be minimal, and antibody dependent enhancement may lead to cases of dengue hemorrhagic fever (DHF), the severe and sometimes fatal clinical manifestation of dengue.[18, 21] Currently, there are no approved vaccines against dengue viruses and vector control is the only means to control disease outbreaks.[20, 21, 26, 27] However, mosquito control interventions take time to mobilize, and should be applied early in the outbreak in order to be most effective. Currently, vector control efforts are often initatied late during an outbreak, perhaps even after the number of human cases has peaked. Early detection of dengue outbreaks combined with characterization of the circulating dengue viruses is also beneficial to rapidly identify introductions of new virus serotypes or genotypes, which may cause an increase in the total number of cases of dengue fever (DF) and DHF. Early identification of an increase in cases also serves to alert the necessary medical services to better ensure that sufficient capacity for diagnostics and supportive care are in place.

There has been relatively little development of early detection systems to aid in the timely detection of dengue outbreaks.[13-15] Rigau-Perez and colleagues developed and

evaluated a Centers for Disease Control and Prevention deviation chart for its utility in detecting dengue outbreaks in Puerto Rico and found that it provided a specific and timely signal for dengue control efforts.[14] Efforts in Thailand identified a significant deviation from the monthly average incidence and termed these periods "epidemic months" thereby indicating the need for launch of intervention activities.[13] Our literature review revealed no publications on early detection systems from Mexico.

In southern Mexico, *Ae. aegypti* adults may be present and dengue cases may occur throughout the year but abundance of both typically peak within the rainy season[28] from July to October.[29-32] This distinct seasonality provides an opportunity for increased vigilance during a relatively short time period when case numbers can be expected to diverge between low- and high-transmission years. This can be used to help determine projected needs in coming months for clinical services as well as vector control, which often requires labor-intensive methods such as physical elimination of larval development sites (disposable containers, debris, tires, etc), application of larvicides to kill immatures in water-holding containers that cannot be removed, and control of adult *Ae. aegypti* through space spraying of chemical insecticides in and around homes.[25]

The objective of this research was to evaluate the feasibility of an early detection system capable of identifying elevated dengue case loads, potentially indicating an impending dengue outbreak. This was done using retrospective data on laboratory-confirmed dengue cases in the city of Merida, Mexico. Within the overall objective, specific aims were to:

1) Identify, for the city of Merida, the period of time within a year when the number of dengue case reports typically begin to increase indicating the early stages of an oncoming outbreak.

2) Develop thresholds based on historical data that characterize the expected weekly dengue incidence rates for small areas (basic geostatistical areas, [BGSAs]) within the city of Merida.

3) Compare recent incidence levels by week and BGSA to historically-derived thresholds in order to identify when observed values are exceeding expected incidence levels alerting to a possible outbreak.

4) Develop a Google Earth™ animation of the spatial and temporal location of alerts indicating impending outbreaks that can be used by vector control personnel to target interventions.

Materials and Methods

Study Area.

The study focused on the city of Merida, which is located on the Yucatan Peninsula of southern Mexico (Fig. VI-1) and has a population of ~730,000 (Instituto Nacional de Estadística y Geografía [INEGI], 2005). This region has a subtropical climate (annual average temperature in the 26°—27°C range) with two distinct seasons; a rainy season from May to October and a dry season from November to April.[32] The abundance of *Ae. aegypti* adults is largely correlated with rainfall, with a peak from July—October.[29, 30]

Merida is divided into 329 basic geo-statistical areas (BGSAs), each of which represent 20—80 city blocks, with an average population of ~2,000.[33-35] BGSAs have been used in Mexico by INEGI since 1980 to group demographic and cartographic information.[33, 34] We used BGSAs versus other potential spatial units such as neighborhoods (*colonias*) as the geographic unit of analysis because the population size of BGSAs is better defined. We also aggregated the BGSAs into 5 regions to improve the clarity of presentation of the study results. BGSAs were assigned to the central region if their centroids fell within Colonia Centro (the central/downtown area of Merida). North-south and east-west axes were then constructed with their vertices at the centroid of Colonia Centro. BGSAs located outside the central region were assigned to the quadrant (northwest, northeast, southwest or southeast) in which their centroid was located (Fig. VI-1).

Dengue case data.

A total of 3,438 laboratory confirmed dengue cases were reported in Merida from 1997—2007. These cases were confirmed by Laboratorio Estatal de Referencia Epidemiological using IgM-capture ELISA or by Laboratorio de Arbovirología, Centro de Investigaciones Regionales Dr. Hideyo Noguchi, at Universidad Autónoma de Yucatán, using IgM capture ELISA and RT-PCR. Dengue cases were georeferenced based on address of residence. Initial georeferencing efforts were conducted using Sistema de Información Geographica Municipal, an online mapping tool provided by the Gobierno Municipal of Merida, Yucatan (SIG Merida; http://www.merida.gob.mx/sig/). Cases whose geographic locations were unresolvable with SIG Merida were located on

the ground during the summer of 2008 by driving to the reported location of residence and collecting the geographic coordinates of the residence using a global positioning system (GPS; Garmin GPSmap 60CSx, Olathe, KS). Geographic coordinates for 5.5% (132) of the cases were not attainable as the address was incomplete or incorrect and these cases were not included in the study. The remaining georeferenced cases were aggregated to BGSAs. A number of BGSAs (n = 58) lacked INEGI population data. Dengue cases that were georeferenced to these BGSAs were excluded from the study as incidence values could not be calculated. The final epidemiological database consisted of 2,720 georeferenced cases (Table VI-1). Dengue incidence per 10,000 population was calculated by epidemiological week and by BGSA for 1997—2007 and was included in the epidemiological database. For cases reported during 1997—2002, INEGI census data from 2000 was used to calculate incidence; 2005 INEGI census data was used to calculate incidence for cases reported from 2003—2007. Additionally, incidence rates (incidence per 100,000 population) were aggregated to the city of Merida and plotted by epidemiological week for years 1997 through 2007 and using all reported dengue cases. Data for locations of dengue cases were provided by Servicios de Salud de Yucatan to Drs. Barry J. Beaty and Lars Eisen of Colorado State University under protocols approved by the Institutional Review Board at Colorado State University.

Threshold setting -- Overview.

Multi-year (1998—2005) incidence rate data, based on a three-week unweighted moving average (UWMA), were used to characterize the expected weekly values for dengue incidence by BGSA. We considered 1997 an outbreak year as the number of

cases reported during that year was approximately 13 times greater than the average number of cases reported during 1998—2005. Inclusion of 1997 would inflate the expected weekly values of dengue incidence and decrease the sensitivity of the early detection system. The expected value for each epidemiological week was calculated using incidence data for the previous, current and following epidemiological weeks. This was done to smooth the incidence rate data, and to account for slight variations in the timing of meteorological events (e.g. rainfall) from year to year that may affect the seasonal patterns of abundance of mosquito vectors and intensity of dengue virus transmission.[36] Thresholds derived from historic incidence rates were calculated for each BGSA and epidemiological week. To evaluate whether prospective incidence rates would exceed the thresholds (and thus produce an alert), incidence data from 2006 and 2007 were withheld and compared to the threshold values of dengue incidence by BGSA and week.

Threshold-setting method.

Cantelli's inequality theorem was applied to determine the k number of standard deviations necessary to add to the historical incidence average to approximate thresholds that would indicate a 95% (p = 0.05) and 75% (p = 0.25) probability that future dengue incidence observations would be less than or equal to the threshold values within that BGSA and for the specific week (Fig. VI- 2). Cantelli's inequality theorem assumes:[37]

$$\Pr(X - \mu \geq k\sigma) \leq \frac{1}{1+k^2}$$

where X is the observed value (incidence value for a specific week and BGSA for 2006 or 2007) and μ is the average incidence for the specific week. Cantelli's inequality theorem

was solved for k to indicate the number of standard deviations that must be added to the mean (μ) in order to arrive at the threshold in which there is a 75% ($p = 0.25$) or 95% ($p = 0.05$) probability of capturing future observations for the specific week and BGSA (Fig. VI-2). These cutoffs were termed Threshold 1 and Threshold 2, respectively. Cantelli's inequality theorem was used for these calculations to account for the nonparametric distribution of the incidence data.[37-40] We also calculated the maximum incidence rate reported within years 1998—2005 to compare to current incidence rate data.

