GEOCODING OF PLACE NAMES FOR COLLECTING, MAPPING, MANAGING, AND COMMUNICATING PUBLIC HEALTH INFORMATION

by

ROBERT RYAN LASH

(Under the Direction of Marguerite Madden)

ABSTRACT

Despite significant advances and widespread adoption of geospatial information science and technology (GIS&T), place names continue to be important in public health. This dissertation includes three different applications of GIS&T to public health problems, and finds that place names remain a dominant way of collecting, mapping, managing, and communicating public health information, despite the ubiquity of GPS and mobile computing in our everyday lives. The first study examines the variation in geocoding results of historic monkeypox surveillance data stored as village place names depending on the type of geocoding methodology employed. This study highlights that digital gazetteers remain limited for geocoding in many international locations, and that archival maps can improve the accuracy of geocoding results considerably. The second study examines the types of place names travelers and clinicians use to describe international travel itineraries, and shows that there is a need for more innovative GIS&T-based applications to enable travelers and clinicians to more easily find location-based travel health recommendations. The third study takes up the challenge of developing and deploying a GIS&T-enabled travel recommendation service so that United States travelers and clinicians can find accurate and up-to-date Zika virus travel health recommendations for any international destination. This case study demonstrates that interactive mapping technology which utilizes the latest web-based geospatial data, software, and services can address the needs described in the second study above. More specifically, that web-based geocoding services can enable people to easily search for and find relevant travel health recommendations using a range of place names. Though the solution described meets the current needs for the Zika virus outbreak, this study points out that geospatial capacity needs to be more broadly distributed and improved within public health programs if similar types of applications are to be developed for other diseases. This dissertation contributes to the understanding of the current GIS&T capabilities and needs within applied public health, and should serve to encourage others with expert GIS&T knowledge to explore further collaboration and research opportunities within applied public health.

INDEX WORDS: geospatial information science and technology; applied public health; health communication; place names; geocoding; interactive web maps

GEOCODING OF PLACE NAMES FOR COLLECTING, MAPPING, MANAGING, AND COMMUNICATING PUBLIC HEALTH INFORMATION

by

ROBERT RYAN LASH

BA, The University of Kansas, 2003

MA, The University of Kansas, 2007

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

GEOCODING OF PLACE NAMES FOR COLLECTING, MAPPING, MANAGING, AND COMMUNICATING PUBLIC HEALTH INFORMATION

by

ROBERT RYAN LASH

Major Professor: Marguerite Madden
Committee: Thomas R. Jordan
Lan Mu
William L. Nicholson

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2017

DEDICATION

This work is dedicated to all those who make and use maps in pursuit of the mission of public health.

ACKNOWLEDGEMENTS

This work would not have been possible were it not for the support and patience of my dissertation committee members. I would specifically like to thank my committee chair, Dr. Marguerite Madden, for her unwavering encouragement to seeing me through the PhD Program from the first day until the last, and for allowing me the time to explore each new opportunity that I discovered while working towards a career in public health. I would also like to thank Dr. Thomas Jordan and Dr. Lan Mu for their continued support of my dissertation research and career interests.

Next, I would like to thank my supervisors at the Centers for Disease Control and Prevention (CDC), in Atlanta, GA, who have allowed me to use my work there as the focus of my dissertation research. These supervisors include Dr. Darin Carroll and Dr. Inger Damon, both formerly in the Poxvirus and Rabies Branch; Dr. William Nicholson in the Rickettsial Zoonoses Branch; and Kelly Holton, Dr. Allison Walker, and Dr. Gary Brunette in the Travelers' Health Branch. There are also too many colleagues in these branches to list, who have either contributed to this research or supported me while it was conducted, who are also deserving of my heartfelt thanks. Dr. Ginny Lee, in the Travelers' Health Branch, stands out among these colleagues because of the amount of time she has personally contributed to assisting me with some of the more tedious parts of this dissertation research, as well as her willingness to listen or give me a hug when either or both were needed.

It is important to note here that although my research focuses on the applications of geospatial science and technology at CDC, the views expressed in this dissertation are solely my own and do not represent those of the CDC, U.S. Government, or any other entity with which my co-authors may be affiliated.

I would also like to acknowledge the support of my close friends and fellow students, Shadrock Roberts, Sergio Bernardes, and Josh Campbell, whom have both inspired and encouraged me as the years went by,. Similarly, I would like to thank the members of the CRMS/CGR Safety Committee for their comradery—although our meetings frequently ran late, I am proud that everyone completed their assigned tasks without incident or injury.

Finally, I would like to thank my family, and most importantly, my wife, Johanna, and our dog Sancho. Johanna and Sancho had to endure my PhD stress for many years, and yet they always remained confident that I was capable, and would eventually earn this degree. We will always have fond memories of our short time together in Athens, GA. Though taking eight years to complete this degree is longer than we all expected, we are fortunate that they have all been good years, and I remain optimistic that more will follow.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Geographers have long been developing a set of tools for answering questions about what, where, and why things exist on the surface of the earth. For many centuries, the geographer's tool kit was limited to the technology of the human eye (what could be seen), paper and pencil (what could be written), and algebra (what could be counted, compared, and summarized). Subsequent efforts at accelerating and increasing the capacity for production, revision, and analysis of maps to answer geographical questions were termed geographical information systems (GIS), a name which gained wider acceptance when used to name newly formed peer-reviewed journals and commercial software packages (Foresman 1998). Though a considerable amount of GIS research was focused on developing spatial information and spatial analysis capabilities, Goodchild (1992) sought to situate the "systems" based research focus on a small subset of what he proposed as a much broader research field called geographic information science. He argued that geographic research in geographic information science, later given the shorthand of GIScience, should include the following: data collection and measurement; data capture; spatial statistics; data modeling and theories of spatial data; data structures, algorithms and processes; display; analytical tools; institutions, managerial, and ethical issues.

Though many geographers have been involved in pursuing the GIScience research agenda Goodchild described, the original GIScience agenda did not account for the impact that the internet, software design, and mobile computing would have on our information systems and tools for understanding the world (Goodchild 2009). DiBiase et al. (2006), redefined the domain of geography techniques yet again, coining the term "Geographic Information Science and Technology", or GIS&T, to account for the ever evolving impacts that new computing and information technologies are having on geographic knowledge and research. The GIS&T body of knowledge defines the domain as being comprised of three sub-domains: geographic information science; geospatial technology; and application of geographic information science and technology. As such, this dissertation research seeks to make a modest contribution to knowledge about GIS&T in public health research and operations.

In this dissertation, I will use the term "map" to refer to visualizations of spatial data according to native arrangement on the surface of the earth. A map may be a static data display in print or digital form, and it can also include dynamic or interactive digital displays. The use of maps to provide novel insight into the study of human health and disease has a rich history spanning Europe and the America's over more than a century (Koch 2005). The use of GIS&T in public health has been broadly surveyed in two editions of textbooks written by Cromley & McLafferty (2002, 2012). In the first edition, they described it as a field "in its infancy (pg. ix)," and 10 years later it had flourished with "hundreds of articles… in the research literature each year." While these characterizations of the field of GIS&T and public health are factually sound and well supported, their textbooks do not attempt to critically analyze nor address why some parts of public health were early adopters of GIS while others are just now adopting, or have yet to adopt, this technology tool.

This dissertation is organized in the format of three different manuscripts that each explore the different application of GIS&T to a different public health problem. The common thread of geocoding and place names ties each of these three manuscripts together. The first manuscript explores the available methods for geocoding historical disease surveillance data in

Africa, which is an important and necessary first-step to enable further analysis of geographic patterns of disease risk. The second manuscript is a descriptive study of the way that travelers and clinicians use place name information when interpreting travel health recommendations produced by public health agencies, and the knowledge derived from this study is important for developing new GIS&T travel medicine applications. Finally, the third manuscript seeks to develop and implement a GIS&T travel medicine application which provides dynamic place name search capabilities so that travelers and clinicians can more easily find Zika virus travel health recommendations.

Research Objectives and Questions

The broad question this dissertation aims to answer is can the implementation of geographic information science and technology (GIS&T) improve international disease surveillance and prevention activities when applied within a national public health institution?

To investigate this question more thoroughly, three specific objectives are sought:

1) Demonstrate GIS&T methods for geocoding legacy disease surveillance data under challenging conditions, namely in foreign locations where geographic reference information is poor, thereby transforming the legacy surveillance data into digital geospatial data which are amenable to modeling and risk assessment. The following questions addressed this objective.
    a. What disease surveillance data elements are important for comparing data across different locations?
    b. What GIS&T tools are available, or can be developed, to organize, distribute, and analyze disease surveillance data compiled from disparate sources?

   c. How can the spatial precision of the original surveillance data be preserved to document the difficulties, limitations, and associated uncertainties in the available methods for geocoding historic disease data in foreign locations when high quality geographic reference information are required?.

2) To assess the need for new ways to report and identify travel health recommendations using GIS&T by evaluating the types of place names used by travelers and clinicians to describe the travelers' intended itinerary we ask the following.
   a. How frequently are country, state, and county place names reported by travelers when describing their intended travel itinerary?
   b. How frequently are other types of place names used to describe travel itineraries?
   c. If the place names used to report travel health recommendations are not the most commonly used, what type of new GIS&T applications could be developed that would assist clinicians with the task of identifying location specific travel heath recommendations?

3) Develop and deploy a novel GIS&T application, an interactive web map, to enable travelers and travel health clinicians to identify location-specific Zika virus travel health recommendations. Research questions include the following.
   a. What are the design goals for an interactive web map application for communicating travel health recommendations?
   b. What types of software, hardware, data, and workflows are needed to support interactive web-mapping applications for the CDC's Travelers' Health website?
   c. What are the demographics of users for this type of interactive web map application?

Significance of Study

This dissertation contributes uniquely to the existing GIS&T body of knowledge about the realities of trying to improve GIS&T infrastructure and applications within the confines of a large federal government agency – the U.S. Centers for Disease Control and Prevention (CDC). Such applied knowledge is necessary so as to test, as well as inform, theory about the utility and applicability of GIS&T in achieving the mission of public health. The Centers for Disease Control and Prevention's current mission statement is to collaborate "to create the expertise, information, and tools that people and communities need to protect their health – through health promotion, prevention of disease, injury and disability, and preparedness for new health threats (Centers for Disease Control and Prevention 2012)." Since it was founded in 1942, the Centers for Disease Control and Prevention (CDC) has always maintained renowned expertise in the two core areas of public health, epidemiology and laboratory science (Etheridge 1992). While rudimentary mapping has long been a part of the epidemiologists tool kit, I argue that advances in GIS&T are enabling novel types of collaborative tools and relationships which are useful for a wider range of research including investigations of disease etiology or causes, developing and implementing intervention strategies, and informing policy decisions, as well as improving the knowledge and awareness of the general public.

A recent literature review of the types of public health applications for which geographic information system (GIS) has been employed revealed four predominant themes: disease surveillance (n = 227), risk analysis (n = 189), health access and planning (n = 138), and community health profiling (n = 115) (Nykiforuk and Flaman 2011). Geographic Information System driven applications have been a part of various activities within CDC for roughly 20 years now (Croner et al. 1996), and a number of CDC GIS&T applications are currently available (Table 1.1). Croner describes GIS as a “much-awaited tool” for public health professionals, and predicts that “in the years ahead, GIS will have a profound impact on public health strategies involving surveillance, risk assessment, analysis, and the control and prevention of human disease.”

The CDC is comprised of multiple national centers, and the study of GIS&T within each single center is likely worthy of an individual dissertation. As such, my dissertation research will focus solely on application of GIS&T to meet the needs and goals of the National Center for Emerging Zoonotic and Infectious Diseases (NCEZID). Diseases studied within this center are caused by infectious pathogens, which require either an animal host or vector to maintain the pathogen and transmit it to humans. Data available from the last four years show a trend of increasing investment in GIS based on the number of GIS Software licenses paid for within NCEZID, and a doubling in the number of software licenses in 2012 compared to 2009 (Figure 1.1). These data suggest that interest in applying GIS&T to the public health activities within NCEZID is continuing to grow.

Specifically, this dissertation will try to address the challenges that still exist for transforming legacy disease data into a geospatial format; using geospatial analysis to map public health risk; and effectively communicating knowledge about public health risks to others. This dissertation aims to generate and contribute knowledge about: 1) how to maximize the geocoding accuracy of legacy disease surveillance data in areas of poor geographic reference data; 2) assess the need for new GIS&T applications for help clinicians to identify location-specific travel health recommendations during pretravel consultations; and 3) report on the development and deployment of a novel GIS&T application, an interactive web-map, for the identification and reporting of location-based travel health recommendations to travelers and clinicians.

References

Centers for Disease Control and Prevention. (2012). "About CDC: vision, mission, core values, and pledge." Retrieved 19 November, 2012, from http://www.cdc.gov/about/organization/mission.htm.

Cromley, E. K. and S. McLafferty (2002). GIS and Public Health. New York, NY, Guilford Press.

Cromley, E. K. and S. McLafferty (2012). GIS and Public Health. New York, NY, Guilford Press.

Croner, C. M., J. Sperling and F. R. Broome (1996). "Geographic information systems (GIS): new perspectives in understanding human health and environmental relationships." Statistics in Medicine **15**(18): 1961-1977.

DiBiase, D., M. DeMers, A. Johnson, K. Kemp, A. T. Luck, B. Plewe and E. Wentz, Eds. (2006). Geographic information science and technology body of knowledge. Washington, Association of American Geographers.

Etheridge, E. W. (1992). Sentinel for health : a history of the Centers for Disease Control. Berkeley, University of California Press.

Foresman, T. W. (1998). The History of Geographic Information Systems: Perspectives from the Pioneers. Upper Saddle River, Prentice Hall PTR.

Goodchild, M. F. (1992). "Geographical information science." International Journal of Geographical Information Systems **6**(1): 31-45.

Goodchild, M. F. (2009). "Geographic information systems and science: today and tomorrow." Annals of GIS **15**(1): 3-9.

Koch, T. (2005). Cartographies of disease: maps, mapping, and medicine. Redlands, ESRI Press.

Nykiforuk, C. I. J. and L. M. Flaman (2011). "Geographic information systems (GIS) for health promotion and public health: a review." Health Promotion Practice **12**(1): 63-73.

Figures

Table 1.1 Survey of CDC web-based GIS&T applications.

A variety of interactive mapping applications have been created and supported by various groups at CDC. The applications are often called interactive maps, atlases, or data portals.

Atlas Name	URL
Interactive Atlas of Heart Disease and Stroke	http://nccd.cdc.gov/dhdspatlas/Default.aspx
Diabetes Data and Statistics	http://www.cdc.gov/diabetes/atlas/countydata/atlas.html
Dengue Map	http://www.healthmap.org/dengue/en/
Fatal Injury Mapping	http://wisqars.cdc.gov:8080/cdcMapFramework/
Interactive Lyme Disease Map	http://www.cdc.gov/lyme/stats/maps/interactiveMaps.html
Vectorborne disease maps	http://diseasemaps.usgs.gov/mapviewer/
National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, (NCHHSTP) Atlas	http://gis.cdc.gov/GRASP/NCHHSTPAtlas/main.html
Interactive Cancer Atlas (InCA)	http://www.cdc.gov/cancer/npcr/about_inca.htm
Data.CDC.gov	https://data.cdc.gov/browse?limitTo=maps&s_cid=cdc-data-001&utf8=%E2%9C%93
Social Vulnerability Index	http://svi.cdc.gov/map.aspx

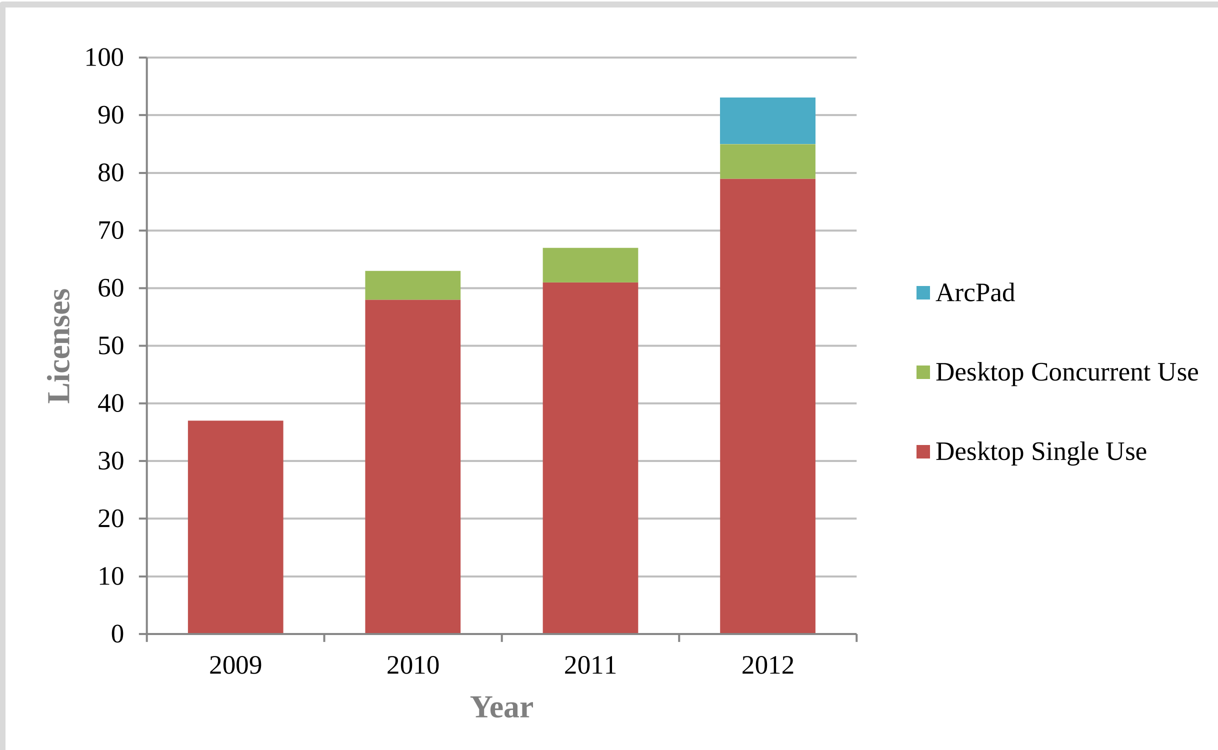


Figure 1.1 Trends in ESRI ArcGIS Software usage in the National Center for Emerging and Zoonotic Infectious Diseases (NCEZID), CDC from 2009-2012.

CHAPTER 2

LITERATURE REVIEW

The story of John Snow's map of a London cholera epidemic has become lore through the popular press (Johnson 2006), and is retold in an inspirational way by many from various academic and professional disciplines. However, that story, where John Snow gets the idea to map where cholera patients lived to discover which water pump was the source of cholera. The map showed a clustering of cases around the Broad Street pump, and thus, Snow had the pump handle removed an no one else got ill; that story is more myth than fact (McLeod 2000). The truths to the Snow map and the mythology which are now associated with it in public health and epidemiology have been more rigorously studied by Koch (2004). Koch's historical research shows that Snow's map was not an early example of exploratory data visualization providing novel epidemiological insight, but rather a purposefully designed diagram meant to persuasively advance Snow's previously published theory that cholera was not an airborne illness as most people at the time believed, but rather a waterborne illness (Koch 2011).

Geographers Cliff & Haggett (1988) summarize various quantitative spatial analytical approaches for mapping and modeling disease occurrence through time and space based on public health data stored as both spatial points and areas. Spatial statisticians have also published books summarizing quantitative methods for studying disease risk and transmission (Waller and Gotway 2004, Lawson 2009). Many of the analytical approaches described have been incorporated into commercial GIS software packages, and cookbooks for using these software tools are available (Kurland and Gorr 2012). Cromley & McLafferty (2012), both long active in the academic discipline of geography, have surveyed and summarized the dominant trends in the application of GIS&T to public health over the last 20 years. This textbook provides a reasonable introduction to GIS&T, geospatial data, sources of public health information, and different approaches to spatially and temporally analyzing public health data.

The idea that GIS&T can be more broadly utilized in public health will be explored. Specifically, this dissertation will consider more deeply the application of GIS&T within the public health sub-fields of epidemiology, public health surveillance, and public health communication. Epidemiology is the study of the distribution and determinants of disease within a specified population, and the application of this knowledge to control disease. Public health surveillance systems are the primary way in which we gather epidemiologic data to construct and refine our understanding of disease. These surveillance systems are principally designed to capture information about what groups of people are at risk for disease; what types of pathogens are responsible for causing disease; and the places where peoples and pathogens converge to enable disease occurrence or further transmission. Public health communication is a field that aims to improve the access and delivery of accurate and relevant public health information to the public. Within these subfields, patterns in the adoption of GIS&T can already be seen within the literature, with variations occurring due to differences in public health program aims, disease system, geographic location, and political scale.

Public Health Surveillance Systems

A principal and primary concern when beginning any disease mapping study is to understand the nature of the disease and the means by which it has been observed, that is, a thorough understanding of how the disease data are collected and produced. When focusing on infectious diseases of international importance, access to appropriate disease data for mapping and modeling remains a challenge. Cliff & Haggett (1988) provide a series of diagrams reprinted here which help explain the multiple processes required for generating accurate disease data (Figure 2.1). These processes begin at the point of infection at the cellular level (Panel A), reporting of the disease at the primary care provider (Panel B), data collection networks at the local, national, and international levels (Panel C), and finally reporting of the case data to the general public (Panel D). Public health map makers need to have a complete understanding of the data flow responsible for generating disease data so that they can account for potential bias and uncertainty in the available data.

The idea of disease surveillance is a relatively modern one, showing some signs of evolution over time. Most of the discussion about what a diseases surveillance system was and how it was designed to operate has been controlled by with the earliest uses of the terminology ascribed to Alex Langmuir (Langmuir 1963). To him, the term meant "the continued watchfulness over the distribution and trends of incidence through the systematic collection, consolidation and evaluation of morbidity and mortality reports and other relevant data… with regular dissemination of this data and its interpretation to all who need to know." According to Langmuir's line of reasoning, by studying the medical reports of ill persons, public health interventions could be better evaluated and targeted.

Thacker et al. (1989) was one of the first to use the term "science of public health surveillance," which he used to argue for the strengthening of public health surveillance activities that would be more rigorous and methodical such that the public's health would eventually improve. At its core, public health surveillance is about creating a network of individuals and institutions that can reliably collecting individual disease reports, collating them, and then analyzing them at a population scale. The relatively new field of public health informatics seeks to apply new computer and communication technology to grow and strengthen disease surveillance networks by integrating this new technology (Krishnamurthy and St. Louis 2010). Examples of these types of informatics initiatives include the Global Infectious Diseases and Epidemiology Network (GIDEON) (Berger 2005), as well as HealthMap (Brownstein et al. 2008). The research proposed in this dissertation has the opportunity to inform public health informatics development, by providing informatics researchers with information about the challenges faced when trying to accurately link public health events to specific geographic locations.

Geocoding Public Health Data

Geocoding is an essential first step towards enabling GIS-based analyses of public health data (Vine et al. 1997, Rushton et al. 2008). It is the process by which textual descriptions of the geographic provenance of cases and diagnostic specimens are transformed into digital spatial data (longitude and latitude coordinates; “geocoding” is generally used to refer to the simpler process of adding geographic coordinates to postal addresses) (Hill 2009). The geocoding process has been generalized into the following components: input records, reference datasets (e.g., gazetteers), and a geocoder (the algorithm used to normalize, standardize, and match input records to the reference dataset) (Goldberg et al. 2007). Ideally, the process is documented with detailed metadata (Wieczorek et al. 2004).

Geocoding methods and services for public health have evolved over time and their usefulness routinely evaluated. Initially, most geocoding was performed on a fee-for-service bases by spatial data processing businesses. Krieger et al. (2001) evaluated the accuracy of different contract services relative to financial cost, timeliness, and quality of customers service when geocoding 70 household street addresses across Massachusetts and Rhode Island to the census tract and block group level, with the best firm correctly matching 96% of the addresses. Improvements in personal computing hardware, geocoding algorithms in desktop GIS software, and improved accuracy and completeness of reference data (e.g., United States Census TIGER centreline files) made geocoding more accessible to the average GIS users. Ward et al. (2005) compared in-house geocoding accuracy of household street addresses from 234 addresses in Iowa using ESRI ArcView 3.2 geocoding software and United States Census TIGER 2000 reference data to the results of an independent contractor. The match rate for the in-house geocoding was 88% while the match-rate for the contractor was 92%, and the two methods agreed on 84% of the locations. They identified a spatial bias in geocoding accuracy, with better accuracy in urban areas than rural areas.

In general, the quality of the geographic reference data available nationally in the United States is quite high, permitting geocoding to within a kilometer of accuracy in rural areas, and even greater accuracy in suburban and urban areas. Sub-kilometer accuracy can be more reliably achieved when geocoding algorithms can make use of multiple geographic reference datasets, selecting the most accurate reference dataset for a suitable geographic location. This type of geocoding is known as a multi-stage method, and subtle yet quantifiable improvements in both the match rate and spatial accuracy of the geocoding results has been shown (Zhan et al. 2006, Lovasi et al. 2007)

The value of geocoded public health data for research, intervention design, and control measures at the state (MacDorman and Gay 1999) and national (Croner et al. 1996, Boulos

2004) public health systems is clear. However, nearly all research on the efficiency, reliability, and accuracy of geocoding methods has relied on examples of contemporary input records and reference datasets from North America and Europe (Abe and Stinchcomb 2008), possibly because geocoding methods evolve as the availability and accuracy of reference datasets increase (Goldberg et al. 2007).

While the value of geocoded health data to public health agencies and researchers is clear, these organizations and individuals have a legal and ethical obligation to protect the privacy of the individuals who supplied this information (Olson et al. 2006, Rushton et al. 2008, Lee and Gostin 2009). Researchers have shown that simple dot maps of disease cases can accurately reveal the true locations of cases through the relatively simple process of reverse-geocoding (Brownstein et al. 2006). U.S. Cancer registries have studied the risks introduced by geocoding in considerable detail (Gittler 2008), and made recommendations as to the best practices for protecting privacy during the geocoding process, as well as subsequently storing, analyzing, and presenting this information (Goldberg 2008). The common practice for protecting the privacy of individual geocoded data is to “mask” the data using one of three approaches: displacing data; aggregating data; or randomly perturbing the data (Rushton et al. 2006).

Ecological Niche Modeling for Estimating Disease Distributions

A common problem in public health is geographically limited or spatially biased disease occurrence data. As a result, public health researchers are interested in a variety of inferential or predictive techniques that can be used to fill in the spatial and temporal gaps inherent in these disease occurrence datasets. Ecological niche modeling (ENM) is one such inferential modeling technique which is increasingly being used to study a growing range of zoonotic disease distributions, including hantavirus (Wei et al. 2011), Marburg and Ebola viruses (Peterson et al. 2006), monkeypox virus (Ellis et al. 2012), and vector-borne zoonotic diseases such as plague (Holt et al. 2009, Maher et al. 2010), tularemia (Nakazawa et al. 2010), Chagas disease (Gurgel-Gonçalves et al. 2012), dengue and chikungunya (Campbell et al. 2015), leishmaniasis (Chalghaf et al. 2016), Crimean-Congo hemorrhagic fever (Estrada-Peña et al. 2008), and Rift Valley fever (Mweya et al. 2017), though this list is not exhaustive.

A comprehensive overview of the theoretical basis for ENM and association computational techniques have been summarized by Peterson, Soberón et al. (2011). A good working definition of an ecological niche for the purposes of this dissertation is "the set of conditions under which the species can maintain populations without immigration of individuals from other areas (Peterson 2006)." In seeking to more explicitly state the set of conditions where a given species is found, Soberón and Peterson (2005) proposed three conditions which must be met for a species to exist: 1) the physical environment and climate must be suitable; 2) the interactions with other species must permit the species of interest to maintain a stable population level for regular reproduction; 3) the species must be able to physically access the geographic region where conditions 1 and 2 are met. Soberón and Peterson assembled a venn diagram (e.g., the BAM diagram) to enable the ecological space defined by these sets of conditions to be visualized (Figure 2.2). In Panel A, set A, the abiotic conditions, represents the physical geographic environment which is suitable for a species to exist. While all of the physical geography conditions within the abiotic set are suitable for the species in question, biotic conditions (set B) such as the absence of predator species or the presence of prey species, limits the species from inhabiting all of set A. Finally, set M is the range of conditions which are accessible to the species. Given this framework, a species ecological niche is defined as the union of set A, with set B, and set M.

Figure 2.2, Panel B is an example of how the BAM diagram can be used to understand the changing ecological niche of human disease plague. Plague is a vectorborne zoonotic disease caused by the bacterium *Yersinia pestis* found originally in small rodents hosts only in Central Asia, and transmitted to humans most commonly through the bite of infected fleas. As trade between Europe and Asia increased in the middle ages, a series of human pandemics occurred and the bacterium spread into other rodent host species, particularly urban *Rattus* species. Plague arrived in port cities in the United States in the 1900s from Asia, and subsequently spread into native rodent species where it is now maintained. Figure 2.2, Panel B shows the changes in the ecological niche of plague over the last century. Notice that the biotic set (B) has contracted as the sanitation movement successfully broke the plague cycle in urban rodents. Meanwhile, the mobility set (M) has expanded the ecological niche as the continued expansion of international trade has brought plague to new parts of South American, Africa, and Southeast Asia.

The process for building ENMs is relatively simple. Two types of input data are needed, a GIS point file of species occurrence records, and appropriate physical geographic and climatic raster data. Some pre-processing of these two input datasets may be needed if point files are to be subset for training and testing, and raster datasets must all have a common spatial projection, spatial resolution, and digital file format, steps which are easily performed in desktop GIS software. ENM software is comprised of an algorithm for extracting environmental values at species occurrence points, as well as pseudo-absence points, and then an automated algorithm that will seek to optimally fit a model around the range of environmental conditions which distinguish the occurrence data sample from the pseudo-absence data.

