2019

Geospatial Analysis of Care and Mortality in the 2014 Liberia Ebola Outbreak

Marion Carlton Kinkade
Walden University

Walden University

College of Health Sciences

This is to certify that the doctoral study by

Marion Kinkade

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Namgyal Kyulo, Committee Chairperson, Public Health Faculty
Dr. Srikanta Banerjee, Committee Member, Public Health Faculty
Dr. Kai Stewart, University Reviewer, Public Health Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2019

Abstract

Geospatial Analysis of Care and Mortality in the 2014 Liberia Ebola Outbreak

by

Marion Kinkade

MCRP, University of Nebraska-Lincoln, 2000

BSAS, University of Nebraska-Lincoln, 1993

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Public Health

Walden University

February 2019

Abstract

The Ebola outbreak in West Africa from 2014 to 2016—the largest Ebola outbreak in history--had more than 28,000 suspected, probable, and confirmed cases. This study examined the relationship between distance from villages to Ebola treatment units (ETUs) and mortality. Using Geographic Information System (GIS) and statistics framed within the Social Ecological Model and the GIS Framework, this study geolocated the Ebola cases by village, mapped the travel routes, and calculated the distance to the ETUs. A logistic regression was used to determine the relationship between distance and mortality, with and without controlling for age and gender, and to calculate the odds ratio. Results showed there was an association between distance and mortality, and that Ebola patients living within 12 kilometers of the ETU were 1.8 times less at risk of mortality (*OR* = 1.778, 95% CI [1.171 – 2.7]) than those living more than 12 kilometers from the ETU. Males had a 1.4 times lower risk of death due to the Ebola virus disease. These findings can inform responses to future outbreaks and placement of treatment units, and lead to social change with respect to individual understanding of access to care, community expectations, and national health care planning.

Geospatial Analysis of Care and Mortality in the 2014 Liberia Ebola Outbreak

by

Marion Kinkade

MCRP, University of Nebraska-Lincoln, 2000

BSAS, University of Nebraska-Lincoln, 1993

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Public Health

Walden University

February 2019

Dedication

I dedicate this study to the Ebola cases and their families. EVD knows no boundary or economic status and ravaged countries that had a fragile health system and poor infrastructure to deal with it. I also dedicate this study to the thousands of person hours by numerous partners who risked their lives and left their families to support Liberia. I dedicate this study to my Liberia Ministry of Health colleagues who I have grown very close to over the last 5 years and have supported this study. In addition to these, I dedicate this study to my family; Marie, Austin, and Peyton Kinkade who supported my numerous deployments for the Ebola response, their willingness to move to Liberia while I supported the Ministry of Health, and their willingness to give up many weekends while I completed my doctoral work.

Acknowledgments

I would like to thank my family for their continued support and patience as I ruined many weekend plans due to homework and writing requirements. I would also like to thank my chair, Dr. Namgyal Kyulo, who provided timely feedback throughout the study process and my committee member Dr. Srikanta Banerjee, who provided methods input and was a good sounding board. I would also like to thank Luke Bawo and Dr. Masoka Fallah, from the Liberia Ministry of Health and the National Public Health Institute of Liberia for their willingness to share the Ebola case data and provide input for the study. Hopefully, out of the sorrow of many lost lives, we can expand our knowledge and prevent this from happening again.

Table of Contents

List of Tables

List of Figures

Section 1: Foundation of the Study and Literature Review

Introduction

In 1994, author Richard Preston published his book *The Hot Zone* detailing the terrifying true story event when men camping in Central Africa visited Kitum Cave and contracted the Marburg virus, subsequently dying a very painful and gruesome death (Preston, 1994). He continued discussing other events involving viral hemorrhage fever (VHF) cases, such as Ebola virus disease (EVD) and the Reston virus, which brought global attention to these serious pathogens. Fortunately, each outbreak of these viruses, prior to the outbreak in West Africa, have remained limited to certain geographical areas or populations. The EVD outbreaks have ranged from 1 to 318 cases until the outbreak in West Africa, which had 28,616 total suspect, probable, and confirmed cases (Centers for Disease Control and Prevention [CDC], 2017). The outbreak in West Africa shook the world and highlighted how easily EVD can spread, not just in a country, but globally. The fact that a person flew from Liberia to Dallas with this potentially deadly pathogen and extended the reach of an outbreak of EVD to a second continent and drove the need for an even more heightened international response.

EVD was first discovered in the Democratic Republic of Congo (DRC) along the Ebola River (CDC, 2016). The natural host for the virus is still unknown but research is pointing toward bats as the reservoir who then spread the virus though contact with primates and humans, or from bats to primates and then to humans (CDC, 2016). In places like Liberia, bushmeat is a common food source, but also a potential EVD exposure route, leading to strong messaging campaigns during the outbreak to avoid it

(Ordaz-Nemeth, 2017). EVD has five species which include; Zaire ebolavirus (Ebola virus), Sudan ebolavirus, Tai Forest ebolavirus, Bundibugyo ebolavirus, and the Reston ebolavirus which are all part of the family Filoviridae (CDC, 2016). Once in humans, the virus spreads through close personal contact with an infected person or their body fluids (CDC, 2016).

Pathophysiology of Ebola Infection and Treatment

When a person is exposed to EVD, the symptoms will appear within 2 to 21 days and include fever, muscle pain, weakness, fatigue, diarrhea, vomiting, and unexplained hemorrhaging (CDC, 2014). Once symptomatic, the infected person can shed the virus through body fluids or contaminated objects like needles (CDC, 2015). If a person is symptomatic, they should seek immediate medical attention. Even though there is currently no cure for EVD, supportive care should start as soon as possible, and the infected person should be isolated (World Health Organization [WHO], 2017). Once in isolation, a person should be treated and monitored until either death or discharge. After the patient arrived, they were tested by reverse transcription polymerase chain reaction (RT-PCR) and if the result was negative and they had been monitored for 3 or more days with no symptoms, they were discharged (CDC, 2014). On the other hand, if they were polymerase chain reaction (PCR) positive, then they were isolated in a high-risk zone in the Ebola Treatment Unit (ETU) where they remained until they tested PCR negative (CDC, 2014).

EVD in West Africa

The Ebola outbreak in West Africa was first detected in Guinea in December 2013 and then it spread to Sierra Leone and Liberia ultimately resulting in over 28,000 suspected, probable, and confirmed cases (WHO, 2014; CDC, 2016). This outbreak was the largest Ebola outbreak in history and required a sustained response from international partners for over 2 years (WHO, 2016). CDC alone deployed over 1,897 staff to the field, logged 110,000 total workdays, and had over 4,000 staff working on the response (Bell et al., 2016). Liberia has 15 counties (see Figure 1: Liberia and the 15 counties), and the virus was first detected in the northern county of Lofa, which is along the border of the two districts in Guinea where the first cases in the outbreak occurred (WHO, 2015). After entering Liberia in early 2014, the virus soon spread throughout the country resulting in over 10,000 suspected, probable, and confirmed cases and almost 5,000 deaths (CDC, 2016).

The Ebola virus is extremely contagious, with a disease fatality rate up to 90%, and spreads through contact with an infected person or animal; the infected entity remains a risk after death (WHO, 2017). One group that is particularly vulnerable to the Ebola virus are health care workers because they have direct contact with the patients. For example, in Liberia, 345 suspected and confirmed health care worker cases of the Ebola virus were identified between January 2014 and March 2015 (WHO, 2015). During the outbreak, case fatality was the highest for young children and older adults, who typically have a low immune defense mechanism and have the least mobility (Leligdowicz et al., 2016). In addition to the issue of vulnerable populations, Liberia's dilapidated road infrastructure contributed to the challenges of accessing care (The World Bank, 2015).

The Ebola virus spread quickly across Liberia in the first half of 2014 but, in July 2014, only two Ebola Treatment Units (ETUs) had been constructed; one was in northern Liberia, in Lofa County, and the other in capital city, Monrovia (Nyenswah et al., 2016). Each of these locations can take between several hours to several days of driving time depending on the season. This wide range of driving time is because Liberia is one of the wettest countries in the world, and with the poor road infrastructure, where most roads are dirt, they can become impassable (Sieh, 2017). During an outbreak, like Ebola where medical attention is needed immediately, impassable roads became a significant barrier to care.