Alert calculations.

Two alert rules were defined: 1) a "1x alert" was indicated if the current incidence exceeded the threshold values during one week; 2) a "2x alert" was indicated if the current incidence exceeded threshold values for two consecutive weeks (Fig. VI-3). The 2x alert was applied with the intention of improving the specificity of the alert system for any given threshold. Alerts were calculated for both Threshold 1 and 2 (Table VI-1).

Visualization of threshold values and alerts.

A Google Earth™ (Google, Mountain View, CA) animation was constructed that provided visualization of when and where alerts were triggered for each BGSA and by epidemiological week. For example, BGSAs which reported incidence during an epidemiological week that exceeded Threshold 1 values were indicated as alerts. The incidence value which caused the alert as well as the Threshold 1 value were displayed in the animation. A Keyhole Markup Language (KML) file was used for expressing geographic locations and for temporal visualization of the data in Google Earth™.

Results

Descriptive results.

A total of 3,438 laboratory confirmed dengue cases were reported in Merida from 1997—2007. The average dengue incidence (1997—2007) was highest during epidemiological week 35, in mid-September. This coincides with the rainy season when abundance of *Ae. aegypti* also peaks. During 1997 and 2007, dengue incidence deviated sharply from incidence reported from 1998—2006. In 1997, incidence began to diverge in epidemiological week 27 and rose sharply to a peak of 21.28 cases per 100,000 population in epidemiological week 35 (Fig. VI - 4). Dengue incidence in 2007 exhibited divergence from 1997—2006 during epidemiological weeks 1—12 when the 2007 city-wide incidence rate averaged 1.31 reported dengue cases per 100,000 population across these weeks. The second and larger 2007 divergence (epidemiological weeks 33—51) had an average of 8.86 reported dengue cases per 100,000 population across the weeks compared to 1.07 per 100,000 population during 1997—2006. The incidence rate for 2007 during the weeks leading up to the second outbreak (epidemiological weeks 31-33) were similar to the rates reported during 1998—2006. By epidemiological week 34, however, the incidence rate in 2007 started to deviate from previous years peaking during epidemiologic week 43 (15.26 cases per 100,000 population, Fig. VI-4).

After excluding cases that were reported to BGSAs which lacked census data, the final epidemiological database used for development of threshold values consisted of 2,720 georeferenced cases (Table VI-1). For Threshold 1, among the 13,104 BGSA-

weeks occurring in 2006, 1x alerts were triggered in 133 (1%) and 2x alerts were triggered in 7 (0.05%) of the BGSA-weeks (Table VI-2). During 2007, the number of alerts was considerably higher, with 851 BGSA-weeks (6.5%) considered 1x alerts and 231 BGSA-weeks (1.8%) identified as 2x alerts. Applying Threshold 2 decreased the total number of alerts in 2006 (except 2x alerts, which remained the same) and 2007 compared to Threshold 1 (Table VI-2).

Spatial distribution of alerts.

Figure VI-5 displays the spatial and temporal distribution of alerts, indicating when and where dengue incidence from 2006 and 2007 exceeded Threshold 1 over two consecutive weeks (2x alert). The most alerts occurred in 2007 (n = 231), with the majority during epidemiological weeks 35—49, peaking in week 40 (n = 23). The southeast region of Merida reported the greatest number of alerts per BGSA for Threshold 1 and 2, for both 1x and 2x alert algorithms. For the Threshold 1 and 2x alert algorithm, in 2007 the southeast region reported 1.28 alerts per BGSA followed by the central region (n = 1.16 alerts per BGSA), northwest region (n = 0.75 alerts per BGSA), southwest region (n = 1.95 alerts per BGSA) and northeast region (n = 0.70 alerts per BGSA, Fig. VI-6).

Google Earth Animation.

Figure VI-7 provides a snapshot of the Google Earth™ animation. Threshold 1 values of dengue incidence are displayed along with alert indications if incidence values from 2006 and 2007 exceeded threshold values by week and BGSA. The animation

shows when and where alerts occurred and provides landmarks helpful to vector control personnel for navigation to BGSAs needing application of preventative measures (Fig. VI-7).

Discussion

We evaluated the feasibility of an early detection system based on historically-derived thresholds to identify dengue outbreaks in a timely fashion. The early detection of incipient outbreaks can aid public health practitioners in the timely application of interventions to reduce peaks in the number of dengue cases.

City-wide window of vigilance in Merida.

A "window of vigilance" was defined, during which outbreaks may be detected early by monitoring the number of dengue cases reported within Merida compared to historical data. We aggregated incidence rates to the city level and compared them to historic city-wide data by epidemiological week (Fig. VI-4). During 1997 and 2007, elevated numbers of dengue cases were reported with the largest incidence deviation from non-outbreak years beginning during epidemiological week 24 in 1997 and epidemiological week 34 in 2007. Public health practitioners should be extra vigilant during this window of time ("window of vigilance") when the city-wide incidence rate in the current year may diverge from historic rates. Within this "window of vigilance", dengue virus transmission may quickly increase to outbreak levels. The current data from Merida indicate that the "window of vigilance" should include epidemiological weeks

24—36 which encompasses the period of time when incidence rates from endemic and outbreak activity years diverge, with case reports rapidly increasing in outbreak years. If city-wide incidence rates rapidly increase during the "window of vigilance", the likelihood of averting or alleviating the outbreak peak is increased if interventions are mobilized promptly, versus waiting until case numbers are excessive and dengue virus transmission is more widespread and intense. Actions triggered during a "window of vigilance" including mobilization of extra laboratory personnel and resources to assist in prompt laboratory confirmation of suspected dengue cases would further enhance the early detection of outbreaks.

BGSA-specific alerts.

Use of a city-wide "window of vigilance" may indicate if an outbreak is incipient, but does not provide a targeted approach to mobilize an intervention within the city. Therefore, historically-derived thresholds specific to BGSAs were developed and BGSA-specific alerts were triggered when current incidence exceeded thresholds. Vector control in Merida is managed by Servicios de Salud de Yucatan (SSY) which allows for overall monitoring of city-wide events (monitoring total case reports during the "window of vigilance"), as well as directing response scenarios specific to individual BGSAs within the city. Centralized, city-wide vector control management is preferential to management conducted at sub-city levels from the perspective that it allows for easier integration of an early detection system that monitors both city-wide events and BGSA-specific events.

Alert algorithms within the early detection system considered two variables: 1) threshold value (Threshold 1 or 2); and 2) number of threshold breaches (1x or 2x alerts).

As expected, Threshold 1, the lower, more sensitive threshold, caused more alerts compared to Threshold 2 for both 1x and 2x alert algorithms. Results indicated that both sets of thresholds and algorithms may be useful for strategic planning purposes in Merida and they will best serve Merida public health programs if they are customized to fit changing resource and capacity concerns. The sensitivity of Threshold 1 would more readily indicate an alert if incidence levels are above historic levels. If Threshold 1 approximations place too great a strain on resource allocation efforts, Threshold 2, the more specific threshold, may be used as an alternate. Furthermore, Threshold 2 will more nearly approximate the upper limit of incidence rates for a given year, and thus may be suitable for long-term capacity and resource planning purposes. For alert triggers, public health authorities could apply Threshold 1 with the 1x alert algorithm if the current objective is to prevent as many dengue cases as possible, regardless of resource availability. During periods when resources are constrained and alerts must be as efficient as possible (e.g. outbreak periods), Threshold 2 with the 2x alert algorithm may be applied as this algorithm is most specific. Additionally, the alert mode (e.g., threshold applied and with the 1x or 2x algorithm) may be adjusted over the season, for example to be more sensitive during the window of vigilance. The use of a hybrid of Threshold 1 and 2 and the ability to adjust the alert modes as needed, will potentially inject an increased level of efficiency into current vector control strategies by informing resource allocation decisions and providing a targeted approach to vector control applications.