Map Communication within Public Health Communication

This research will contribute uniquely to research on how maps are used to communicate public health information, specifically how maps and GIS can be employed to deliver individually tailored travel health recommendations. As such, it aims to demonstrate how methods and techniques found in the literature on cartographic design and map communication can inform health communication work such that public health information users receive for tailored information. This type of research is notably absent from currently published applications of GIS&T to public health, which has been well surveyed and recently summarized (Cromley and McLafferty 2012).

The study of map communication began with Robinson's book "The Look of Maps (1952)," which was the first of its kind to persuasively argue that the scientific study of map design was needed to support the growth in map production, due to the then recent developments in photographic techniques for map production and reproduction. For the next half-century, understanding the map communication process and defining optimal cartographic design principles were the focus of much cartographic research. Robinson et al. (1995) described methods for standardizing static map production, while Dent (1999) focused on the cartographic design principles for effective static thematic map production. MacEachren (1995) explored issues of cartographic representation, visualization, and design, and his cartographic visualization cube provided conceptual clarity to how changes in the audience of a map alters cartographic decision making. More recently, Muehlenhaus (2014) summarized map communication principles and methods useful for designing static and interactive maps on the web. Methods for 3-dimensional and 4-dimensional data visualization have also been developed for cartography

Health communication is a growing area of importance within public health which aims to produce and deliver information promoting healthy behaviors in individuals. It is broadly defined as "the art and technique of informing, influencing, and motivating individual, institutional, and *public audiences* about important health information (U.S. Department of Health and Human Services 2000; emphasis added by this author)." Most public health policies and interventions are guided by public health research which is vetted through the peer-reviewed publishing process. However, most peer-reviewed writing is accessible only to a select audience who are knowledgeable in the technical details and specialized vocabulary of these fields, so in practices, health communicators aim to transform public health information to make it more accessible to the general public where its relevance and importance can be easily understood.

Health communicators frequently customize public health messaging to help them reach the intended audience(s) through either targeted or tailored approaches. Targeting is the process of transforming public health information to reach a sub-population of the intended audience, while tailoring is the process of developing health communication messages specific to an individual audience member (Kreuter and Wray 2003). The concept of tailoring health communication was first proven successful in studies of printed health communication materials which showed that messages are more easily read and remembered when tailored (Skinner et al. 1999), because tailoring enhances the relevance of the content of health message in the eyes of the individual.

Research on the role of pictures and graphics in health communication found that pictures improve individual understanding of health education information when compared to written or spoken text alone (Houts et al. 2006). However, little research has been published in the health communication literature about the role of maps as communication devices. Parrott et al. (2007)

studied the communication function and effects of static and interactive maps within cancer control plans used for policy making. They examined maps used to present information as evidence for decision making, but doubted how reliably users could correctly interpret correlations between geographic location and disease incidence without more careful instruction and training.

Outside the health communication literature, maps have been studied principally as tools used by public health researchers to answer questions about epidemiology, or health communication to an existing technically literate peer audience. The National Center for Health Statistics (NCHS) has performed an extensive amount of research on the design, production, and use of county level cancer rate maps for studying the epidemiology of cancer. These statistical rate maps were first made in the 1960s and provided novel insights into localized risk patterns associated with the rate of cancer incidence for different types of cancers (Hoover et al. 1975). While these early maps proved useful for generating new hypotheses and identifying “hotspots” for further study, little empirical data existed which could be used to inform how best to symbolize statistical rate data. For example, how did the map maker’s decisions on the use of color gradients, fill patterns, or graduated symbols to represent different classes of data impact how the mapped data were interpreted? In a study which examined how map users perceived spatial clustering of health data depending on the use of monochrome, diverging color, categorical colors, dot density, and pie maps, Lewandowsky et al. (1993), found that classifying the data using monochrome symbol classes was optimal for enabling users to reliably detect the location of a common disease cluster. Pickle (2009) has recently published a historical account of the 40 years of design, production, and use of these national mortality atlases, and concludes that though the maps have some design limitations, they have led to import etiologic findings and interventions to reduce cancer rates and health disparities.

References

Abe, T. and D. G. Stinchcomb (2008). Geocoding practices in cancer registries. Geocoding health data: the use of geographic codes in cancer prevention and control, research, and practice. G. Rushton, M. P. Armstrong, J. Gittler et al. Boca Raton, FL, CRC Press**:** 111-125.

Berger, S. A. (2005). "GIDEON: a comprehensive web-based resource for geographic medicine." International Journal of Health Geographics **4**(1): 10.

Boulos, M. N. (2004). "Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom." International Journal of Health Geographics **3**(1): 1.

Brownstein, J. S., C. A. Cassa and K. D. Mandl (2006). "No place to hide--reverse identification of patients from published maps." New England Journal of Medicine **355**(16): 1741-1742.

Brownstein, J. S., C. C. Freifeld, B. Y. Reis and K. D. Mandl (2008). "Surveillance sans frontières: internet-based emerging infectious disease intelligence and the HealthMap project." PLoS Medicine **5**(7): e151.

Campbell, L. P., C. Luther, D. Moo-Llanes, J. M. Ramsey, R. Danis-Lozano and A. T. Peterson (2015). "Climate change influences on global distributions of dengue and chikungunya virus vectors." Philosophical Transactions of the Royal Society B: Biological Sciences **370**(1665).

Chalghaf, B., S. Chlif, B. Mayala, W. Ghawar, J. Bettaieb, M. Harrabi, G. B. Benie, E. Michael and A. B. Salah (2016). "Ecological niche modeling for the prediction of the geographic distribution of cutaneous leishmaniasis in Tunisia." American Journal of Tropical Medicine and Hygiene **94**(4): 844-851.

Cliff, A. D. and P. Haggett (1988). Atlas of Disease Distributions: Analytic Approaches to Epidemiological Data. Cambridge, MA USA, Blackwell Publishers.

Cromley, E. K. and S. McLafferty (2012). GIS and Public Health. New York, NY, Guilford Press.

Croner, C. M., J. Sperling and F. R. Broome (1996). "Geographic information systems (GIS): new perspectives in understanding human health and environmental relationships." Statistics in Medicine **15**(18): 1961-1977.

Dent, B. D. (1999). Cartography thematic map design. St. Louis, WCB/McGraw-Hill.

Ellis, C. K., D. S. Carroll, R. R. Lash, A. T. Peterson, I. K. Damon, J. Malekani and P. Formenty (2012). "Ecology and georaphy of human monkeypox case occurences across Africa." Journal of Wildlife Diseases **48**(2): 335-347.

Estrada-Peña, A., Z. Zatansever, A. Gargili, M. Aktas, R. Uzun, O. Ergonul and F. Jongejan (2008). "Modeling the spatial distribution of Crimean-Congo hemorrhagic fever outbreaks in Turkey." Vector-borne and zoonotic diseases **7**(4): 667-678.

Gittler, J. (2008). Cancer registry data and geocoding: privay, confidentiality, and security issues. Geocoding health data: the use of geographic codes in cancer prevention and control, research, and practice. G. Rushton, M. P. Armstrong, J. Gittler et al. Boca Raton, FL, CRC Press**:** 195-223.

Goldberg, D. W. (2008). A geocoding best practices guide. Springfield, IL, North American Association of Central Cancer Registries.

Goldberg, D. W., J. P. Wilson and C. A. Knoblock (2007). "From text to geographic coordinates: the current state of geocoding." URISA Journal **19**(1): 33-46.

Gurgel-Gonçalves, R., C. Galvão, J. Costa and A. T. Peterson (2012). "Geographic distribution of Chagas disease vectors in Brazil based on ecological niche modeling." Journal of Tropical Medicine **2012**.

Hill, L. L. (2009). Georeferencing : the geographic associations of information. Cambridge, MA, MIT.

Holt, A. C., D. J. Salkeld, C. L. Fritz, J. R. Tucker and P. Gong (2009). "Spatial analysis of plague in California: niche modeling predictions of the current distribution and potential response to climate change." International Journal of Health Geographics **8**(1): 38.

Hoover, R., T. Mason, F. McKay and J. Fraumeni (1975). "Cancer by county: new resource for etiologic clues." Science **189**(4207): 1005-1007.

Houts, P. S., C. C. Doak, L. G. Doak and M. J. Loscalzo (2006). "The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence." Patient Education and Counseling **61**(2): 173-190.

Johnson, S. (2006). The ghost map : the story of London's most terrifying epidemic--and how it changed science, cities, and the modern world. New York, Riverhead Books.

Koch, T. (2004). "The map as intent: variations on the theme of John Snow." Cartographica **39**(4): 1-14.

Koch, T. (2011). Disease maps : epidemics on the ground. Chicago, University of Chicago Press.

Kreuter, M. W. and R. J. Wray (2003). "Tailored and targeted health communication: strategies for enhancing information relevance." American Journal of Health Behavior **27**(1): S227-S232.

Krieger, N., P. Waterman, K. Lemieux, S. Zierler and J. W. Hogan (2001). "On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research." American Journal of Public Health **91**(7): 1114-1116.

Krishnamurthy, R. S. and M. E. St. Louis (2010). Informatics and the management of surveillance data. Principles & Practice of Publice Health Surveillance. L. M. Lee, S. M. Teutsch, S. B. Thacker and M. E. St. Louis. New York, Oxford University Press**:** 65-87.

Kurland, K. S. and W. L. Gorr (2012). GIS Tutorial for Health. Redlands, CA, ESRI Press.

Langmuir, A. D. (1963). "The surveillance of communicable diseases of national importance." New England Journal of Medicine **268**(4): 182-192.

Lawson, A. (2009). Bayesian disease mapping : hierarchical modeling in spatial epidemiology. Boca Raton, CRC Press.

Lee, L. M. and L. O. Gostin (2009). "Ethical collection, storage, and use of public health data: A proposal for a national privacy protection." JAMA: The Journal of the American Medical Association **302**(1): 82-84.

Lewandowsky, S., D. J. Herrmann, J. T. Behrens, S.-C. Li, L. Pickle and J. B. Jobe (1993). "Perception of clusters in statistical maps." Applied Cognitive Psychology **7**(6): 533-551.

Lovasi, G., J. Weiss, R. Hoskins, E. Whitsel, K. Rice, C. Erickson and B. Psaty (2007). "Comparing a single-stage geocoding method to a multi-stage geocoding method: how much and where do they disagree?" International Journal of Health Geographics **6**(1): 12.

MacDorman, M. F. and G. A. Gay (1999). "State initiatives in geocoding vital statistics data." Journal of Public Health Management and Practice **5**(2): 91-93.

MacEachren, A. M. (1995). How maps work: representation, visualization, and design. New York, Guilford Press.

Maher, S. P., C. Ellis, K. L. Gage, R. E. Enscore and A. T. Peterson (2010). "Range-wide determinants of plague distribution in North America." American Journal of Tropical Medicine and Hygiene **83**(4): 736-742.

McLeod, K. S. (2000). "Our sense of Snow: the myth of John Snow in medical geography." Social Science & Medicine **50**(7-8): 923-935.

Muehlenhaus, I. (2014). Web Cartography: Map Design for Interactive and Mobile Devices. Boca Raton, FL, CRC Press.

Mweya, C. N., L. E. G. Mboera and S. I. Kimera (2017). "Climate influence on emerging risk areas for Rift Valley Fever epidemics in Tanzania." American Journal of Tropical Medicine and Hygiene **97**(1): 109-114.

Nakazawa, Y., R. A. J. Williams, A. T. Peterson, P. S. Mead, K. J. Kugeler and J. M. Petersen (2010). "Ecological niche modeling of *Francisella tularensis* subspecies and clades in the United States." American Journal of Tropical Medicine and Hygiene **82**(5): 912-918.

Olson, K. L., S. J. Grannis and K. D. Mandl (2006). "Privacy protection versus cluster detection in spatial epidemiology." American Journal of Public Health **96**(11): 2002-2008.

Parrott, R., S. Hopfer, C. Ghetian and E. Lengerich (2007). "Mapping as a visual health communication tool: promises and dilemmas." Health Communication **22**(1): 13-24.

Peterson, A. T. (2006). "Ecological niche modeling and spatial patterns of disease transmission." Emerging Infectious Diseases **12**(12): 1822-1826.

Peterson, A. T., R. R. Lash, D. Carroll and K. M. Johnson (2006). "Geographic potential for outbreaks of Marburg hemorrahagic fever." American Journal of Tropical Medicine and Hygiene **75**(1): 9-15.

Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martinez-Meyer, M. Nakamura and M. B. Araujo (2011). Ecological niches and geographic distributions. Princeton, Princeton University Press.

Pickle, L. W. (2009). "A history and critique of U.S. mortality atlases." Spatial and Spatio-temporal Epidemiology **1**(1): 3-17.

Robinson, A. H. (1952). The look of maps : an examination of cartographic design. Madison, WI, University of Wisconsin Press.

Robinson, A. H., J. L. Morrison, P. C. Muehrcke, A. J. Kimerling and S. C. Guptill (1995). Elements of Cartography. New York, John Wiley & Sons, Inc.

Rushton, G., M. P. Armstrong, J. Gittler, B. R. Greene, C. E. Pavlik, M. M. West and D. L. Zimmerman (2006). "Geocoding in cancer research: a review." American Journal of Preventive Medicine **30**(2 Suppl): S16-24.

Rushton, G., M. P. Armstrong, J. Gittler, B. R. Greene, C. E. Pavlik, M. M. West and D. L. Zimmerman, Eds. (2008). Geocoding health data: the use of geographic codes in cancer prevention and control, research, and practice. Boca Raton, FL, CRC Press.

Skinner, C., M. Campbell, B. Rimer, S. Curry and J. Prochaska (1999). "How effective is tailored print communication?" Annals of Behavioral Medicine **21**(4): 290-298.

Soberón, J. and A. T. Peterson (2005). "Interpretation of models of fundamental ecological niches and species' distributional areas." Biodiversity Informatics **2**: 1-10.

Thacker, S. B., R. L. Berkelman and D. F. Stroup (1989). "The science of public health surveillance." Journal of Public Health Policy **10**(2): 187-203.

U.S. Department of Health and Human Services (2000). Healthy People 2010: Understanding and Improving Health and Objectives for Improving Health. Washington, D.C., U.S. Government Printing Office.

Vine, M. F., D. Degnan and C. Hanchette (1997). "Geographic information systems: their use in environmental epidemiologic research." Environmental Health Perspectives **105**(6): 598-605.

Waller, L. A. and C. A. Gotway (2004). Applied spatial statistics for public health data. Hoboken, John Wiley & Sons.

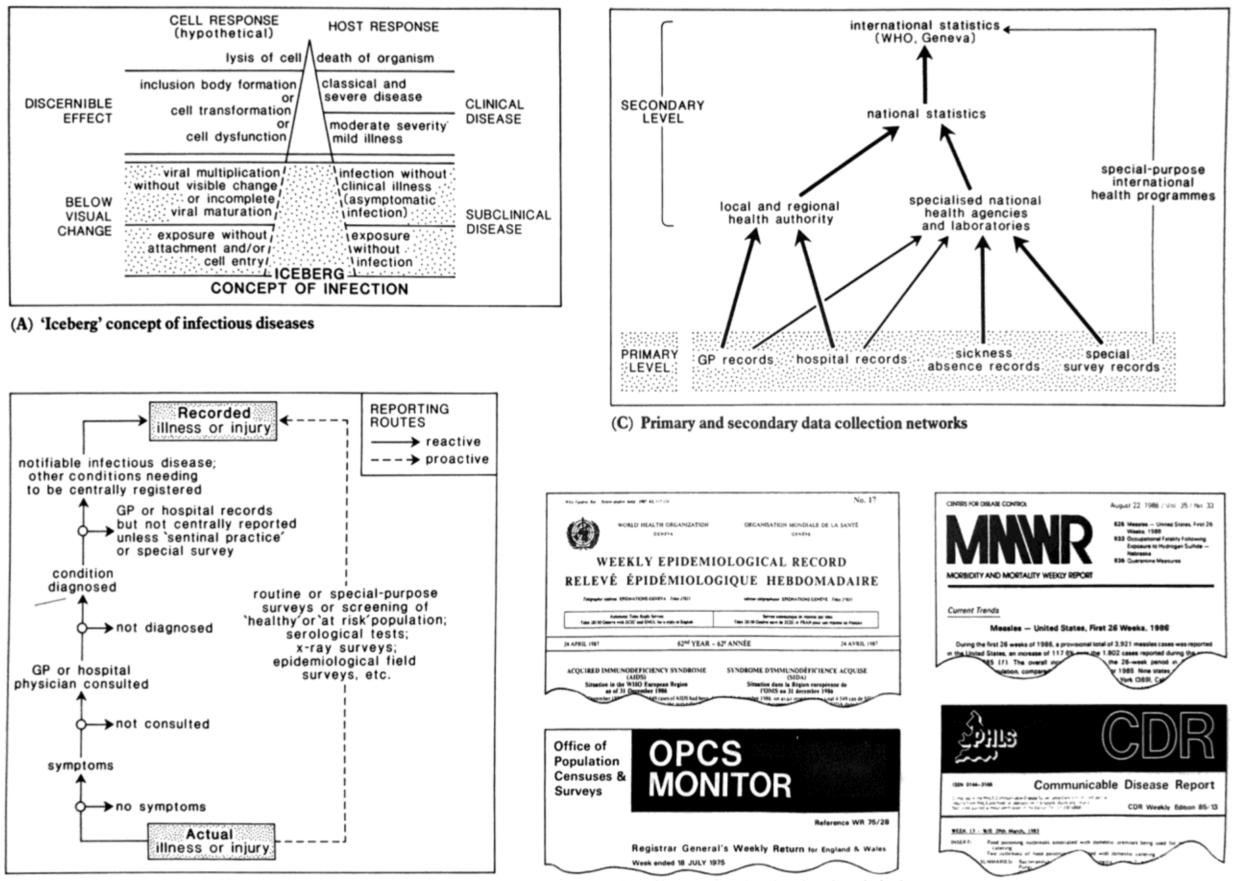
Ward, M. H., J. R. Nuckols, J. Giglierano, M. R. Bonner, C. Wolter, M. Airola, W. Mix, J. S. Colt and P. Hartge (2005). "Positional accuracy of two methods of geocoding." Epidemiology **16**(4): 542-547.

Wei, L., Q. Qian, Z.-Q. Wang, G. E. Glass, S.-X. Song, W.-Y. Zhang, X.-J. Li, H. Yang, X.-J. Wang, L.-Q. Fang and W.-C. Cao (2011). "Using geographic information system-based ecologic niche models to forecast the risk of hantavirus infection in Shandong Province, China." American Journal of Tropical Medicine and Hygiene **84**(3): 497-503.

Wieczorek, J., Q. Guo and R. J. Hijmans (2004). "The point-radius method for georeferencing locality descriptions and calculating associated uncertainty." International Journal of Geographical Information Science **18**(8): 745-767.

Zhan, F. B., J. D. Brender, I. De Lima, L. Suarez and P. H. Langlois (2006). "Match rate and positional accuracy of two geocoding methods for epidemiologic research." Annals of Epidemiology **16**(11): 842-849.

Figures



(A) 'Iceberg' concept of infectious diseases
(B) Morbidity reporting at the primary level
(C) Primary and secondary data collection networks
(D) Examples of international and national statistical sources

Figure 2.1 Summary of the disease infection, reporting, collection networks, and finally public reporting

Cliff & Haggett (1988) provide a series of diagrams reprinted here which help explain the multiple processes required for generating accurate disease data. These processes begin at the point of infection at the cellular level (Panel A), reporting of the disease at the primary care provider (Panel B), data collection networks at the local, national, and international levels (Panel C), and finally reporting of the case data to the general public (Panel D). Public health map makers need to have a complete understanding of the data flow responsible for generating disease data so that they can account for potential bias and uncertainty in the available data.

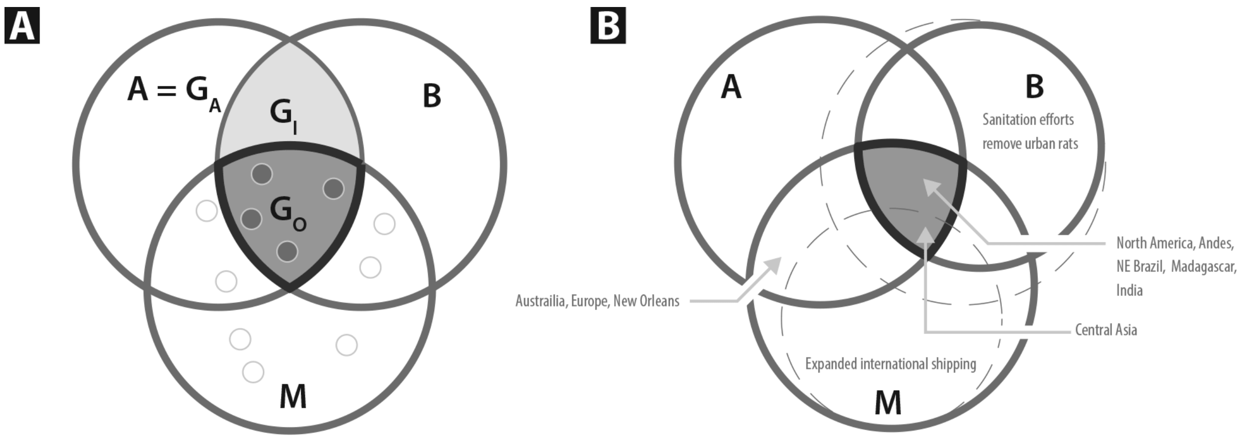


Figure 2.2. Venn diagram showing the conditions necessary for a species to be present according to ecological niche theory

This Venn diagram shows the three conditions necessary for a species to be present according to ecological niche theory, which are defined set A as abiotic physical environmental conditions; set B as biotic conditions, or favorable relations with other species occupying the same space; and set M as mobility conditions, meaning that the portions of set A which the species occupies must be physically accessible to the species of interest. Figure B applies this ecological niche theory to the disease plague, with the dashed outlines of set B and M showing the conditions present around 1900, and the solid outlines representing present day conditions. G_O represents the actual area of distribution of the species, where abiotic and biotic conditions are favorable and within reach to dispersing individuals. G_AG_I is a potential area of distribution, invasible if the structure of M changes. Redrawn from Peterson et al. (2011).

CHAPTER 3

EFFECTS OF GEOREFERENCING EFFORT ON MAPPING MONKEYPOX CASE DISTRIBUTIONS AND TRANSMISSION RISK[1]

[1] Lash, R., D. Carroll, C. Hughes, Y. Nakazawa, K. Karem, I. Damon and A. Peterson. 2012. *International Journal of Health Geographics* 11(1): 23.

Reprinted here with permission of the publisher.

Abstract

Background:

Maps of disease occurrences and GIS-based models of disease transmission risk are increasingly common, and both rely on georeferenced diseases data. Automated methods for georeferencing disease data have been widely studied for developed countries with rich sources of geographic referenced data. However, the transferability of these methods to countries without comparable geographic reference data, particularly when working with historical disease data, has not been as widely studied. Historically, precise geographic information about where individual cases occur has been collected and stored verbally, identifying specific locations using place names. Georeferencing historic data is challenging however, because it is difficult to find appropriate geographic reference data to match the place names to. Here, we assess the degree of care and research invested in converting textual descriptions of disease occurrence locations to numerical grid coordinates (latitude and longitude). Specifically, we develop three datasets from the same, original monkeypox disease occurrence data, with varying levels of care and effort: the first based on an automated web-service, the second improving on the first by reference to additional maps and digital gazetteers, and the third improving still more based on extensive consultation of legacy surveillance records that provided considerable additional information about each case. To illustrate the implications of these seemingly subtle improvements in data quality, we develop ecological niche models and predictive maps of monkeypox transmission risk based on each of the three occurrence data sets.

Results:

We found macrogeographic variations in ecological niche models depending on the type of georeferencing method used. Less-careful georeferencing identified much smaller areas as

having potential for monkeypox transmission in the Sahel region, as well as around the rim of the Congo Basin. These results have implications for mapping efforts, as each higher level of georeferencing precision required considerably greater time investment.

Conclusions:

The importance of careful georeferencing cannot be overlooked, despite it being a time- and labor-intensive process. Investment in archival storage of primary disease-occurrence data is merited, and improved digital gazetteers are needed to support public health mapping activities, particularly in developing countries, where maps and geographic information may be sparse.

Background

Georeferencing is an essential first step towards enabling GIS-based analyses of public health data (Vine et al. 1997, Rushton et al. 2008). It is the process by which textual descriptions of the geographic provenance of cases and diagnostic specimens are transformed into digital spatial data (longitude and latitude coordinates; "geocoding" is generally used to refer to the simpler process of adding geographic coordinates to postal addresses) (Hill 2009). The georeferencing process has been generalized into the following components: input records, reference datasets (e.g., gazetteers), and a georeferencer (the algorithm used to normalize, standardize, and match input records to the reference dataset) (Goldberg et al. 2007). Ideally, the process is documented with detailed metadata (Wieczorek et al. 2004).

The value of georeferenced public health data to state (MacDorman and Gay 1999) or national (Croner et al. 1996, Boulos 2004) public health systems is clear, as it enables all spatial data analysis. However, nearly all research on the efficiency, reliability, and accuracy of georeferencing methods has relied on examples of contemporary input records and reference datasets from North America and Europe (Abe and Stinchcomb 2008), possibly because georeferencing methods evolve as the availability and accuracy of reference datasets increase (Goldberg et al. 2007). In contrast, our study compares three georeferencing approaches to legacy monkeypox data from villages across Central and West Africa.

Qualitative assessments of different georeferencing methods for public health data have been developed previously (Krieger et al. 2001, Ward et al. 2005, Zhan et al. 2006, Lovasi et al. 2007, Zandbergen et al. 2012). Efforts aimed at georeferencing public health data in data-poor parts of the world include trypanosomiasis in Africa (Guerra et al. 2007) and malaria globally (Cecchi et al. 2009). However, although these studies acknowledge the challenges faced during the georeferencing process for locations where reference data are sparse or of poor quality, they do not provide a comparison of various georeferencing methods that could guide future studies needing georeferenced disease data.

Monkeypox Background

Monkeypox (MPX) virus was first identified as an agent of human disease in 1970 in the Democratic Republic of Congo (“DRC,” then Zaire) (Ladnyj et al. 1972). Prior to that date, MPX virus had been isolated only from captive cynomologous monkeys (McConnell et al. 1962). MPX presents clinically in a manner nearly indistinguishable from smallpox, and thus was cause for great concern among public health officials trying to eradicate smallpox (Foster et al. 1972).

During 1970–1986, human MPX cases were identified from seven countries across Central and West Africa as a result of localized active disease surveillance efforts (summarized in Figure 3.1). MPX cases have since been identified in Gabon (Meyer et al. 1991) and the Republic of Congo (Learned et al. 2005). Even more recently, a limited outbreak of human MPX in the United States was linked to rodents imported from Ghana (Reynolds et al. 2010), and human MPX cases have been identified in South Sudan (Formenty et al. 2010).

An MPX-specific research agenda was outlined in 1969 to address the problems that MPX posed to the smallpox eradication campaign (Fenner et al. 1988). Under this plan, World Health Organization (WHO) Collaborating Centers in the United States and the former Soviet Union, the Centers for Disease Control (CDC), and the Moscow Research Institute for Viral Preparations, respectively, provided laboratory diagnostic services, enabling new information on MPX to be assembled. This collaborative work supported serological studies during the 1970s and into the 1980s (World Health Organization 1982): surveillance activities intensified during

1981–1986 (World Health Organization 1982, 1984, 1986), when 21,994 specimens were tested from Congo, Ivory Coast, Sierra Leone, and Zaire (Ježek and Fenner 1988). During this period of intensified surveillance, 228 cases were confirmed by electron microscopy or virus culture; only 99 cases were confirmed based on serology alone, while 11 additional cases died before specimens could be collected. In all, during 1970–1986, 404 cases of human MPX disease were documented and confirmed (Ježek and Fenner 1988).

Collection of diagnostic specimens from suspected cases of MPX followed a system established by WHO during the smallpox eradication campaign (Fenner et al. 1988). Staff at local health facilities were responsible for completing semi-standardized case forms at the time diagnostic specimens were collected from patients. Specimens and forms were sent to WHO Headquarters in Geneva, Switzerland, where they were divided and sent on to the two collaborating centers. After diagnostic testing, a diagnostic result form was generated by the lab; results were either cabled to WHO Headquarters, or sent directly to personnel in the field.

During the active surveillance period, summary information from the case forms for the 404 confirmed cases was organized in data tables. Later, WHO researchers generated a digital spreadsheet of individual case information; the geographic information in this spreadsheet enabled subsequent MPX research (Levine et al. 2007). The spreadsheet contains five hierarchical place name fields for each case: country, region, district/zone, town, and locality. Unfortunately, details of the provenance of the data on the WHO spreadsheets are not known. In 2007, CDC researchers discovered that in the late 1980s, after much of the initial research agenda regarding orthopoxviruses had been completed, many of the CDC laboratory diagnostic records were converted to microfilm and the originals likely destroyed. The microfilm has since been scanned digitally, and converted to PDF formats. Preliminary comparisons of data from a few case forms against the information in the WHO spreadsheet identified several inconsistencies, which served as a motivation for this study.

An active area of recent MPX ecology and epidemiology research is based on GIS mapping and modelling techniques used to search for patterns between the locations of case occurrences and geographic and environmental variables (Ježek and Fenner 1988, Levine et al. 2007, Fuller et al. 2010, Rimoin et al. 2010). Historically, broad association of MPX virus and tropical forest was observed in early MPX research (Arita et al. 1985, Ježek et al. 1987, Khodakevich et al. 1987b); later, continental-scale ecological niche models showed that disease occurrence had stronger association with mean annual precipitation than with land cover (Levine et al. 2007). Subsequent analyses at finer spatial scales constrained to within the Congo Basin, however, pointed back to proximity to dense forest (Fuller et al. 2010), probably reflecting different scales and resolutions. However, studies to date have not considered the quality of the georeferencing of the case occurrence data used as model inputs—this point, although seemingly a simple methodological step, ends up being quite important.