In addition, when there were only two ETUs in the country, the distance to care and the road conditions may have contributed to higher mortality rates of the Ebola cases because of the inability to seek immediate medical attention. Similar research by Kenny et al. (2015) determined that distance from health facilities was a barrier for maternal and child health for routine care in Liberia but this need for quick access to health facilities is even more critical during a large outbreak, such as the recent Ebola outbreak. Distance to care barriers are well documented in literature but there is a gap in literature focused on large outbreak responses, and specifically where to place treatment units in relation to populations (Kadobera, Sartorius, Masanja, Mathew, & Waiswa, 2012; Hanson, Cox, Mbaruku, Manzi, Gabrysch, Schellenberg, …& Schellenberg, 2015; Kenny, Basu, Ballard, Griffiths, Kentoffio, Niyonzima, …& Kraemer, 2015). Current studies have focused on barriers to health care in typical health care settings and where to construct clinics based on the catchment area but no studies were identified that focused on an emergency setting where new treatment units are constructed to meet the temporary demand for an outbreak response (Kadobera, Sartorius, Masanja, Mathew, & Waiswa, 2012; Hanson, Cox, Mbaruku, Manzi, Gabrysch, Schellenberg, …& Schellenberg, 2015; Kenny, Basu, Ballard, Griffiths, Kentoffio, Niyonzima, …& Kraemer, 2015). By studying the 2014 – 2016 Liberia Ebola case data, an understanding of the association of distance to ETUs on mortality rates can provide critical guidance for future planning for large-scale outbreaks.

Figure 1. Liberia and the fifteen counties.

Problem Statement

The Ebola outbreak in Liberia rapidly spread through the country for several reasons; a weak healthcare system, shortage of healthcare workers, poor hygiene practices, inability to limit population movement across the borders, lack of understanding the disease, poor infrastructure, and cultural practices (United Nations Children's Fund [UNICEF], 2014). Furthermore, the healthcare system in Liberia was not prepared for this type of outbreak due to limited health-care worker experience, supplies, Ebola-specific treatment units, basic health service limitations, and a mistrust of the health care system (UNICEF, 2014). In addition, cultural practices of traditional burials, that involve washing the body, added to the risk of disease spread (UNICEF, 2014). The poor infrastructure contributed to the access to care challenges and specimen delivery to the laboratory for EVD confirmation (Shoman, Karafillakis, and Rawf, 2017). In Liberia, over 60% of the population live more than 5 km from the closest health facility and this lack of fast access to care has led to one of the highest maternal mortality rates in the world. Basic health care delivery in non-emergency settings is challenging much less during an emergency (Lucklow et al., 2017). According to the Liberia 2013 Demographic and Health Survey, 49.9% of the rural population travel over 60 minutes to the nearest health facility and over 75% walk (Liberia Institute of Statistics and Geo-Information Services [LISGIS], 2013). In urban Liberia, 16.7% of the population own a car or motorcycle whereas in rural areas, only 6.2% owns a car or motorcycle (LISGIS, 2013). This low vehicle ownership means that most people either walk or rely on public transportation. Therefore, during the Ebola outbreak, a sick person had to either rely on someone to take them to a clinic or walk over an hour.

In a large outbreak, such as the Ebola outbreak, the placement of treatment units (TUs) becomes a critical decision by the government and especially, in a country like Liberia, where the road conditions and available transportation can become a significant barrier to care. Treatment in an ETU can significantly increase the odds of survival for a patient but with limited ETUs in the country, it became challenging for many patients (Leligdowicz et al., 2016; Nyenswah et al., 2016). In addition to access to care challenges, there was resistance by some Liberians toward using health facilities; therefore, they remained in their community, which led to community leaders creating their own support structure for suspect cases (Abramowitz et al., 2015). Health messaging supported behavior changes during the response to limit direct contact, quarantine, and safe burials in the community but there was still limited access to ETUs early in the response (Wagner, 2015). The initial ETUs were placed in Lofa County, where the initial cases occurred, and in Monrovia because it is the largest city and where most of the health care resources are located (Nyenswah et al., 2016). This decision made sense due to the initial cluster and then the shifting of the epicenter to Monrovia, but for cases in the distant counties, it made receiving care challenging. Distance to care may have contributed to higher mortality for those cases that traveled to either of these ETUs and therefore warrant further investigation. Ideally, ETUs could have been in each county or near each population center to allow easier access to care. This systematic placement of ETUs was later part of the plan for the United States Army (Beaubien, 2014). The United States Army had planned to construct 18 ETUs but when the case count started to decline, most of the facilities were put on hold (Beaubien, 2014).

The locations of two of the additional ETUs were near the airport outside of Monrovia and in Tubmanburg, which is between Monrovia and Lofa County (Beaubien, 2014). The locations were near population centers and national resources but not necessarily data driven. If the Ebola case data were mapped and analyzed by cluster, analysis could have been conducted to determine the best location for the ETUs. In addition, by analyzing the case data related to the ETU locations, distance-related mortality could have been determined which would have provided data driven guidance on the most effective locations. This analysis process can be used in future responses to change the behavior of response planners and community leaders when planning the location of treatment units.

No studies were found that analyzed the impact of distance from villages to ETUs on mortality rates but Ly et al. (2016) found that facility-based delivery declined by 30% during the outbreak and distance was a factor. Lori et al. (2017) expanded on this analysis and found that distance was a barrier for women to use maternity waiting homes during the outbreak; therefore, they tended to deliver at home. Carter et al. (2017) determined that not only did distance impact seeking treatment in Sierra Leone but the perception of distance was also a barrier. The perception was that people left in ambulances and didn't return, which fueled rumors and inhibited their understanding of the critical care centers (Carter et al., 2017).

Purpose of the Study

Distance between home and healthcare facilities has been shown to create both physical and mental barriers to care (Kenny et al, 2015; Carter et al., 2017). By understanding the impact of these barriers during the 2014 – 2016 Ebola response in Liberia, guidelines can be developed for future responses. These guidelines can then be used during outbreak responses where additional treatment units are needed to serve the population at risk. Additionally, these guidelines can be used to guide health care delivery in non-emergency settings, such, as when new clinics are being constructed, for mobile health clinics, creating waiting houses for mothers, and a broader understanding of access to care barriers.

The physical distance to care centers impacts non-emergency related basic health services to the extent that the Liberia government established maternity waiting homes with midwives to reduce the maternal mortality rates (Lori, Munro-Kramer, Shifman, Amarah, and Williams, 2017). By analyzing distance as a physical barrier across Liberia, it has created social change toward basic services, like maternal health services. The purpose of this study was to understand the association between mortality and distance to ETUs for the Ebola patients in Liberia. This understanding can be used to plan the placement of treatment units (TUs) in future responses, change the way healthcare leaders plan emergency preparedness responses, and influence social change from the individual, community, and societal levels as it pertains to public health preparedness and response.

By understanding the impact of appropriately locating treatment units based on both physical and mental distance to health barriers, public health leadership can both plan and communicate appropriately. This process requires social change at all levels of the social ecological model (SEM). Social change at the policy level to understand the impact of distance to care barriers in national plans. At the community level, distance to care barriers need to be addressed with community leaders to ensure appropriate messaging is in place during an outbreak and that TUs are available for their residents.

This social change aspect will influence community leaders to become more engaged in the planning process, communicate with national leadership, and communicate with community residents. At the interpersonal and personal levels, individuals need to understand the impact of distance on health outcomes and plan for it. When there are challenges to care, due to distance, for normal basic health services, how will they get to care or take loved-ones to care when they are seriously ill? Also, the need to trust medical services leads to social change for the willingness to go to a TU. Just because a TU is farther away, the distance barrier needs to be overcome, both physically and mentally, for the sake of the patient.

Research Question(s) and Hypothesis

Research Question 1: Is there an association between distance from Ebola-affected villages and ETUs and the mortality rates during the 2014 – 2016 Ebola outbreak in Liberia?

$H_0$1: There is no association between distance from Ebola-affected villages and ETUs and the mortality rates during the 2014 – 2016 Ebola outbreak in Liberia.

H_a1: There is an association between distance from Ebola-affected villages and ETUs and the mortality rates during the 2014 – 2016 Ebola outbreak in Liberia.

Research Question 2: Is there an association between the distance of Ebola-affected villages and the ETUs on mortality rates during the 2014 – 2016 Ebola outbreak in Liberia when controlling for sex and age?

$H_0 2$: There is no association between distance from Ebola-affected villages and ETUs and the mortality rates when controlling for sex and age during the 2014 – 2016 Ebola outbreak in Liberia.

$H_a 2$: There is an association between distance from Ebola-affected villages and ETUs and the mortality rates when controlling for sex and age during the 2014 – 2016 Ebola outbreak in Liberia.