Thresholds may also be further adjusted to better suit strategic planning efforts. For example, if Merida lacks the capacity and/or resources to fully account for a disease burden represented by Threshold 1, adjustments of the thresholds can be achieved by

applying Cantelli's Inequality Theorem to calculate the k necessary to achieve any user-defined threshold level. For example, a higher threshold level may be calculated that would effectively increase the specificity of the system and the number of alerts. Current incidence may also be compared to historic maximum incidence rates for each BGSA to identify if they are unprecedented compared to historic reports. Additionally, if threshold alerts are too frequent for current vector control capacity, the alert algorithm visualized on the Google Earth™ weekly map may be adjusted from 2x alerts to alerts defined only if the threshold is exceeded for more than 2 consecutive weeks. Furthermore, weekly Google Earth™ maps may be automatically generated on a prospective basis to indicate how incidence rates from the current week compare to the historic threshold values. These maps may be used by vector control practitioners to plan weekly intervention activities, and to identify the necessary resources for that week.

Nonparametric data.

The application of Cantelli's inequality theorem is a novel approach for threshold setting and was applied because the spatial distribution of the incidence data did not assume a normal distribution.[37] A number of threshold-setting methodologies are similar to the approach used here as they aim to identify points in a disease time series that are outside confidence intervals determined from historical case loads.[10, 14, 15] These methods, however, typically set thresholds by assuming a normal distribution and adding the mean of the incidence data plus 1.96 standard deviations to capture 95% of the cases.[14, 41, 42] This is problematic, however, if the distribution of the incidence data is nonparametric as was the case with the incidence data used in this study and is a common

occurrence in many epidemiological datasets. Alert rules as applied in this analysis are not novel; a number of studies specific to dengue,[13] malaria[9] and other diseases apply alert rules when current incidence values exceed thresholds. For example Barbazan and colleagues (2002) defined clusters of dengue hemorrhagic fever if there were two or more successive epidemic months in the same province in Thailand and found that the occurrence of two consecutive epidemic months in a given area had a high probability (66%) of being followed by further epidemic months.[13]

Considerations.

BGSA was the geographic unit used for calculation of threshold levels and identification of alerts as it was the only geographic subdivision available that was larger than city block while also having associated population figures. The use of BGSAs is somewhat problematic, however. Although used by INEGI for demographic and cartographic information since 1980,[33, 34] BGSAs are not used or referenced by vector-control practitioners in their day-to-day work activities. Rather, vector-control workers use *colonias* as the common unit of reference for navigation purposes. These units, however, are not associated with any official population data, which effectively precluded this unit from use in this study. The *colonia* layer may be added to the Google Earth™ map outputs to provide landmarks to assist vector control practitioners in locating BGSAs under alert.

Dengue surveillance in Merida is passive and depends on case reports from physicians who recognize dengue-like illness.[19] Passive surveillance can be insensitive to dengue virus transmission activity as the clinical picture includes a large number of

asymptomatic and mild cases.[15,43] This is especially problematic during periods of low dengue virus transmission activity when clinicians' index of suspicion also may be low. Furthermore, the length of time from onset of symptoms to laboratory confirmation and reporting can be lengthy (for cases included in this study, the time period from onset of symptoms to laboratory confirmation averaged 11 days; these data were available for 85% of the dengue cases used in this analysis). If this early detection system is applied to guide vector control application, real-time alerts triggered by reported dengue cases may be delayed because of the length of time from onset of symptoms to reporting, meaning that substantial dengue virus transmission may have already occurred, and even peaked, and control measures may have a limited effect. This time period may be shortened by incorporating a "window of vigilance" to encourage heightened index of suspicion in the clinics, and to allocate additional laboratory personnel and resources to aid in quick laboratory confirmation of suspected dengue cases during this critical time period.

Specific to this study, underreporting of dengue may have led to lower historically-derived thresholds as well as decreased the total number of alerts in 2006 and 2007. This is not necessarily problematic if the assumption that the surveillance system remained consistent from 1997 to 2007 is indeed true. Although the case definition for dengue remained the same throughout the years included in the epidemiological database, measurement bias may have occurred from year to year and clinic to clinic based on differences in clinical practices and reporting procedures.[44] However, because individual thresholds were constructed for each BGSA and alerts considered specific to each BGSA (versus calculating an overall city-wide threshold) each BGSA acts as its own reference. Hence, locally specific incidence rate aberrations are detectable, even considering the

limitations of passive surveillance and/or bias associated with access to care, changing case definitions, or other factors within the city of Merida.

Future research should investigate the effectiveness of the approach outlined here compared to interventions directed only by dengue case occurrences (e.g., dengue cases prompt vector control regardless of how this relates to historically-derived threshold values for expected case numbers). Using the historically-derived threshold approach, BGSAs that have historically reported no cases may prompt an alert if even a small number of cases are reported. If these alerts prompt vector control, space sprays or other methods may be applied to the BGSA, when in actuality, there were very few cases that were reported. In comparison, BGSAs that historically report high numbers of dengue cases may be overlooked as alerts would occur only if current incidence exceeds threshold values.

The conditions favorable to an outbreak in Merida are numerous, but typically include rainfall as an important factor. Additional research would be helpful to describe the specific conditions which increase the probability of an incipient outbreak for example temperature, rainfall or introductions of new dengue virus serotypes. Further, use of these conditions within an early warning system may increase the lead time prior to outbreak events and provide additional time for mobilization of appropriate interventions.[1]

To allow vector control practitioners to visualize current and historic incidence that prompted each alert, we included these values within the Google Earth™ animation. These data are presented in a clear and understandable fashion and informs vector control practitioners of the actual incidence data associated with an alert to incorporate into their

decision of whether to initiate vector control applications following an alert. The clear visualization of these data in space and time via Google Earth™, a freely dowloadable mapping program, provides vector control practitioners with data that previously was not readily available. This can improve preparedness and allocation of public health resources in Merida, particularly during outbreak periods. The early detection system may, in the future, be added to the dengue decision support system (DDSS), an informatic tool under development at Colorado State University that will include entomological, epidemiological, and related data and software to aid in decisions of best approaches for local dengue control.[45, 46] Inclusion of an early detection system into the DDSS has potential to not only provide guidance to public health administrators and vector control practitioners for strategic planning, but also to assist in early and targeted allocation of vector control interventions, especially helpful during outbreak periods.

Acknowledgements

I thank the following persons for their contributions to this chapter with regards to providing data or assistance with data management or analysis: Sanika Chitari, Marlize Coleman, Lars Eisen, Saul Lozano-Fuentes, David Penhall, and Dominique Piche from Colorado State University; Jose Arturo Farfan-Ale, Julian Garcia-Rejon, and Maria Alba Loroño-Pino from Universidad Autónoma de Yucatán; Salvador Gomez-Carro, Rosario Najera-Vazquez, and Victor Lira-Zumbardo from Servicios de Salud de Yucatan; Mark Delorey of the Centers for Disease Control and Prevention; and, finally, Miguel Orlans and Benjamin Winters. The study was funded, in part, by the Innovative Vector Control Consortium.

Tables

Table VI-1 Laboratory confirmed dengue cases in Merida, Mexico, 1997-2007.

Year	Laboratory Confirmed Dengue Cases*
1997	861
1998	6
1999	7
2000	1
2001	69
2002	321
2003	11
2004	13
2005	94
2006	201
2007	1,136
Total	2,720

*Includes only cases georeferenced to address of residence and to BGSAs with INEGI census data available.

Table VI-2 Alerts for early detection of dengue outbreaks

Year	Alert Type	Number of Alerts
Threshold 1[a]		
2006	1x	133
2006	2x	7
2007	1x	851
2007	2x	231
Threshold 2[b]		
2006	1x	126
2006	2x	7
2007	1x	808
2007	2x	202

[a] Threshold 1 indicated a 75% probability ($p = 0.25$) of capturing future observations for the specific week and BGSA.