Here, we test the hypothesis that different levels of effort invested in the georeferencing process can introduce considerable biases into geographic models of disease transmission. Specifically, we produce three georeferencing data sets for the MPX disease occurrences based on the same original WHO data, but differing in the detail and care with which they were derived. The first was based on automated georeferencing modules developed to facilitate the georeferencing process for biodiversity data ("automated data set"). Such automated approaches approximate the level of care and attention that many researchers pay to this step, and indeed exceed greatly the standards of some studies, which have depended on Internet search engines such as Google, Bing, and Yahoo maps, along with Open Street Map. The second data set, or "worked data set," was developed by consulting a broader suite of geographic data sources to refine the first. This method explores the results one might obtain if not intimately aware of the nuances of a set of disease data. The final data set, or "researched data set," was developed by consulting both geographic datasets and legacy CDC records ("researched data set"). This method represents the product of exhaustive searches for the greatest number of highest-quality georeferences could produce for our study system.

To compare the results of these methods, we developed ecological niche models and maps of potential MPX distributions based on each of the three occurrence data sets, and thereby can assess the effects of the different georeferencing methods on maps of MPX transmission risk (this latter defined for the purposes of this particular example as the potential for transmission at a site, given its environmental characteristics and geographic position).

Methods

Georeferencing

We used the point-radius approach (Wieczorek et al. 2004) and implemented the recommended metadata architecture (Chapman and Wieczorek 2006) to document the georeferencing process in the production of all three data sets. This approach captures (1) the original data, such that the lineage of information is preserved back to its source; (2) all decisions and assumptions made in the course of the georeferencing process; (3) the georeferenced coordinates, in a specified format and datum; and (4) a summary of uncertainty associated with the georeference. This summary of uncertainty represents an integration of uncertainty inherent in the geographic reference (e.g., an incomplete description), uncertainty in components of the geographic reference (e.g., "5 miles east" may be anything between 4.5 and 5.5 miles, and anything between northeast and southeast), and uncertainty in the underlying geography (e.g., the spatial footprint of the site referred to, distances among 'multiple hits' in matching gazetteer data). It is expressed as the radius of a circle that sums the diverse sources of uncertainty in the georeference. We relied on the MaNIS georeferencing calculator for estimating positional uncertainty (Wieczorek 2001) and excluded any locality with an uncertainty greater than 10 km.

Automated data set

The methods for producing the automated data sets are similar to the single-stage georeferencing methods described elsewhere (Lovasi et al. 2007, Wilson et al. 2010), and are summarized in Figure 3.2. We reduced the initial set of input data to unique textual locality records, and submitted the resulting table of country, state, district, municipality, and locality records to Biogeomancer for automated georeferencing. We used the automated georeferencing facility implemented in the Biogeomancer workbench (Biogeomancer Workbench 2012). This free, web-based platform automates georeferencing by taking the WHO spreadsheet input data, and searching for matching localities in the National Geospatial-Intelligence Agency's (NGA) GEOnet Names Service (GNS) database (National Geospatial-Intelligence Agency 2012), and then automatically calculating and populating the MaNIS metadata fields (Guralnick et al. 2006).

Worked data set

The methods for producing the worked data sets are akin to multi-stage georeferencing methods described elsewhere (Lovasi et al. 2007, Wilson et al. 2010), wherein we attempted to match manually input data for which satisfactory georeferences were not produced by the automated method (Figure 3.3). Here, the initial Biogeomancer output was processed further by a person knowledgeable in African geography, but without access to the case reports. Using the automated output from the Biogeomancer Workbench facility (see above) as a starter, the data were explored further, refining initial automated results using locality information on the Biogeomancer site, and incorporating additional information from additional sources: gazetteer data (Falling Rain Genomics 2010), Google Earth, and general Internet searches. The objective was to ascertain the location of each record with greater precision, and to describe uncertainty (Wieczorek et al. 2004) more accurately. This step involved 5–30 minutes of work per locality, and the result is referred to as our "worked" dataset.

Researched data set

The method used for georeferencing the researched data departs considerably from the previous two methods, and may be characterized as an iterative, detailed clerical review (Boscoe 2008), and is summarized in Figure 3.4. It is distinguished from the previous two methods because it utilized legacy primary disease data to refine the input data, and it consulted a broader range of geographic reference material than those used in the automated and researched methods. The CDC legacy case form provided the basis for modifying and refining the input data, based on the assumption that the WHO spreadsheet contained transcription and other typographical errors. Additional legacy data was used to enrich the available geographic reference material, by compiling all available historic maps of MPX case locations into a common GIS map document to easily overlay and compare geographic information from different sources (Centers for Disease Control and Prevention 1971, Smallpox Eradication Program 1971, Foster et al. 1972, Ladnyj et al. 1972, Breman et al. 1977, World Health Organization 1981, Mutombo et al. 1983, World Health Organization 1984, Arita et al. 1985, Ježek et al. 1987, Khodakevich et al. 1987a, Khodakevich et al. 1987b, Ježek and Fenner 1988, Ježek et al. 1988, Khodakevich et al. 1988, Breman et al. 1999). GNS geographic reference data was further supplemented with Joint Operation Graphics (JOG) topographic reference maps (KU Humanitarian Demining 2007, Lee 2007).

The workflow used to produce this dataset for MPX cases was iterative, as persistent and repeated searches sometimes turned up additional useful information. The initial step was to identify and resolve discrepancies between the input data from the WHO spreadsheet and the available case forms. Next, we examined all information available about individual cases to construct a sound spatial logic for identifying locations. When discrepancies were encountered, information from different sources had to be prioritized. We deemed original case forms as the most authoritative, but these records were not available for all cases. If original case forms were unavailable, the earliest published journal article reports were prioritized. If these two sources proved unhelpful, then information in review articles or marginal annotations was considered.

Once we had verified the geographic information for a given case, we began the search for a matching reference location. Our general strategy for assigning a georeference was to consult the JOG maps first, which had the finest spatial resolution, using all available information sources to find the locality on JOG maps (sometimes including preliminary GNS searches). If no location could be found or inferred there, then less-detailed data resources were used in order of decreasing precision. To expedite locating areas of interest within the JOG maps, GNS was consulted because it could be queried electronically. If a single GNS match was found, then the location could frequently be confirmed on the JOG maps and more precise coordinates recorded. If no probable match was found in GNS, or if more than one location had the same place name, then information from alternative data sources was used to guide searches. In all cases, prior to model development (see below), we discarded localities for which the uncertainty radius exceeded 10 km.

We evaluated the quality of results for each of the georeferencing methods based on completeness, positional accuracy, concordance, and repeatability (Zandbergen et al. 2012). Completeness is determined by the number of locations which could be matched to latitude and longitude coordinates. Positional accuracy is determined here by the spatial resolution of the geographic reference dataset. Concordance is difficult to quantify in this study, as it assesses whether the georeferenced coordinates match truthfully those referenced by the locality place name. Because this study is based on historical data for which it is impossible to revisit, our measure of concordance is the number of localities falling within the political geography boundary cited in the original data record. Repeatability is largely determined by the georeferencing methodology.

Ecological Niche Model Comparisons

Ecological niche modeling is a methodology that has seen extensive use in recent years (Peterson et al. 2011), and that has seen increasing applications to understanding disease geography (Peterson 2008a). We used a simple application of the methodology, as the purpose of these analyses was only to test whether different georeferencing methodologies identify different areas as "at risk" of MPX transmission. In particular, we developed models using the Genetic Algorithm for Rule-set Prediction, or GARP (Stockwell and Noble 1992), based on default settings, save for generating 100 random replicate models instead of 20, and derived a consensus model that summed the 10 models with lowest omission error out of the original 100 models.

We analyzed known MPX occurrences for each of the three georeferencing approaches in the context of 7 dimensions of climate drawn from the WorldClim climate data set (Hijmans et al. 2005). Specifically, we used annual mean temperature, mean diurnal range, maximum temperature of warmest month, minimum temperature of coldest month, annual precipitation, and precipitation of the wettest and driest months, which represent a diverse and relatively uncorrelated environmental space in which to calibrate models (Jiménez-Valverde et al. 2009). All analyses were conducted at 2.5' spatial resolution, which is equivalent to ~6.5 km near the

Equator. The niche model results were summarized as maps of putative suitable conditions, and compared by means of calculation of difference maps on a pixel-by-pixel basis.

Results

Differences in Georeferencing Methods

The 404 recorded MPX cases in the WHO spreadsheet came from 231 unique localities, a figure which may vary slightly depending on whether spelling variations are interpreted as valid entries or human error. The automated method successfully georeferenced only 69/231 localities (30% match rate); the worked method successfully georeferenced 116/231 localities (50% match rate), while the researched method successfully georeferenced 106/231 localities (match rate = 46%). Match rates for each method are broken down geographically in Table 3.1.

The georeferencing process for the researched data set is of particular interest. During this process, 48 locations were georeferenced using the input data as listed in originally in the WHO spreadsheet; georeferencing remaining localities involved careful checking against primary records and/or alternative sources of geographic information. Table 3.2 summarizes the relative utility of the additional data resources used: CDC legacy records and JOG maps provided the most valuable information, followed by a coarse-scale (1:1,000,000) map that provided information on 7 localities (World Health Organization 1981); several useful articles came from Ebola virus outbreak investigations, which covered many of the same villages.

The above discussions of development of georeferenced public health data sets may all be inconsequential if the additional precision and documentation that they provide make no tangible difference to the outcome of analyses. That is, if the results of analyses are qualitatively the same with such high-quality data as with less-carefully-prepared data, then no reason exists to invest time in the processes outlined above. Comparing the distribution of localities of these three datasets (Figure 3.5A), no MPX occurrences along the eastern, southeastern, and northeastern limits of the known distribution of the pathogen were reliable, as none could be substantiated in the researched data set.

The spatial projections of the three niche models identified areas that differed consistently. In brief, the researched data set identified broader areas throughout West Africa, as well as broader areas to the southwest and east in the Congo Basin (Figure 3.5B). Visualizing the occurrence points in a simple environmental space (annual mean temperature X annual precipitation; Figure 3.5E), we see that, although researched points define most of the extremes of the distribution of the pathogen, the points with lowest annual rainfall come from the automated dataset only. Additionally, only the worked dataset includes areas of both high temperature and high precipitation.

Discussion

The method with the best match rate overall was the worked dataset (50% match rate overall), followed by the researched dataset (46%), and finally the automated dataset (30%) (Table 3.1). Comparing match rates by country shows that the worked dataset achieved 100% success only in Ivory Coast, whereas the researched dataset achieved 100% success in Ivory Coast, Liberia, Nigeria, and Sierra Leone; the automated dataset did not achieve 100% success in any country. The researched data set was successful, for example, in Liberia, because a detailed map and set of site descriptions (Smallpox Eradication Program 1971) were among the materials that it used. A previous study (Levine et al. 2007) georeferenced 156 of 231 locations (68% match rate), but the georeferencing methods were not documented in detail.

While comparing match rates across each country provides a metric of how well different georeferencing methods performed broadly across the continent, 220/231 (95%) of MPX cases occurred in the DRC. In the DRC, the worked method achieved a match rate of 51%, the researched method 45%, and the automated method only 30%. Issues of concordance arise, however: for example, consider numbers of cases georeferenced in the DRC regions of Bas Zaire, Haut Zaire, and Shaba. The worked method identified 9 localities in Haut Zaire, but the WHO spreadsheet indicated only three (marked with an asterisk in Table 3.1). The automated method had even lower concordance, identifying 8 localities in Haut Zaire, one in Bas Zaire, and one in Shaba, when the WHO spreadsheet showed three in Haut Zaire and none in the other two regions.

Additional issues of concordance may go undetected in these automated and worked datasets, as it is not entirely clear how these methods dealt with multiple ‘hits,’ i.e., several places having the same name. In the researched processing, localities were only entered into the database if the locations fell within the indicated political geographic unit, which reduced match rates by excluding some questionable localities that did have valid returns; however, it minimized the probability of including sites falsely. Under the other two methods, this conflicting evidence was clearly viewed subjectively (worked data) or managed in unknown ways depending on distances among the multiple localities (automated data).

Information Resources for Georeferencing

When georeferencing historical disease data for foreign locations, this study shows that georeferencing results are improved by both supplementing geographic reference information, and consulting a variety of information sources to check and validate input data. The overall match rate improved considerably between the automated method and the worked and researched methods because the latter two methods utilized additional geographic reference information beyond a single gazetteer (e.g., GNS). While the overall match rate between the worked and researched methods were similar, the researched method used more authoritative geographic information resources. The worked method included the Falling Rain digital gazetteer (Falling Rain Genomics 2010) for which there is no metadata about its data sources or standards. In comparison, the researched method made extensive use of the JOG maps, which have very detailed standards and specifications (Defense Mapping Agency 1995).

The CDC legacy case forms were a unique and informative resource that illuminated and modified the information in the WHO spreadsheet which has previously been available to MPX specialists. These records allowed us to seek details of geographic reference in several dimensions—place of residence, location of the reporting clinic, etc. Such information may frequently not be available for other disease systems, but their utility in this study pointed clearly to the importance of tracking down all levels of documentation for disease case occurrences in such studies.

The legacy case forms posed challenges, though. They were not available for all 404 cases; four different variations of the typed form had been used; and forms were almost always completed by hand. In theory, cases for which CDC provided confirmatory testing ($n = 193$) should have been available; however, not all of these case forms could be located. Generally, forms captured important information, including patient identification, patient history, health facility contact information, examining physician, and regional surveillance team, and each patient was assigned a unique identification number. Specific to the geographic information on the form, a case's place of residence was captured using a hierarchy of place names, including the following fields: name of region (e.g., administrative level-1), sub-region (e.g., administrative level-2), zone (e.g., administrative level-3), collectivité (a French term for a local government administrative unit, e.g., administrative level-4), and locality (e.g., village of residence). Only one of the four versions of the case form included the sub-region field. Two versions of the form included separate zone, collectivité, and locality fields for where the affected person was when illness began, and where the case had resided two weeks prior to onset of symptoms; however, this information was most commonly identical. One version of the form did not have separate fields for each of the hierarchical place names; rather, it asked for the "complete address" of the case, and the person completing the form filled in abbreviated field names for collectivité, zone, and region.

The JOG maps also proved useful for overcoming the limited precision of the GNS data. It is worth noting that when localities from the GNS data are overlaid on the JOG maps in ArcGIS, the village locations between the two do not align perfectly, apparently owing to the higher spatial precision of the JOG maps (Figure 3.6). In GNS, nearly all Congo Basin localities have been truncated to the nearest 1' (~2.6 km near the Equator), whereas the scale of the JOG maps provides geographic precision finer than 1 km. A limitation of both the GNS and JOG maps, is the fact that little information is known about the temporal provenance of the information in either resource. Similar temporal problems with georeferenced data have been noted elsewhere (Krieger et al. 2002), and potential end users of the data must be aware that no solution is readily available.

While the GNS data set provides a helpful textual search functionality, JOG maps (which must be inspected visually by the user) allow more accurate georeferencing. Operationally, using GNS and the JOGs in tandem was the most efficient process. If a locality could be found using the text-based search in GNS, it could frequently be found and georeferenced with greater precision using the JOG maps. When a record could not be identified in GNS at the locality level, the next-higher unit place name (county, district, etc.) could frequently be found on the JOG maps, which then guided visual searches of the JOG maps for the locality—many of the place names found on JOG maps have not been captured in the GNS database. Because JOG maps were not available for our entire study area, some potential exists for spatial bias in the resulting georeferencing database. However, such areas were not omitted completely because some records could be georeferenced via other information resources, so we neglect this source of bias in our results.

The following provides an example of one of the unique and more complex instances of the georeferencing process, for the locality "Libela." Libela was recorded as a MPX occurrence locality from a case in 1972, but was not found in either the GNS database or the JOG maps. Likely alternative spellings (e.g., Libella, Lebella, etc.) were considered, but again no matching records were found. After an Internet search using Google, a reference to Libela was identified in the proceedings of a conference on Ebola virus held in 1977, where the author notes a fatal case of possible hemorrhagic fever "in Libela (38 km south of Yambuku) (Van der Groen et al. 1978)." Figure 3.6 shows a portion of a JOG map near Yambuku Mission (not labeled on the map, but noted with a church symbol, and included in the GNS database). Following the only road south from Yambuku for 38 km leads to an unlabeled populated place symbol, which we inferred to be Libela. Hence, in this example, we had to use the conjunction of GNS and JOG to identify Yambuku, and then non-standard Internet resources to find the relationship of Libela to Yambuku.

Monkeypox Transmission Geography

The extra effort invested in the 'researched' data set impacted the results of the ecological niche models. As the data in Table 3.1 shows, the researched dataset matched all of the West African locations (Nigeria, Ivory Coast, Liberia, Sierra Leone), but both the automated and worked datasets failed to locate many of the cases in this region (Figure 3.5A). Ecological niche models generated from the results of the researched method (Figure 3.5B) therefore include more area in West Africa as part of their predictions. However, models generated from the results of automated (Figure 3.5C) and worked (Figure 3.5D) georeferencing methods largely do not include much of these West African locations in their predicted distribution. The ecological conditions represented by the West African locations are different than much of the rest of the MPX ecological niche, as shown in the highlighted portion of Figure 3.5E. Areas along the northern and southern edges of the Congo Basin were more variable in the effects of researching data points, as the signals from the worked and automated data sets differed for these areas.

Even without the modeling step, the exercise of investigating each occurrence record in great depth was illuminating, and the linking of individual diagnostic results with each unique location proved insightful. No researched data point fell in the eastern quarter of the Congo Basin. Biologically more importantly, however, no researched data point comes from the Republic of the Congo, on the west side of the Congo River above Kinshasa. This latter area has not seen massive political conflicts, so this absence may in fact be real; research is underway into the causes of this lack of records from the region. Because the relational database created was able to incorporate data on confirmatory lab test as well, we can state that laboratory confirmation of MPX by viral culture occurred in 70 (66%) of the 106 localities in the researched data set, a higher standard for disease confirmation than serology testing alone. Hence, earlier studies based on the less carefully researched WHO spreadsheet (Levine et al. 2007) must be taken with a grain of salt: quite simply, different georeferencing have very-real implications for results of mapping exercises.

Conclusions

This paper contributes uniquely because we document the difficulties and limitations in the available methods for georeferencing under challenging conditions, namely historic disease data in foreign locations with poor geographic reference information. We demonstrate the utility of institutional legacy data and importance of consulting a variety of geographic data resources to the process of georeferencing. We show meaningful differences in the resulting MPX distribution depending on the georefrencing method chosen. While other studies have encountered and identified similar difficulties to georeferencing historic public health data from developing countries (Guerra et al. 2007, Cecchi et al. 2009), the MPX data used in this study are even older; we believe that our results may help other researchers in the future to plan strategically for georeferencing other historic public health data sets. Elsewhere, analyses are appearing in the literature using ecological niche modeling or other related GIS based modeling methods to examine disease distributions in various locations and at various spatial scales e.g., (Thomson et al. 1999, Peterson et al. 2006, Lash et al. 2008, Fichet-Calvet and Rogers 2009). Too often, however, occurrence data are used without careful introspection or the georeferencing process is executed without detailed attention.

Such concerns have seen considerable discussion and development in the biodiversity informatics world (Chapman 1999, Peterson et al. 2004, Wieczorek et al. 2004, Chapman 2005). In public health, a clear and robust argument of the need for georeferenced health data was put forth nearly 15 years ago (Krieger et al. 1997). Since then, a large amount of research has focused on georeferencing domestic disease occurrences (Krieger et al. 2001, Goldberg 2008, Goldberg et al. 2008, Henry and Boscoe 2008, Rushton et al. 2008). The work herein, like that of Serebriakova (2005), suggests that greater investment in georeferencing resources for international public health research is needed, and that legacy map library collections should be used to fill gaps in digital gazetteer data (Cromley 2011). In this vein, automated approaches to extracting information from scanned maps (Chiang and Knoblock 2011) may offer even greater efficiency than manual digitizing. Discussions have begun as regards alternative formats for capture of human disease occurrence data (Eisen and Eisen 2007, Peterson 2008b), but much more contemplation is needed, owing to differences in disease surveillance systems and geographic information infrastructure around the world. Emerging technologies may be one way of strengthening public health surveillance capacity, such as monitoring Twitter feeds (Davis Jr. et al. 2011), and other types of mobile communications (Freifeld et al. 2010). In light of the ongoing threat posed by emerging and re-emerging infectious diseases (Jones et al. 2008), it seems most advantageous to initiate a focus on constructing high-quality, well-documented geographic summaries of primary disease data.

Acknowledgements

The authors acknowledge the assistance of several individuals who contributed to this research at various stages. Most notably, we thank the World Health Organization (WHO) for curating and sharing the 1976–1986 monekypox database, and also Rebecca Levine, of Emory University; Timothy Fleistra of the Special Pathogens Branch, CDC; Scott McEathron of the T.R. Smith Map Library; Christine Ellis of the U.S. Department of Agriculture; and Jerome Dobson and Matt Dunbar of the Department of Geography at the University of Kansas. Shannon Keckler of the Poxvirus Program, CDC, provided useful comments on earlier drafts of this manuscript. The authors also thank two anonymous reviewers for numerous constructive comments. The views expressed in this paper are solely those of the authors and do not represent those of the CDC, U.S. Government, or any other entity with which the authors may be affiliated.

References

Abe, T. and D. G. Stinchcomb (2008). Geocoding practices in cancer registries. Geocoding health data: the use of geographic codes in cancer prevention and control, research, and practice. G. Rushton, M. P. Armstrong, J. Gittler et al. Boca Raton, FL, CRC Press**:** 111-125.

Arita, I., Z. Jezek, L. Khodakevich and K. Ruti (1985). "Human monkeypox: a newly emerged orthopoxvirus zoonosis in the tropical rain forests of Africa." American Journal of Tropical Medicine and Hygiene **34**(4): 781-789.

Biogeomancer Workbench. (2012). "Biogeomancer Workbench." from http://bg.berkeley.edu/latest/.

Boscoe, F. P. (2008). The science and art of geocoding: tips for improving match rates and handling unmatched cases in analysis. Geocoding health data: the use of geographic codes in cancer prevention and control, research, and practice. G. Rushton, M. P. Armstrong, J. Gittler et al. Boca Raton, CRC Press**:** 95-109.

Boulos, M. N. (2004). "Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom." International Journal of Health Geographics **3**(1): 1.

Breman, J. G., K. M. Johnson, G. van der Groen, C. B. Robbins, M. V. Szczeniowski, K. Ruti, P. A. Webb, F. Meier and D. L. Heymann (1999). "A search for ebola virus in animals in the Democratic Republic of the Congo and Cameroon: ecologic, virologic, and serologic surveys, 1979-1980. Ebola virus study teams." Journal of Infectious Diseases **179 Suppl 1**: S139-147.

Breman, J. G., J. H. Nakano, E. Coffi, H. Godfrey and J. C. Gautun (1977). "Human poxvirus disease after smallpox eradication." American Journal of Tropical Medicine and Hygiene **26**(2): 273-281.

Cecchi, G., M. Paone, J. R. Franco, E. M. Fevre, A. Diarra, J. A. Ruiz, R. C. Mattioli and P. P. Simarro (2009). "Towards the atlas of human African trypanosomiasis." International Journal of Health Geographics **8**: 15.

Centers for Disease Control and Prevention (1971). "Human infection with monkeypox Vvrus." Morbidity and Mortality Weekly Report **20**(8): 65-66.

Chapman, A. D. (1999). Quality control and validation of point-sourced environmental resource data. Spatial accuracy assessment: Land information uncertainty in natural resources. K. Lowell and A. Jaton. Chelsea, MI, Ann Arbor Press**:** 409-418.

Chapman, A. D. (2005). Principles of data quality. Report for the Global Biodiversity Information Facilty. Copenhagen**:** 58.

Chapman, A. D. and J. Wieczorek, Eds. (2006). Guide to best practices for georeferencing. Copenhagen, Global Biodiversity Information Facility.

Chiang, Y.-Y. and C. Knoblock (2011). "A general approach for extracting road vector data from raster maps." International Journal on Document Analysis and Recognition: 1-27.

Cromley, E. K. (2011). "The role of the map and geographic information library in medical geographic research." Journal of Map and Geography Libraries **7**(1): 37-41.

Croner, C. M., J. Sperling and F. R. Broome (1996). "Geographic information systems (GIS): new perspectives in understanding human health and environmental relationships." Statistics in Medicine **15**(18): 1961-1977.

Davis Jr., C. A., G. L. Pappa, D. R. R. de Oliveira and F. de L. Arcanjo (2011). "Inferring the location of Twitter messages based on user relationships." Transactions in GIS **15**(6): 735-751.

Defense Mapping Agency (1995). [Report] Military Specifications: joint operation graphics series 1501A and 1501 (JOG Air/Ground). U.S. Department of Defense. Washington D.C. **:** 333.

Eisen, L. and R. J. Eisen (2007). "Need for improved methods to collect and present spatial epidemiologic data for vectorborne diseases." Emerging infectious diseases **13**(12): 1816-1820.

Falling Rain Genomics, I. (2010). "Global Gazetteer 2.0." from http://www.fallingrain.com.

Fenner, F., D. A. Henderson, I. Arita and Z. Ježek (1988). Smallpox and its eradication. Geneva, World Health Organization.

Fichet-Calvet, E. and D. J. Rogers (2009). "Risk maps of Lassa Fever in West Africa." PLoS Neglected Tropical Diseases **3**(3): e388.

Formenty, P., M. O. Muntasir, I. Damon, V. Chowdhary, M. L. Opoka, C. Monimart, E. M. Mutasim, J. C. Manuguerra, W. B. Davidson, K. L. Karem, J. Cabeza, S. Wang, M. R. Malik, T. Durand, A. Khalid, T. Rioton, A. Kuong-Ruay, A. A. Babiker, M. E. Karsani and M. S. Abdalla (2010). "Human monkeypox outbreak caused by novel virus belonging to Congo Basin clade, Sudan, 2005." Emerging Infectious Diseases **16**(10): 1539-1545.

Foster, S. O., E. W. Brink, D. L. Hutchins, J. M. Pifer, B. Lourie, C. R. Moser, E. C. Cummings, O. E. K. Kuteyi, R. E. Eke, J. B. Titus, E. A. Smith, J. W. Hicks and W. H. Foege (1972). "Human monkeypox." Bulletin of the World Health Organization **46**(5): 569-576.

Freifeld, C. C., R. Chunara, S. R. Mekaru, E. H. Chan, T. Kass-Hout, A. Ayala Iacucci and J. S. Brownstein (2010). "Participatory epidemiology: use of mobile phones for community-based health reporting." PLoS Medicine **7**(12): e1000376.

Fuller, T., H. Thomassen, P. Mulembakani, S. Johnston, J. Lloyd-Smith, N. Kisalu, T. Lutete, S. Blumberg, J. Fair, N. Wolfe, R. Shongo, P. Formenty, H. Meyer, L. Wright, J.-J. Muyembe, W. Buermann, S. Saatchi, E. Okitolonda, L. Hensley, T. Smith and A. Rimoin (2010). "Using remote sensing to map the risk of human monkeypox virus in the Congo Basin." EcoHealth **8**(1): 14-25.

Goldberg, D. (2008). A geocoding best practices guide. Springfield, IL, North American Association of Central Cancer Registries.

Goldberg, D., J. P. Wilson, C. Knoblock, B. Ritz and M. Cockburn (2008). "An effective and efficient approach for manually improving geocoded data." International Journal of Health Geographics **7**.

Goldberg, D. W., J. P. Wilson and C. A. Knoblock (2007). "From text to geographic coordinates: the current state of geocoding." URISA Journal **19**(1): 33-46.

Guerra, C. A., S. I. Hay, L. S. Lucioparedes, P. W. Gikandi, A. J. Tatem, A. M. Noor and R. W. Snow (2007). "Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project." Malaria Journal **6**: 17.

Guralnick, R. P., J. Wieczorek, R. Beaman and R. J. Hijmans (2006). "BioGeomancer: automated georeferencing to map the world's biodiversity data." PLoS Biology **4**(11): e381.

Henry, K. A. and F. P. Boscoe (2008). "Estimating the accuracy of geographical imputation." International Journal of Health Geographics **7**: 3.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones and A. Jarvis (2005). "Very high resolution interpolated climate surfaces for global land areas." International Journal of Climatology **25**(15): 1965-1978.

Hill, L. L. (2009). Georeferencing : the geographic associations of information. Cambridge, MA, MIT.

Ježek, Z., I. Arita, M. Mutombo, C. Dunn, J. H. Nakano and M. Szczeniowski (1986). "Four generations of probable person-to-person transmission of human monkeypox." American Journal of Epidemiology **123**(6): 1004-1012.

Ježek, Z. and F. Fenner (1988). Human monkeypox. New York, Karger.

Ježek, Z., B. Grab, K. M. Paluku and M. V. Szczeniowski (1988). "Human monkeypox: disease pattern, incidence and attack rates in a rural area of northern Zaire." Tropical and Geographical Medicine **40**(2): 73-83.

Ježek, Z., J. H. Nakano, I. Arita, M. Mutombo, M. Szczeniowski and C. Dunn (1987). "Serological survey for human monkeypox infections in a selected population in Zaire." Journal of Tropical Medicine and Hygiene **90**(1): 31-38.

Jiménez-Valverde, A., Y. Nakazawa, A. Lira-Noriega and A. T. Peterson (2009). "Environmental correlation structure and ecological niche model projections." Biodiverstiy Informatics **6**: 28-35.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman and P. Daszak (2008). "Global trends in emerging infectious diseases." Nature **451**(7181): 990-993.

Khodakevich, L., Z. Jezek and D. Messinger (1988). "Monkeypox virus: ecology and public health significance." Bulletin of the World Health Organanization **66**(6): 747-752.