Theoretical Foundation for the Study

To address the physical and mental barriers to care during an outbreak, a combination of the social ecological model (SEM) and a Geographic Information System (GIS) framework guided the understanding of barriers to healthcare, the spatial relationships between village locations and treatment units, and the behavior change when responding to an event (Kadobera et al., 2012). The SEM is a multi-level model that frames society at the individual, interpersonal, community, organizational, and policy levels (UNICEF, 2016). This framework can guide the understanding of the personal and environmental factors that affect behavior and guide interventions to influence social change during a large-scale outbreak. This multilevel framework can frame the geospatial analysis, statistical analysis, and provide guidance for future outbreak response planning.

A GIS framework guides the data collection that will enable a model, that represents the real world, and identify spatial patterns relevant to the topic under investigation (Environmental Systems Research Institute [Esri], 2017). By using this framework, the variables were identified that are relevant to physical barriers to care during the Ebola outbreak. GIS is used for spatial analysis, analysis of health service planning, and public health (Lyseen et al., 2014). This process focuses on using a series of steps to map the data, create models, and then conduct spatial analysis (Wang, Shen, and Tang, 2014). For this specific study, a GIS framework was used to map the Ebola-affected villages, road and travel networks, and locations of the ETUs, analyzed clusters of Ebola cases, modeled distances between the Ebola-affected villages and the ETUs, and calculated mortality rate gradients by distance to determine if distance affected mortality during the outbreak. By using the individual, community, and societal levels of the SEM, it helped to frame the GIS analysis and direct the outcome.

Nature of the Study

The study was quantitative in nature, utilizing the retrospective Ebola case data from Liberia from 2014 – 2016. The analysis was a combination of geospatial analysis and statistical analysis using ArcGIS 10.3 and SPSS software. The first step was to analyze the location of the villages, with cases, compared to the location of the ETUs. To determine actual distance traveled, a Manhattan model was used. A Manhattan model can estimate actual distance traveled in health service research (Shahid, Bertazzon, Knudston, and Ghali, 2009). The Manhattan model analyzes the actual travel pattern to get from point A to B compared to Euclidean models that analyze distance based on straight-line distance between point A and B. Using the output of a Manhattan model integrated with the status of cases (dead or alive), an odds ratio by distance was calculated. Using this information, there were certain distances that were a higher risk of mortality. The case data was then divided into categories, based on distance between the village of symptom onset and the ETU where they received treatment. Using these distance categories and the final status of the case, alive or dead, an odds ratio was calculated to identify the odds of mortality by distance category.

The key independent variable was the distance between the Ebola-affected villages and the ETUs. Final Status (dead or alive) was the dependent variable. The home village for each case was identified and mapped using GIS. From the mapped locations, network distance to the ETU can be analyzed for patterns of higher mortality. These distances were then categorized to identify mortality rates per segment distance from the ETU that was used and thus determined if distance from home locations of cases to the ETU effected the health outcome.

To understand how the distance can impact mortality rates, both geospatial analysis and statistical analysis was conducted. The statistical analysis was an odds ratio used after the geospatial analysis which calculated the network distance from the village of symptom onset to the ETU. The calculation provided the odds of surviving when the Ebola patient was beyond a certain distance from an ETU when they experienced the onset of symptoms. In addition, geospatial analysis was conducted to understand the impact of distance on mortality rates examining the Manhattan model distances to care. Unlike Euclidean distance, which is the straight-line distance between two points (as the bird flies between the village and the ETU), the Manhattan distance use the road networks (actual traveling distance) (Shahid, Bertazzon, Knudtson, & Ghali, 2009). Between the two analysis methods, the Manhattan model is more accurate but also more challenging due to the requirement of a complete transportation network to calculate the distance. The outcome provided statistically significant breakdowns of whether distance effected mortality during the outbreak or not. An example of the potential outcome is: a distance of greater than 1 km resulted in X mortality rate, greater than 2 km resulted in X mortality rate, etc. In addition, there may be clean cut points, such as more than 10 km resulted in a mortality rate double that of less than 10 km therefore distance would have affected the health outcome.

Literature Review

Literature Search Strategy

The literature search strategy was combination of searches in the Walden University Library, Google Scholar, ProQuest, and Google. Using ProQuest, I conducted a search for recent theses and dissertations to find related studies on spatial analysis, access to care challenges, distance analysis, GIS Framework, and SEMs. Similar searches were conducted using Google and Google Scholar, but additional themes were used to include Ebola, public health disaster response, spatial cluster analysis, and distance to care analysis. The key term used to conduct the search included: *Ebola, community-based Ebola response, spatial cluster analysis, Euclidean analysis, Manhattan model, access to health care spatial analysis, GIS framework, social ecological model, disaster management analysis, distance to care barriers, spatial clusters analysis of Ebola,* and *Liberia Ebola response*.

The search time frame was between 2012 and 2017 but since most of the Ebola related studies were conducted within in the last 2 years, most of the Ebola-related publications were recent. In addition to publications, information was referenced from WHO, CDC, USAID, and African newspapers.

Literature Review Related to Key Variables and/or Concepts

The literature review focused on five concepts: Ebola-related spatial analysis from the recent outbreak in West Africa, distance to care analysis, geospatial analysis of distance, GIS Framework, and the social ecological framework. The reason I focused on these concepts is that they are related to the core components of the study. There was a gap in the literature related to the exact study but each of these concepts illustrate how related research has been conducted and was applied to this current study. Each of these concepts was reviewed and leveraged for this study for variables and analysis.

Ebola-Related Spatial Analysis

The Ebola outbreak in Liberia began in Lofa County before spreading to the Monrovia and then the rest of the country (Nyenswah et al., 2016). The knowledge of where it started and where it spread is important, but it doesn’t provide the full picture. To understand the full picture of the outbreak, a broader understanding of the spatial analysis across the outbreak in West Africa is important, beginning with the spatiotemporal transmission of the virus. Fang et al. (2016) mapped spatiotemporal transmission patterns, identified influential factors, and assessed the effects of the interventions at the local level. This analysis focused on the transmission patterns in Sierra Leone and the diffusion corridors using a Poisson model between transmission routes, imported cases, and transmission between chiefdoms. The geospatial surface analysis identified two diffusion corridors; one from the border of Guinea into the

Kailahun District and the other from districts in the west. Each of these corridors spread the virus across Guinea quickly. They identified an association between chiefdoms with primary and secondary roads being affected by the Ebola virus. Fang et al. (2016) also found an association between chiefdoms with hospitals and the number of Ebola cases but suggested a surveillance bias may be the cause due to proximity to care and rapid diagnoses and reporting.

The role of geography in transmission is an important factor for preparedness planning and effective responses. Kramer et al. (2016) focused analysis on the individual level using a generalized gravity model to understand the transmission between locations. The results illustrated how strongly geographic features influenced the spread of EVD but were influenced by distance, population density, and national borders (Kramer et al., 2016). The focus was on the spread of Ebola, not specific distance to care spatial analysis, but the understanding of how the model was used is helpful. Gleason et al., (2017) used GIS to conduct spatial-temporal analysis within a severely-affected village in Sierra Leone. This study focused on household level transmission within the village, inter-household distances, space-time cluster analysis, and a generalized estimating equation (GEE) to understand the spread of the EVD. The results indicated that the temporal aspect outweighed the spatial effects, with the results of the GEE showing 4% increased odds of a household getting EVD with each additional person and 2.6% each day after onset of symptoms. This analysis identified the need for prompt localized interventions to stop the spread.

Spatial determinants can also be used to assess disease risk. Zinszer, Morrison, Verma, and Brownstein, (2017) used district level Ebola case data to model population-level predictors of the EVD. By combining the district level Ebola data with demographic data, temperature, rainfall, and land cover estimates, Zinszer et al. (2017) estimated the EVD risk to evaluate spatial variability using a Bayesian hierarchical Poisson model. This analysis determined there was a heighted risk for EVD for households "not possessing a radio (*RR* = 2.79, 95% CI [0.90-8.78]; *RR* = 4.23, 95% CI [1.16-15.93], increasing rainfall (*RR* = 2.18; 95% CI [0.66-7.20]; *RR* = 5.34, 95% CI [1.20-23.90], and urban land cover (*RR* = 4.87, 95% CI [1.56-15.40]; *RR* = 5.74, CI [1.68-19.67]." Health messaging, via radio, was a critical component of the response during the outbreak, which would help to explain the households at higher risk if they did not have a radio (Wagner, 2015).