[b] Threshold 2 indicated a 95% probability ($p = 0.05$) of capturing future observations for the specific week and BGSA.

Figures

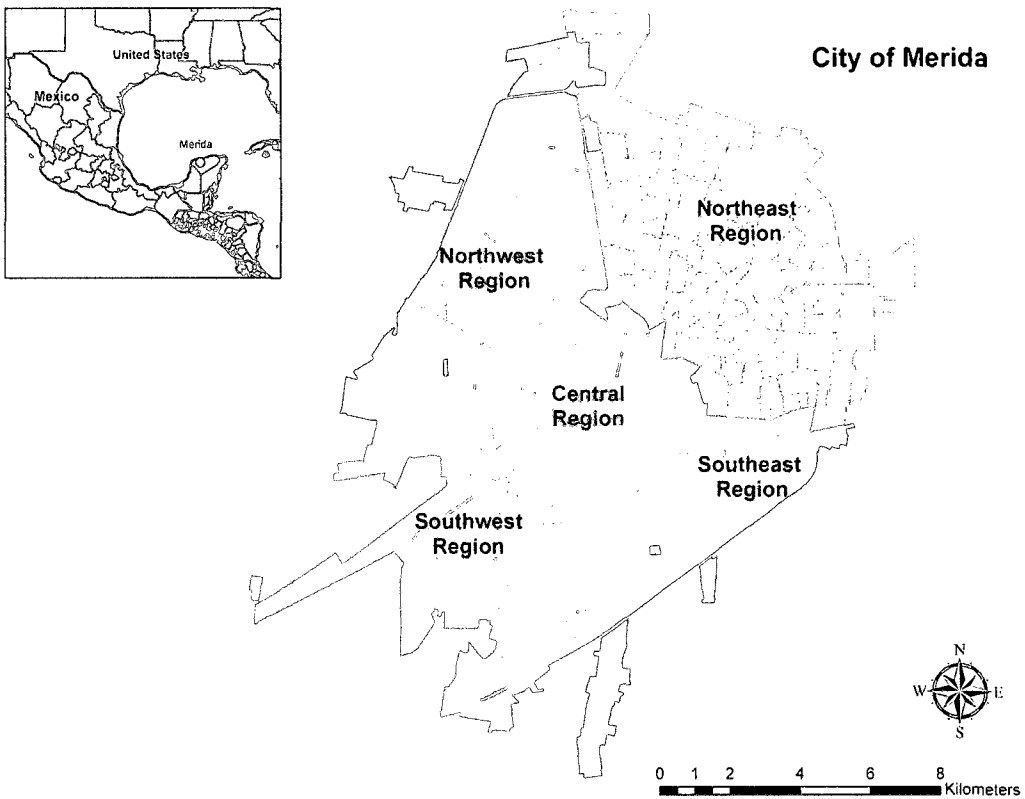

Figure VI-1 Study area
The city of Merida is located on the Yucatan Peninsula of southern Mexico. The city was divided into 5 regions (central, northwest, northeast, southwest, southeast) for purposes of aggregating results of the study. Areas within each region are basic geostatical areas (BGSAs) designated by the Instituto Nacional de Estadística y Geografía (INEGI).

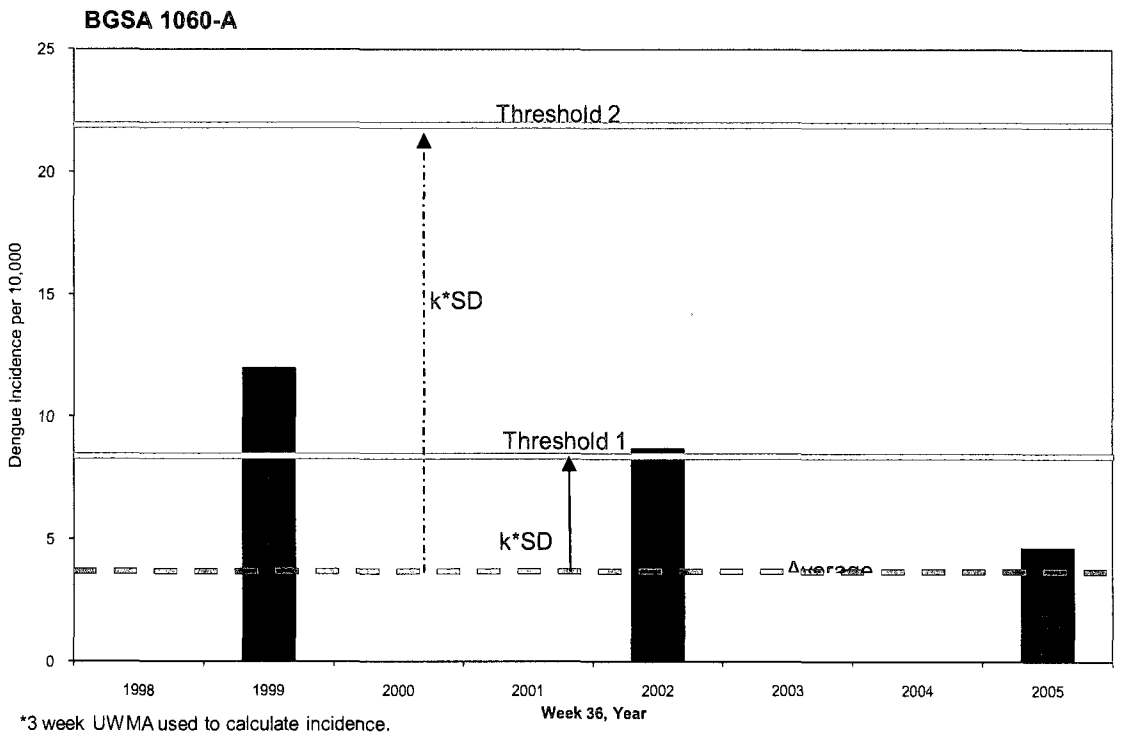

Figure VI-2 Threshold calculation schematic
This graph portrays dengue incidence for 1998—2005 during week 36 for BGSA 1060-A. The average incidence for epidemiological week 36 was calculated and is indicated here by a wide dashed line. Threshold 1 and 2 were calculated by adding k*SD to the average incidence (Threshold 1, k = 4.359, p = 0.05; Threshold 2, k = 1.732, p = 0.25). Note, incidence for each epidemiological week was calculated using a three-week unweighted moving average.

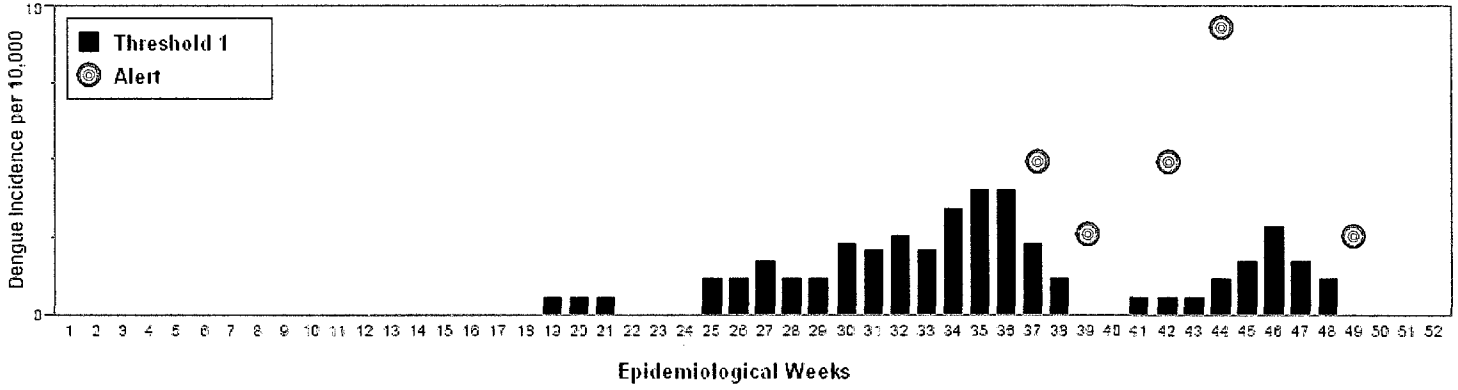

Figure VI-3 Comparison of current dengue incidence with historical threshold
Comparison of 2007 dengue incidence with Threshold 1 by epidemiological week for BGSA 1180-5 located in the northwest region of Merida. Threshold 1 is indicated by black bars. Alerts are indicated when 2007 dengue incidence exceeded Threshold 1 for two consecutive weeks.