Khodakevich, L., M. Szczeniowski, Manbu-ma-Disu, Z. Jezek, S. Marennikova, J. Nakano and D. Messinger (1987a). "The role of squirrels in sustaining monkeypox virus transmission." Tropical and Geographical Medicine **39**(2): 115-122.

Khodakevich, L., M. Szczeniowski, D. Nambu ma, Z. Jezek, S. Marennikova, J. Nakano and F. Meier (1987b). "Monkeypox virus in relation to the ecological features surrounding human settlements in Bumba zone, Zaire." Tropical and Geographical Medicine **39**(1): 56-63.

Krieger, N., J. T. Chen and G. Ebel (1997). "Can we monitor socioeconomic inequalities in health? A survey of U.S. health departments' data collection and reporting practices." Public Health Reports **112**(6): 481-491.

Krieger, N., P. Waterman, J. T. Chen, M. J. Soobader, S. V. Subramanian and R. Carson (2002). "Zip code caveat: bias due to spatiotemporal mismatches between zip codes and US census-defined geographic areas--the Public Health Disparities Geocoding Project." American Journal of Public Health **92**(7): 1100-1102.

Krieger, N., P. Waterman, K. Lemieux, S. Zierler and J. W. Hogan (2001). "On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research." American Journal of Public Health **91**(7): 1114-1116.

KU Humanitarian Demining (2007). Joint Operation Graphics: Democratic Republic of Congo country mosaic [DIGITAL DATA FILE]. T.R. Smith Map Collection, University of Kansas Libraries. Lawrence.

Ladnyj, I. D., P. Ziegler and E. Kima (1972). "A human infection caused by monkeypox virus in Basankusu Territory, Democratic Republic of the Congo." Bulletin of the World Health Organization **46**(5): 593-597.

Lash, R. R., N. A. Brunsell and A. T. Peterson (2008). "Spatiotemporal environmental triggers of Ebola and Marburg virus transmission." Geocarto International **23**(6): 451 - 466.

Learned, L. A., M. G. Reynolds, D. W. Wassa, Y. Li, V. A. Olson, K. Karem, L. L. Stempora, Z. H. Braden, R. Kline, A. Likos, F. Libama, H. Moudzeo, J. D. Bolanda, P. Tarangonia, P. Boumandoki, P. Formenty, J. M. Harvey and I. K. Damon (2005). "Extended interhuman transmission of monkeypox in a hospital community in the Republic of the Congo, 2003." American Journal of Tropical Medicine and Hygiene **73**(2): 428-434.

Lee, K. D. (2007). "Access to aeronautical charts and data by the U.S. National Geospatial-Intelligence Agency: an ominous trend." Journal of Map and Geography Libraries **3**(1): 5-37.

Levine, R. S., A. T. Peterson, K. L. Yorita, D. Carroll, I. K. Damon and M. G. Reynolds (2007). "Ecological niche and geographic distribution of human monkeypox in Africa." PLoS One **2**(1): e176.

Lovasi, G., J. Weiss, R. Hoskins, E. Whitsel, K. Rice, C. Erickson and B. Psaty (2007). "Comparing a single-stage geocoding method to a multi-stage geocoding method: how much and where do they disagree?" International Journal of Health Geographics **6**(1): 12.

MacDorman, M. F. and G. A. Gay (1999). "State initiatives in geocoding vital statistics data." Journal of Public Health Management and Practice **5**(2): 91-93.

McConnell, S. J., Y. F. Herman, D. E. Mattson and L. Erickson (1962). "Monkey pox disease in irradiated cynomologous monkeys." Nature **195**(4846): 1128-1129.

Meyer, A., J. J. Esposito, F. Gras, T. Kolakowski, M. Fatras and G. Muller (1991). "First appearance of monkey pox in human beings in Gabon." Medecine Tropicale **51**(1): 53-57.

Mutombo, M., I. Arita and Z. Jezek (1983). "Human monkeypox transmitted by a chimpanzee in a tropical rain-forest area of Zaire." Lancet **1**(8327): 735-737.

National Geospatial-Intelligence Agency. (2012). "GEOnet Names Server (GNS)." Retrieved 30 May, 2012, from http://earth-info.nga.mil/gns/html/.

Peterson, A. T. (2008a). "Biogeography of diseases: a framework for analysis." Naturwissenschaften **95**(6): 483-491.

Peterson, A. T. (2008b). "Improving methods for reporting spatial epidemiologic data." Emerging Infectious Diseases **14**(8): 1335-1336.

Peterson, A. T., R. R. Lash, D. Carroll and K. M. Johnson (2006). "Geographic potential for outbreaks of Marburg hemorrahagic fever." American Journal of Tropical Medicine and Hygiene **75**(1): 9-15.

Peterson, A. T., A. G. Navarro-Siguenza and R. Scachetti Pereira (2004). "Detecting errors in biodiversity data based on collectors' itineraries." Bulletin of the British Ornithologists' Club **124**(2): 143-150.

Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martinez-Meyer, M. Nakamura and M. B. Araujo (2011). Ecological niches and geographic distributions. Princeton, Princeton University Press.

Reynolds, M. G., D. S. Carroll, V. A. Olson, C. Hughes, J. Galley, A. Likos, J. M. Montgomery, R. Suu-Ire, M. O. Kwasi, J. J. Root, Z. Braden, J. Abel, C. Clemmons, R. Regnery, K. Karem and I. K. Damon (2010). "A silent enzootic of an orthopoxvirus in Ghana, West Africa: evidence for multi-species involvement in the absence of widespread human disease." American Journal of Tropical Medicine and Hygiene **82**(4): 746-754.

Rimoin, A. W., P. M. Mulembakani, S. C. Johnston, J. O. Lloyd Smith, N. K. Kisalu, T. L. Kinkela, S. Blumberg, H. A. Thomassen, B. L. Pike, J. N. Fair, N. D. Wolfe, R. L. Shongo, B. S. Graham, P. Formenty, E. Okitolonda, L. E. Hensley, H. Meyer, L. L. Wright and J.-J. Muyembe (2010). "Major increase in human monkeypox incidence 30 years after smallpox vaccination campaigns cease in the Democratic Republic of Congo." Proceedings of the National Academy of Sciences of the United States of America **107**(37): 16262-16267.

Rushton, G., M. P. Armstrong, J. Gittler, B. R. Greene, C. E. Pavlik, M. M. West and D. L. Zimmerman, Eds. (2008). Geocoding health data: the use of geographic codes in cancer prevention and control, research, and practice. Boca Raton, FL, CRC Press.

Serebriakova, T. N. (2005). Mapping radiation from the external and internal sources in Belarus after the Chernobyl disaster: Implications for epidemiological analysis PhD thesis, University of Connecticut.

Smallpox Eradication Program (1971). [Special Report] Human infections with monkeypox virus: Liberia and Sierra Leone. Atlanta, U.S. Center for Disease Control. **5:** 34-49.

Stockwell, D. R. B. and I. R. Noble (1992). "Induction of sets of rules from animal distribution data: a robust and informative method of data analysis." Mathmatics and Computers in Simulation **33**(5-6): 385-390.

Thomson, M. C., D. A. Elnaiern, R. W. Ashford and S. J. Connor (1999). "Towards a kala azar risk map for Sudan: mapping the potential distribution of *Phlebotomus orientalis* using digital data of environmental variables." Tropical Medicine and International Health **4**(2): 105-113.

Van der Groen, G., K. M. Johnson, F. A. Webb, H. Wulff and J. V. Lange (1978). Results of ebola antibody surveys in various populations groups. Proceedings of an International Colloquium on Ebola Virus Infection and other Haemorrhagic Fevers Antwerp, Belgium, Elsevier / North-Holland Biomedical Press.

Vine, M. F., D. Degnan and C. Hanchette (1997). "Geographic information systems: their use in environmental epidemiologic research." Environmental Health Perspectives **105**(6): 598-605.

Ward, M. H., J. R. Nuckols, J. Giglierano, M. R. Bonner, C. Wolter, M. Airola, W. Mix, J. S. Colt and P. Hartge (2005). "Positional accuracy of two methods of geocoding." Epidemiology **16**(4): 542-547.

Wieczorek, J. (2001). "MaNIS georeferencing calculator." Retrieved 23 January, 2011, from http://manisnet.org/gc.html.

Wieczorek, J., Q. Guo and R. J. Hijmans (2004). "The point-radius method for georeferencing locality descriptions and calculating associated uncertainty." International Journal of Geographical Information Science **18**(8): 745-767.

Wilson, J., D. Shepherd, M. Rosenman and A. Kho (2010). "Identifying risk factors for healthcare-associated infections from electronic medical record home address data." International Journal of Health Geographics **9**(1): 47.

World Health Organization (1978). "Ebola haemorrhagic fever in Zaire, 1976." Bulletin of the World Health Organization **56**(2): 271-293.

World Health Organization (1981). Report of meeting on the implementation of post-smallpox eradication policy, Geneva, 3-5 February 1981 [WHO/SE/81.159]. Geneva, Switzerland, World Health Organization**:** 21.

World Health Organization (1982). Report of the first meeting of the Committee on Orthopoxvirus Infections, Geneva, 3-5 March 1982 [WHO/SE/82.160]. Geneva, Switzerland, World Health Organization**:** 10.

World Health Organization (1984). Report of the third meeting of the Committee on Orthopoxvirus Infections, Geneva, 28-30 March 1984 [WHO/SE/84.162]. Geneva, Switzerland, World Health Organization**:** 18.

World Health Organization (1986). Report of the fourth meeting of the Committee on Orthopoxvirus Infections, Geneva, 24-26 March 1986 [WHO/SE/86.163]. Geneva, Switzerland, World Health Organization**:** 18.

Zandbergen, P. A., T. C. Hart, K. E. Lenzer and M. E. Camponovo (2012). "Error propagation models to examine the effects of geocoding quality on spatial analysis of individual-level datasets." Spatial and Spatio-temporal Epidemiology **3**(1): 69-82.

Zhan, F. B., J. D. Brender, I. De Lima, L. Suarez and P. H. Langlois (2006). "Match rate and positional accuracy of two geocoding methods for epidemiologic research." Annals of Epidemiology **16**(11): 842-849.

Figures

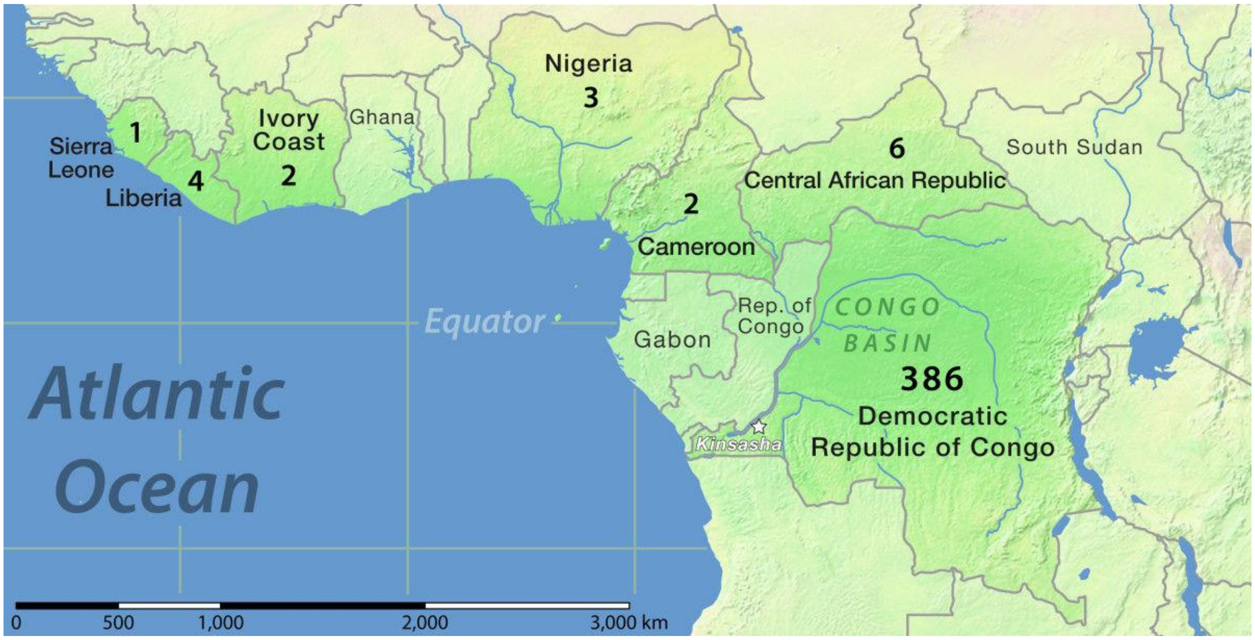


Figure 3.1. Total reported monkeypox case distribution across Central and West Africa, 1970–1986.

The distribution of monkeypox (MPX) cases in seven countries where MPX cases were reported through the joint WHO/CDC surveillance efforts, including the total number of cases identified within each county (Ježek and Fenner 1988). Countries labeled in gray without numbers indicate locations where additional MPX or MPX-related disease have occurred since 1986 (Meyer et al. 1991, Learned et al. 2005, Formenty et al. 2010, Reynolds et al. 2010).

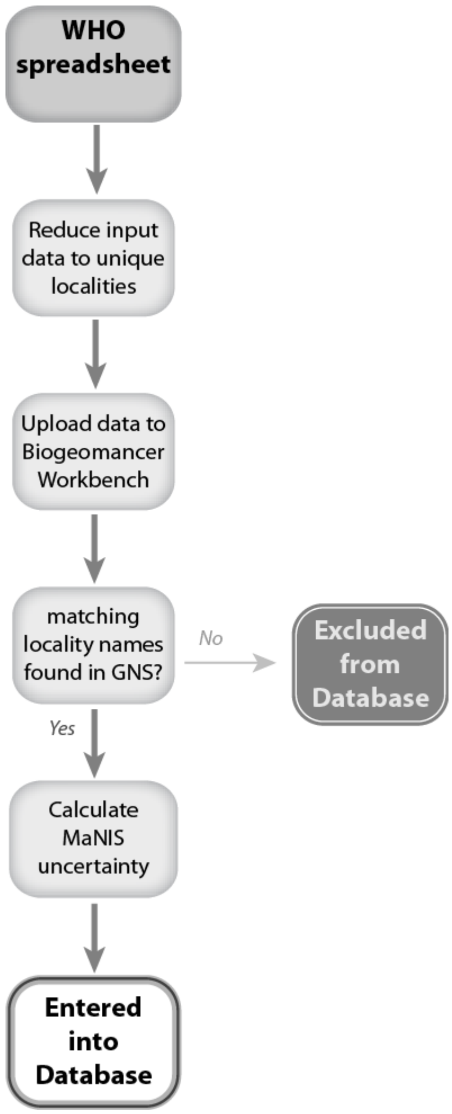


Figure 3.2 Flow diagram of the geocoding process used to generate the Automated data set.

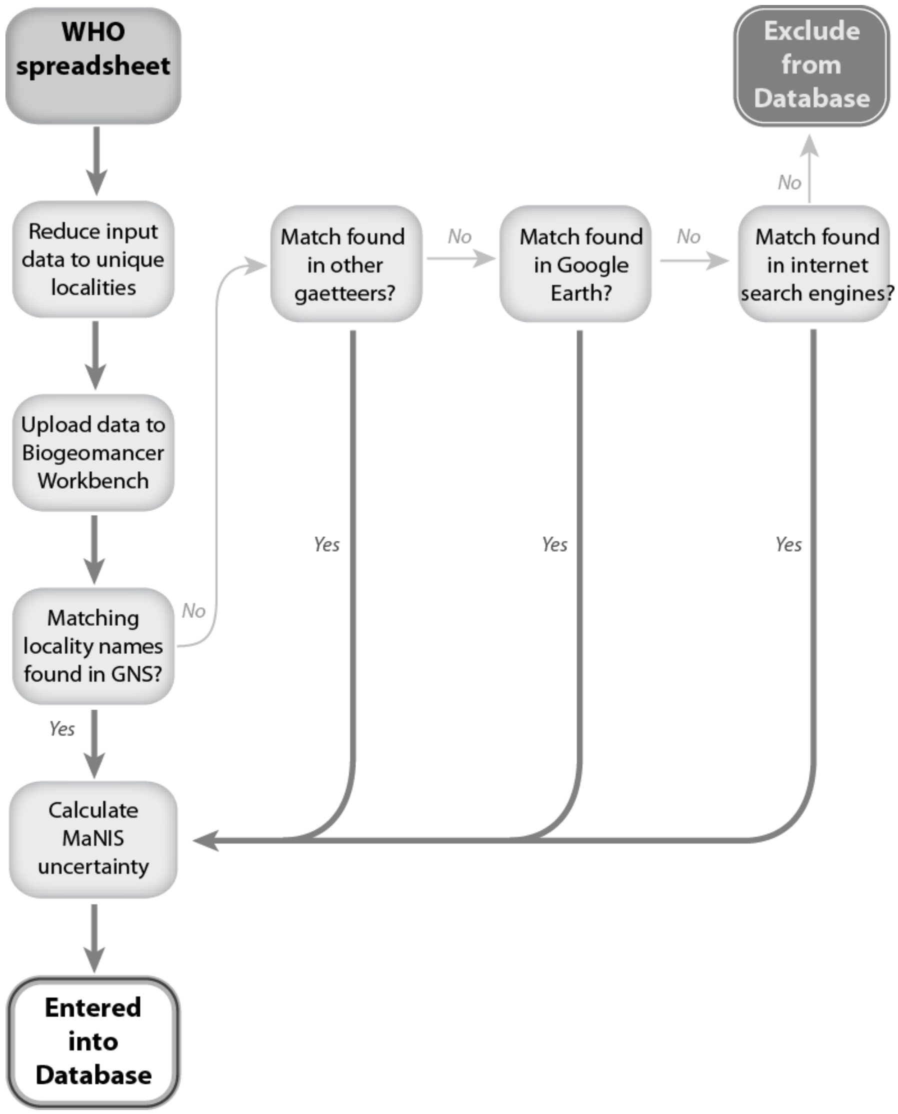


Figure 3.3 Flow diagram of the geocoding process used to generate the Worked data set.

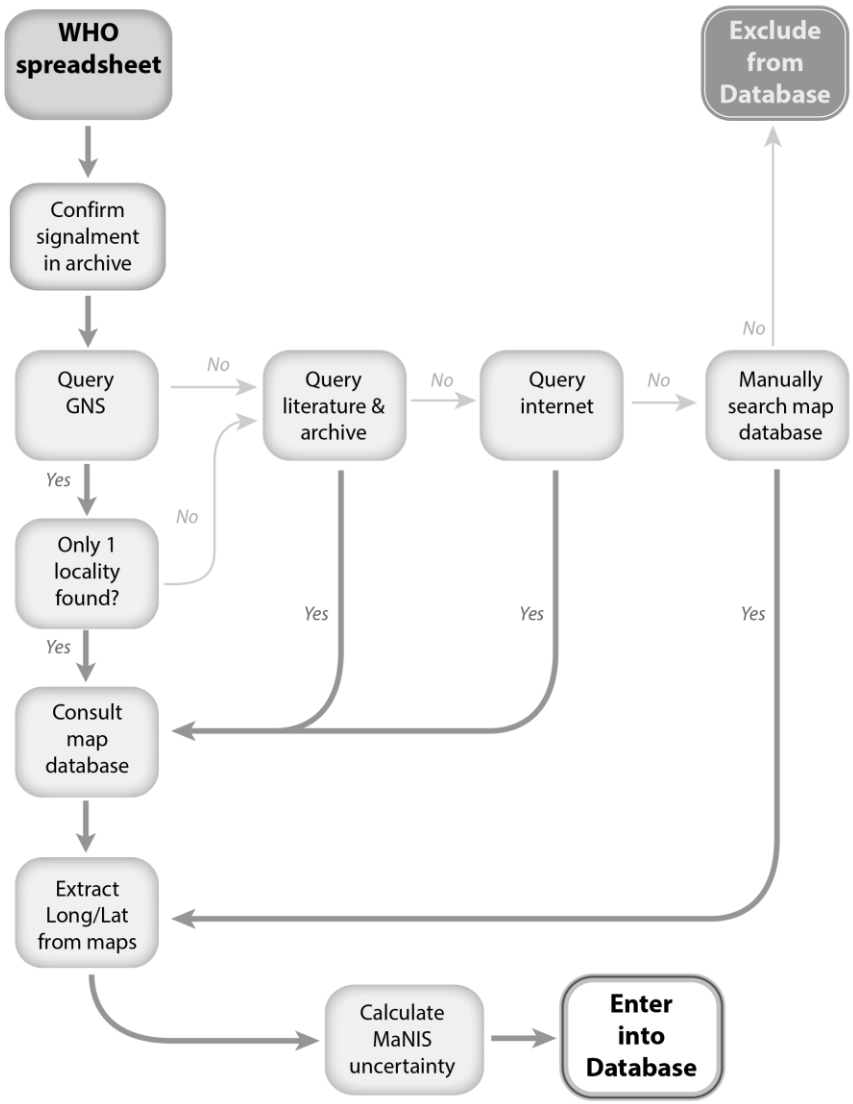


Figure 3.4 Flow diagram of the geocoding process used to generate the Researched data set.

Table 3.1 Comparison of georeferencing match rates across countries and sub-national units for each different method

The number of monkeypox case localities were matched at different rates in different national and sub-national units (i.e. state or province), which are expressed as fractions and percentages, relative to numbers of unique localities reported there in the WHO spreadsheet. Bolded regions in the DRC represent likely errors of commission, where more localities were georeferenced than would be expected based on the WHO spreadsheet. Asterisks identify probable specific instances of this type of error, such that calculating match rate percentages are not useful.

		Researched		**Worked**		**Automated**	
Country							
Sub-national unit	**WHO Locations**	**Matched**	**%**	**Matched**	**%**	**Matched**	**%**
Cameroon	2	0/2	0	1/2	50	0/2	0
Centre	2	0/2	0	1/2	50	0/2	0
Central African Republic	2	0/2	0	0/2	0	0/2	0
Sangha	2	0/2	0	0/2	0	0/2	0
Democratic Republic of the Congo	220	99/220	45	112/220	51	67/220	30
Bandundu	37	14/37	38	23/37	62	12/37	32
Bas Zaire	0	0/0	n/a	0/0	n/a	1/0*	n/a
Equateur	143	62/143	43	71/143	50	38/143	27
Haut Zaire	3	2/3	67	9/3*	n/a	8/3*	n/a
Kasai Occidental	3	2/3	67	1/3	33	2/3	67
Kasai Oriental	31	19/31	61	6/31	19	5/31	16
Kivu	3	0/3	0	2/3	67	0/3	0
Shaba	0	0/0	n/a	0/0	n/a	1/0*	n/a
Ivory Coast	2	2/2	100	2/2	100	1/2	50
Abengourou	1	1/1	100	1/1	100	0/1	0
Haut-Sassandra	1	1/1	100	1/1	100	1/1	100
Liberia	2	2/2	100	0/2	0	0/2	0
Grand Gedeh	2	2/2	100	0/2	0	0/2	0
Nigeria	2	2/2	100	1/2	50	1/2	50
East Central	1	1/1	100	0/1	0	0/1	0

		Researched		Worked		Automated	
Country							
Sub-national unit	**WHO Locations**	**Matched**	**%**	**Matched**	**%**	**Matched**	**%**
Oyo	1	1/1	100	1/1	100	1/1	100
Sierra Leone	1	1/1	100	0/1	0	0/1	0
Southern	1	1/1	100	0/1	0	0/1	0
Overall	**231**	**106/231**	**46**	**116/231**	**50**	**69/231**	**30**

Table 3.2 Geographic information resources consulted for "researched" dataset

The number of monkeypox case localities which benefited from more detailed CDC legacy data and other historic materials, by resource name.

Localities	Name	Reference
43	Joint Operation Graphic's (JOG's)	(KU Humanitarian Demining 2007)
18	Legacy CDC case forms	
	Reports	
4	Report of Meeting on the implementation of Post-Smallpox Eradication Policy	(World Health Organization 1981)
3	Human infections with monkeypox virus: Liberia and Sierra Leone	(Smallpox Eradication Program 1971)
	Articles	
3	The role of squirrels in sustaining monkeypox virus transmission.	(Khodakevich et al. 1987a)
2	Ebola haemorrhagic fever in Zaire, 1976.	(World Health Organization 1978)
4	A search for Ebola virus in animals in the Democratic Republic of the Congo and Cameroon: ecologic, virologic, and serologic surveys, 1979–1980.	(Breman et al. 1999)
1	Human monkeypox.	(Foster et al. 1972)
1	Human poxvirus disease after smallpox eradication.	(Breman et al. 1977)
1	Four generations of probable person-to-person transmission of human monkeypox.	(Ježek et al. 1986)
1	Results of Ebola antibody surveys in various populations groups	(Van der Groen et al. 1978)

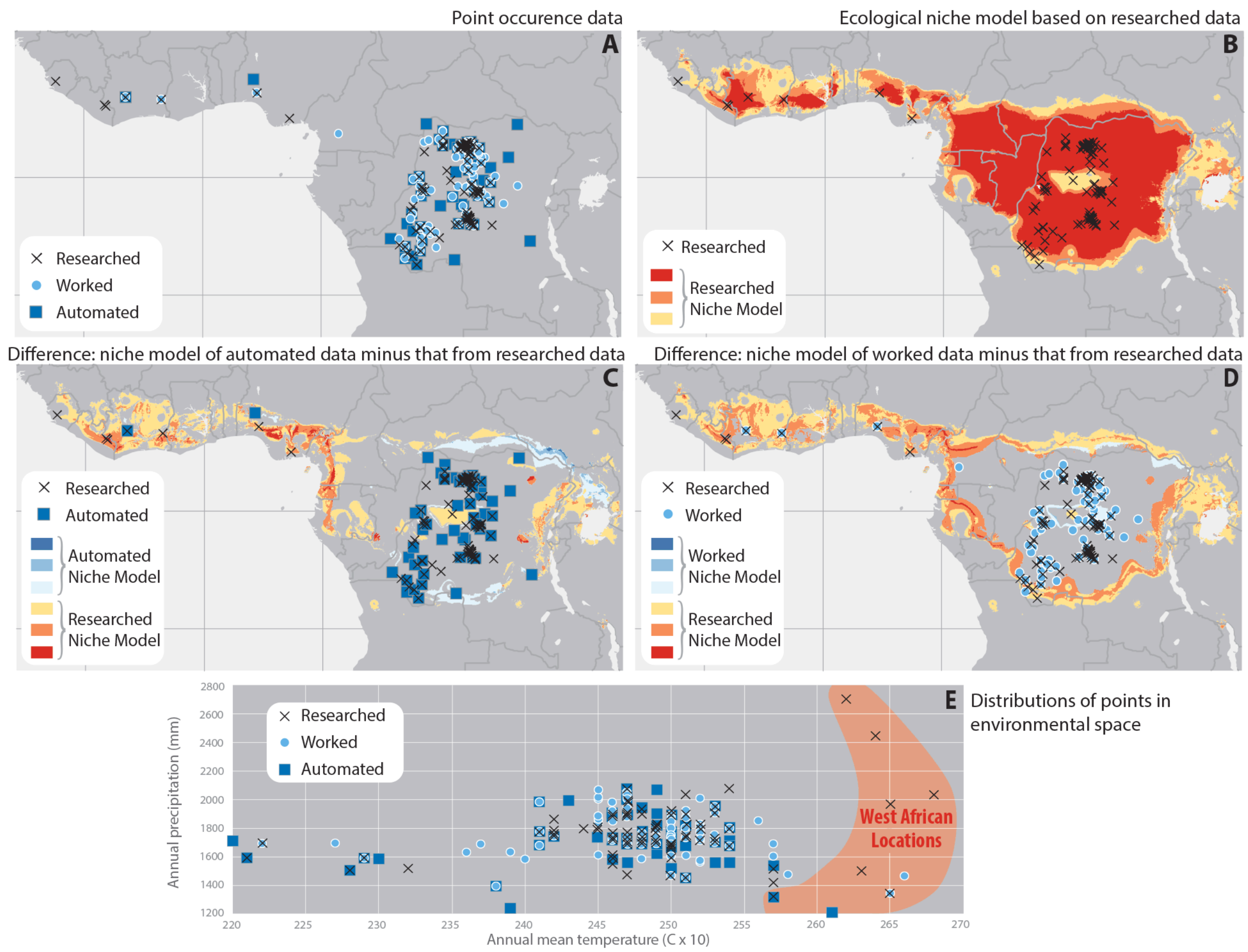


Figure 3.5 Exploration of effects of different levels of care and detail in georeferencing of human monkeypox cases on derivative transmission risk maps.

Models derived from the automated and worked occurrence data differ in environmental and geographic dimensions from those based on the carefully researched occurrence data points. See text for additional detail. Red and orange areas in panels C and D are those that are more extensive in the researched data set, while blue areas are those that are less extensive. Panel E highlights portions of the ecological niche unique to the West African countries (Nigeria, Ivory Coast, Liberia, Sierra Leone) which were located using the researched method, but largely missed by the other two methods.