Carter et al., (2017) examined barriers and enablers for seeking treatment in Sierra Leone. The analysis included interviews, focus groups, and questionnaires from approximately 350 individuals in more than 200 villages to tease out what affected community members from seeking care and the three primary barriers were fear and limited information, concern about unknown outsiders, and prohibitive distance to treatment facilities (Carter et al., 2017). Respondents commented that the need to travel 2 to 3 hours to health care was prohibitive, in addition to the fear that if they died in the location, cultural burial practices may not be followed. The distance from health facilities also contributed to limited knowledge about EVD, which led many to believe rumors and myths (Carter et al., 2017). When new critical care centers were constructed in some communities, the distance barrier was reduced which led to an increase in treatment-seeking behavior and a reduction in rumors and myths (Carter et al., 2017).

Distance to Care Analysis

In a non-outbreak setting, distance is a well-documented barrier to care. Kenny et al. (2015) conducted analysis on the impact of distance to health care in rural Liberia for maternal and child health care. This study conducted a cluster-sample survey in rural southeastern Liberia to determine the association between health facility service use and communities for maternal and child services (Kenny et al., 2015). The analysis consisted of logistic regression models using distance quartiles and the odds of outcomes of interest. This was a non-geospatial approach but still yielded statistically significant results, "living in the farthest quartile was associated with lower odds of attending 1–or–more ANC checkup ($AOR = 0.04$, $p < 0.001$), 4–or–more ANC checkups ($AOR = 0.13$, $p < 0.001$), delivering in a facility ($AOR = 0.41$, $p = 0.006$), and postnatal care from a health care worker ($AOR = 0.44$, $p = 0.009$)", indicated that distance was associated to reduced health care uptake (Kenny et al., 2015, p. 1). This is an important finding that illustrates the impact of distance as a barrier to care during a non-outbreak setting. If women are unwilling to seek maternal or child care services due to distance, the same behavior may occur during an outbreak putting both her and her family at risk.

A similar study was conducted in rural Tanzania analyzing maternal mortality and distance to facility-based obstetric care. Hanson et al., (2015), conducted a georeferenced cross-sectional secondary census data analysis across five rural districts in Tanzania in 2007. The analysis involved a multilevel logistic regression to determine the effects of distance to health facilities providing delivery care and included 818,583 people (Hanson et al., 2015). A Global Positioning System (GPS) receiver was used by fieldworkers to collect the geographical coordinates for the household and Euclidean distance (straight line) to the nearest health facility was calculated. This distance was then used in the regression to calculate the effect on mortality. The results indicated a strong association between distance and deaths due to direct pregnancy-related causes. “Deaths due to direct causes of maternal mortality were strongly related to distance, with mortality increasing from 111 per 100,000 livebirths among women who lived within 5 km to 422 deaths per 100,000 livebirths among those who lived more than 35 km from a hospital (AOR = 3.68, 95% CI [1.37 – 9.88])” which aligns to the impact of distance on mortality in other studies (Hanson et al., 2015, p. e387).

Another study in Tanzania examined the impact of distance on childhood mortality from 2005 – 2007. Kadobera et al. (2012), examined data from the Ifakara Health and Demographic Surveillance System for secondary analysis of 28,823 children under the age of 5. This study calculated both Euclidean distance (straight line) and Manhattan distance (networked) to analyze the impact of distance. The Cox proportional hazard regression model was used to analyze the effect of distance on child mortality. ArcGIS 9.1 software was used to analyze the Manhattan and Euclidean distances from the home to the closest facility. These variables represented the explanatory variables in the Cox proportional hazard regression model while adjusting for other confounding factors. The results indicated that “children who lived in homes with networked distance >5 km experienced approximately 17% increased mortality risk (HR=1.17, 95% CI

[1.02–1.38] compared to those who lived <5 km networked distance to the nearest health facility" (Kadobera et al., 2012). This study represented similar results to the maternal mortality study where 5 km seems to be a strong cutoff for positive health outcomes. In addition to being similar to the maternal mortality study, it is the most similar to the analysis proposed in this study.

In Liberia, approximately 60% of the population lives more than 5 km from the nearest health facility which lead the country to focus the community health worker (CHW) program on these populations (Lucklow et al., 2017). Kentoffio et al. (2016) compared rural and remote healthcare utilization in Liberia and found that 59.6% of the women surveyed walked more than an hour to the nearest health facility. In their multivariable logistic regression models, they focused on demographic variables versus distance to care, but it is a significant finding that women would spend more than two hours walking to and from a health facility. This would be a significant barrier for a person with EVD.

In addition to distance, transportation can be a barrier when many Liberians do not have vehicle. They are dependent on public transportation or a relative, which can also put the drivers at risk during an outbreak. Gortland, Taryor, Norman, and Vermund, (2012) conducted a study in north-central Liberia on maternal access to health facilities. The convenience sample of 307 women focused on health care utilization through interviews and identified transportation barriers due to cost and transportation (Gortland et al., 2012). Distance was not analyzed in this study and all communities involved were within 15 km of a health facility, but the issue of transportation is important and was a barrier during the EVD outbreak. The transport of patients to health facilities was significantly delayed due to poor roads and transportation services (WHO, 2015).

These studies focused on the effect of distance on mortality or health outcomes but during non-outbreak settings. The methods of analysis, for this study, would be similar but there is a gap in the literature discussing this analysis in an outbreak and response setting. Similar methods could be used for infectious disease preparedness and response.

Geospatial Analysis of Distance

The understanding that distance from health care impacts healthy outcomes is important but there also needs to be an understanding of how distance is calculated. There are several models used to model distance in GIS. One method, that is the most straight forward, is the Euclidean or straight-line distance, or as the bird flies. This method is easy to calculate, if the two end point locations are known, and are often used to create catchment areas (Zinszer et al., 2014). A common framework for catchment areas uses a 5 km straight line distance from the health facility which is about an hour walk from the homes to the health care facility (Ueberschar, 2010). This method of analyzing access to care and health care coverage areas has been successfully applied in numerous studies but has some limitations. It assumes health facilities are near population centers and that there are no geographic barriers. Macharia et al. (2017) demonstrated that in a country like South Sudan, Euclidean distance and the standard 5 km catchment areas only captured about 28.6% of the population because the population is so spread out. There is also a difference between capturing a population number for catchment areas within 5 km and then ensuring that population can actually walk to the facility within an hour. If the goal of the straight-line distance delineation is for the population served to be within an hour's walk, then the actual geography is important. Noor et al., (2006) showed how the Euclidean distance can be used to successful "capture" the population but due to geographic barriers, actually getting to some of the clinics within an hour was challenging. They went on to suggest that a network model better captures actual travel time (Noor et al., 2006).

Euclidean distance calculation provides a straight forward method of capturing population within a given distance of a health facility but when discussing outbreak response, travel time to a clinic and the placement of treatment units to serve a population become just as important as the population numbers. To calculate actual travel times from a home to a clinic, it requires a network or Manhattan model. A Manhattan model uses the road and pathway network to create the actual route taken and impedances can be added to the model to represent the impact of terrain (Noor et al., 2006). For example, Noor et al., (2006) used the Naismith-Langmuir rule for walking times which state that on a flat road a person can travel 100 m in 1.2 minutes but for ascents it adds 0.1 minutes per 1 m in ascent. This is not accounted for in a Euclidean model, but a network model can become very complex. A Manhattan or network model requires a geometric network but in countries, like Liberia, only the road system exists in geospatial format, however, foot paths are used as much as roads for walking to clinics which are not mapped (Eros, 2016). Where a Euclidean distance tends to underestimate travel distance, the Manhattan model can over-estimate (Shahid, Bertazzon, Knudtson, and Ghali, 2009). Shahid et al.

(2009) compared the Euclidean, Manhattan, and Minkowski models to estimate travel distances and found that the Minkowski model had the best results but is more complex.

Literature Review related to Theoretical Framework

GIS Framework

When considering a geospatial approach to determine the best placement of treatment units during an outbreak, the analysis should be framed within an GIS Framework. A GIS Framework identifies the variables to be analyzed, defines the construct of their use, the geospatial analysis method to be used, and the expected outcome (Bower, 2015). This process outlines the methods and the expected outcomes that can be used within a SEM to address the needs of the study at each level. To understand what variables are needed and how they would fit in the SEM, an understanding of space-time patterns is important. Lafreniere and Gilliland, (2015) conceptualized a GIS Framework for everyday life. This framework utilized two environmental stages, social and built, by integrating qualitative and quantitative data from the individual, community, and societal levels (Lafreniere and Gilliland, 2015). In this example, Lafreniere and Gilliland (2015), created stages of development; the built environment stage which focused on physical data layers such as topographic maps, geodetic surveys and fire insurance plans, and the social environment stage, which focused on social indicators such as census data, demographics, relationships, and employment. Each of these were combined in a model to address different studies, to include, the exploration of "potential exposures to noxious industrial environments during children's daily journey to school in the 1880s" (Lafreniere and Gilliland, 2015).