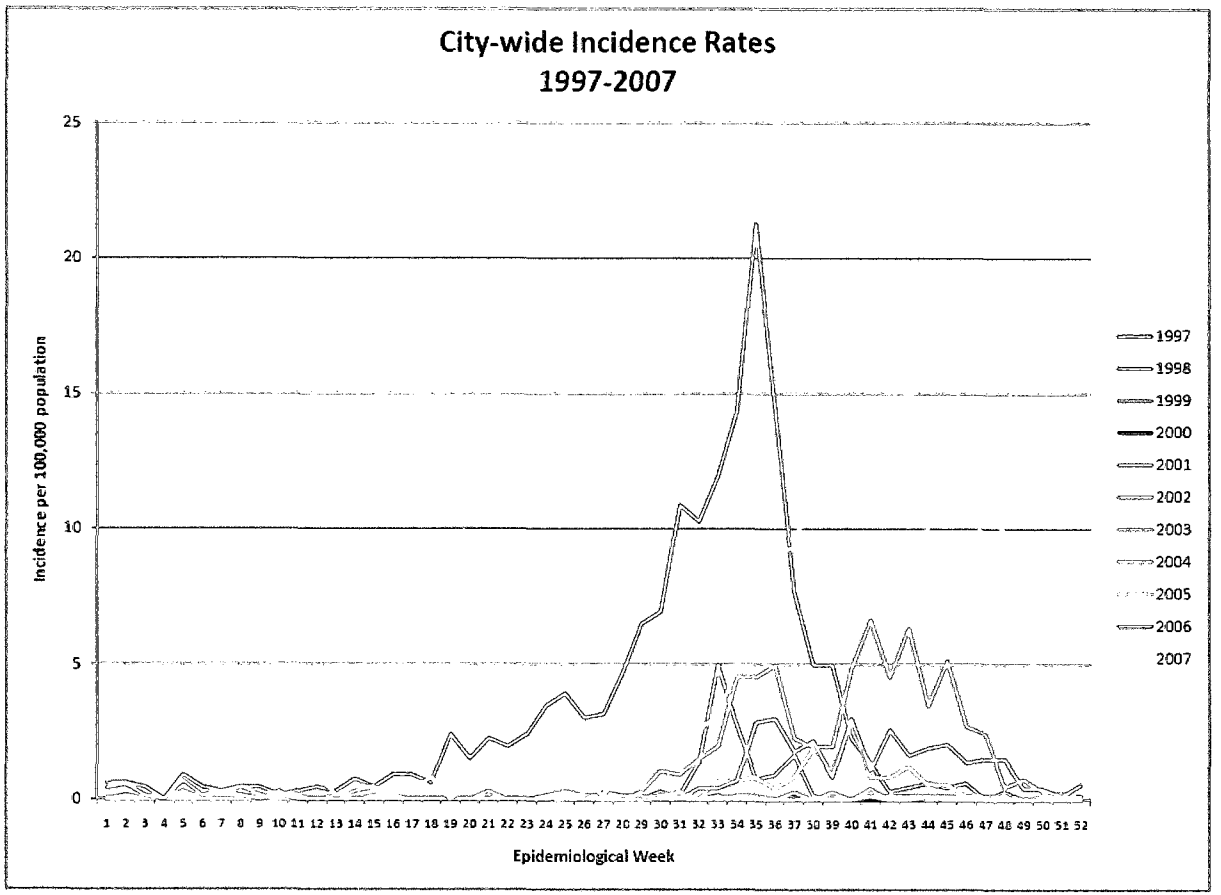

Figure VI-4 City-wide dengue incidence rates for the city of Merida
Dengue incidence rates per 100,000 population were plotted for 1997—2007 by epidemiological week. Between epidemiologic weeks 31—33, the incidence rate for 2007 followed rates reported in previous years. By epidemiological week 34, however, the incidence rate in 2007 deviated from previous years and then peaking during epidemiologic week 43.

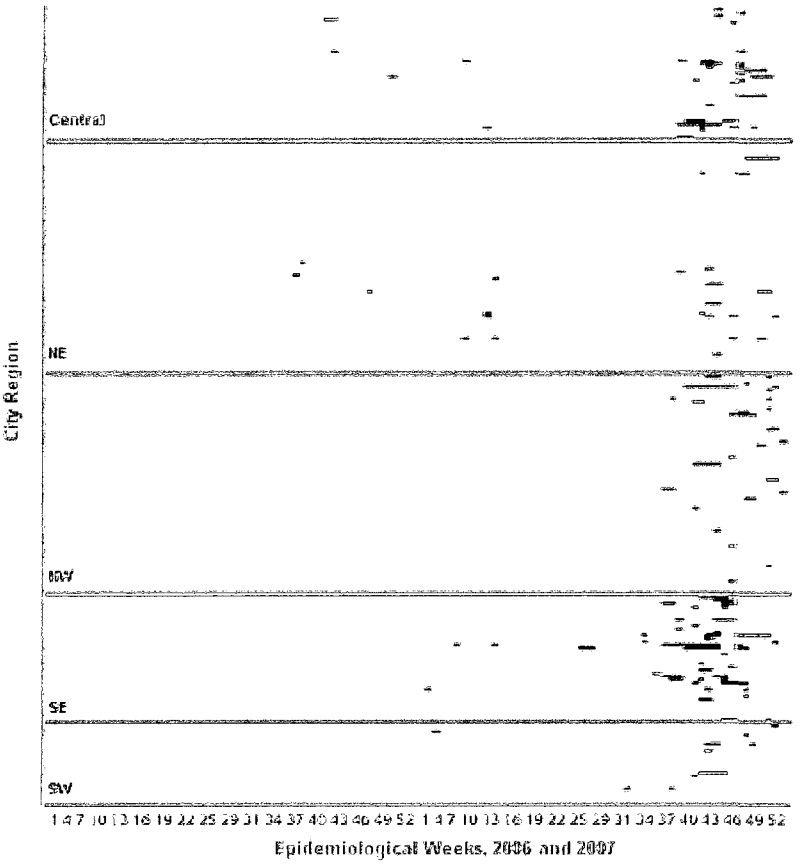

Figure VI-5 Spatial and temporal distribution of alert signals
Alert signals (2x alerts) by week and BGSA are indicated for 2006 and 2007. BGSAs are categorized by region of the city in which they are located including central, northeast (NE), northwest (NW), southeast (SE), and southwest (SW). Alerts were signaled if current incidence (2006 and 2007) exceeded Threshold 1 calculated from dengue incidence data from 1998 – 2005. BGSAs are along the y-axis by distance; weeks for 2006 and 2007 are arranged along the x-axis.