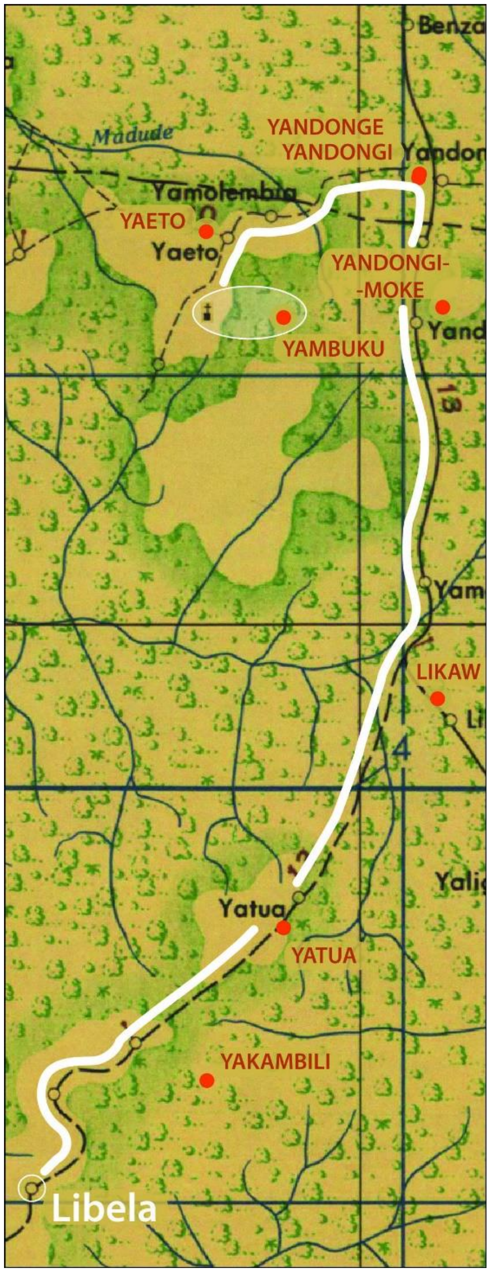


Figure 3.6 Example of application of complex spatial logic to georeferencing a difficult locality.

A portion of a JOG map is shown, with GNS gazetteer data overlaid as orange dots with orange labels. The village of Libela did not appear on either the JOG map or in the GNS database, but anecdotal reference was made to it as being 38 km south of Yambuku (Van der Groen et al. 1978). Using ArcGIS, a 38 km distance (solid white line) from Yambuku Mission (church symbol on JOG map highlighted in white) to the south led to an unnamed village on the JOG map 38 km away, which could reasonably be inferred to be Libela.

CHAPTER 4

ENABLING CLINICIANS TO EASILY FIND LOCATION-BASED TRAVEL HEALTH RECOMMENDATIONS-- IS INNOVATION NEEDED?[1]

[1] Lash, R. R., A. Walker, C. V. Lee, R. C. LaRocque, S. R. Rao, E. T. Ryan, G. Brunette, N. K. Holton and M. J. Sotir. To be submitted to the *Journal of Travel Medicine.*

Abstract

Background

The Centers for Disease Control and Prevention (CDC) publishes travel health recommendations to enable clinicians to perform risk assessments during pretravel consultations. Most recommendations describe risk at the country level; however, recommendations for subnational areas are also made when disease risk varies within a country, as they currently do for malaria and yellow fever. It is unknown whether the types of place names CDC uses to describe disease risk (e.g., country name, state name, city name, etc.) match the types found in travel itinerary descriptions during pretravel consultations. Understanding place name that types clinicians must search for would be valuable for developing new search tools and display formats to help clinicians find more targeted recommendations on a multitude of travel-related illnesses.

Objectives

Analyze the types of place names currently found in travel itinerary descriptions and evaluate how these terms can be used to develop new tools for clinicians to use.

Methods

Data analyzed were individual travel itineraries collected using a standard pretravel consultation form at GlobalTravEpiNet (GTEN) travel clinic sites. We selected a stratified random sample of itineraries from records which contained responses to an open-ended question from 18 GTEN clinics. Place names were extracted and classified as one of 8 different types: multi-national area; country; state/province; county/district; physical geographic area (e.g., island, mountain region); vague subnational area; populated places (e.g., cities, villages); tourist destination (e.g., national park, historic site). Itineraries could include multiple place names and place name types. Summary statics were generated.

Results

Of 1,756 itineraries sampled, 1,570 (89%) itineraries included one or more place name, for a total of 3,377 place names. More frequently found types were: 2,119 (63%) populated places; 336 (10%) tourist destinations, 283 (8%) physical geographic area, and 206 (6%) vague subnational areas. The types used by CDC were found less frequently, including 163 (5%) state, 153 (5%) country, and 48 (1%) county.

Conclusions

This study shows that the type of place names most frequently used to describe travel itineraries during pretravel consultations are rarely the ones used currently by CDC to describe national and subnational health recommendations. This means that clinicians must use additional maps, atlases, or online search tools to cross reference the provided place names to the available health recommendations. Clinical tools using geographic information technology to directly identify health recommendations would make it easier for clinicians to find recommendation information.

Introduction

The Centers for Disease Control and Prevention (CDC) publishes travel health recommendations to enable clinicians to perform risk assessments during pretravel consultations. Most recommendations describe risk at the country level; however, recommendations for subnational areas are also made when disease risk varies within a country, as they currently do for malaria and yellow fever (Hay et al. 2013). As disease surveillance improves in the future, subnational recommendations may be needed for other diseases. Subnational place name (e.g., state name, city name, etc.) search terms regularly used by clinicians are largely undocumented. Obtaining place name search data would be useful in developing new search tools and display formats to help clinicians find more targeted recommendations on a multitude of travel-related illnesses.

Objectives

This study has three objectives. The first objective is to classify the types of place names used to describe a travelers' itinerary during pretravel consultations, as recorded in a representative sample of patient intake forms obtained from the Global Travel EpiNetwork (GTEN). The second objective is to summarize types of place names as compared to the currently used terms to report travel health recommendations. If there are discrepancies between the place names types used in pretravel consultations, and the place names used in travel health recommendations, then this suggests clinicians may experience a burden when trying to find and interpret travel health recommendations. If a discrepancy is found, then the final objective is to make recommendations about how new geospatial technology (e.g., interactive maps and place name search services) could be used to develop new tools for clinicians to use to find relevant travel health information more efficiently.

Background

There is a rich history of mapmaking and spatial analysis in public health (Koch 2005), and increasingly the application of geographic information systems (GIS) within epidemiology (Cromley and McLafferty 2012). Bauer and Puotinen (2002) suggest different ways that GIS can contribute to travel medicine, including the use of GIS for querying spatial databases of disease presence. At the time that Bauer and Puotinen published their article, GIS primarily consisted of data and software on a personal computer that required considerable training to be able to use. Since then, global positioning systems (GPS), as well as web-based mapping and mobile technologies have revolutionized the way that we use and consume map information (Batty et al. 2010, Longley et al. 2011).

One GIS application that Bauer and Puotinen (2002) envisioned was the ability for travel health clinicians to quickly and easily search global GIS databases of travel related diseases and disease risks. Such tools could save valuable time during pre-travel consultations, which can often take more of the clinicians time than what their appointment schedule permits (Hatz and Chen 2013). The individual risk assessment is an important and potentially time consuming part of the consultation. The risk assessment requires travelers to provide important details about their medical history and planned trip, including a detailed travel itinerary, while the clinician must process this information to determine the relevant health risks the traveler will likely encounter.

To process destination information, clinicians usually use a travel medicine reference book or website. The CDC publishes a clinical medicine textbook called the CDC's Yellow Book (Brunette 2017), and also makes this information available through CDC's Travel Health website (Centers for Disease Control and Prevention 2017). In addition to CDC's reference materials, clinicians may use other reference sources, such as the World Health Organization's

International Travel and Health website (World Health Organization 2012a), other countries' travel health recommendations, or those of commercial travel information providers (e.g. Shoreland®, Global Infectious Disease and Epidemiology Online Network (GIDEON), TropiMed®, etc.). However, recommendations for certain diseases may vary within a country, such as yellow fever and malaria. Because these recommendations are made based on political boundaries within countries, clinicians need sufficiently detail maps in order to locate the travel destinations and described risk areas (Deye and Magill 2013), and have been advised to acquire separate atlases, world maps, or globes (Hill and Rosselot 2013) to assist them. An international study of nurses suggested that requiring clinicians to rely on their own ability to locate and interpret these numerous mapping resources remains overly burdensome (Bauer et al. 2013).

To try to address this need, the CDC Yellow Book has included some country-specific yellow fever and malaria maps (Figure 4.1). The scale and format of these maps can only show a limited number of labeled places on the map, so should a clinician not find the destination they are looking for, they must still rely on separate mapping resources to be able to interpret the published travel health recommendations. Some CDC map users indicate that the map's ease of use may be questionable (Kohl 2016).

Previous research has shown that the itinerary information provided during pretravel consultations can be problematic. It is subject to changing after the pre-travel consultations (Rossi and Genton 2012), or may be ill-defined at the time of the pretravel consultation (Flaherty and Md Nor 2016). Rossi and Genton (2012) noted that the impact of the differences between reported and actual travel plans only altered their recommendations for the rabies-vaccine, and all other vaccine and malaria prophylaxis recommendations remained the same.

Data and Methods

Global TravEpiNet (GTEN) Patient Intake Forms

Global TravEpiNet (GTEN) is a network of 29 travel clinics from across the United States (Figure 4.2) organized to improve the health of those traveling internationally (LaRocque et al. 2012). Each clinic uses a standardized electronic data collection form. Although travelers are sked to complete the form online prior to their appointment, clinicians may add additional details during the pre-travel consultation. The form captures the intended travel itineraries in two ways: 1) a mandatory question that asks users to select one or more places from a list of country names; and 2) an optional question allowing users to write-in “additional details regarding their destinations.” After reviewing the proposed research and data request, GTEN data managers executed the desired stratified random sampling scheme, and furnished a deidentifed data table which included the two itinerary variables, as well as demographic variables describing traveler age, sex, length of travel, and purpose for travel.

A total of 35,119 forms from 29 GTEN clinics contained responses to the second “additional destination detail” question. 11 of the 29 GTEN clinics were excluded from the eligible sample set (n=110 form) because the optional question was not regularly completed at these clinics, reducing the total number of eligible forms down to 35,009 from 18 clinics. From those clinics with greater than 100 eligible forms (n=15), 100 forms were randomly selected for review. From the remaining clinics that had fewer than 100 eligible forms but greater than 50 (n=3), all available forms were selected (256 forms). In summary, 18 out of 29 (62%) GTEN sites met our sampling requirements and resulted in a sample size of 1,756 intake forms.

Classifying Itinerary Place Names by Type

A manual process was used to identify, extract, and classify individual place names found within the intake forms that included responses to the “additional destination details” question. The processing methodology is summarized as follows:

- Examine additional destination information and identify possible place names
- Record possible place names in place name data table
- Look up possible place names in online resources that travelers likely use (e.g., Google Maps, Wikipedia, and other travel websites) to identify place name type and correct any misspellings
- Classify place name type according to Table 4.1

Table 4.2 shows an example of how the manual process was executed. In this entry, the additional destination details response was a list of place names separated by commas, and ending with a period, followed by a shorthand clinician’s note. The first location can be easily recognized as the city of Kolkata, in India, which is assigned the “populated place” type. Similarly, Thimpu is the capital city of Bhutan, and also a populated place type. “Mountain areas” would be considered a physical geographic area. Jaipur is recognized as another populated place in India. A search for “Ramthambore” using the Google Search Engine did not yield a match, but the search results identify Ranthambore National Park as a likely spelling itinerary, and national parks are a type of tourist destination. Varanasi is found to be another populated place in India. The remainder of the information “. Guided, OAT” is interpreted as a shorthand note entered by the clinician meaning “guided overseas adventure travel.” This example is representative of the types of semi-structured free-text responses found within the intake forms.

A Microsoft Access (Microsoft Office 2013, Redmond, WA) relational database with custom data entry forms was used to input and manage data during manual processing contained in the “additional destination details” field, and for creating summary data tables. During the process of identifying individual place names, their type, and spelling, some assumptions were needed to be made to handle the variation within the unstructured data. These assumptions were:

1) If names such as Singapore, or Hong Kong, or Sao Paulo were encountered, they were coded as the type populated place, although there are also administrative areas with the same name;
2) If a hierarchical place name list was provided, such as "Cancun, Quintana Roo, Mexico", it was interpreted that the traveler was visiting only one destination "Cancun," and that the other place names were simply descriptors of which "Cancun" the traveler was going to.
3) Country place names were not counted if they were already selected as a response to the mandatory country list question.

Results

Table 4.3 shows a demographic breakdown of the entire GTEN dataset compared to the sample of 1,756 intake forms selected through the stratified random sampling procedure. Overall the sample appears representative of the larger GTEN dataset. The sample is comprised of 14% more women than men, and is largely comprised of adults born before 1990. The highest proportion of travelers were traveling to the African Region

Of all the itineraries sampled, 1,570 (89%) itineraries included one or more place name, for a total of 3,366 place names (Table 4.4). By comparing the total number of place names per GTEN site to the number of itineraries, we see that two are not strongly correlated. Table 4.5 shows that the most frequently found types were: 2,119 (63%) populated places; 336 (10%) tourist destinations, 283 (8%) physical geographic area, and 206 (6%) vague subnational areas.

The types used in CDC recommendations were found less frequently, including 163 (5%) state, 153 (5%) country, and 48 (1%) county.

Figure 4.3 shows that when you compare the proportion of place name types described in aggregate above (column 1) to the individual proportions from each GTEN site, the trends are similar. This figure shows that the proportion of populated places varies between GTEN site, though it consistently makes-up the largest proportion of all place names. GTEN Site R has the highest proportion of country type place names of all the GTEN sites, and conversely the lowest proportion of populated place types. Further examination of the individual records for this site shows that one intake form reports an itinerary that includes 16 different countries.

Discussion

This data shows that populated place types were listed 6 times more often than tourist destinations and 10 times more often than state/province and country types. It is clear that the types of place names commonly used to describe travel itineraries during pretravel consultations are different than the place names commonly used to define travel health recommendations. There are some limitations to these results. Because the “additional destination detail” question on the GTEN form is optional and users are given essentially no instruction on what type of information to input, little is known about why some forms contain this information and others do not. Similarly, because both the traveler and the clinician are able to enter information into the intake form, it is difficult to associate patterns in the data with the knowledge, attitudes, or practices of either the traveler or the clinician.

These results are consistent with results from related tourism and hospitality research, which has sought to understand the way that travelers use location to search for travel information (e.g., tourist activities, hotels, etc.) to inform tourist destination website development. Hwang et al. (2009) studied the transcripts of U.S. based domestic travelers calling and Illinois state tourism information call-center. As the study reported herein, they classified the different place names reported in the queries as either the state, region within the state, county, and city place name types, for single destination searches and multi-destination searches. They found that cities were the most frequent place name type used overall, with 83% of the single destination searched and 75% of the multi-destination searches, while counties and states were the least common and second-least common place name types, respectively. Similar research was done on a sample of internet search engine queries to study travelers' accommodation search query, and found that cities were again the most common type of place name type used, and were used four times more frequently than country and state place names (Pan et al. 2007).

Based on the evidence we have presented, as well as the data reported by tourism related research, it is reasonable to conclude that travelers plan their travel itineraries at a local scale, planning their itineraries in terms of "Where is the airport located?", "Where am I going to sleep?", "What types of activities do I want to do while I am there?" These behaviors are unlikely to change in such a way that it would make it quicker and easier for travel health clinicians to identify the appropriate travel health recommendations. It is equally important to acknowledge that public health agencies generate public health recommendations based upon the best available surveillance data and the smallest political geographic units at which the data are reported (Shlim and Magill 2017). It is conceivable that improved surveillance and reporting systems may be implemented in the future however, and that subnational travel health recommendations may become more common (Hay et al. 2013), and thus more burdensome to clinicians.

Conclusions

Using new geospatial technology such as interactive maps to display travel health information is not new, (Centers for Disease Control and Prevention 2009, World Health Organization 2012b), including more recent examples from commercial entities (Sanofi Pasteur Austrailia 2017, Travax® 2017). However, our study is the first to provide data to document the clinical need for these tools, specifically the specific need for place name search capabilities. Future studies should continue this line of inquiry to enable new and better clinical tools to be developed so that doctors, nurses, and travelers may be confident that they are using the best and most accurate information available.

This study shows that the type of place names most frequently used to describe travel itineraries during pretravel consultations are rarely the ones by public health authorities use to report national and subnational travel health recommendations. As a result, clinicians must undertake a time-consuming process of cross-referencing the place names in the itinerary against additional maps, atlases, or online search tools to find out where these places are located relative to the description of the travel health recommendations. This time consuming process for clinicians could be made less burdensome by developing a travel health recommendation search application which incorporates geospatially formatted travel health recommendation data, interactive web maps, and a place name search services.

References

Batty, M., A. Hudson-Smith, R. Milton and A. Crooks (2010). "Map mashups, Web 2.0 and the GIS revolution." Annals of GIS **16**(1): 1-13.

Bauer, I., S. Hall and N. Sato (2013). "Providing travel health care - The nurses' role: An international comparison." Travel Medicine and Infectious Disease **11**(4): 214-224.

Bauer, I. L. and M. Puotinen (2002). "Geographic information systems and travel health." Journal of Travel Medicine **9**(6): 308-314.

Brunette, G., Ed. (2017). CDC Yellow Book 2018: Health Information for International Travel. New York, Oxford University Press.

Centers for Disease Control and Prevention. (2009). "CDC Malaria Map Application." Retrieved 2 March, 2016, from http://www.cdc.gov/malaria/map/.

Centers for Disease Control and Prevention. (2017). "CDC Travelers' Health Website." Retrieved 4 April, 2017, from www.cdc.gov/travel.

Cromley, E. K. and S. McLafferty (2012). GIS and Public Health. New York, NY, Guilford Press.

Deye, G. A. and A. J. Magill (2013). Malaria: epidemiology and risk to the traveler. Travel Medicine. D. O. Freedman, P. E. Kozarsky, B. A. Connor and H. D. Nothdurft. London, Content Repository Only!**:** 135-142.

Flaherty, G. and M. N. Md Nor (2016). "Travel itinerary uncertainty and the pre-travel consultation—a pilot study." Journal of Travel Medicine **23**(1): tav010-tav010.

Hatz, C. and L. H. Chen (2013). Pre-travel consultation. Travel Medicine. D. O. Freedman, P. E. Kozarsky, B. A. Connor and H. D. Nothdurft. London, Content Repository Only!**:** 31-36.

Hay, S. I., D. B. George, C. L. Moyes and J. S. Brownstein (2013). "Big data opportunities for global infectious disease surveillance." PLOS Medicine **10**(4): e1001413.

Hill, D. R. and G. Rosselot (2013). Starting, organizing, and marketing a travel clinic. Travel Medicine. D. O. Freedman, P. E. Kozarsky, B. A. Connor and H. D. Nothdurft. London, Content Repository Only!**:** 13-24.

Hwang, Y.-H., Z. Xiang, U. Gretzel and D. R. Fesenmaier (2009). "Assessing Structure in Travel Queries." Anatolia **20**(1): 223-235.

Koch, T. (2005). Cartographies of disease: maps, mapping, and medicine. Redlands, ESRI Press.

Kohl, S. (2016). "Shoreland® vs. TropiMed® vs. Travel Care®: Which is the Best Travel Medicine Software?" TravelReadyMD https://travelreadymd.com/reveiw-travel-medicine-software/ 2017.

LaRocque, R. C., S. R. Rao, J. Lee, V. Ansdell, J. A. Yates, B. S. Schwartz, M. Knouse, J. Cahill, S. Hagmann, J. Vinetz, B. A. Connor, J. A. Goad, A. Oladele, S. Alvarez, W. Stauffer, P.

Walker, P. Kozarsky, C. Franco-Paredes, R. Dismukes, J. Rosen, N. A. Hynes, F. Jacquerioz, S. McLellan, D. Hale, T. Sofarelli, D. Schoenfeld, N. Marano, G. Brunette, E. S. Jentes, E. Yanni, M. J. Sotir, E. T. Ryan and t. G. T. Consortium (2012). "Global TravEpiNet: A national consortium of clinics providing care to international travelers—analysis of demographic characteristics, travel destinations, and pretravel healthcare of high-risk US international travelers, 2009–2011." Clinical Infectious Diseases **54**(4): 455-462.

Longley, P., M. Goodchild, D. J. Maguire and D. W. Rhind (2011). Geographic information systems & science. Hoboken, NJ, Wiley.

Pan, B., S. W. Litvin and T. E. O'Donnell (2007). "Understanding accommodation search query formulation: The first step in putting 'heads in beds'." Journal of Vacation Marketing **13**(4): 371-381.

Rossi, I. A. and B. Genton (2012). "The reliability of pre-travel history to decide on appropriate counseling and vaccinations: a prospective study." Journal of Travel Medicine **19**(5): 284-288.

Sanofi Pasteur Austrailia. (2017). "Vaccine Hub." Retrieved 20 Jne, 2017, from https://vaccinehub.com.au/travel-vaccination-advice.

Shlim, D. R. and A. J. Magill (2017). Perspective: why guidelines differ. CDC Yellow Book 2018: Health Information for International Travel. G. W. Brunette. New York, Oxford University Press**:** 9-13.

Travax®, S. (2017). "Travax provides a wealth of information for 239 countries." Retrieved 13 June, 2017, from http://www.shoreland.com/services/travax/destination-information#maps.

World Health Organization. (2012a). "International Travel and Health." Retrieved 19 July, 2017, from http://www.who.int/ith/.

World Health Organization. (2012b). "International Travel Health Interactive." Retrieved 20 June, 2017, from http://apps.who.int/ithmap/.

World Health Organization. (2017). "Alphabetical List of WHO Member States." Retrieved 1 June, 2017, from http://www.who.int/choice/demography/by_country/en/.

Figures

Recommendations: None.

MALARIA

No malaria transmission.

ECUADOR, INCLUDING THE GALÁPAGOS ISLANDS

(see Maps 3-25 and 3-26.)

YELLOW FEVER

Requirements: Required if traveling from a country with risk of YFV transmission and ≥1 year of age, including transit >12 hours in an airport located in a country with risk of YFV transmission.[1]

Recommendations:

Recommended for all travelers ≥9 months of age traveling to areas <2,300 m in elevation[2] in the following provinces east of the Andes Mountains: Morona-Santiago, Napo, Orellana, Pastaza, Sucumbios, and Zamora-Chinchipe (see Map 3-25).

Generally not recommended for travelers whose itinerary is limited to areas <2,300 m in elevation[2] in the following provinces west of the Andes mountains: Esmeraldas, Guayas, Los Rios, Santa Elena, Santo Domingo de los Tsachilas, and designated areas of Azuay, Bolivar, Canar, Carchi, Chimborazo, Cotopaxi, El Oro, Imbabura, Loja, Pichincha, and Tungurahua (see Map 3-25).

Not recommended for travelers whose itineraries are limited to all areas >2,300 m in elevation,[2] the cities of Guayaquil and Quito, or the Galápagos Islands (see Map 3-25).

MALARIA

Areas with malaria: Areas at altitudes <1,500 m (4,921 ft) in the provinces of Carchi, Esmeraldas, Morona Santiago, Orellana, and Pastaza. Rare cases in other provinces in areas <1,500m (4,921 ft). Not present in the cities of Guayaquil and Quito or the Galápagos Islands (see Map 3-26).

Estimated relative risk of malaria for US travelers: Very low.

Drug resistance[4]: Chloroquine.

Malaria species: *P. vivax* 66%, *P. falciparum* 34%.

Recommended chemoprophylaxis: Areas with malaria in Carchi, Esmeraldas, Morona Santiago, Orellana, and Pastaza Provinces: Atovaquone-proguanil, doxycycline, or mefloquine.

Other areas with rare cases of malaria: Mosquito avoidance only.

MAP 3-25. Yellow fever vaccine recommendations in Ecuador[1]

[1,2] See footnotes on page 424.

detained in quarantine for up to 6 days after departure from an area at risk of yellow fever transmission.

Recommendations: None.

MALARIA

No malaria transmission.

EL SALVADOR

YELLOW FEVER

Requirements: Required if traveling from a country with risk of YFV transmission and ≥1 year of age, including transit >12 hours in an airport located in a country with risk of YFV transmission.[1]

Recommendations: None.

EQUATORIAL GUINEA

YELLOW FEVER

Requirements: Required if traveling from a country with risk of YFV transmission and ≥6 months of age.[1]

Recommendations: ***Recommended*** for all travelers ≥9 months of age.

MALARIA

Areas with malaria: All.

Estimated relative risk of malaria for US travelers: High.

Drug resistance[4]: Chloroquine.

Malaria species: *P. falciparum* 85%; *P. malariae, P. ovale,*

Figure 4.1 Photograph of Yellow Fever and Malaria travel health recommendations for Ecuador from 2018 CDC Yellow Book

This photograph shows the way that yellow fever and malaria travel health recommendation information for Ecuador are presented in the current CDC Yellow Book (Brunette 2017). The Ecuador country information is described in words beginning in the upper-left, and a map visualizing this information can be seen in the upper-right. Recommendations are defined by provinces in the text, so all provinces are shown and labeled on the map, though the density of these labels prevents other places from being shown and labeled on the map. A separate Ecuador malaria risk map appears on the following page of the book.

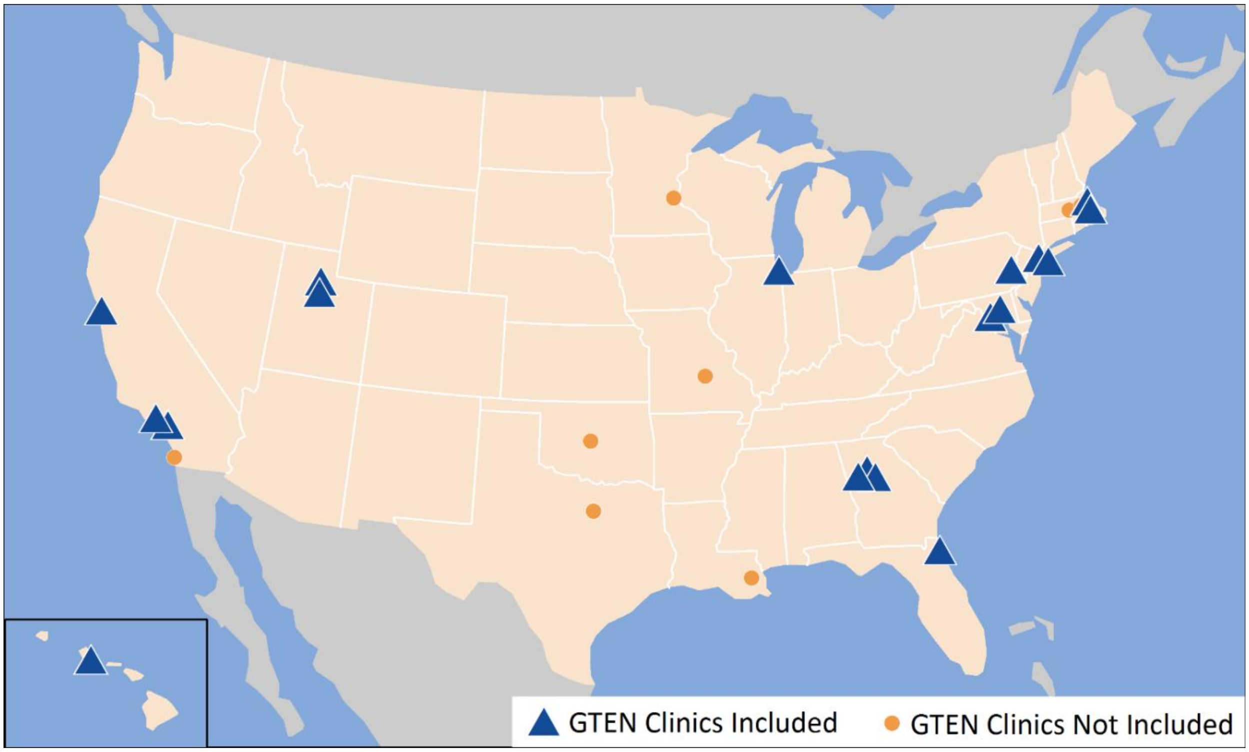


Figure 4.2. Distribution of Global Travel EpiNet (GTEN) clinic sites included and not included in the final analysis

This map shows the geographic distribution of Global Travel EpiNet (GTEN) clinic sites across the United States that use the standardized patient intake form. Blue triangles show the GTEN clinics that were included in this study, and orange circles show the clinics which were not included in the study (based up criteria defined in the text).