This framework outlines the potential for analysis because it incorporates qualitative and quantitative variables for space and time. When conducting a study, that examines individual level variables, such as demographic data, with location data, and time, it emphasizes the need for a well-designed framework and interaction between these variables.

Makanga et al. (2016) developed guidelines for creating a GIS data framework focused on low and middle-income countries. Using a similar structure of grouping data in themes to analyze geographical access to health services, data was grouped physical data, such as infrastructure, socio-cultural variables, such as financial support and emergency transport, and spatio-temporal data, such as survey data (Makanga et al., 2016). The results are a well thought out process for utilizing spatial data to analyze access to care services, in addition, to the importance of a GIS Framework approach to solving spatial problems.

Social Ecological Framework

A study that has individual, community, and societal level variables would need a SEM approach to understand the relationship and impact of the variables. Figueroa (2017) used a SEM-based approach to examine communication and behavior to contain the Ebola outbreak in Liberia. The conceptual framework studied the effect of individual and community Ebola prevention and control messaging, service delivery systems, and the social political environment on the Ebola response (Figueroa, 2017). The approach outlined the importance of a multi-level response for a crisis, like the Ebola outbreak, to change behavior, inform interventions, and to determine behavior determinants which are critical when "time is of the essence and lives are on the line" (Figueroa, 2017).

In non-emergency related health care delivery, a social ecological approach has been used across numerous disease specific access to care studies. Gombachika et al., (2012) identified barriers to HIV care at five levels of the social ecological model. At the individual level, literacy, beliefs, and attitudes about HIV-related care was a barrier and at the interpersonal level miss-information from family and friends caused barriers to seeking care (Gombachika et al., 2012). At the organizational level, service times and locations limited access and then at the community and policy levels cultural norms were barriers. This study illustrated the importance a multi-level approach to barriers to care. Beyond just health care delivery, Liberia has focused on health system resiliency and it's their ability to bounce back after the Ebola outbreak. Cretney, (2014) states the importance of an integrated social and ecological approach to resilience. This approach creates a multilevel definition of resilience that includes ecological, social, socio-ecological, community, and urban data to model the geographies of resilience.

Significance

The significance of this study is to change the way public health preparedness and response are currently conducted, as it pertains to locating assets. By modifying the policy decisions around the placement of treatment units, effected populations can be served better, and thus social change can be accomplished. This modification can then influence behavior change in populations to expect less burdensome access to care and understand the importance of using it. Literature suggests that distance affects health outcomes for routine care, but little research has been conducted on emergency-related access to care. This study builds on lessons learned from the 2014 – 2016 Ebola outbreak in Liberia but focusing on the placement of the Ebola treatment units and its effect on mortality. The analysis focused on the association of the placement of the ETUs on Ebola mortality, which can guide future response efforts when treatment units are being constructed to support an outbreak.

Summary

The EVD is a deadly pathogen than can have mortality rates as high as 90% but outbreaks have, for the most part, remained small clusters until the outbreak in West Africa (WHO, 2017; CDC, 2017; CDC, 2016). The outbreak in West Africa consisted of over 28,000 suspect, probable, and confirmed cases (CDC, 2016). There is no cure for EVD, but immediate supportive care has been shown to successfully overcome the virus in most cases (WHO, 2017). While in the community, a person who is contagious, should be isolated in an ETU until they are tested PCR-negative and, therefore, virus free (CDC, 2014).

When EVD arrived in Liberia from Guinea in Lofa County, it then spread to Monrovia, the capital of Liberia (WHO, 2015). The virus then spread across the country resulting in over 10,000 suspect, probable, and confirmed cases (CDC, 2016). Early in the response, there were only two ETUs to treat patience and quickly filled up (Nyenswah et al., 2016). With only two ETUs and the poor road system with a lack of transportation, traveling to an ETU was challenging (Shoman, Karafillakis, and Rawf, 2017). This poor infrastructure created physical barriers to care that may have resulted in higher mortality for those living a long distance from care but there were also mental barriers due to distance and a concern about the “real” role of the clinics due to rumors (Carter et al., 2017). To understand the association between distance and mortality, geospatial analysis was conducted to analyze the distances from villages to ETUs. This multilevel analysis was framed using both a GIS framework and a SEM to understand the association from the individual level to the policy level.

The output of the multilevel geospatial analysis can influence social change across the SEM by providing guidance for individual messaging, community and organizational planning and response, and national preparedness planning and response.

Before and during an outbreak response, an understanding of the physical and mental barriers to care is important. One aspect of barriers to care is the physical distance to treatment which can impact mortality rates. Using the 2014 – 2016 Ebola outbreak data in Liberia, the Ebola case data was used to model distances to ETUs to identify a potential association between distance and mortality during the outbreak. This knowledge can be used to guide further TU locations during an outbreak but also to guide social change at the individual, interpersonal, community, organizational, and policy levels for health care delivery. Significant analysis has been conducted to analyze the spatial spread of Ebola transmission, but no spatial analysis has been conducted on the impact of distance to care on mortality. In related analysis on the association between distance and both maternal and child mortality, distance has been shown to impact healthy outcomes, but these studies have been limited to non-emergency health care delivery.

Section 2 – Research Design and Data Collection

Introduction

The purpose of this study was to examine the association of distance between Ebola-affected village locations and ETUs on mortality during the Ebola outbreak in Liberia. By understanding this association, this knowledge can guide future response efforts during large-scale outbreaks and create social change around health care delivery in rural areas. This study used GIS to analyze the spatial clusters of Ebola by village and then modeled the distances to ETUs with the focus on whether the patient lived or died. In addition, by using the binary logistic regression to calculate an odd ratio, the effect of distance to the ETU can be investigated (Kenny et al., 2015). This section provides the overview of the research design, data description, data analysis plan, and their use to answer the research questions.

Research Design and Rationale

A quantitative retrospective study was conducted to examine the relationship between distance (the independent variable) and mortality (the dependent variable). This study used the secondary data set from the Ebola outbreak in Liberia consisting of all the Ebola case data provided by the Liberia Ministry of Health. The key independent variable is the distance between the Ebola-affected villages and the ETUs. Final Status (dead/alive) is the dependent variable. In similar analysis, studies have shown a correlation between mortality and distance to trauma centers due to car accidents (Hu, Dong, & Huang, 2017). The village where the onset of symptoms for each case was identified and mapped using GIS. From the mapped locations of Ebola cases, network (travel) distances were calculated between cases and ETUs using ArcGIS's Network Analyst Extension. These distances were then grouped by categories and analyzed in SPSS to identify the odds ratio for each category. This odds ratio identified the risk of mortality by each distance to care category. The key variables for the analysis were village of symptom onset, location, gender, age, and final status of the case (dead or alive), and the ETU locations. The final status was the dependent variable, distance between the village and the ETU was the independent variable, and the analysis controlled for gender and age.

Methodology

Population

The study population was the suspect, probable, and confirmed Ebola cases during the outbreak from 2014 to 2016 in Liberia. The data was provided by the Liberia Ministry of Health and represented the data collected during the Ebola outbreak response. This was a case-based dataset and included names, age, gender, village location, epi classification, final status of case, date of onset, data of death, exposure history, symptoms, laboratory results, contact information, profession, and the location patient became sick, but was scrubbed to remove all of the data except county of symptom onset, district of symptom onset, village location of symptom onset, gender, age, date of symptom onset, ETU name, and final status of case.

Sampling and Sampling Procedures

The Liberia Ebola case data consists of 8,440 cases that were documented between January 2014 and December 2015. The dataset was merged between the Ebola

EpiInfo Viral Hemorrhagic Fever (VHF) module, spreadsheets, and the District Health Information System (DHIS2) by the Ministry of Health.