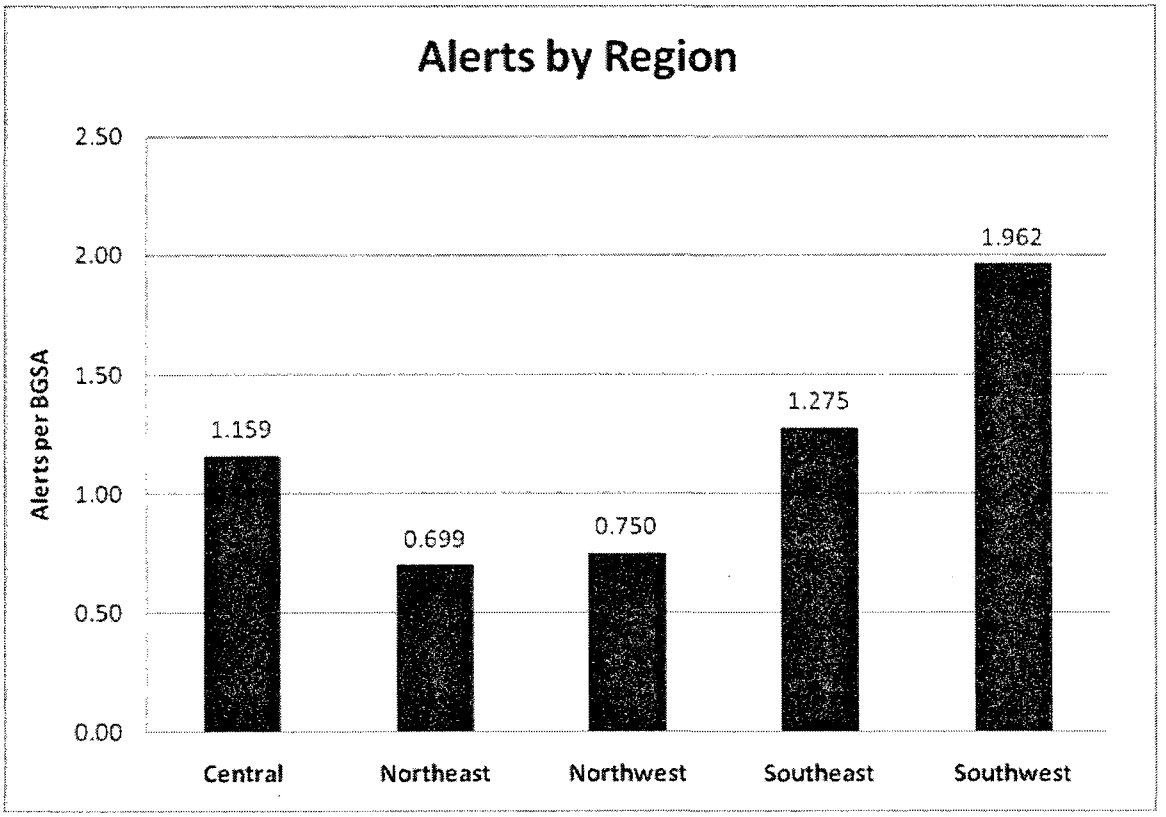

Figure VI-6 Alerts by city region
Alerts were indicated if 2007 dengue incidence exceeded Threshold 1 for two consecutive weeks (2x alert). City regions are mapped in Fig. VI-1.

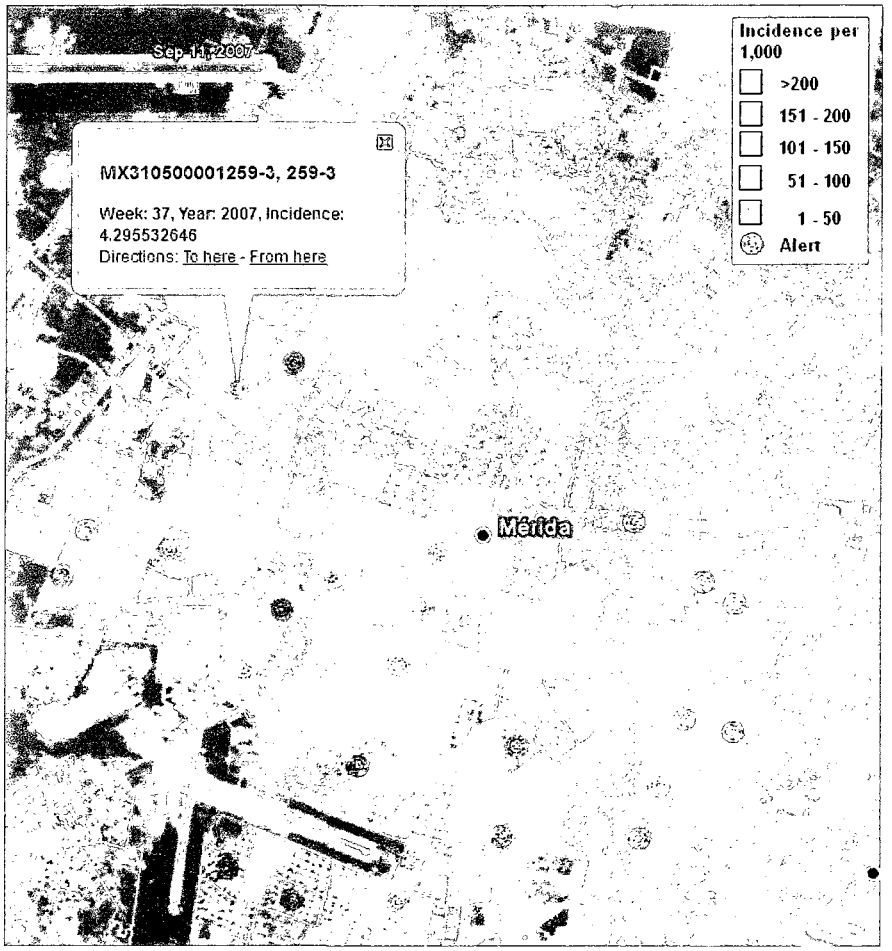

Figure VI-7 Google Earth™ animation
A Google Earth™ animation was developed which provided visualization of Threshold 1 values of dengue incidence (1998—2005) by epidemiological week and BGSA. Alerts were also indicated in the animation when current incidence (from 2006 and 2007) exceeded Threshold 1 values for each epidemiological week and by BGSA. This screenshot displays epidemiological week 37 (September 9-15) in 2007.

References Cited

1. Abeku TA, Hay SI, Ochola S, Langi P, Beard B, de Vlas SJ, Cox J, 2004. Malaria epidemic early warning and detection in African highlands. *Trends Parasitol 20:* 400-405.

2. Kuhn K, Campbell-Lendrum DH, Haines A, Cox J: Using climate to predict infectious disease epidemics. Geneva: World Health Organization, 2005.

3. Lewis R, Nathan N, Diarra L, Belanger F, Paquet C, 2001. Timely detection of meningococcal meningitis epidemics in Africa. *Lancet 358:* 287-293.

4. Moore M, Chan E, Lurie N, Schaefer AG, Varda DM, Zambrano JA, 2008. Strategies to improve global influenza surveillance: a decision tool for policymakers. *BMC Public Health 8:* 186.

5. Yang WZ, Xing HX, Wang HZ, Lan YJ, Sun Q, Hu SX, Lu W, Yuan ZA, Chen YX, Dong BQ, 2004. [A study on early detection for seven infectious diseases]. *Zhonghua Liu Xing Bing Xue Za Zhi 25:* 1039-1041.

6. MMWR, 2006. Summary of notifiable diseases -- United States, 2006. *MMWR 55:* 1-100.

7. Thacker SB, Choi K, Brachman PS, 1983. The surveillance of infectious diseases. *JAMA 249:* 1181-1185.

8. Mueller DH, Abeku TA, Okia M, Rapuoda B, Cox J, 2009. Costs of early detection systems for epidemic malaria in highland areas of Kenya and Uganda. *Malar J 8:* 17.

9. Teklehaimanot HD, Schwatrz J, Teklehaimanot A, Lipsitch M, 2004. Alert threshold algorithms and malaria epidemic detection. *Emerg Infect Dis 10:* 1220-1226.

10. Hay SI, Simba M, Busolo M, Noor AM, Guyatt HL, Ochola SA, Snow RW, 2002. Defining and detecting malaria epidemics in the highlands of western Kenya. *Emerg Infect Dis 8:* 555-562.

11. Coleman M, Coleman M, Mabuza AM, Kok G, Coetzee M, Durrheim DN, 2008. Evaluation of an operational malaria outbreak identification and response system in Mpumalanga Province, South Africa. *Malar J 7:* 69.

12. World Health Organization Roll Back Malaria, 2001. Malaria Early Warning Systems: Concepts, Indicators and Partners. A Frameword for Field Research in Africa. *WHO/CDS/RBM/2001.32.*

13. Barbazan P, Yoksan S, Gonzalez JP, 2002. Dengue hemorrhagic fever epidemiology in Thailand: description and forecasting of epidemics. *Microbe Infect 4:* 699-705.