Table 4.1. Defining different types of place names

Place Name Type	Definition
Multi-national area	An area covering more than one country
Country/territory	A sovereign political entity found on GTEN country list
State/province	A subnational first-order administrative region within a country
County/district/municipality	A subnational second-order administrative region within a country, nested within a state/province
Populated place	A city, village, or airport
Tourist destination	Any specific park, resort, or cultural heritage site
Physical geographic area	A mountain, mountain range, river, ocean, ecological zone
Vague subnational area	An area clearly within a country but for which the location or boundaries are ill-defined and not clearly demarcated on any available map
Undefined	A named location which cannot be found in any of the online resources consulted

Table 4.2. Example of data processing scenario

Additional destination details:		
Kolkata, Thimpu, Mountain Areas, Jaipur, Ramthambore, Tented Camp, Varanasi. Guided, OAT		
Country	**Place Name**	**Place Name Type**
India	Kolkata	Populated place
Bhutan	Thimpu	Populated place
Bhutan	Mountain Areas	Physical geo area
India	Jaipur	Populated place
India	Ranthambore National Park	Tourist destination
India	Varanasi	Populated place

Table 4.3. Demographic and Travel-Related Characteristics of Travelers Included in Sample

Characteristic		All Travelers n= 40,810		Travelers Sampled n=1,756	
Sex					
	Female	22,987	56%	1,001	57%
	Male	17,823	44%	755	43%
Birth cohort					
	1957-1989	35,053	86%	1,417	81%
	1990 or after	5,757	14%	339	19%
Region of Travel [1,2]					
	Africa	14,471	35%	667	38%
	Americas	11,562	28%	537	31%
	Eastern Mediterranean	2,156	5%	86	5%
	Europe	2,025	5%	112	6%
	Southeast Asia	10,090	25%	431	25%
	Western Pacific	7,052	17%	339	19%
Reason for travel [2]					
	Business	8,775	22%	299	17%
	Humanitarian service work	3,180	8%	298	17%
	Leisure	20,507	50%	977	56%
	Other	4,935	12%	169	10%
	Research/Education	5,272	13%	188	11%
	Visiting Friends and Relatives	3,208	8%	139	8%
Duration of Travel					
	<14 day	17,014	42%	677	39%
	>= 14 days	23,766	58%	1,077	61%

[1] Region of travel defined according to WHO Administrative Regions (World Health Organization 2017)
[2] Not mutually exclusive groups

Table 4.4. Sampled intake forms with additional place names recorded by GTEN site

GTEN Site ID	Sampled Intake Forms	Itineraries with Place names	Place names
Site A	100	98	138
Site B	100	83	240
Site C	90	80	245
Site D	100	85	307
Site E	100	94	325
Site F	100	93	176
Site G	100	89	215
Site H	100	91	135
Site I	100	89	249
Site J	83	76	152
Site K	100	81	130
Site L	100	95	195
Site M	100	91	125
Site N	100	91	168
Site O	100	96	219
Site P	100	71	88
Site Q	100	92	161
Site R	83	75	98
TOTAL	**1,756**	**1,570**	**3,366**

Table 4.5. Frequency of different place name types

Type of place name	Count of References	Percentage of total references
Populated place	2,119	63%
Tourist destination	336	10%
Physical geographic area	283	8%
Vague subnational area	206	6%
State/province	163	5%
Country	142	4%
Multi-national area	57	2%
County/district/municipality	48	1%
Undefined	12	0%
Total place names	**3,366**	

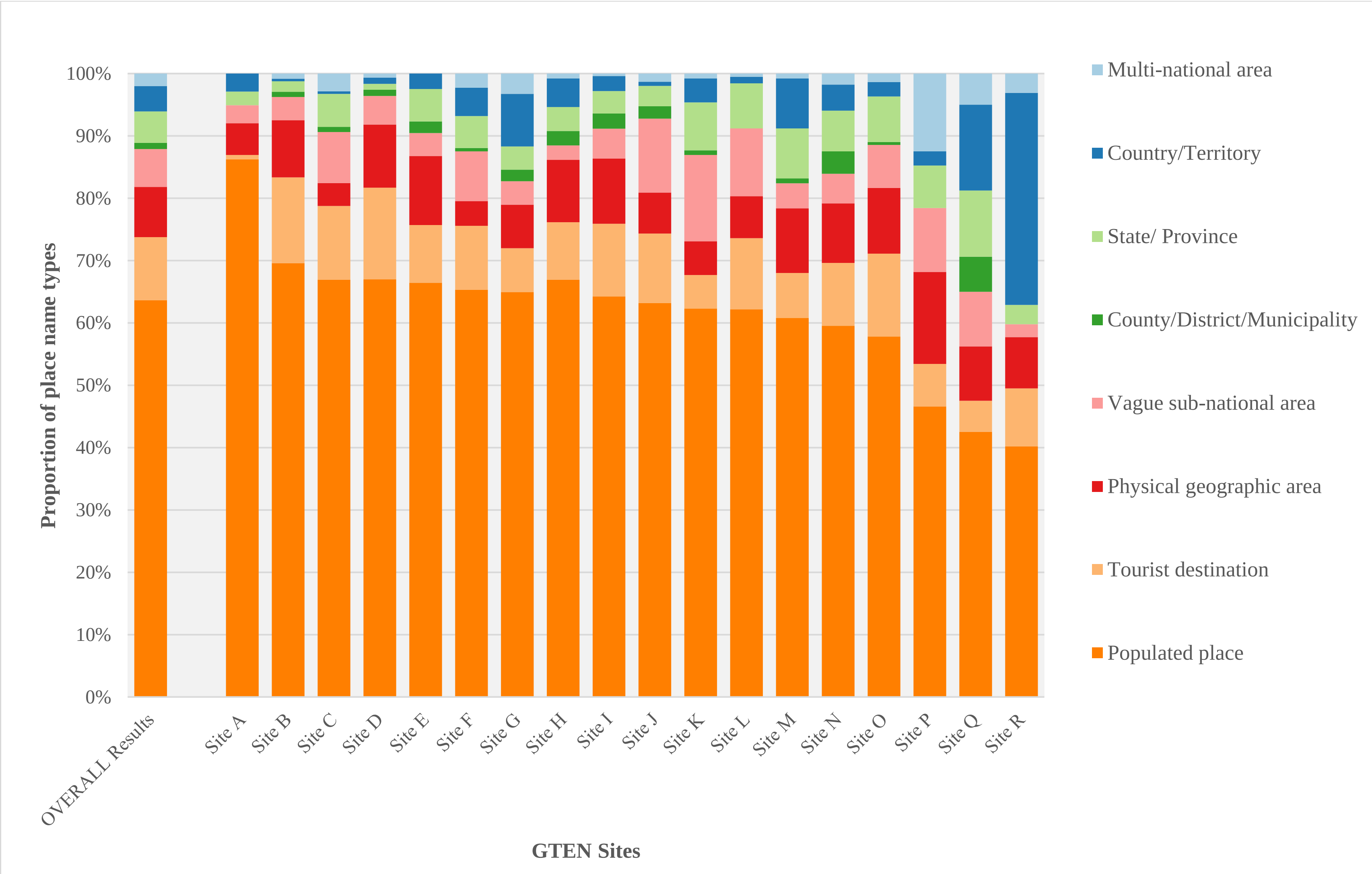


Figure 4.3 Trends in place name type variation across GTEN Sites

CHAPTER 5

COMMUNICATING NATIONAL ZIKA VIRUS TRAVEL HEALTH RECOMMENDATIONS THROUGH AN INTERACTIVE WEB MAP: DESIGN, DEVELOPMENT, AND DEPLOYMENT[1]

[1] Lash, R., C. Lee, R. Henry, W-L Juang, T. Kim. A. Walker, Y. Haber, J. Wages, M. O'Sullivan, S. Schuafenhauer, A. Friedman, H. Stroud, M. Dessecker, K. Holton, and G. Brunette. To be submitted to the *International Journal of Health Geographics.*

Abstract

Web-based interactive mapping technology was used to build a novel interactive map to query and display global Centers for Disease Control and Prevention (CDC) travel health recommendations during the 2016-2017 Zika virus outbreak. The map enables users to quickly and reliably search for and identify location-based Zika virus recommendations for travelers by integrating three principal features: a place name search service; a detailed global Zika virus travel recommendation map built from OpenStreetMap administrative boundary data and Earth Environment 90-m Digital Elevation Model ; and customized pop-up messages delivering authoritative travel health recommendations for any area of the world. This map was deployed on March 10, 2017, and has been viewed over 750,000 times within three months. It represents a significant increase in CDC's Traveler's Health Branch (THB) cartographic capacity, and is an important case study of how interactive mapping technology can be used to make the process of communicating location-based travel health recommendations less burdensome to travelers and travel health clinicians who provide medical education and care.

Introduction

The Centers for Disease Control and Preventions (CDC) Travelers' Health Branch (THB) has the mission of reducing illness and injury in U.S. residents traveling internationally or living abroad. In support of this mission, THB has regularly made static maps (i.e., traditional fixed and non-interactive maps) to help communicate important international travel health risk and recommendation information to the general public and medical clinicians. These maps are primarily published in the CDC Yellow Book (Brunette 2017). The CDC encourages travelers to visit and use the CDC Travelers' Health Website (Centers for Disease Control and Prevention 2017c) when planning their trips to learn about country-specific health recommendations to keep them healthy and safe while traveling abroad. To inform travelers about the latest health issues related to specific destinations, CDC will post travel notices using a three tiered system based on the type of recommendations which are provided: Level-1 (Watch); Level-2 (Alert); or Level-3 (Warning) (Centers for Disease Control and Prevention 2017f). A travel notice will contain a brief description of the unique health risks, and then the recommendations for how a traveler can eliminate or mitigate the risks.

While outbreak related travel notices can garner a lot of attention, the most frequently accessed material on the THB website are usually the country destination pages. These pages provide country specific health recommendations, including: vaccine and medicine recommendations; preventing foodborne diseases; preventing mosquitoborne disease; how to avoid injuries; a healthy travel packing list; and post-travel health information. Most of the recommendations mentioned above apply universally across all parts of a country; however, yellow fever vaccine and malaria prophylaxsis recommendations are two important exceptions. Both diseases can be fatal, and some countries require proof of yellow fever vaccine to allow travelers to enter the country. Identifying the appropriate location-specific recommendations for yellow fever vaccine and malaria prophylaxis can be difficult because the specific recommendations may vary geographically within a country and the place names used to describe the geographic variation may be foreign to the user, as was described in the previous chapter. To make the recommendation information on the Travelers' Health Website easier to interpret and understand, THB has created more disease and country specific static maps with information at a subnational level (Centers for Disease Control and Prevention 2017a).

Despite the popularity of these maps, the map scale and format can make it difficult to interpret the health information on the maps, particularly if a traveler is going to a destination not labeled on the map. Because of these limitations, THB mapping experts have been interested in developing interactive web map versions of these maps. They believe this would make the information easier to use and interpret, thereby improving the communication of health risk and prevention to travelers and clinicians. As part of CDC's response to the 2016-2017 Zika virus outbreak, additional mapmaking resources became available for the development of interactive travel health information maps, and these maps were eventually deployed in March of 2017. Because developing and deploying new interactive mapping technology within a United States federal government agency is a unique challenge, the intent of this article is to document this accomplishment as a case study for how other government and non-governmental agencies may consider adopting similar mapping technologies to support health communication efforts in the future.

Background

Web-based Maps in Travel Medicine

Croner (2003) predicted that internet and geographic information systems (GIS) based applications in public health would see important and exciting developments in the new millennium. The topic of web-based GIS use in public health surveillance systems has been recently reviewed (Luan and Law 2014), and they found that most examples of web-based GIS in public health today were of the "health atlas" format, and primarily focused on data sharing and visualization for expert audiences. Koenig et al. (2011) found that public health practioners had difficulty accurately interpreting exemplar health atlases, and suggested that both practioners needed more basic map interpretation training and that health atlas systems needed improved design.

Bauer and Puotinen (2002) predicted internet-based GIS technology could eventually reach travel medicine providers. While this prediction appears true, the examples are limited. CDC produced one of the earliest examples of interactive maps for travel medicine with their Malaria Map Application launched in 2009, though the application has since been abandoned (Centers for Disease Control and Prevention 2009). In 2012, the World Health Organization released their International Travel Health Interactive map application (World Health Organization 2012). While this application is still available on the web, this application does not incorporate any place name search capabilities, nor a very rich interactive map or pop-up message content. In summary, though there have been some examples of web-base GIS applications in travel medicine, none have taken full advantage of the recent technological improvements in web-based geospatial data, software, and services described by Batty et al. (2010).

CDC's Travelers' Health Website

The CDC's THB became aware of isolated Zika virus outbreaks in Brazil beginning in June of 2015, and issued a Level-1 travel notice for the country (Table 5.1) to inform travelers going there that they were at risk of acquiring Zika virus infection and recommended that travelers prevent mosquito bites to avoid infection. As the outbreak spread, CDC issued additional travel notices in December 2015 and January 2016 for affected countries in Central, North, and South America (Centers for Disease Control and Prevention 2015, 2016). As a result of new research that showed that there was a link between Zika virus infection in pregnant women and severe fetal birth defects, the World Health Organization designated the Zika virus outbreak a Public Health Emergency of International Concern. CDC subsequently issued revised travel recommendations that stated pregnant women should not travel to any of the affected Zika virus outbreak countries and territories (Table 5.1) (Centers for Disease Control and Prevention 2016). In March 2016, CDC issued revised travel recommendations stating that pregnant women should continue to abstain from travel to Zika virus outbreak countries, unless they were traveling only to a high-elevation area within those countries where the mosquitoes transmitting Zika virus were not likely to be found (Cetron 2016). High-elevation areas were defined high-elevation areas as greater than 2,000 meters above sea level. The new elevation-based recommendation was added to the country-specific travel notices, and the Mexico travel notice shown in Figure 5.1 is representative of the way this information was published on the THB website.

To determine whether it was safe for a pregnant woman to travel to a particular location, travelers or their clinician needed to know whether the elevation of their intended travel destination is above or below 2,000 meters in elevation. The CDC, therefore, produced and published country-specific elevation maps on the THB website. Each of these maps classified the country into two categories, either below 2,000 meters elevation, or above 2,000 meters (Figure

5.2). Maps were only made for countries containing regions with elevations above 2,000 meters, of which there were eventually 19 countries (denoted with asterisks in Table 5.1). However, like the yellow fever and malaria maps previously described, the usability of these Zika virus elevation maps were limited due to their map scale, which constrained the number of destinations which could be labeled on the static map and the accuracy with which users could discern the elevation-based recommendation for their destination.

As the number of Zika virus-affected countries and territories grew, users looking for Zika virus-free travel destinations were faced with an additional geographic information navigation problem. The THB was publishing a global map of countries and territories with active Zika virus transmission (Figure 5.3), with the intention to show all of the locations of Zika virus occurrence. However, it was not easy to identify which countries were Zika virus-free on these maps because only locations with Zika virus were labeled. If you were someone seeking to find the few remaining Zika virus-free island vacation destinations at the time, you likely would not realize that the Cayman Islands and The Bahamas were both Zika virus free unless you noticed on your own that they were missing from the map and list of locations. Simply put, static mapping technology could not meet the requirements of being an effective graphic communication device for all possible travel destinations to for both, users who were interested in finding information either for a specific set of countries, or users who wanted to find the destinations which remained Zika virus free. Though the limitations of static maps were apparent, the capacity to produce and publish interactive web-maps on the CDC website did not exist at the time. It was believed that such mapping technology could make locating and using Zika virus-related travel health recommendation information easier and more efficient for website users.

Objectives

The study has multiple objectives. The first objective was to design and implement an interactive map that would clearly communicate location-specific Zika virus travel health recommendations to international travelers. The second objective was for this interactive map to be flexible enough to accommodate the demands and uncertainties of an outbreak response environment, meaning that the recommendations messaging or areas of risk would likely change rapidly. The third objective was to deploy this system and assess how users respond.

Data and Methods

The interactive map was developed using an iterative approach, and some slight modifications have been made since it deployed as a result of user feedback received. The focus of the methods section will be to describe the interactive map as it currently exists at the time of writing, not as it was when it originally deployed. The methods section consists of describing three different parts of the interactive map: the necessary data inputs; the mapping software and services used to build the interactive map; and the system architecture. This section will end with a brief summary of the different web analytics and user feedback tools available for monitoring the interactive map and gauging its success as a health communication tool.

Non-Spatial Data

The most important part of the interactive map is the Zika virus travel health recommendation messages that display as pop-up messages when users click on the map, for which there are many permutations (Table 5.2). The goal of these messages was to provide the most important actionable information using words which would be easily and widely understood, and also to provide a hyperlink to other CDC web content where the users could find additional health information. Nearly all message text included some dynamic content, and how this content was stored will be covered in the spatial data section below. How it was queried and used, will be described in the system architecture section.

Each pop-up message is tied to an individual data layer in the map. Zika virus risk areas are defined based on an assessment of the local risk to travelers from the best available data and expert judgement. For countries, this assessment of risk has been performed and published by the World Health Origination (WHO) (World Health Organization 2017). The classification of domestic Zika virus risk areas is determined by CDC (Centers for Disease Control and Prevention 2017e). The remainder of the data section will explain the different types of geographic data used to represent the Zika virus risk areas within the interactive map.

Spatial Data Representing Zika Virus Risk Areas

The spatial representation of the different Zika virus risk areas and related data were compiled from different sources (Table 5.3), and their attributes were customized to meet the need for storing some of the dynamic content of the pop-up messages. A schematic of the map application system architecture is shown in Figure 5.5, where these custom datasets are shown as inputs for custom tilesets hosted in the Mapbox cloud. Once in the Mapbox cloud, this data is combined with Mapbox Streets base data to create the custom Mapbox Map style seen by the users. We used OpenStreetMap (OSM) as the source for nearly all of the custom datasets (OpenStreetMap Wiki 2017a) because OSM is also the source for the Mapbox Streets base data (Mapbox 2017b), and we wanted our custom datasets to align with Mapbox Streets.

Country polygons with territorial waters

All international Zika virus risk areas were represented as individual country polygons, so that whenever a user clicked on any part of the country’s land area, they would see the appropriate message for that country. Country polygons for the entire globe based on OSM were downloaded using the OSM Boundaries service (Nordmann 2017). This is a web service where users can select their desired national, state, or country boundaries, for free download as polygon shapefiles.

When downloading country polygons using OSM Boundaries users are given the choice to have the coastal boundaries extend and include all of the territorial water claims of that country, or to have them follow the physical coastline. We chose to use the territorial water boundary extent for two reasons: first, this generalized the shape of the coastline considerably and reduced the overall file size for this dataset; and second, small islands which occur along the coast were often not included in the physical coastline files.

Country polygons were downloaded as shapefiles. Due to limitations of the OSM Boundaries servers, we were limited to only downloading 15 countries at a time. These downloaded files were then combined into a single global shapefile using ArcGIS Desktop software (ESRI. Redlands, CA).

The national administrative units defined by OSM Boundaries did not exactly align with the country and territorial administrative areas CDC was using to issue Zika virus travel recommendations, and these issues were addressed through manual editing of the files. For example, the multi-polygon feature for the French Republic needed to be separated into individual features for mainland France, French Guiana, Martinique, Guadeloupe, etc. Similar problems existed with China, Chile, Ecuador, The Netherlands, Portugal, Spain, United Arab Emirates, and the United States of America.

The attributes for the data file were reduced to only two text fields. The first attribute was called “CDCName” and contained the name CDC was using to reference that location. The second attribute was the unique text string used to create the URL for the hyperlink in the pop-up messages.

To ensure that the shapefiles were successfully converted into vector tiles using Mapbox Studio Classic, several additional file formatting requirements had to be met. The files needed to be reprojected into the WGS 84 geographic projection (EPSG: 4326). All multi-polygon features needed to be split into single polygon features. Finally, the geometry of the polygons needed to be valid, which involved checking and potential repair using tools in ArcGIS Toolbox.

CDC Country Labels

While the Mapbox Streets base data includes country labels, it was necessary to create our own label tileset because of the way CDC was issuing travel recommendations to countries and territories alike. We wanted to be able to label countries and territories with the same label format, and we also needed to make sure that the names used to identify areas conforms with United States Government norms (e.g., Burma is labeled as Myanmar). The OSM Boundaries polygon data was converted into a point dataset, reprojected, and converted into a GeoJSON files using QGIS. This GeoJSON file was then uploaded into Mapbox Studio, and the Mapbox Studio data editor was used to position the labels in the same location as existing Mapbox Streets labels.

High Elevation Areas

Raster digital elevation model (DEM) data were processed to generate the polygons representing high-elevation areas (>=2,000 meters) of countries. Because the interactive map enabled users to zoom in to zoom level-11, (map scale of ~1:250,000; resolution of ~75meters per pixel (OPenStreetMap Wiki 2017b)) 90-m DEM was sought. The EarthEnv-DEM90 dataset, which merges ASTER GDEM 2 data and CGIAR-SCI v4.1 products, was chosen because it had the desired resolution, near global coverage (Robinson et al. 2014). These data were downloaded from www.EarthEnv.org as band interleaved by line (BIL) files, and were then imported into an

ArcGIS Raster Mosaic for ease of analysis. ArcGIS Model Builder was used to make that data processing easier and more efficient. A reclassification operation was performed on the raster mosaic to identify all areas greater than 2000m in elevation, and these areas were then converted into a vector polygon with simplification enabled. Extremely small area polygons were deleted, which represented isolated individual mountain peaks, to further reduce the overall file size. Finally this polygon layer was intersected with the OSM Country Polygon layer to create individual country-specific high-elevation polygons that had the same attribute fields as the previous country polygon layer.

Continental United States polygon with detailed coastline

Once the Zika virus outbreak spread to the continental United States, it was necessary to symbolize this information on the map. It was decided that this area should be highlighted, and a polygon feature with a detailed coastline was downloaded from the OSM Boundaries service (Nordmann 2017). This file was processed as previously described.

United States State polygons with coastal waters

As mentioned above, once the Zika virus outbreak spread to the state of Florida, it was necessary to symbolize this information on the map. State level data was needed in the same format as the country polygons with territorial waters described above. State polygons were downloaded from the OSM Boundaries service (Nordmann 2017). Because this data would be used to highlight the states when a user zoomed in close, the territorial boundary extended further off the coast than was desirable for visualization purposes, and had to be redrawn more closely to the coastline. This file was processed as previously described.

Domestic Zika virus areas

The location and extent of domestic Zika virus areas was highly variable, as these had to be drawn on a case-by-case basis between the CDC State and Local Task Force, and the state and local public health authorities, in accordance with the CDC Zika virus Interim Response Plan (Centers for Disease Control and Prevention 2017e). We received these boundaries from the CDC Emergency Operation Center's Situational Awareness Team.

Software and Development Approach

The interactive mapping application was built primarily using interactive mapping technology from commercial mapmaking software and service company, Mapbox (Washington, D.C.). ESRI ArcGIS Desktop Advanced Software with Spatial Analyst extension was used for processing raster elevation data and polygon boundary data. Mapbox Studio Classic software was used to convert large polygon shapefiles into the Mapbox Tiles format (MbTiles). Smaller shapefiles could be directly uploaded into Mapbox Studio, where the application would automatically convert the files into the MbTiles format. In addition to limited file conversion, Mapbox Studio was used to host tilesets; and create and host custom Mapbox Map Styles. Mapbox GL JS is a JavaScript library published by Mapbox, and was used to load our Mapbox Studio hosted map onto a CDC webpage, and controls additional user functionalities of the interactive map. Finally, the Mapbox Geocoding service was used to enable a placename search bar which allowed users to type in their desired travel destination, select the destination name from an auto-complete list, and have the map automatically pan and zoom into that location.

System Architecture

Figure 5.5 shows how the different types of data, software, and services were integrated into the final map application. Although the custom datasets listed at the bottom of the figure have already been described above,the Mapbox Map Style deserves greater explanation. The map style can be thought of as analogous to ESRI ArcMap's map document file (.MXD). The map style is a JSON file that stores all of the necessary information needed for an application user's web browser to connect to the map data and style it accordingly. In our application, the map style is responsible for storing the file paths to each of the data layers within the map, filtering this data based on SQL code, if so desired,, and symbolizing the data according to the graphic style parameters chosen. As shown in the diagram, the map style seamlessly integrates the previously described custom datasets with data from the Mapbox Streets base data.

The HTML Map Document is the next important part of the interactive map application. This HTML file loads the Mapbox GL JS library into the map user's web browser, and then uses the defined methods, objects, and properties to support all of the necessary functionalities supported by the application. In general terms, the interactive map uses GL JS to load the custom Zika virus map style, create a search bar control inside of the map, and connect that search bar to the Mapbox Geocoder database via the Mapbox Geocoding API.

The GL JS also enables an on-click event to query the map data beneath the application user's pointer, and identify which geographic features are being displayed. The result of the on-click event returns information about that particular feature, including the name of the data layer the feature resides in, as well as any attributes of that feature. The name of the data layer name is then used to select the appropriate pop-up message text, and a pop-up message is then displayed on the screen where the other properties from the map feature are used to populate the dynamic content (e.g., country name, hyperlink, etc.) of the pop-up message text.

Finally, the HTML Map Document is loaded as an iFrame object onto the CDC webpage (https://wwwnc.cdc.gov/travel/page/world-map-areas-with-Zika virus) that represents the application interface. It is on this page that additional text and graphic content is contained, including the name of the map, instructions on how to use the application, and the map legend that appears beneath the map.

Mapbox Studio Style Editor

As stated previously, the map style is important for managing which data are shown on the map and in which data layers they appear. Table 5.4 shows a list of the different Zika virus data layers used in the interactive map. For example, should Zika virus be found in a new country where previously there was no known risk of Zika virus, said country would be added to the corresponding layer for recommendation area "Area with risk of Zika virus [epidemic country]" layer, and removed from the "Area with no known risk of Zika virus [all other countries]" layer. The data features shown on a given layer is a function of: 1) the data source for the layer; and 2) the data filter.

<u>User Analytics and Metrics</u>

A variety of metrics were available after the map was launched to track how the application was being used and how users were discovering the map. CDC's web analytics platform Adobe Analytics (Adobe System, San Jose, CA), as well as information about the number of place name searches processed by the Mapbox place name search service were monitored on a daily basis to try to understand patterns in traffic volumes. The effects of promotional efforts were captured by analytics from THB accounts on Facebook and Twitter. Sysomos (Toronto, Canada), a social media marketing analytics service, was used to assess the reach of news media reports which directly mentioned and linked to the map application. Finally, directed user feedback was received when users emailed us directly with questions or comments, or when they completed a generic CDC website customer satisfaction survey.

Results

The interactive web map was launched on the CDC THB website (https://wwwnc.cdc.gov/travel/page/world-map-areas-with-Zika virus), Friday, March 10, 2017. To demonstrate the functionalities of the interactive map application for travelers, consider the following hypothetical scenario:

> *A young pregnant woman and her husband are considering a babymoon to a beach location before their first child arrives in November. This woman heard a NPR story during her morning commute that mentioned that Zika virus was still a concern, and that travelers should look at a CDC map before booking any plane tickets this summer (Doucleff 2017). She goes to the internet and finds the CDC map. Her friend just came back from Saint Lucia and loved it, so she decides to start her search there.*

Figure 5.6 shows how this traveler can use the functionalities of the new interactive mapping application to find location-based Zika virus travel recommendations. Beginning with Panel A, the traveler arrives at the CDC World Map of Areas with Zika virus page. She looks at the map, and does not see Saint Lucia labeled anywhere, so she tries the destination search bar (Panel B). As she starts to type, the suggested results appear. When she sees Saint Lucia she clicks on it, and the map responds by panning and zooming to the extent of the island (Panel C). She then clicks on the island, and a pop-up message for “Areas with risk of Zika virus” appears telling her that CDC does not recommend that pregnant women travel there and directs her to a hyperlink for the Saint Lucia travel notice (Panel D). She decides to zoom out on the map to see if there are any other Caribbean Islands that would be Zika virus free, and as she zooms-out, she sees that the island of Martinique is a different color. She clicks on Martinique, and the pop-up message states that there is “No known Zika virus” in Martinique, and directs her to the Martinique destination page for additional information about how to stay healthy and safe while traveling in Martinique.

Assessing Application Usage Metrics

Adobe Analytics continuously captures some basic information about each of the users who visit the interactive mapping application page, such as the daily number of page views, which is a count of the number of times that the webpage is loaded by a user. Figure 5.7, Panel A shows daily page views from March 12 through June 30, 2017. Over this time period, the webpage has had 1,103,137 views, and the average daily page views ranged from a minimum of 5,279 to a maximum of 16,693, with a mean of 9,880. A strong weekly periodicity is apparent, meaning that the the day-of-the-week heavily impacts the traffic volume. Mondays averaged the highest number of page views (13,243), while Saturdays averaged the lowest number of page views (6,789).

Figure 5.7, Panel B smooths the daily variation by summarizing page views per week. Looking at the changes in the height of the stacked bars each week shows that though there has been slight variation in total number of page views week-to-week, with a range from roughly 62,000 to 76,000 views a week.

The stacked bars are divided into four categories based on count of visits per individual. This additional information enables us to consider how many first-time visitors (blue part of chart) come to the page each week, and whether this number is trending in a particular direction. A total of 704,277 users visit the webpage at least once, with an average of 43,508 first time visits each week. As can be seen, the total number of visits decreases with each increase in the visit count. It is noteworthy that 174,905 users return to the web-page for a second visit, and 76,197 users for a third visit. Though not shown in this figure, the number of users who have returned 10 or more times is 5,675, and the number of users who have returned 20 or more times is 1,231.

Figure 5.8 shows where users of the webpage are located geographically. As expected, these data are skewed towards the United States (69%), followed by the other large English speaking countries of Canada (10%), United Kingdom (7%), and Australia (2%).

Results of Promotion Effort

The new map was announced in a CDC Media Statement on March 10, 2017(Centers for Disease Control and Prevention 2017d), and again through a CDC Emergency Partners newsletter on March 15 (Centers for Disease Control and Prevention 2017b). THB also maintains a Twitter account (@CDCTravel, 23,900 followers) and Facebook page (@CDCTravelersHealth, 21,346 followers) which were used to promote the interactive map. The map was mentioned in five different tweets between the deployment date and March 18, and this generated 42,715 twitter impressions, but only 274 click-throughs. There was only one Facebook post promoting the map, which had 6,550 impressions and reached 4,165 people.

The map was also mentioned and linked to by a number of articles in English-langugage print, radio, and internet news sites, according to a search of the the Sysomos database. Through June 23, 2017, there were 76 such articles, with an estimated reach of 7.4 million people. The single most influential article was published by National Public Radio on June 11, titled “Is Zika Virus Still a Problem in Florida and the Caribbean?” The original article reached 3.1 million people when posted to the NPR.org website, and then was republished across 49 other state and local public radio websites, reaching another 172,000 people.

A separate but related search of the Adobe Analytics database shows the diversity of other websites not covered by the Sysomos database summarized above. Included in this list were Government, Healthcare, Blog and Trade News, Message Boards, and other non-English or non-United States media organization websites with webpages that included links to the map (Table 5.5). The single largest group of website referrals came from the United States State Department's websites, and those of United States Embassies, who maintain webpages for American citizens traveling or living abroad about the latest international health and security risks.

The next largest group of websites represented the Healthcare Providers seeking to educate and inform their clients, as well as other websites targeting this information to doctors, nurses, physicians, and other healthcare providers needing to the know the latest Zika virus information. Notable among the Healthcare category were websites from Canada and Switzerland, which show that although CDC's stated mission is to serve the American public, our resources are made available and found useful to citizens of other countries.

A third group of website referrals came from Blogs and Trade News sites, particularly those sites targeting expectant mothers, new mothers, and women seeking fertility assistance. Travel Industry websites focused on travel medical insurance. Noteworthy websites in this category included Spanish-language pregnancy blogs, as well as a lone blog directed at paternity, Fatherly.com.

The fourth group was message boards supporting a number of the various communities described above in the Blogs and Trade News groups. These message boards are unique because oftentimes the posters are community members themselves, and thus these opinions represent more of a lay audience than that of the reporter or writer responsible for blog posts or websites.