Sampling Frame

The sampling frame was all suspect, probable, and confirmed cases in the Ebola case dataset as enumerated in final status variable. Excluded from the sampling frame were cases that were later classified as non-cases but left in the dataset, in addition to, case data that were missing the location of symptom onset, missing gender, age, ETU, or missing the Final Status (Dead or Alive). The Ebola case data was collected via the Ebola case investigation form during the outbreak and entered into either the EpiInfo Viral Hemorrhagic Fever (VHF) module or into the District Health Information System (DHIS2). Early in the response, the EpiInfo VHF was used to store and analyze the case data but in January 2015 case data was stored in DHIS2 and excel for analysis, see Figure 2 for the data collection process (McNamara, et al., 2016). At the end of 2016, the data was merged in Excel and maintained by the MoH for analysis. The MoH provided the data, as is, for the purposes of this study and therefore required cleaning and scrubbing to de-identify the cases. Access to the data were granted by the MoH with the expectation that it would not be distributed, the analysis and study would be reviewed by the MoH, and it would receive IRB approval (permission letters located in the appendix).

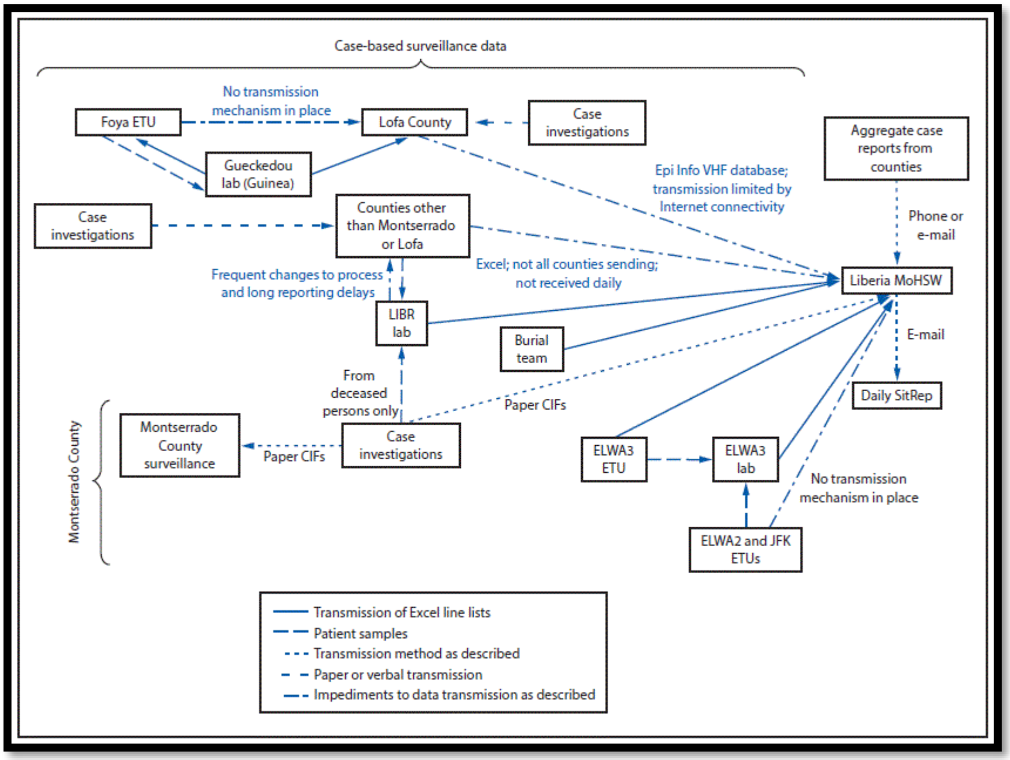

Figure 2. Ebola outbreak response data flow (McNamara, et al., 2016).

Sample Size

An alpha level of 0.5 was used to reduce type 1 error, a minimum power level of 90% to reduce type 2 error, and a moderate effect size of 0.3. The minimum sample size required is 109 cases which was calculated using G*Power 3.1.9.2, however, because all data from the outbreak is available, and to understand the distribution across Liberia, the full dataset was used.

Data Analysis Plan

Data Description and Cleaning

Prior to analysis, the Ebola case data was cleaned and sorted. The original Ebola dataset consisted of 8,440 cases but many were either duplicates or had missing values.

The key variable required for spatial analysis is the location of symptom onset. The dataset also contains the location of residence but, between the two, the location of symptom onset is the most important due to the need to seek care after symptom onset. The case data consists of 92 variables but was trimmed down to only nine, which included: county of symptom onset, district of symptom onset, location of symptom onset, gender, age, ETU, geographic coordinates, and final status. The distance was calculated based on a Manhattan model in GIS using the location of symptom onset and ETUs. To conduct this analysis, two variables are required; location of symptom onset and the location of the ETUs. Each of the variables were mapped in GIS using a village GIS layer, satellite imagery, and open street map.

Each of the Ebola case variables are described in the table below:

Table 1

Ebola Case Data Set

Variable	Type	Variable Details	Description
ID	Numeric	Number	Sequential ID given to each case
Age	Numeric	Number	Age of the Ebola case
Gender	Categorical	Male, Female	Gender of the Ebola case
Date of Onset	String	Date, Unknown = -99	Date of onset of symptoms
County_Sick	Categorical	Text, Unknown = -99	County where the case became symptomatic
District_Sick	Categorical	Text, Unknown = -99	District where the case became symptomatic
Location_Sick	Categorical	Text, Unknown = -99	Village or location where the case became symptomatic
Xcoordinates	Numeric	Decimal Degrees, Unknown = -99	Longitude coordinates of the location where the case became symptomatic
Ycoordinates	Numeric	Decimal Degrees, Unknown = -99	Latitude coordinates of the location where the case became symptomatic
ETU	String	Text, Unknown = -99	Ebola treatment unit where the case was admitted, if they were admitted into an ETU
Final_Status	Categorical	Dead, Alive	Final status of the case was either alive or dead

After the data was trimmed down to only the required variables, the data was cleaned to remove records that would not support the analysis. Figure 3 shows the data cleaning process and the datasets that will be used in the different analysis phases. The initial step was to remove all records in the dataset that are missing the location of symptom onset or any duplicate records. This output dataset was then used to map all cases by county, N = 8,240, and district, N = 8,107 to visualize the distribution of all cases. The next step was to remove any records that were missing the final status, gender, age, village of symptom onset, and final status, resulting in an N = 916. Last, univariate analysis and boxplots were conducted to determine if there were any outliers in the dataset and 31 outliers were removed. The final dataset of 885 cases provided the input for the statistical analysis.

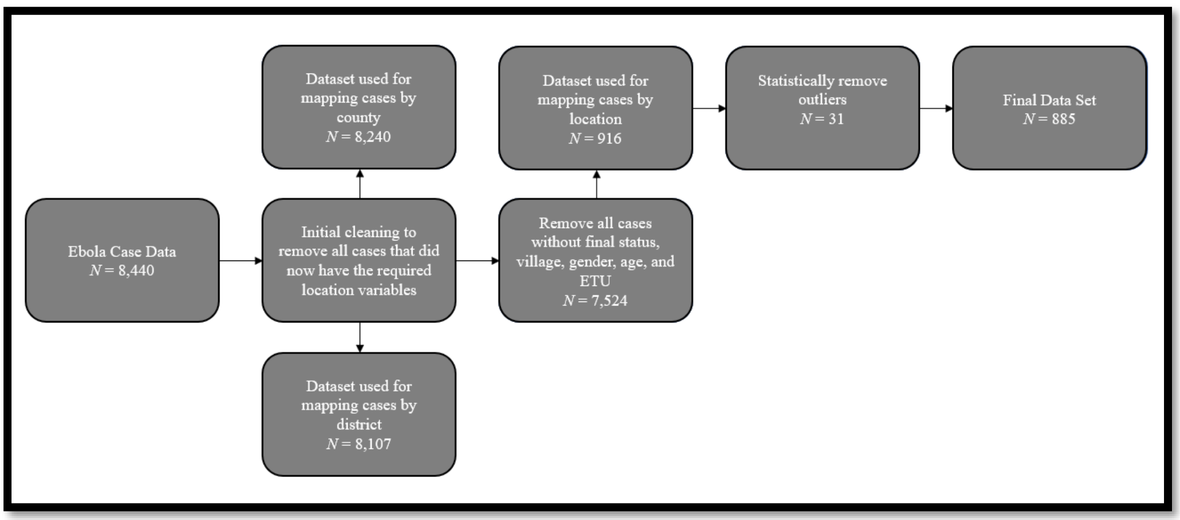

Figure 3. Data cleaning process.