14. Rigau-Perez JG, Millard PS, Walker DR, Deseda CC, Casta-Velez A, 1999. A deviation bar chart for detecting dengue outbreaks in Puerto Rico. *Am J Public Health 89:* 374-378.

15. Runge-Ranzinger S, Horstick O, Marx M, Kroeger A, 2008. What does dengue disease surveillance contribute to predicting and detecting outbreaks and describing trends? *Trop Med Int Health 13:* 1022-1041.

16. Watkins RE, Eagleson S, Veenendaal B, Wright G, Plant AJ, 2008. Applying cusum-based methods for the detection of outbreaks of Ross River virus disease in Western Australia. *BMC Med Inform Decis Mak 8:* 37.

17. Ryan P, Gatton M, Kay B, 2006. Web based surveillance systems for mosquito-borne diseases: An example using Ross River virus disease in Queensland, Australia. *Mosquito Bites in the Asia Pacific Region 1:* 17-21.

18. Dengue Haemorrhagic fever - Diagnosis, treatment, prevention and control, 2nd ed. Geneva: World Health Organization, 1997.

19. Pan American Health Organization: Dengue and dengue hemorrhagic fever in the Americas: guidelines for prevention and control. Washington DC: PAHO, 1994.

20. Halstead SB: Dengue virus-mosquito interactions. Annu Rev Entomol 2008; 53: 273-91.

21. Gubler DJ: Resurgent vector-borne diseases as a global health problem. Emerg Infect Dis 1998; 4(3): 442-450.

22. Heukelbach J, de Oliveira FA, Kerr-Pontes LR, Feldmeier H, 2001. Risk factors associated with an outbreak of dengue fever in a favela in Fortaleza, north-east Brazil. *Trop Med Int Health 6:* 635-642.

23. Hayes JM, Garcia-Rivera E, Flores-Reyna R, Suarez-Rangel G, Rodriguez-Mata T, Coto-Portillo R, Baltrons-Orellana R, Mendoza-Rodriguez E, De Garay BF, Jubis-Estrada J, Hernandez-Argueta R, Biggerstaff BJ, Rigau-Perez JG, 2003. Risk factors for infection during a severe dengue outbreak in El Salvador in 2000. *Am J Trop Med Hyg 69:* 629-633.

24. Reiter P, Lathrop S, Bunning M, Biggerstaff B, Singer D, Tiwari T, Baber L, Amador M, Thirion J, Hayes J, Seca C, Mendez J, Ramirez B, Robinson J, Rawlings J, Vorndam V, Waterman S, Gubler D, Clark G, Hayes E, 2003. Texas lifestyle limits transmission of dengue virus. *Emerg Infect Dis 9:* 86-89.

25. Focks D, Barrera R: Dengue transmission dynamics: assessment and implications for control. Report of the Scientific Working Group meeting on Dengue. Geneva: World

Health Organization, Special Programme for Research and Training in Tropical Diseases, 2007.

26. Gubler D, Clark G, 1995. Dengue/Dengue Hemorrhagic Fever: The Emergence of a Global Health Problem. *Emerg Infect Dis 1:* 55-57.

27. Halstead SB, 2002. Dengue. *Current Opinion in Infectious Diseases 15:* 471-476.

28. Focks DA, Daniels E, Haile DG, Keesling JE, 1995. A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. *Am J Trop Med Hyg 53:* 489-506.

29. Lorono-Pino MA, Farfan Ale JA, Rosado Paredes EP, Kuno G, Gubler DJ, 1993. Epidemic dengue 4 in the Yucatan, Mexico, 1984. *Revista do Instituto de Medicina Tropical de Sao Paulo 35:* 449-455.

30. Garcia-Rejon J, Lorono-Pino MA, Farfan-Ale JA, Flores-Flores L, Del Pilar Rosado-Paredes E, Rivero-Cardenas N, Najera-Vazquez R, Gomez-Carro S, Lira-Zumbardo V, Gonzalez-Martinez P, Lozano-Fuentes S, Elizondo-Quiroga D, Beaty BJ, Eisen L: Dengue virus-infected *Aedes aegypti* in the home environment. Am J Trop Med Hyg 2008; 79(6): 940-50.

31. 2003. Official Mexican Norm NOM-032-SSA2-2002 for the epidemiologic surveillance, prevention and control of vector-borne diseases. *Official Journal of the Federation.*

32. Servicio Meteorologico Nacional. *Available at: http://smn.cna.gob.mx/SMN.html. Accessed May 31, 2009.*

33. Instituto Nacional de Estadistica y Geografia (INEGI). *Available at: www.inegi.org.mx. Accessed May 22, 2009.*

34. INEGI: Il Conteo de Poblacion y Vivienda 2005, Manual de Cartografia Censal. Mexico City, 2005.

35. Manrique-Saide P, Davies CR, Coleman PG, Rebollar-Tellez E, Che-Medoza A, Dzul-Manzanilla F, Zapata-Peniche A, 2008. Pupal surveys for *Aedes aegypti* surveillance and potential targeted control in residential areas of Merida, Mexico. *J Am Mosq Control Assoc 24:* 289-298.

36. Gubler DG, Kuno G: Dengue and Dengue Hemorrhagic Fever. Oxfordshire: CABI, 1997.

37. Walpole R, Myers R: Probability and Statistics for Engineers and Scientists. New York: Macmillan Publishing Company, 1985.

38. Sprent P, Smeeton N: Applied Nonparametric Statistical Methods. New York: Chapman and Hall/CRC, 2001.

39. Dickenson Gibbons J, Chakraborti S: Nonparametric Statistical Inference. New York: Marcel Dekker, Inc, 1992.

40. Tong Y: Probability Inequalities in Multivariate Distributions. New York: Academic Press, 1980.

41. Cullen JR, Chitprarop U, Doberstyn EB, Sombatwattanangkul K, 1984. An epidemiological early warning system for malaria control in northern Thailand. *Bull World Health Organ 62:* 107-114.

42. Albonico M, De Giorgi F, Razanakolona J, Raveloson A, Sabatinelli G, Pietra V, Modiano D, 1999. Control of epidemic malaria on the highlands of Madagascar. *Parassitologia 41:* 373-376.

43. Bandyopadhyay S, Lum LC, Kroeger A, 2006. Classifying dengue: a review of the difficulties in using the WHO case classification for dengue haemorrhagic fever. *Trop Med Int Health 11:* 1238-1255.

44. Rothman K: Epidemiology -- an Introduction. New York: Oxford University Press, 2002.

45. Hemingway J, Beaty BJ, Rowland M, Scott TW, Sharp BL, 2006. The Innovative Vector Control Consortium: improved control of mosquito-borne diseases. *Trends Parasitol 22:* 308-312.

46. 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 Negl Trop Dis 3:* e411.

VII. Conclusion and Future Directions

The objective of this dissertation was to explore methods to generate information allowing for improved targeting of vector control and disease prevention resources in developed and developing countries. This included development of spatial risk models for WNV disease (Chapter III and IV) and plague (Chapter V). These models were based on associations between epidemiological (Chapter IV and V) or entomological data (Chapter III and IV) and Geographic Information System- or Remote Sensing-derived environmental data to identify high-risk areas for infection. The WNV disease spatial risk models based on epidemiological data (Chapter III) were robust and correctly classified risk areas for presence of WNV disease and for high incidence of WNV disease, providing information helpful for targeted allocation of public health resources including vector control application and education campaigns in Colorado. The objective of Chapter III was met by first improving the knowledge of risk of exposure to WNV in Colorado, as a state-wide spatial risk model had not previously been developed. Furthermore, Chapter III determined that spatial risk models based on data from the plains landscape of eastern Colorado were not applicable to the mountainous western part of the state and vice versa.