The final category was foreign language and international media publishing articles with content similar to the United States English-speaking media, but representing unique minority communities for CDC information and materials.

Anecdotal User Feedback

The final set of results is anecdotal feedback from map users capture in one of two ways, and although these data do not come from a representative sample of map users, they nonetheless do represent a unique type of direct user feedback. The first set of results comes from 48 users who responded to the Foresee customer satisfaction survey, which is randomly targeted to users of any THB website to get general demographic information an overall satisfaction with content on THB webpages. Unfortunately, it was not possible to tailor the questions in the survey to provide more direct and pertinent questions about the interactive map. Respondents were asked to respond to the following questions using a scale from 1-10 to the following questions:

- What is your overall satisfaction with this site?
- How well does this site meet your expectations?
- How does this site compare to your idea of an ideal website?

The average score for the first two questions was 8.2, the average score for the last question was an 8.3

The second dataset includes 39 people who emailed the CDC's THB directly using the email address posted on the webpage, and these messages have been summarized according to the topic of the reported comment or question (Table 5.6). A total of 15 users reported that they could not view the map in their browser. Ten users asked sent questions seeking additional explanation or clarification about what the different categories on the map meant or how it should be interpreted. Five users requested Zika virus testing recommendation information, and 2 users requested Zika virus epidemiologic case information. Two other users were healthcare providers, and they asked for large format printable versions of the map for display in their clinic.

Discussion

The fact that this project was able to develop and deliver a fully functional interactive mapping application in the middle of a major outbreak response activation is noteworthy. That over 700,000 people have seen this mapping application since it was first deployed has exceeded the expectations of all involved, as has the fact that more than 75,000 people have visited the map 3 times. These numbers, along with the numerous websites that direct their users to visit the map, suggest that the map is successfully communicating important Zika virus prevention information to people who need to have this information.

Despite the success of the application, there has been a need to make improvements to the application since it was deployed and even still today. These changes will be the focus of the rest of the discussion. As noted in the background section, the interactive map application replaced a static global map of countries with Zika virus travel notices (Figure 5.3), and included on this page was a list of all of the country names. When the interactive map application was deployed, it was decided that this list would be removed from the page, as it was assumed that this list was strictly there to aid those users who could not easily find their country of interest on the static map. However, within days of deploying the new map, we received one email from a user stating that the interactive map would not load on their computer. We then received a second email from a user who stated that they were a blood donation center who had been checking that list on a daily basis to know whether someone returning from international travel should be able to donate blood. Based on this user feedback, we quickly modified the interactive map page to bring back the country list, and also to create a separate PDF file (https://wwwnc.cdc.gov/travel/files/zika-areas-of-risk.pdf) which include an updated static map and country list (Figure 5.9).

Another important change was made to the map legend after the map was deployed. At the time that the map was initially deployed, the labels for the dark purple and light purple color swatches in the legend read “Low elevation” and “High elevation” respectively (Figure 5.10). These labels were referring to the elevation-based Zika virus travel recommendations described in the background section, and the term elevation was meant to refer to the physical geographic elevation of areas on the map. However, some users reported that these labels were confusing to them, because without additional explanation, they were inclined to think that the term“elevation” was referring to the level of Zika virus risk in an area, and that areas shaded in dark purple represented areas of low Zika virus risk. To resolve this confusion, the decision was made to update the map legend on June 15, 2017, to change the labels of the dark purple areas to “Area with risk of Zika”, and the light purple areas to “Area with minimal risk of Zika.”

<u>System Architecture in Operation</u>

The system architecture that the interactive map was built upon appears stable and flexible as desired. Despite the fact that 15 users have reported that their browsers would not load the interactive map, to the best of our knowledge the application has been online and operational 24 hours a day, 7 days a week since it deployed. It is our assumption that the reasons these 15 users were unable to view the map was because they were either using an unsupported web browser (Mapbox 2017a), or their browser settings did not permit the execution of the Mapbox GL JS javascript required for displaying the map.

Since the map deployed, there has been a need to update the Zika virus risk areas shown on the map. These types of changes were anticipated, given that the Zika virus outbreak had the potential to spread to new locations and possibly cease in others. The changes that were made in accordance with CDC's included the following.

- American Samoa, Guadalupe, Martinique, Saint Barthelemy, were moved from the "Area with risk of Zika virus [Epidemic country]" category to the "No known Zika virus" category.
- Brazil was moved from the "Area with risk of Zika virus [Epidemic country]" to the "Area with risk of Zika virus [Endemic country]".
- Pakistan was moved from the "No known Zika virus" category to the "Area with risk of Zika virus [Endemic country]".

For each of these changes, we were able to simply use Mapbox Studio to change the data selection filters for each of these layers-- removing the country name from the filter query for one data layer, and adding it to the filter query for the new layer. Once the changes were made,, the revisions were published to the map style, and the revisions were visible on the published map within minutes. Although this approach to managing the data in the map by manually changing selection queries may appear technologically inelegant, as opposed to directly managing all of these changes in a single database with a transactional log, it is an approach that we have found is adequate for the current map and our current human and technical infrastructure. The unfortunate part about this approach is there is no automated transactional log for recording what changes occur when, although we are recording this information in the MS Word Document.

Similar data changes have taken place in the domestic Zika virus risk areas, although these changes included creating an entirely new risk category and pop-up messages. At the time that the map first deployed, Miami-Date County was still categorized as a “Zika virus cautionary yellow area”. However, on June 2, 2017, the county was categorized into a new risk category called “Area previously designated as Zika virus cautionary (yellow) area”, with the new recommendation that this area was now safe for pregnant women to travel to, among other non-travel related recommendation changes (Centers for Disease Control and Prevention 2017g). To implement these changes, a new data layer had to be created in Mapbox Studio to support the new symbolization style of this risk category and a revised map style published. Also, new pop-up message text had to be drafted, and incorporated into the HTML Map Document. Although the changes within Mapbox Studio were executed quickly, CDC IT policy requires that the revised HTML Map Document undergo a security scan, which typically takes 1-2 weeks.

The geocoding component of the map has been the source of considerable amounts of internal discussion since the application deployed. Most of these criticisms focused on one of two things: 1) Do the auto-complete results returned by the service match the users’ expectations?; and 2) Does the map display respond to show the desired location? Unfortunately, these discussions have not led to any actionable conclusions, because it is difficult to objectively assess whether individual users expectations and experiences of searching specific results are representative of the whole user community. As such, our experience has shown that there remains a need for improved tools for comparing the accuracy, strengths, and limitations of various geocoding services. Although numerous studies can be found of comparisons for geocoding street addresses in developed countries (Krieger et al. 2001, Whitsel et al. 2006, Zhan et al. 2006, Wey et al. 2009, Duncan et al. 2011, Goldberg et al. 2013), there does not appear to be a single study or tool available for assessing how well currently available geocoding services handle the complexity of global geocoding.

Challenges Posed by the Outbreak Response Environment

A unique aspect of the development and deployment of this interactive mapping application is the fact that all of this took place during the ongoing CDC Zika virus Outbreak Response from January 2016 to present. During major outbreak response activities, additional funding becomes available to those CDC programs and projects that are supporting the response activities. Large-scale responses to infections disease outbreaks such as the West African Ebola Outbreak and the current Zika virus outbreak pose risks to United States citizens traveling abroad, and as such, the CDC's THB has played a larger role. During outbreak responses, the THB is responsible for standing-up and staffing the Global Migration Task Force (GMTF), which is a crosscutting team representing the various areas of expertise within the Division of Global Migration and Quarantine. These areas of expertise include the science, communication, and policies regarding travel medicine, immigrant and refugee health, travel health screenings at US ports of entry, and risk analysis for the spread of infectious diseases. It is for this reason that the THB mapping experts were involved in making Zika virus maps, and subsequently put forth the idea of using an interactive map to communicate Zika virus health information. Once the preliminary application demonstrated a proof-of-concept, additional Zika virus response funds were allocated to secure the necessary Mapbox services. Similarly, the outbreak response demanded that developing and supporting the application be a high-priority for all of the personnel involved.

It should be noted that the outbreak response operations also brought some additional burdens to the production process. For example, the outbreak response structure creates a number of temporary positions, which are filled by a variety of CDC staff often rotating through the position on 30-day cycles. As a result, the map development team leads had to routinely brief these new individuals about the project goals, previous design decisions, and current project needs.

Also, because the map covered high-profile communication material, specifically, advising the general public and healthcare providers about important Zika virus related health decisions, the final map had to receive a number of layers of approval. This included:

- GMTF Epidemiology and Communication Team Leads;
- GMTF Lead;
- Joint Information Center Lead, with cross-clearance to the Epidemiology and Surveillance Task Force, Pregnancy and Birth Defects Task Force;
- Zika virus Response Incident Manager; and
- Director of the Centers for Disease Control and Prevention.

Although it is not unusual for CDC information materials to undergo a number of rounds of review and a number of layers of approval (Centers for Disease Control and Prevention 2005), the layers of approval named above is unique for maps.

Conclusions

The work presented herein describes a novel design and implementation of an interactive mapping application to facilitate the communication location-specific Zika virus travel health recommendations to international travelers. This application has proven flexible enough to accommodate the demands and uncertainties of an outbreak response environment, as has been demonstrated by the fact that Zika virus risk areas have continued to change since the map was deployed, and each time the map has been updated within hours to reflect these changes. The user analytics presented here show that this interactive mapping application has had sustain levels of new and returning users since it was deployed.

More generally, this interactive mapping application shows great promise for increasing CDC’s Traveler’s Health Branch (THB) map-making capacity, and is an important case study of how interactive mapping technology can be used to make the process of communicating location-based travel health recommendations less burdensome to travelers and travel health clinicians who provide medical education and care. The sustained level of new and returning visitors to the map has been surprising, and as such, has invigorated more interest in using this interactive mapping technology to communicate other types of geographically complicated travel health information.

To further support future interactive mapping applications for communicating public health information and more accurately characterize the user experience and satisfaction with interactive maps, more extensive user evaluation studies should be performed. These studies should try to better characterize whether there are different information needs between users who are medical clinicians and users who do not have any public health training.

Additional studies of the user interface are also needed, because a major limitation of the current application is that developers were limited to only communicating recommendation information via pop-up boxes. It is possible that other interface designs could be more useful, particularly for desktop users. These studies should also consider more fully if there are alternative design solutions which would be more useful for mobile device users, because the proportion of mobile web users is continually growing, and mobile-specific opportunities or challenges were not considered when the current application was developed.

This study contributes to the literature on the use of interactive mapping technology within public health, and offers a unique perspective given that the work was performed within a national public health agency during an outbreak response. The evidence presented herein clearly demonstrates that this type of mapping technology is useful for communicating geographically specific health information to the masses. To make these types of maps available for other diseases in the future though, a renewed focus should be placed on the development of organizational geospatial capacity and public health spatial data infrastructure. In this vein, we hope most of all that the current interactive map proves to be both inspirational and motivational.

References

Batty, M., A. Hudson-Smith, R. Milton and A. Crooks (2010). "Map mashups, Web 2.0 and the GIS revolution." Annals of GIS **16**(1): 1-13.

Bauer, I. L. and M. Puotinen (2002). "Geographic information systems and travel health." Journal of Travel Medicine **9**(6): 308-314.

Brunette, G., Ed. (2017). CDC Yellow Book 2018: Health Information for International Travel. New York, Oxford University Press.

Centers for Disease Control and Prevention. (2005). "Clearance of information products disseminated outside CDC for public use." Retrieved 20 July, 2017, from http://www.cdc.gov/maso/Policy/PublicUse.pdf.

Centers for Disease Control and Prevention. (2009). "CDC Malaria Map Application." Retrieved 2 March, 2016, from http://www.cdc.gov/malaria/map/.

Centers for Disease Control and Prevention (2015). Media Statement: First case of Zika virus reported in Puerto Rico. Atlanta, GA**:** 1.

Centers for Disease Control and Prevention (2016). Media Statement: CDC issues interim travel guidance related to Zika virus for 14 Countries and Territories in Central and South America and the Caribbean. Atlanta, GA.

Centers for Disease Control and Prevention. (2017a). "2018 Yellow Book: List of Maps." Retrieved 10 June, 2017, from https://wwwnc.cdc.gov/travel/yellowbook/2018/list/maps.

Centers for Disease Control and Prevention (2017b). CDC Emergency Partners newsletter, March 15 2017. Atlanta, GA.

Centers for Disease Control and Prevention. (2017c). "CDC Travelers' Health Website." Retrieved 4 April, 2017, from www.cdc.gov/travel.

Centers for Disease Control and Prevention (2017d). CDC updates Zika travel guidance for pregnant women to not travel to any area with Zika risk. Atlanta, GA, US Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention (2017e). CDC Zika Interim Response Plan. Atlanta, GA**:** 22.

Centers for Disease Control and Prevention. (2017f). "Travel Health Notices." Retrieved 5 July, 2017, from https://wwwnc.cdc.gov/travel/notices.

Centers for Disease Control and Prevention. (2017g). "Update: Advice for people living in or traveling to South Florida." Retrieved 9 July, 2017, from https://www.cdc.gov/zika/intheus/florida-update.html.

Centers for Disease Control and Prevention. (2017h). "World Map of Areas with Risk of Zika." Retrieved 30 June, 2017, from https://wwwnc.cdc.gov/travel/page/world-map-areas-with-zika.

Cetron, M. (2016). "Revision to CDC's Zika travel notices: minimal likelihood for mosquito-borne Zika virus transmission at elevations above 2,000 meters." MMWR Morb Mortal Wkly Rep **65**(10): 267-268.

Croner, C. M. (2003). "Public health, GIS, and the Internet." Annual Review of Public Health **24**(1): 57-82.

Doucleff, M. (2017). Is Zika still a problem in Florida and the Caribbean? Your Health. Washington, DC, National Public Radio.

Duncan, D. T., M. C. Castro, J. C. Blossom, G. G. Bennett and S. L. Gortmaker (2011). "Evaluation of the positional difference between two common geocoding methods." Geospat Health **5**.

Goldberg, D. W., M. Ballard, J. H. Boyd, N. Mullan, C. Garfield, D. Rosman, A. M. Ferrante and J. B. Semmens (2013). "An evaluation framework for comparing geocoding systems." International Journal of Health Geographics **12**(1): 50.

Koenig, A., E. Samarasundera and T. Cheng (2011). "Interactive map communication: Pilot study of the visual perceptions and preferences of public health practitioners." Public Health **125**(8): 554-560.

Krieger, N., P. Waterman, K. Lemieux, S. Zierler and J. W. Hogan (2001). "On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research." American Journal of Public Health **91**(7): 1114-1116.

Luan, H. and J. Law (2014). "Web GIS-based public health surveillance systems: a systematic review." ISPRS International Journal of Geo-Information **3**(2): 481.

Mapbox. (2017a). "Browser Support." Retrieved 9 July, 2017, from https://www.mapbox.com/help/mapbox-browser-support/.

Mapbox. (2017b). "Mapbox Streets v7." Retrieved 30 June, 2017, from https://www.mapbox.com/vector-tiles/mapbox-streets-v7/.

Nordmann, W. (2017). "OSM Boundaries." 4.2. Retrieved 2 July, 2017, from https://wambachers-osm.website/boundaries.

OpenStreetMap Wiki. (2017a). "Using OpenStreetMap." Retrieved 30 June, 2017, from http://wiki.openstreetmap.org/wiki/Using_OpenStreetMap.

OPenStreetMap Wiki. (2017b). "Zoom Levels." Retrieved 30 June, 2017, from http://wiki.openstreetmap.org/wiki/Zoom_levels.

Robinson, N., J. Regetz and R. P. Guralnick (2014). "EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data." ISPRS Journal of Photogrammetry and Remote Sensing **87**: 57-67.

Wey, C. L., J. Griesse, L. Kightlinger and M. C. Wimberly (2009). "Geographic variability in geocoding success for West Nile virus cases in South Dakota." Health Place **15**.

Whitsel, E. A., P. M. Quibrera, R. L. Smith, D. J. Catellier, D. Liao, A. C. Henley and G. Heiss (2006). "Accuracy of commercial geocoding: assessment and implications." Epidemiologic Perspectives & Innovations **3**.

World Health Organization. (2012). "International Travel Health Interactive." Retrieved 20 June, 2017, from http://apps.who.int/ithmap/.

World Health Organization (2017). Zika virus country classification scheme: interim guidance [WHO/ZIKV/SUR/17.1]. Geneva, Switzerland**:** 5.

Zhan, F. B., J. D. Brender, I. De Lima, L. Suarez and P. H. Langlois (2006). "Match rate and positional accuracy of two geocoding methods for epidemiologic research." Annals of Epidemiology **16**(11): 842-849.

Figures

Table 5.1. Date when a country's Zika virus travel notice was issued by CDC

Date	Country
6/9/2015	Brazil*
11/6/2015	Colombia*, Suriname
12/10/2015	El Salvador*, Guatemala*, Panama*, Mexico*, Paraguay, Venezuela*
12/18/2015	Honduras*, French Guiana
12/31/2015	Puerto Rico
1/15/2016	Haiti*, Martinique
1/22/2016	Cape Verde, Barbados, Guadeloupe, Saint Martin, Samoa, Bolivia*, Ecuador*, Guyana*
1/26/2016	Dominican Republic*, United States Virgin Islands
2/1/2016	Curacao, Costa Rica*, Nicaragua*, American Samoa
2/3/2016	Jamaica*, Tonga
2/18/2016	Aruba, Bonaire
2/23/2016	Trinidad and Tobago, Marshall Islands
2/29/2016	Saint Vincent and the Grenadines, Sint Maarten
3/9/2016	New Caledonia
3/19/2016	Cuba
3/22/2016	Dominica
4/1/2016	Micronesia (Kosrae)
4/4/2016	Fiji
4/13/2016	Saint Lucia
4/18/2016	Belize
4/29/2016	Papau New Guinea*
5/5/2016	Peru*
5/9/2016	Saint Barthelemy
5/12/2016	Grenada
5/25/2016	Argentina*
6/28/2016	Anguilla
7/14/2016	Sint Eustatius
7/25/2016	Saba
8/3/2016	Antigua and Barbuda, Turks and Caicos Islands
8/11/2016	Cayman Islands
8/23/2016	Bahamas, The
8/30/2016	Singapore, British Virgin Islands
9/23/2016	Saint Kitts and Nevis
11/16/2016	Palau
11/21/2016	Montserrat
3/10/2017	Angola*, Guinea-Bissau, Maldives, Salomon Islands

* These countries contain regions with elevations greater than 2,000 meters

Zika Virus in Mexico

Warning - Level 3, Avoid Nonessential Travel

Alert - Level 2, Practice Enhanced Precautions

Watch - Level 1, Practice Usual Precautions

What is the current situation?

Local mosquito transmission of Zika virus infection (Zika) has been reported in Mexico. Local mosquito transmission means that mosquitoes in the area are infected with Zika virus and are spreading it to people.

Because Zika virus is primarily spread by mosquitoes, CDC recommends that travelers to Mexico protect themselves from mosquito bites. The mosquitoes that spread Zika usually do not live at elevations above 6,500 feet (2,000 meters) because of environmental conditions. Travelers whose itineraries are limited to areas above this elevation are at minimal risk of getting Zika from a mosquito. The following map shows areas of Mexico above and below 6,500 feet.* For more information, see Questions and Answers: Zika risk at high elevations.

*The categories shown on this map are intended as a general guideline and should not be considered to indicate absolute risk. Elevation may vary within an area to a larger extent than this map can depict. The presence of mosquitoes may change seasonally, with increasing temperatures or rainfall, and may change over time. Travelers to destinations that cross or are near an elevation border may wish to consider the destination as an area of lower elevation. Travelers to high elevations are still at risk of getting Zika from sex.

Click to enlarge

Sexual transmission of Zika virus from a male partner is also possible, so travelers are also encouraged to use condoms or not have sex.

Most people infected with Zika virus do not get sick. Among those who develop symptoms, sickness is usually mild, with symptoms that last for several days to a week. Zika may also be linked to Guillain-Barré syndrome (GBS), a rare disorder that can cause muscle weakness and paralysis for a few weeks to several months. Most people fully recover from GBS, but some have permanent damage.

As more information becomes available, this travel notice will be updated. Please check back frequently for the most up-to-date recommendations.

Zika Virus in Pregnancy

Zika virus can be spread from a pregnant woman to her fetus, and infection is linked to a serious birth defect of the brain called microcephaly and other poor pregnancy outcomes. CDC recommends special precautions for the following groups:

- Women who are pregnant:
  - Should not travel to any area of Mexico below 6,500 feet (see map).
  - If you must travel to one of these areas, talk to your doctor first and strictly follow steps to prevent mosquito bites during your trip. If your itinerary is limited **entirely** to areas above 6,500 feet, there is minimal risk of getting Zika from a mosquito.
  - If you have a male partner who lives in or has traveled to an area with Zika, either use condoms or do not have sex (vaginal, anal, or oral) during your pregnancy.
- Women who are trying to become pregnant:
  - Before you or your male partner travel, talk to your doctor about your plans to become pregnant and the risk of Zika virus infection.
  - You and your male partner should strictly follow steps to prevent mosquito bites.
- Men who have traveled to an area with Zika and have a pregnant partner should use condoms or not have sex (vaginal, anal, or oral) during the pregnancy.

Figure 5.1 Mexico Travel Notice from March 21, 2016, after elevation-based travel recommendations were issued.

* The categories shown on this map are intended as a general guideline and should not be considered to indicate absolute risk. Elevation may vary within an area to a larger extent than this map can depict. The presence of mosquitoes may change seasonally, with increasing temperatures or rainfall, and may change over time. Travelers to destinations that cross or are near an elevation border may wish to consider the destination as an area of lower elevation. Travelers to high elevations are still at risk of getting Zika from sex.

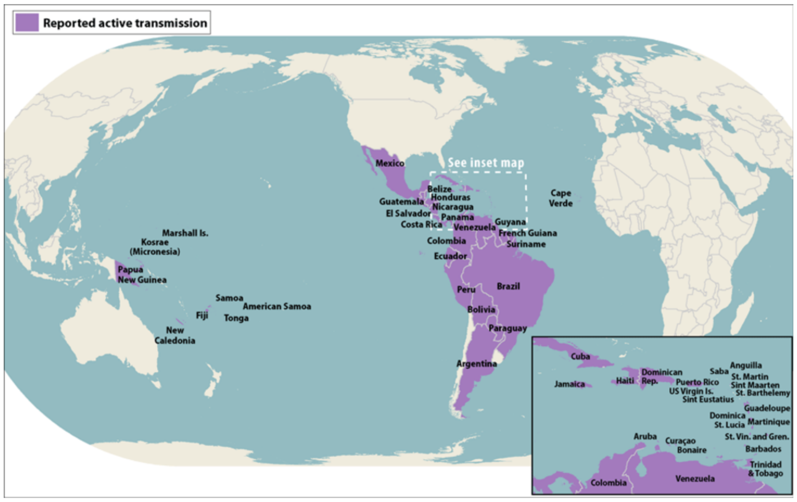
Figure 5.2. Example of the country-specific static elevation classification maps used to help communicate elevation-based travel health recommendations

CDC > Zika Virus Home > Areas with Zika

All Countries & Territories with Active Zika Virus Transmission

Language: English

As of July 26, 2016



See CDC's Country Classification Technical Guidance

Americas

- Anguilla
- Argentina
- Aruba
- Barbados
- Belize
- Bolivia
- Bonaire
- Brazil
- Colombia
- Commonwealth of Puerto Rico, US territory
- Costa Rica
- Cuba
- Curacao
- Dominica
- Dominican Republic
- Ecuador
- El Salvador
- French Guiana
- Grenada
- Guadeloupe
- Guatemala
- Guyana
- Haiti
- Honduras
- Jamaica
- Martinique
- Mexico
- Nicaragua
- Panama
- Paraguay
- Peru
- Saba
- Saint Barthélemy
- Saint Lucia
- Saint Martin
- Saint Vincent and the Grenadines
- Sint Eustatius
- Sint Maarten
- Suriname
- Trinidad and Tobago
- U.S. Virgin Islands
- Venezuela

Oceania/Pacific Islands

- American Samoa
- Fiji
- Kosrae, Federated States of Micronesia
- Marshall Islands
- New Caledonia
- Papua New Guinea
- Samoa
- Tonga

Africa

- Cape Verde

Figure 5.3. CDC Webpage showing all countries and territories with active Zika virus transmission (as of July 26, 2016)

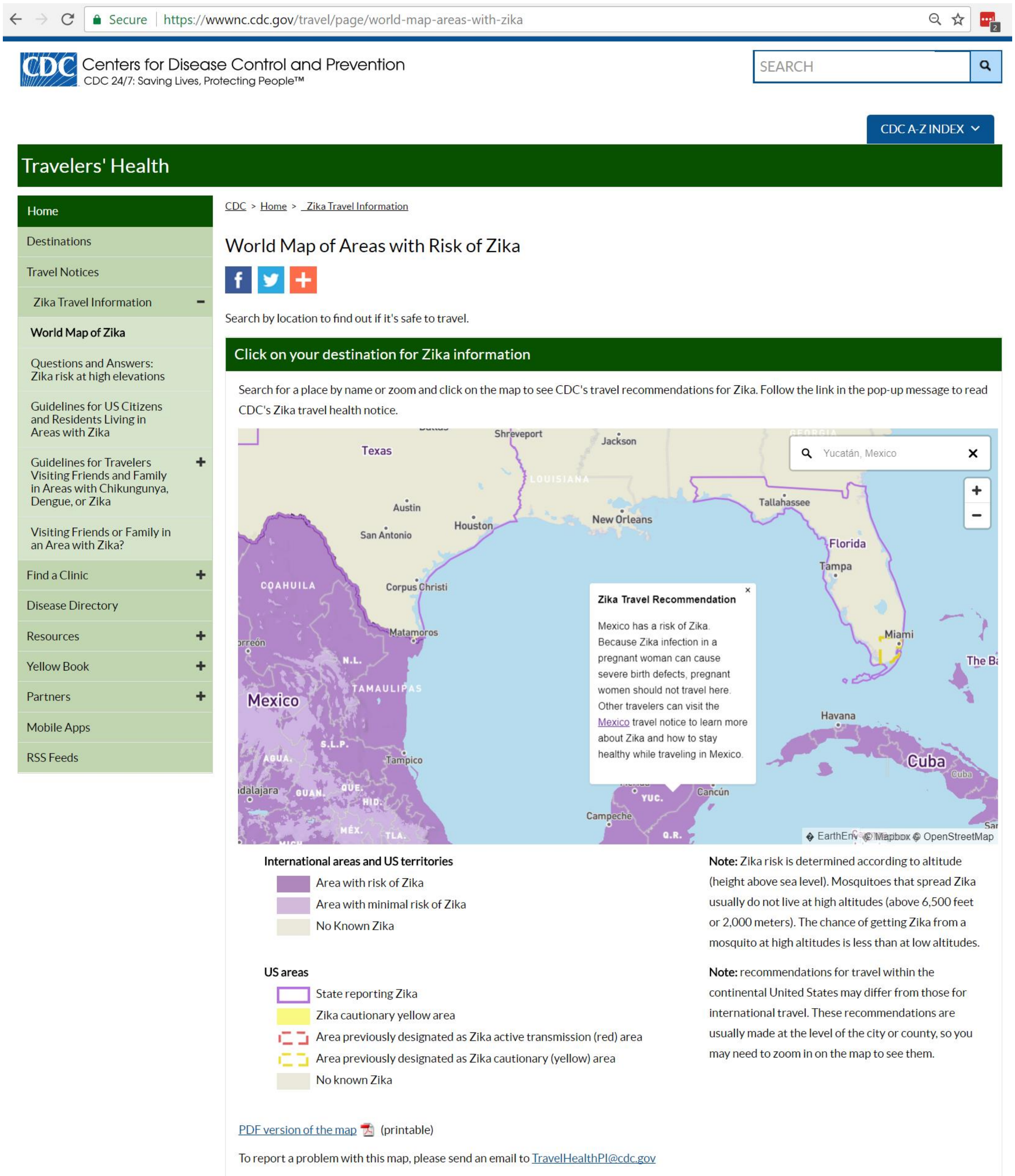


Figure 5.4 Interactive map application on the CDC Travelers' Health Webpage (Centers for Disease Control and Prevention 2017h).

Table 5.2. Pop-up message content for the map

Location	Recommendation Area	Example Text
International	Area with risk of Zika virus [Epidemic country]	Mexico has a risk of Zika virus. Because Zika virus infection in a pregnant woman can cause severe birth defects, pregnant women should not travel here. Other travelers can visit the Mexico travel notice to learn more about Zika virus and how to stay healthy while traveling in Mexico.
International	Area with minimal risk of Zika virus [epidemic]	Mexico has a risk of Zika virus. However, this is a high-elevation area where mosquitoes that can spread Zika virus may not live. Travelers, including pregnant women, who never go below 2,000 meters elevation are at lower risk of getting Zika virus from a mosquito. Visit the Mexico travel notice to learn more about Zika virus and how to stay healthy while traveling in Mexico.
International	Area with risk of Zika virus [endemic country]	Kenya has a risk of Zika virus. Because Zika virus infection in a pregnant woman can cause severe birth defects, pregnant women should not travel here. Other travelers can visit the Kenya page to learn more about Zika virus and how to stay healthy while traveling in Kenya.
International	Area with minimal risk of Zika virus [Endemic country]	Kenya has a risk of Zika virus. However, this is a high-elevation area where mosquitoes that can spread Zika virus may not live. Travelers, including pregnant women, who never go below 2,000 meters elevation are at lower risk of getting Zika virus from a mosquito. Visit the Kenya page to learn more about Zika virus and how to stay healthy while traveling in Kenya.
International	[country with] No known Zika virus	Mozambique has no known risk of Zika virus from mosquitos. Visit the Mozambique page to learn more about staying healthy while traveling in Mozambique.
Domestic	State reporting Zika virus	Texas has reported cases of Zika virus spread by local mosquitoes. Visit CDC's Advice for people living in or traveling to Brownsville, Texas page to learn more.
Domestic	Zika virus active red area	The intensity of Zika virus transmission in this area presents a significant risk to pregnant women. Visit CDC's Advice for people living in or traveling to [State Name] page to learn more.
Domestic	Zika virus cautionary yellow area	Local spread of Zika virus has been identified here, but there is no current evidence of widespread transmission. Although the specific level of risk in yellow areas is unknown, there is still a risk to pregnant women. Visit CDC's Advice for people living in or traveling to South Florida page to learn more.
Domestic	Area previously designated as Zika virus active transmission (red) area	For more information, visit Zika virus in Florida.