Geospatial Data Cleaning and Creation

In addition to the Ebola case data, several geospatial data sets are required for mapping and analysis. These geospatial data sets include road, paths, ETU locations, villages, districts, and counties for GIS. The counties and district layers were used to map the number of cases by counties and districts. Additionally, the location of symptom onset was used to map the actual location of the cases which were used to calculate the distance to the ETU. The road and path layers will be used to determine approximate travel distances from the location of symptom onset to the ETUs. These geospatial layers will be obtained from LISGIS, UN, and open street map. In addition, satellite imagery was used to manually digitize roads and pathways that did not exist in the roads layer and to finalize the village locations for some villages in the Ebola data that did not match the village layer.

Data Analysis

The data analysis consisted of multiple phases, which include; mapping the case data by county and location of symptom onset, mapping Ebola clusters, analyzing the distances between location of symptom onset and ETUs, determining the mortality rates by distance, conducting the odds ratio between distance and mortality, dividing the data into categories based on the distance threshold for increased risk of mortality, and then statistically comparing the groups while controlling for gender and age.

Geospatial Mapping and Analysis

ArcGIS 10.3 was used to map the Ebola case data by county and district to understand the case distribution and location of symptom onset to illustrate the distribution of the cases across the country in comparison to the limited number of ETUs. After mapping the individual cases by location using a mixture of the village geospatial layer from the Government of Liberia, satellite imagery, and Open Street Map, manual digitizing of missing roads was conducted using satellite imagery.

To conduct the network distance analysis the road geospatial layer had to be prepared. Unfortunately, the existing roads layer was missing many smaller roads that were needed to connect the location of the cases to the main roads. To solve this issue, satellite imagery was used to manually digitize the many missing roads, see *Figure 4*, Image of Liberia Road Network and Distribution of Cases. The process involved zooming into the satellite imagery to a case that was not near a road, identifying from the imagery the pathway or road the lead from the village to another secondary road or the main road. These pathways were then manually traced using the imagery and snapping to the existing road network.

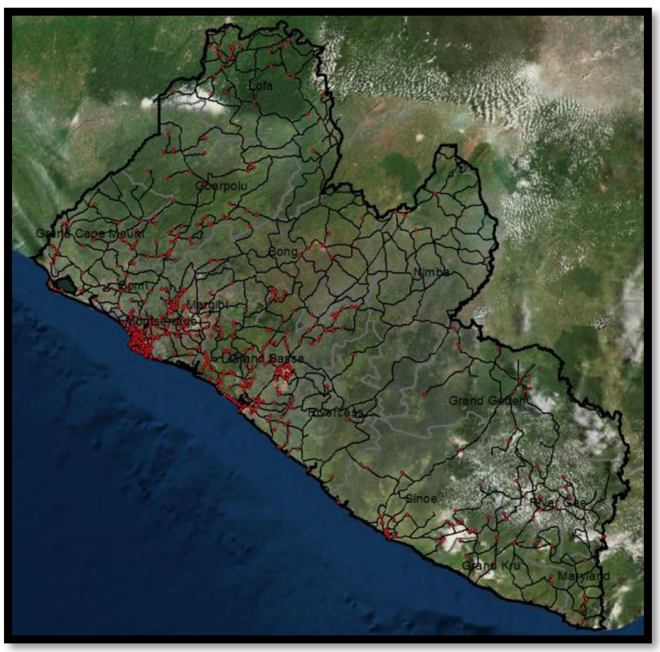

Figure 4. Image of Liberia road network and distribution of cases.

When the road network was complete, the next step was to create a geospatial network in ArcGIS Network Analyst. Using ArcCatalog, a new Network Dataset. This process entailed importing the newly created Liberia Road layer, setting the parameters, and the projection. The initial attempt returned a series of errors due to the poor quality of the original dataset where many roads were broken up into numerous segments and were not connected. Therefore, using the Error log file, each error was identified and fixed using the ArcGIS snapping tools and cleaning up the segments. Then the Network Dataset was created again and was successful. This dataset provided the transportation network to determine the distance from the Ebola case points and the ETUs.

The next step in the geospatial mapping was to map the transportation route from each village to the ETU, where they went for treatment, using the road layer. Distance was calculated ArcGIS Network Analyst and a Manhattan model (route taken to care). To create a Manhattan model, a transportation layer was created using a geometric network of foot paths, motorcycle paths, and roads (Lucklow et al., 2017). The process was a combination of using existing transportation layers and digitizing pathways between the location and the transportation route in the existing GIS layer. After the distances were calculated between the Ebola case locations and the ETU, GIS was used to analyze the distance by mortality using SPSS. The outcome of the statistics was an Odds Ratio by distance category.

To determine network distance using a Manhattan Model, the Network Analyst extension in ArcGIS was used and specifically the Origin-Destination (OD) Cost Matrix Tool. The OD Cost Matrix allows a geospatial point layer to be used as an Origin and this case it was the Ebola Cases by village location layer and another layer to be used as the Destination and in this case, it was the ETU layer. The only challenge with the OD Cost Matrix Tool is that it determines the distance from each Origin to all Destinations which is not what was needed. Therefore, a series of steps were needed to manipulate the tool. ArcGIS allows the user to select a group of records based on an attribute and, in this case, each case was selected by the ETU they received treatment. This was done for each ETU and a subset of the data was created by each ETU. For the Ebola Final Dataset of 916 cases, they went to fourteen ETUs, therefore fourteen subsets were created. Each of these subsets were then separately imported as Origins into the OD Cost Matrix Tool. When a subset for a specific ETU was imported, that ETU was selected in the ETU layer and then imported as a Destination into the OD Cost Matrix Tool. This allowed the network distance to be calculated for each case to the specific ETU. The output is a separate file which does not contain the other case data, therefore, an attribute join was then conducted to extract the distance, in meters, and place the information into the original Ebola case layer as a new Distance variable. This process was completed fourteen times to get the distance for all the cases. Another variable was then created for kilometers and calculated from the meters distance variable. The OD Cost Matrix Tool draws straight lines from the Origin (Ebola Case data) to the Destination (ETUs) for visualization purposes, but the distance stored in the table is network distance, Figure 5.

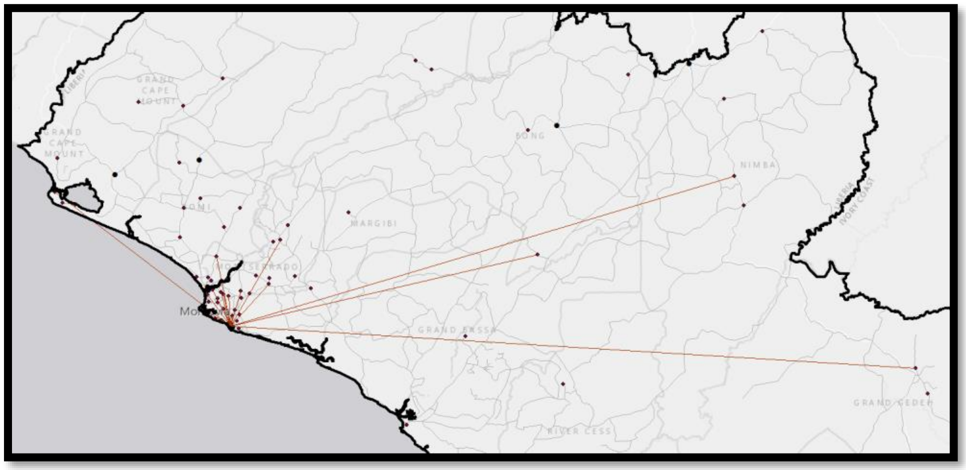

Figure 5. OD cost matrix tool output of distance calculation.

Statistical Analysis

The IBM SPSS Version 25 statistical software was used to calculate the odds of mortality, based on distance between the Ebola-affected location and the ETU, using an Odds Ratio. Using the distance data calculated in GIS and the outcome categorical data for mortality (dead or alive), the odds of mortality by distance was determined. The Chi-square test and binary logistic regression was used to test hypothesis. The Chi-square test provides the significance of observed differences between the groups of comparison, which was used to test hypothesis one (McHugh, 2013). A logistic regression was used to test hypothesis two to control for age and gender.

Research Questions and Hypotheses

Research Question 1: Is there an association between distance from Ebola-affected villages and ETUs and the mortality rates during the 2014 – 2016 Ebola outbreak in Liberia?

$H_0$1: There is no association between distance from Ebola-affected villages and ETUs and the mortality rates during the 2014 – 2016 Ebola outbreak in Liberia.

H_a1: There is an association between distance from Ebola-affected villages and ETUs and the mortality rates during the 2014 – 2016 Ebola outbreak in Liberia.

Research Question 2: Is there an association between the distance of Ebola-affected villages and the ETUs on mortality rates during the 2014 – 2016 Ebola outbreak in Liberia when controlling for sex and age?