Inherent weaknesses are associated with the use of epidemiological data such as that used to build the spatial risk model in Chapter III. For example, low population numbers in some regions may under-estimate or over-estimate the risk of exposure to

vector-borne pathogens. Weaknesses also exist for entomological data including the fact that these data do not account for the importance of human behavior such as use of mosquito repellents. In Chapter IV, we accounted for the inherent weakness of using either entomological or epidemiological risk measures separately by developing a spatial model for entomological risk of exposure to *Cx. tarsalis* and an epidemiological risk map for WNV disease. We then constructed a novel risk classification index combining data for the independently derived measures of entomological and epidemiological risk. This showed that risk of vector exposure was high in the densely populated eastern plains portion of the Colorado Front Range but low in cooler montane areas to the west. The entomological risk model performed well when applied to the western, mountainous part of Colorado and validated strongly when compared to epidemiologic data.

A spatial model for plague in Uganda was also constructed and included in this dissertation (Chapter V). This chapter described the spatial distribution of clinical plague cases in the West Nile region of Uganda, something that had not been done previously. Furthermore, ecological correlates of incidence of plague were identified and incorporated into a predictive model that defined areas of plague risk. Like the WNV disease models, the plague model validated robustly, and also identified regions within the study area where cases may be under-reported and enhanced surveillance and preventative measures would be beneficial to decrease the disease burden.

The spatial modeling chapters demonstrated: (1) the importance of selecting a suitable model development area in an ecologically and climatically diverse environment, and (2) the need for care when extrapolating model results beyond the model development area. For example, spatial models for WNV disease risk developed using

data from the eastern plains in Colorado performed poorly in the mountainous western region and vice versa. Furthermore, spatial models based on entomological data performed well when validated in regions possessing similar environments as those found in the model development area but not in other areas. This finding indicates that spatial models constructed using coarse environmental data and covering large, ecologically diverse areas may not correctly identify risk areas and therefore are not beneficial for precise targeting of vector control and prevention resources. This is especially relevant when spatial models are extrapolated to areas outside the development area and advocates that model extrapolation should be contained to regions with similar climate and ecology.

A basic question in spatial epidemiology is what scale is most appropriate for presentation of spatial assessments of risk for vector-borne pathogens. This question motivated the aim of Chapter II which was to determine: (1) the degree to which estimates of vector-borne disease occurrence is influenced by spatial scale (i.e., county vs. census tract); and (2) the extent of concordance between spatial risk patterns based on case counts versus incidence. The analyses focused on WNV disease in Colorado and revealed that county-scale presentation accounted for only ~50% of the overall variance in incidence compared to census tract level presentation. Thus, sub-county scale more precisely indicates variance in incidence making the presentation of sub-county data to stakeholder communities beneficial.

In the final chapter, the objective was to evaluate an early detection system for dengue outbreaks in Merida, Mexico. Thresholds were derived from historic dengue incidence levels specific to small regions (BGSAs) within the city of Merida to provide indication when current incidence rates exceeded historical norms providing alerts to

vector control practitioners of a possible outbreak situation. This system may also be applicable for resource and capacity planning; Merida public health authorities may use the information to anticipate specific insecticide, capital and human resource requirements and implement a vector-control strategy governed by these historical precedents.

The dissertation chapters relied on existing data sources including epidemiological databases of laboratory or clinically diagnosed disease cases, an entomological database of mosquito catch data, and existing GIS- or RS-based environmental data (e.g. climate variables, landcover variables, population density). Use of existing databases illustrates a benefit of spatial risk models as they can be developed with minimal additional effort outside of pre-existing data collection efforts. The drawbacks of relying on existing data sources for the research projects are many, however, and include lack of point of exposure locations for each disease case which may misrepresent the location of high-risk areas. All epidemiological data included in the dissertation were aggregated to a specific spatial unit [e.g., county, census tract, zip code (Chapters II-IV), parishes (Chapter V) or BGSAs (Chapter VI)]. Environmental data were often of coarse resolution and tended to represent long-term averages (e.g. 30-year averages, Chapters III and IV) versus more recently collected data. The drawbacks of these existing data sources, however, must be weighed against the cost and effort of collecting additional data. Point of exposure locations are especially problematic when dealing with diseases for which the disease agent is transmitted through a common occurrence such as a mosquito bite. Furthermore, the cost and effort of developing situation-specific GIS- and RS-based environmental data is considerable.

Future efforts may focus on continued reliance on existing data (e.g. surveillance data and existing environmental data) to develop spatial models with explicit efforts to integrate these models into existing public health infrastructure. If public health decision makers perceive the cost-effectiveness and applicability of spatial risk models in their programs, enthusiasm to collect more fine-scale epidemiological, entomological and environmental data for incorporation into enhanced models may increase. Furthermore, enhancement of surveillance activities (e.g. moving away from passive surveillance and towards active surveillance) would likely increase the accuracy of the spatial risk models as underreporting due to lack of access to health care or presence of asymptomatic infections would be better accounted for.

The five separate research projects that compose this dissertation all coincide with the shared aim to apply spatial analyses to improve targeting of public health resources to prevent and control vector-borne disease. This dissertation covers three separate disease systems and methods used may be applied to a number of other diseases, including non vector-borne diseases. As is often the case in scientific research, the investigations motivated future directions including:

1) Exploration of methods to enhance the potential for spatial risk models to be applied across a range of climates and environments.
2) Analysis of how potential climate warming over the next 50 years in Colorado will affect the distribution and abundance of *Cx. tarsalis*. The Colorado Front Range is exceptionally well suited for long-term empirical studies on the effect of climate warming on spatial patterns of *Cx. tarsalis* and presence of WNV in local mosquito populations.

3) Development of spatial risk models based on point location data for human disease cases versus models based on aggregated data (e.g. aggregated to zip code, census tract or parish). Future development of spatial risk models based on exposure site locations that reflect actual site of risk would provide fine-resolution outputs that may be more accurate compared to aggregated outputs. Findings in Chapter V have motivated activities that are now underway to develop a fine-scale case and control location-based model for human plague in the West Nile district of Uganda.
4) Further investigation of the validity of spatial models when extrapolated to areas outside of the model development areas. This information would be useful for extrapolation of risk assessment models to areas where consistent epidemiological surveillance activities are lacking and where the burden of disease remains unclear. For example, models from the West Nile region in Uganda very likely are applicable to areas with similar elevations and ecological characteristics in the northeastern part of the Democratic Republic of Congo where surveillance data are lacking, and the distribution of plague is unclear.
5) Investigation of ecological factors predictive of risk, especially for plague, where little information is known regarding risk factors for infection.

Finally, spatial risk modeling has far-reaching applications, but is largely under-used by the public health community. This dissertation investigated best practices for presentation of spatial epidemiologic data and provides guidance and examples of how spatial models and threshold-setting for outbreak detection may be developed. Future

efforts should seek to increase the use of spatial risk models and risk maps and early detection systems by the public health community. Researchers involved in spatial analyses should engage the public health community and communicate the benefits of spatial analyses for cost-effective targeting of intervention activities. A collaborative study amongst economists, public health practitioners, and academics to investigate the actual cost-savings achieved by incorporation of spatial analyses into a public health program should be initiated and results shared with the broader public health community. Such a study has the potential to increase interest in the public health community for use of spatial analyses for cost-effective and efficient intervention approaches. Positive feedback approaches where public health programs share their data with academics involved in spatial analyses that, in turn, benefit the public health programs, would also enhance the application of spatial risk models and early detection systems.

VIII. Appendix – Terminology

Note: This appendix is not an exhaustive list of terminology. Rather, the definitions that follow include terms that are used within the dissertation and may be unclear to the reader.

Term	Definition
Entomological risk	Risk estimate based on the probability of exposure to vectors capable of transmitting a specific pathogen or to infected vectors.
Epidemiological risk	Risk estimate based on occurrence or incidence of a specific disease among a human population.
Outbreak	A sudden increase in the incidence of a disease within a focalized area (e.g., a city). Outbreak is differentiated from epidemic which refers to a sudden increase in the incidence of a disease across a large region and affecting a large population.
Risk factors	External (e.g., rainfall) or internal (e.g., immune status) characteristics that increase the probability of disease occurring in a population.
Risk of exposure	Probability of contact with a specific pathogen, in this case a vector-borne pathogen.