Domestic	Area previously designated as Zika virus cautionary (yellow) area	For more information, visit Zika virus in Florida.
Domestic	No known Zika virus	Reports of local mosquitoes spreading Zika virus and infecting people have been limited to small areas in Florida and Texas. Visit the Areas with Zika virus page to learn more.
International and Domestic	Water	Please click on a land area to see the appropriate Zika virus information message.

Table 5.3 Geographic datasets used to represent Zika virus risk areas

Tileset Name	Original Data Source	Data Description	Feature type: Properties	Spatial extent
Country polygons with territorial waters	OSM Boundaries	Polygon layer of countries (defined by CDC) with very generalized coastal boundaries	Polygon: CDCName (filter); CDCUrl	Global
CDC Country labels	NaturalEarthData.com	Point layer of country labels according the United States State Department/CDC	Point: CDCName (filter)	All Countries
High-Elevation Areas	Earth Environment 90m DEM	90m DEM which combines the best qualities of SRTM and ASTER DEMs	Polygon: CDCName (filter); CDCUrl	Global areas between 40N and 50S
Continental United States polygon with detailed coastline	OSM Boundaries	Polygon layer of continental United States with a detailed coast line	Polygon:	Americas and Oceania
United States' State polygons with coastal waters	OSM Boundaries	Polygon layer of United States' States with very generalized coastal waters	Polygon: StateName	US States
Domestic Zika virus areas	OSM; manual digitization of OSM data	Polygon layer which is an agglomeration of open-source boundaries for various United States administrative areas representing current Zika virus status	Polygon: Zika virusStatus (filter); CDCName; CDCUrl;	Select United States' States

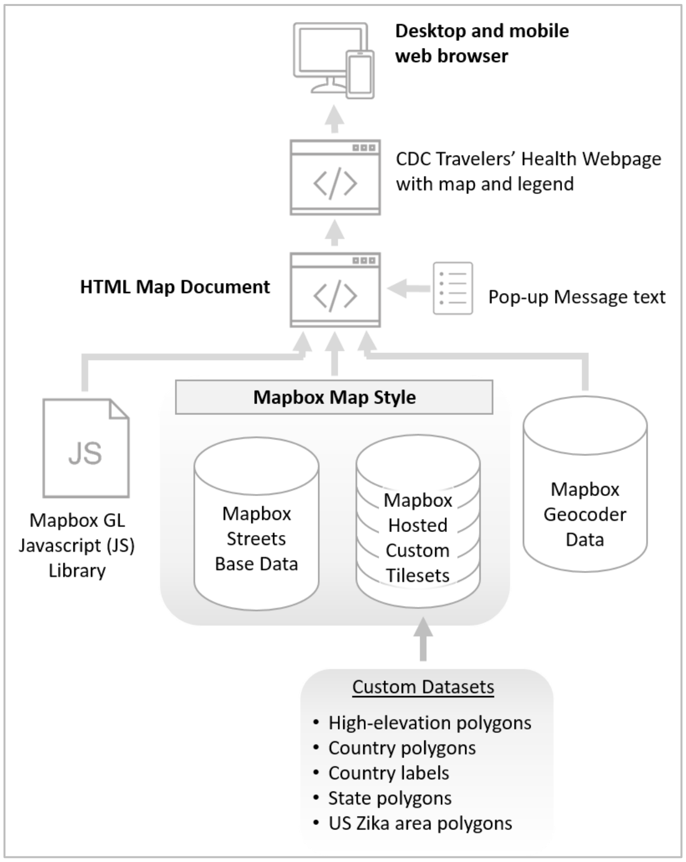


Figure 5.5. Schematic of the application architecture for identifying the individual components of data, software, internet connections, and map functionalities.

Table 5.4 Zika virus data layers used in custom map style

Recommendation area	Data Source Name	Layer Name	Feature type: Properties
Area with risk of Zika virus [epidemic country]	Country polygons with territorial waters	Epidemic	Polygon: CDCName (filter); CDCUrl
Area with minimal risk of Zika virus [epidemic]	High-Elevation Areas	EpidemicElevMask	Polygon:
Area with risk of Zika virus [endemic country]	Country polygons with territorial waters	Endemic	Polygon: CDCName (filter); CDCUrl
High-Elevation areas of endemic countries	High-Elevation Areas	EndemicElevMask	Polygon: CDCName (filter); CDCUrl
Area with no known risk of Zika virus [all other countries]	Country polygons with territorial waters	NoRisk	Polygon: CDCName (filter); CDCUrl
*Outline of the Continental United States	Country polygon with detailed coastline	Us-Outline	Polygon:
*Outline of State reporting Zika virus	US States polygon	UsZika virusStates_outline	Polygon: StateName (filter);
State reporting Zika virus	US States polygon	UsZika virusStates	Polygon: StateName (filter); CDCUrl; RefName
Zika virus active red area	Domestic Zika virus area polygons	UsActive	Polygon: Zika virusStatus (filter); CDCName; CDCUrl; BoudaryType
Zika virus cautionary yellow area	Domestic Zika virus area polygons	UsCaution	Polygon: Zika virusStatus (filter); CDCName; CDCUrl; BoudaryType
Area previously designated as Zika virus active transmission (red) area	Domestic Zika virus area polygons	UsHistoric	Polygon: Zika virusStatus (filter); CDCName; CDCUrl; BoudaryType
Area previously designated as Zika virus cautionary (yellow) area	Domestic Zika virus area polygons	UsHistoric-Outline	Line: Zika virusStatus (filter); CDCName; CDCUrl; BoudaryType
No Know Zika virus	US States polygon	UsNoZika virusStates	Line: Zika virusStatus (filter); CDCName; CDCUrl; BoudaryType
Labels for large area epidemic countries	CDC Country labels	country-label-lg Epidemic	Point: CDCName (filter)
Labels for medium area epidemic countries	CDC Country labels	country-label-md Epidemic	Point: CDCName (filter)
Labels for small area epidemic countries	CDC Country labels	country-label-sm Epidemic	Point: CDCName (filter)
Labels for United States States with Zika virus	CDC Country labels	UsState-label	Point: CDCName (filter)

*These data layers do not have a corresponding Zika virus travel health recommendation pop-up message because they are polygons formatted to be just the outline of the corresponding area.

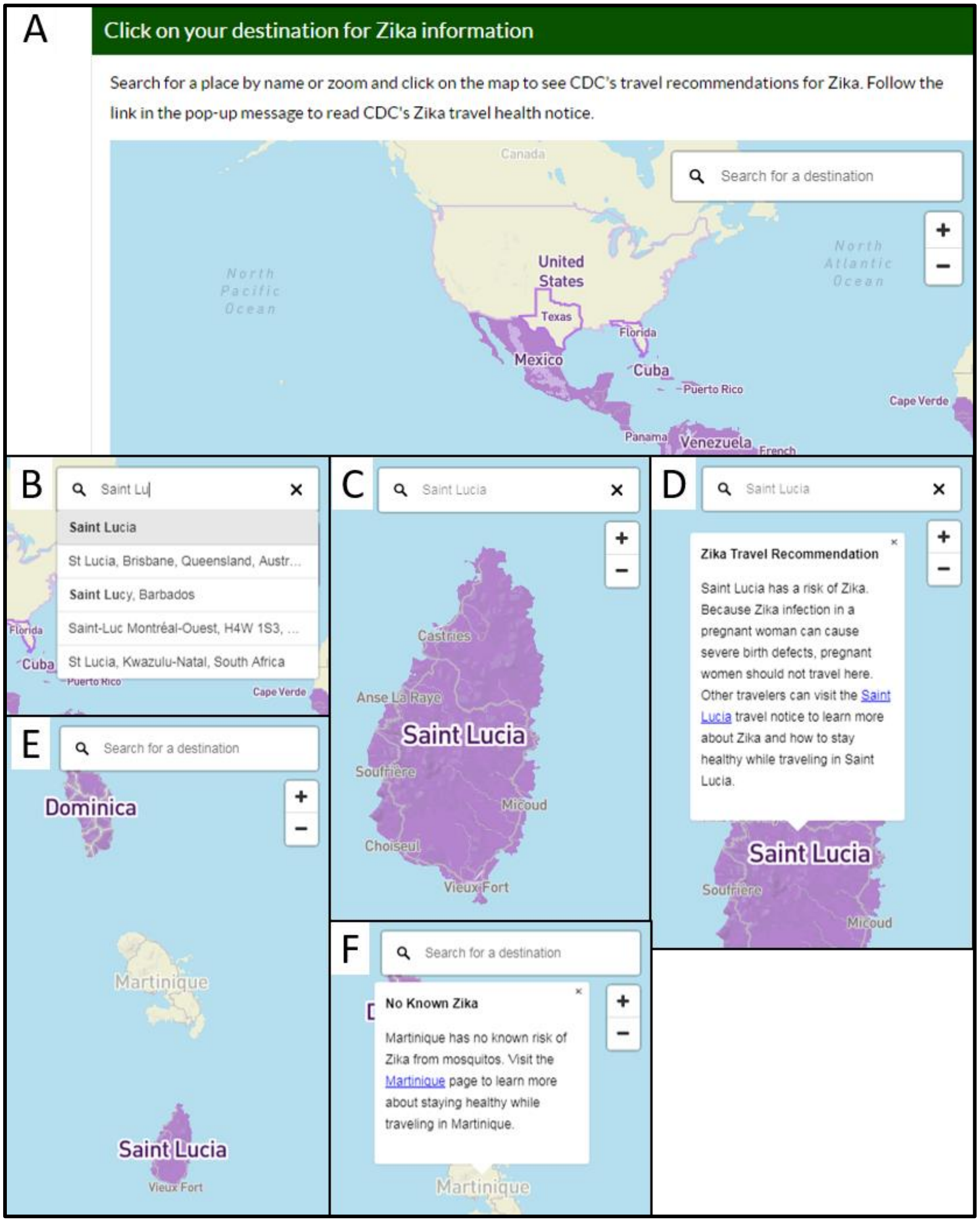


Figure 5.6 Use Case for a traveler searching for a Zika virus information for Saint Lucia, and then Martinique

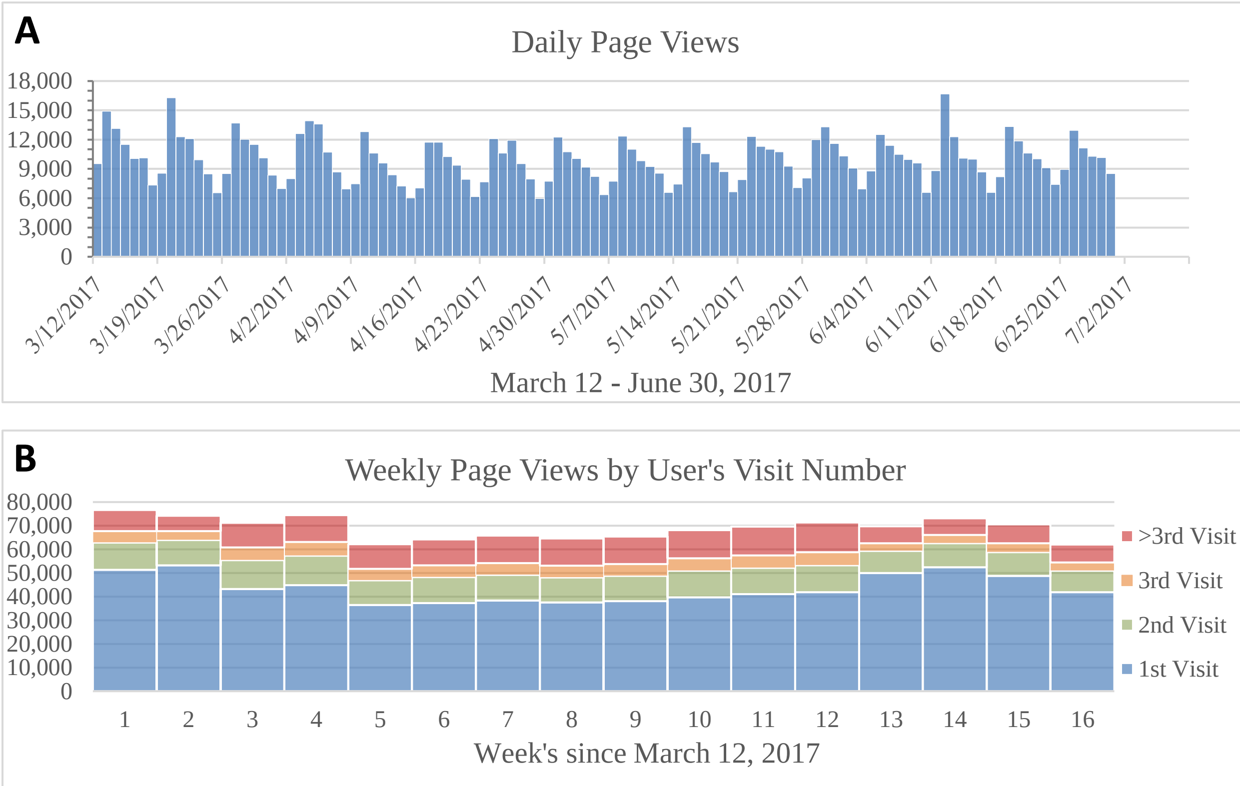


Figure 5.7 Total number of daily page views (Panel A), and total number of weekly page views per visit number (Panel B)

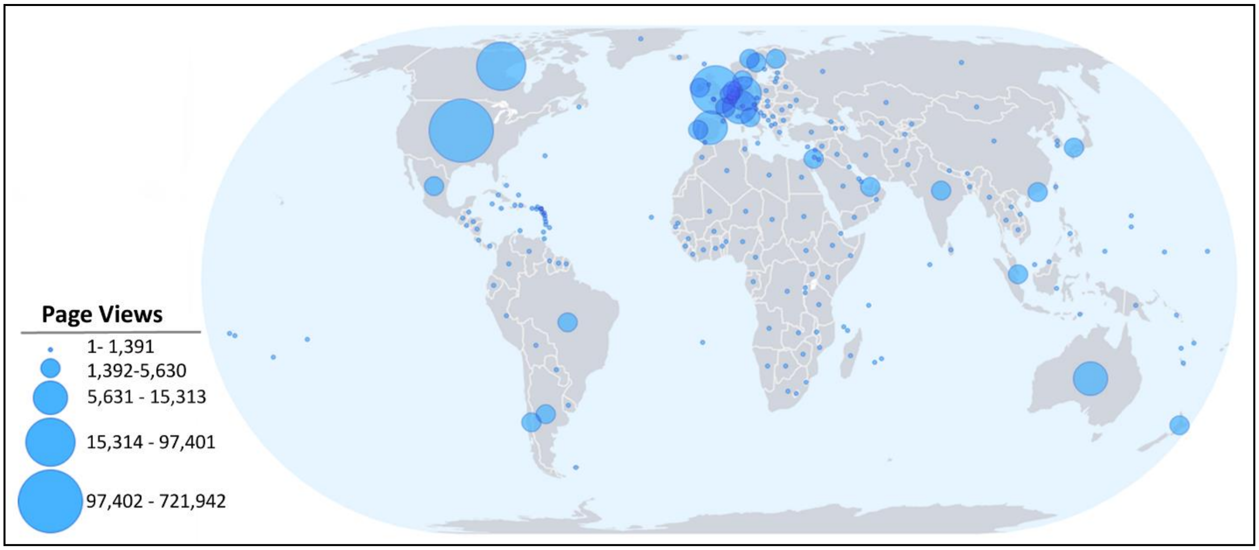


Figure 5.8 Geographic distribution of page views by user country, March 12 - June 30, 2017

Table 5.5. Diversity and frequency of different types of referral web pages

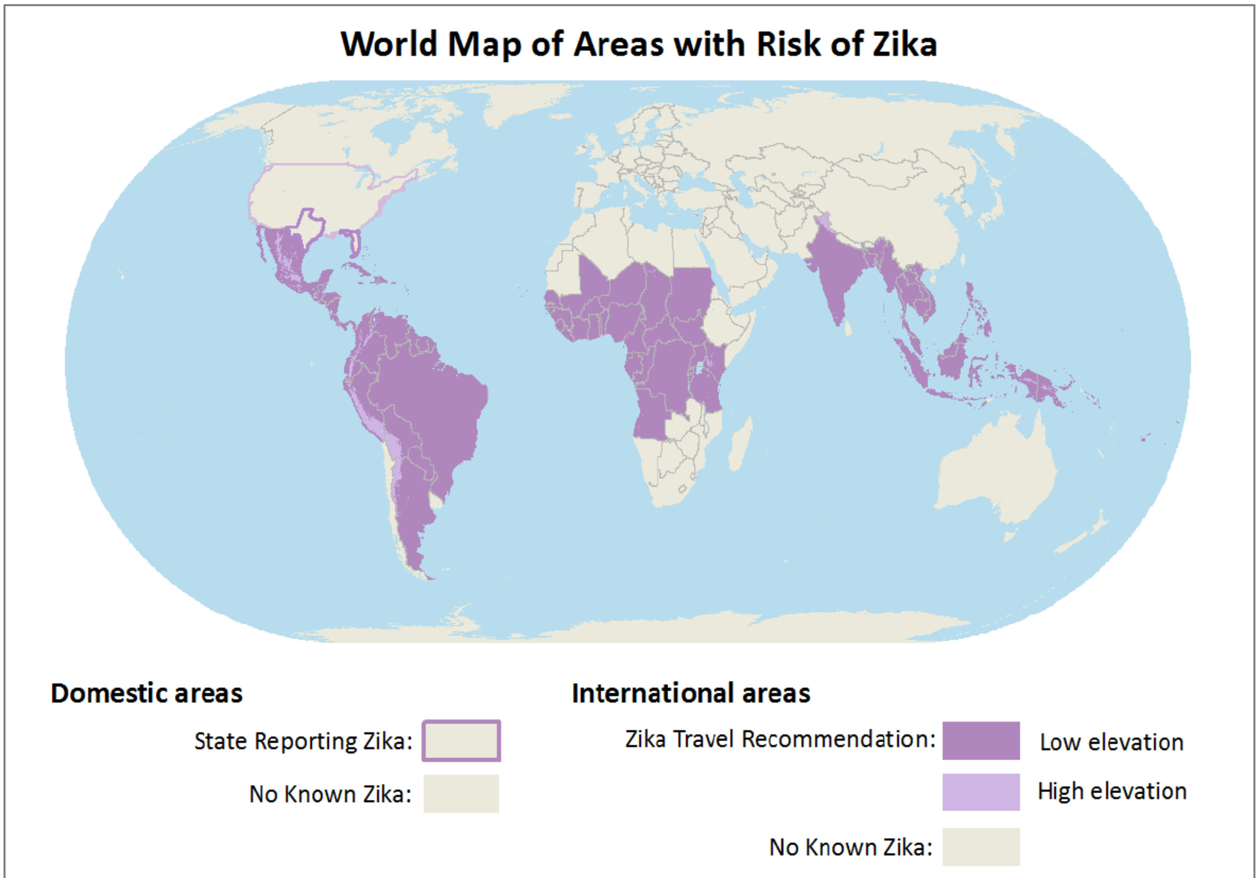
Referrer Categories	Referrer Types	Count of webpages
Government	Government, Federal	22
	Government, County	10
	Government, State	7
	Total	**39**
Healthcare Providers	Healthcare Provider	13
	Clinical Information Provider	10
	Healthcare provider (Canada)	3
	Healthcare provider (Switzerland)	1
	Total	**27**
Blogs and Trade News	Maternal Information Provider	5
	Health Information Provider	4
	Travel Industry	3
	Maternal Information Provider (Spanish)	1
	Paternal Information providers	1
	Trainer for healthcare providers	1
	Total	**15**
Message Boards	Message board (pregnant women)	5
	Message Board (travelers)	4
	Message board (scientists)	3
	Message Board (travelers; Spanish language)	1
	Total	**13**
Other News Media	News Media (French language)	2
	News Media, Africa	1
	News Media, Asia	1
	News Media, Canada	1
	News Media, Europe	1
	News Media, United States (Spanish language)	1
	Total	**7**

Table 5.6 Summary of comments and questions reported from users via email

Topic of comment or question	Count
Reporting that the map is not loading	15
Questions about data classification and interpretation	10
Requesting Zika virus testing recommendation information	5
Requesting Zika virus epidemiologic case information needed	2
Requesting printable poster size map	2
Reporting broken hyperlink	1
Unable to find Hawaii on the map	1
Unable to download map	1
Requesting permission to reuse map	1
Requesting the map be translated into Spanish	1
Total	**39**

Figure 5.9 Example of the PDF map made for users who could not load the interactive map

The full PDF document is available for download at https://wwwnc.cdc.gov/travel/files/Zika virus-areas-of-risk.pdf



Travel Notices

Zika Virus in Cape Verde

Zika Virus in Mexico

Africa: Angola, Guinea-Bissau

Asia: Maldives, Singapore

The Caribbean: Anguilla; Antigua and Barbuda; Aruba; The Bahamas; Barbados; Bonaire; British Virgin Islands; Cayman Islands; Cuba; Curaçao; Dominica; Dominican Republic; Grenada;

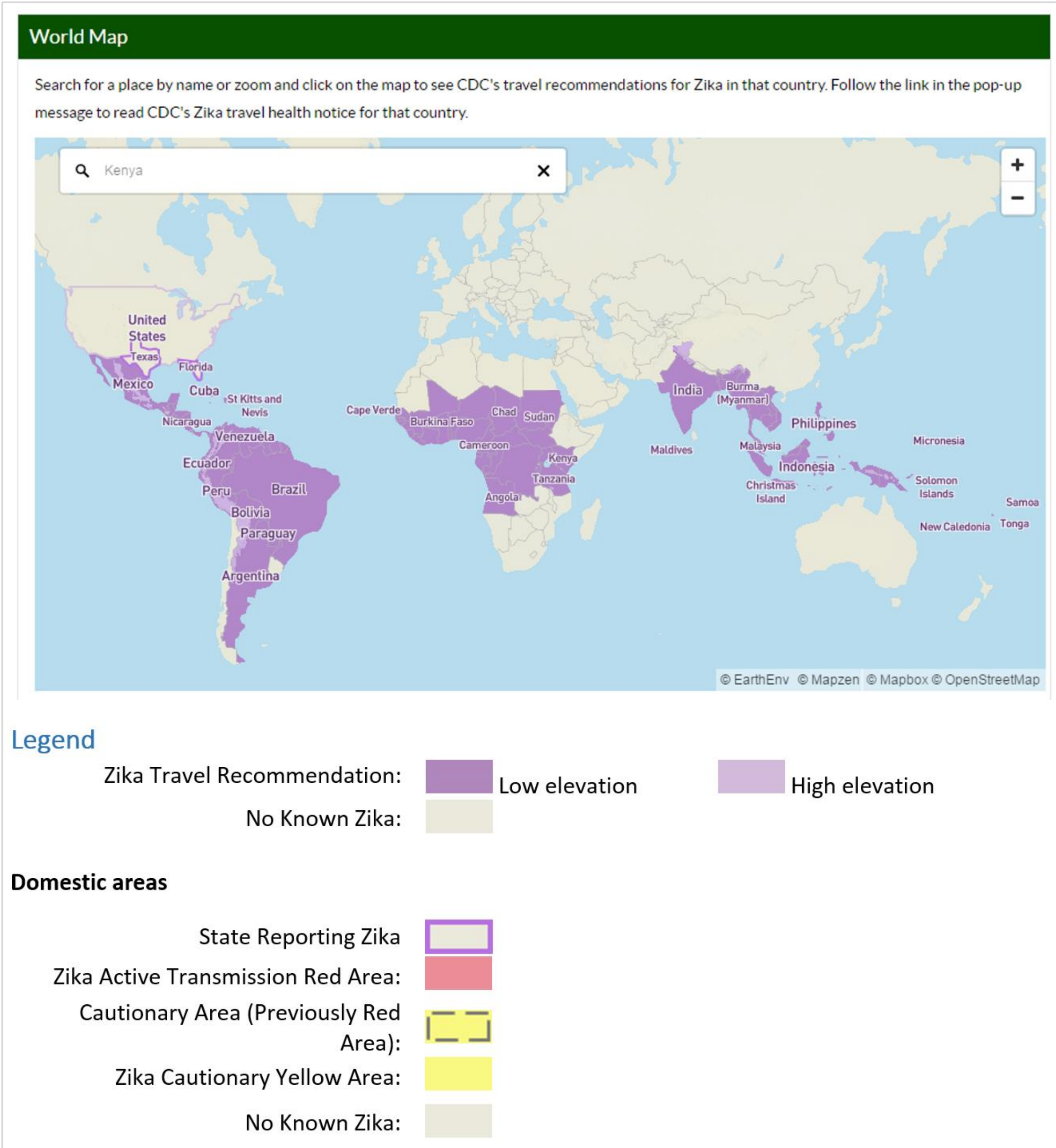


Figure 5.10 Labels for purple areas on the map that were confusing when the map was first deployed, and were subsequently revised on June 15, 2017

The labeling of dark purple and light purple areas as “Low elevation” and “High elevation” referred to the elevation based Zika virus travel recommendations. Some users reported, however, that they thought these labels were referring to Zika virus risk however, and not physical geographic elevation. The label for “Low elevation” was changed to “Areas with risk of

Zika virus”, and the “High elevation” label was changed to “Areas with minimal risk of Zika virus”.

CHAPTER 6

CONCLUSIONS

The broad question that this dissertation aimed to answer was: Can the implementation of geographic information science and technology (GIS&T) improve international disease surveillance and prevention activities when applied within a national public health institution? To investigate this broad question more thoroughly, answers to three more specific questions were sought: 1) Can GIS&T methods be used to accurately geocode legacy disease surveillance data for foreign locations where geographic reference datasets are poor or non-existent?; 2) Is there a need for GIS&T-enabled applications to help international travelers and their clinicians discover location specific health recommendation and disease prevention information?; and 3) Can a GIS&T-based interactive map be developed and deployed to deliver the type of location specific travel health recommendation described above, thereby helping to prevent disease? The conclusions to these questions will be discussed in greater detail below. This section will conclude with recommendations for future research studies on the application of GIS&T within applied public health.

The first study investigates a historic monkeypox surveillance database of human cases in Central and West Africa in the 1980s to see how accurately these surveillance data could be mapped using modern GIS&T. This study found that the spatial accuracy of geocoded data varied as a result of the different ways in which the information was recorded and stored, as well as the availability of accurate and reliable reference data. The locational accuracy and utility of the geocoded data could be improved if the geocoding toolkit was expanded to include not only digital gazetteer data, but archival map data as well. Finally, this study suggests that disease surveillance data could be more accurately preserved if in addition to collecting location data on case forms using a hierarchy of nested place names (e.g., town, district, province, country), the reference maps of the study area are available, stored and preserved at the time the data are collected.

While the first study focused on the challenge of accurately geocoding foreign place names recorded in primary disease surveillance data, the second and third studies examined the ways in which geocoding services of could be used to make searching for and communicating public health recommendations and disease prevention messages easier. In the second study, we sought to better understand the types of place names that travelers and travel health clinicians use when they are searching for travel health recommendations. This question sought to understand if the political geographic approach to reporting this recommendation information by public health agencies was useful, or if the information might be accessed in other ways, namely, geocoding services connected to interactive maps. We found that travelers and clinicians are using city names as the primary means of structuring their travel itineraries, and not state or country names currently used in CDC recommendations. This study is the first of its kind to collect data on the limitations and potential burdens that current health recommendation reporting methods appear to place on travelers and travel health clinicians. The results of this study contributed to the justification of additional investment by public health agencies in improving their GIS&T capacity. Specifically, it is anticipated that public health organizations will improve the effectiveness of outreach and communication of travel recommendations if they begin to generate and publish recommendation information in readily available and interpretable digital geospatial formats. Use of these data to build GIS&T applications that could help deliver this important recommendation information ultimately benefits both travelers and clinicians.

Finally, the third study sought to develop and deploy a novel GIS&T-based application, an interactive web map, to help travelers identify location-specific Zika virus travel health recommendations. Building on the findings of the second study, this third study specifically aimed to compile public health recommendation information into digital geospatial formats, and to use a geocoding service to help users find the location-specific recommendation for their destination of choice. The latest technology in web-based map-making software and data services were used to develop a system architecture which could be maintained with the available personal and technical infrastructure. Once deployed, the user analytics for this interactive map have shown that such a tool is frequently accessed with many visitors returning more than once.

This dissertation research has demonstrated contributions of applied GIS&T within public health, along with the need for further investment of GIS&T implementation, training and incorporation into global policies of public health analysis and information dissemination. Although historical and contemporary examples of the successful use of mapping and map making can be found in public health, these current studies show that there remain numerous opportunities for expanding the GIS&T capacity within public health. Future studies should seek to assess and improve the public health spatial data infrastructure across all levels of government. Doing so requires a greater appreciation and more honest appraisal of the people, policies, standards, and technologies that are in practice today and which may become available in the future. Trained personnel, regular professional development (as in training updates) and appreciation of GIS&T knowledge are critically important across all administrative and programmatic levels. Hopefully in the future, all public health programs will be knowledgeable and capable of collecting, storing, and reporting public health data in commonly accessible digital formats using established and commonly agreed upon best practices. If this is achieved, public health professionals will be better able to monitor, identify, and respond to disease outbreaks, and thereby improve global health.