$H_0$2: There is no association between distance from Ebola-affected villages and ETUs and the mortality rates when controlling for sex and age during the 2014 – 2016 Ebola outbreak in Liberia.

H_a2: There is an association between distance from Ebola-affected villages and ETUs and the mortality rates when controlling for sex and age during the 2014 – 2016 Ebola outbreak in Liberia.

Threats to Validity

The Ebola case data set was collected during the largest Ebola outbreak in history under less than perfect conditions. The data collection methods, processes, analysis, and reporting evolved as the outbreak continued. Because of this, the data is less than perfect, and some records are more complete than others. The village names, ETUs, epi classification, and final status are as accurate as possible, and the data was from the source therefore the data was as valid as possible. The data were cleaned with records removed that did not contain the required variables for analysis, in addition, statistical outliers were removed. The final cleaned data set contained 885 records.

External Validity

External validity describes the studies ability to be generalized to the larger population (Aschengrau & Seage III, 2008). The Ebola Case dataset is based on the specific outbreak in Liberia from 2014 to 2015 and the impact of distance on mortality. Distance and travel times can be calculated using various methods resulting in different results therefore the external validity will depend on the method used. Even though the exact results may vary, the statistical association between distance and mortality in a large outbreak is consistent.

Internal Validity

To maintain internal validity, a study must rule out bias, confounding, and random error (Aschengrau & Seage III, 2008). The Ebola case data examined the association between distance, based on the distance between villages with a case and the ETU, and mortality. This analysis modeled the travel distance and geography but is not exact. Each calculation was based on standard model design and did not introduce a bias affecting the outcome of the association. A potential confounder, such age and gender, could affect the outcome of the case. This is the purpose of second research question to control for them to compare the analysis with the first research question when they are not controlled. The resulting outcome of mortality by distance was calculated based on the variables in the Ebola case data with the results derived by chance thus ruling out random errors in the methodology.

Ethical Procedures

This study required a data use agreement with the MoH and the agreement states that the MoH had final say in the content of the study. An Institutional Review Board (IRB) application was submitted to Walden University IRB for approval but an MoH IRB was not required. Prior to analysis, the data was scrubbed of any personally identifiable information (PII) and the data use agreement expressed that the MoH had the right to withdraw from the agreement at any time. Access to the data was restricted to the researcher and if any other person or organization needed access, they would need to contact the Liberia MoH directly.

Summary

A quantitative retrospective study was conducted to examine the association between distance (between Ebola-affected villages and the ETUs) and mortality using the Ebola case data from Liberia. GIS was used to map the cases by village and then a Manhattan model were used to calculate distance. The Manhattan model calculates actual distances based on the path and road taken to travel from the village to the ETU. Using the most accurate distance calculations, mortality by distance intervals was calculated. Using these distances and the final status of the cases (dead or alive), odds ratios and chi-square were used to analyze the odds of mortality by distance and to test the hypotheses.

In section 3, the results of geospatial mapping and analysis are presented to illustrate the distribution of the cases, transportation networks, distance calculations, and statistical analysis. Additionally, the statistical analysis results will be presented showing the odds-ratio, chi-square, and logistic regression. This analysis will then be combined to present a complete picture of the results and the findings.

Section 3: Presentation of the Results and Findings

Introduction

The purpose of this study was to explore the association between distance from Ebola-affected villages and ETUs and the mortality rates by distance using the Ebola case data from the 2014 to 2016 Ebola outbreak in Liberia. By understanding the impact on mortality of distance from health care during mass outbreaks, guidelines can be developed for future responses. These guidelines can then be used during outbreak responses where additional treatment units are needed to serve the population at risk. Additionally, these guidelines can be used to guide health care delivery in non-emergency settings, such, as when new clinics are being constructed, for mobile health clinics, creating waiting houses for mothers, and a broader understanding of access to care barriers. The study results section includes the results of the chi-square test to answer Research Question 1 (RQ1) and the results of the logistics regression to answer Research Question 2 (RQ2).

My research questions are:

Research Question 1: Is there an association between distance from Ebola-affected villages and ETUs and the mortality rates during the 2014 – 2016 Ebola outbreak in Liberia?

$H_0$1: There is no association between distance from Ebola-affected villages and ETUs and the mortality rates during the 2014 – 2016 Ebola outbreak in Liberia.

H_a1: There is an association between distance from Ebola-affected villages and ETUs and the mortality rates during the 2014 – 2016 Ebola outbreak in Liberia.

Research Question 2: Is there an association between the distance of Ebola-affected villages and the ETUs on mortality rates during the 2014 – 2016 Ebola outbreak in Liberia when controlling for sex and age?

$H_0$2: There is no association between distance from Ebola-affected villages and ETUs and the mortality rates when controlling for sex and age during the 2014 – 2016 Ebola outbreak in Liberia.

H_a2: There is an association between distance from Ebola-affected villages and ETUs and the mortality rates when controlling for sex and age during the 2014 – 2016 Ebola outbreak in Liberia.

In Section 3, the geospatial and statistical analysis methods are described. Esri ArcGIS 10.6 with the Network Analyst extension were used for the geospatial analysis and IBM SPSS Version 25 was used for the statistical analysis. The results include descriptive statistics about the outbreak, mapping of the Ebola cases by administrative boundaries, mortality rates by distance from ETUs, and an odds ratio of the association between distance from Ebola-affected villages and ETUs. These results are described in the conclusion with tables and figures describing the Ebola outbreak.

Description of the Secondary Data

Liberia Ebola Outbreak Data from 2014 – 2016

The data was provided by the Liberia MoH which was collected from the beginning of the Ebola outbreak in 2014 until the final cases were resolved in 2016. The data are a line list of all Ebola cases from the outbreak but have been deidentified. The data were a combination of excel spreadsheets, EpiInfo Viral Hemorrhagic Fever (VHF) module, and the District Health Information System (DHIS2) Tracker module. At the beginning of the outbreak, data were collected in spreadsheets but early in the response, CDC introduced the EpiInfo VHF module (Nyenswah et al., 2016). EpiInfo was used until 2015 when the ministry transitioned to DHIS2 for online data entry. After the response ended, all of the data were combined into one Ebola case dataset, however, because three different methods where used over the two-year period, not all data for each case is complete, therefore, some data were excluded that were missing key required variables.

Study Sample

The Ebola case data was cleaned and analyzed in a series of steps to represent the outbreak a clearly as possible. Liberia has an administrative hierarchy that is village, district, county, and national. Each case has a county of report, county of symptom onset, district of symptom onset, and village of symptom onset but not all records have all this data, however, all the records have the county of report. The 8,440 records were used for the initial descriptive geospatial mapping of the case data by district and county.

Missing Data

The Ebola dataset consisted of 8,440 records but many of these case records where not complete. Some were missing the gender, age, location of symptom onset, ETU, epidemiological classification, and final status of the case. The first step was to remove any records that did not contain the county of symptom onset. This reduced number of records to 8,240 records, and these records were used to map the case data by county. The next step was to remove all records that were missing the district of symptom onset which left 8,107 records to map by district. Finally, all cases removed that did not have the location of symptom onset, ETU, age, gender, and final status. This left 916 records in the dataset for analysis.

Minimum Sample Size

When testing a hypothesis, two types of errors can be committed; Type I and Type II (Sullivan, 2016). A type I error occurs when the H_0 is incorrectly rejected and a Type II error is the H_0 is not rejected when it is false (Sullivan, 2016). To prevent these errors, a level of significance (α) and power ($1 - \beta$) are selected to reduce the chance of these errors (Sullivan, 2016). For hypothesis testing a level of significance should be $\alpha = 0.05$ and high power. To ensure high power, a sufficient sample size is required. The sample size can be calculated with level of significance, desired power, and the effect size (Sullivan, 2016). G*Power 3.1.9.2, can calculate the sample size by providing these variables. To determine the sample size for this study a level of significance of $\alpha = 0.05$, power of .090, and a moderate effect size of 0.3 was used. Based on this input, a minimum sample size required is 109 cases. The final data set of 916, prior to the removal of statistical outliers, provided an adequate sample for analysis. Even with records being removed, due to missing variables, the data set is large enough to maintain a statistical strength.

Data Analysis

Esri ArcGIS version 10.6 with the Network Analyst extension was used to map and determine the network distance between the village of symptom onset and the ETU. A series of maps were created to provide a visual illustration of the distribution of the cases at the county, district, and village levels. In addition, ArcGIS was used to map the locations of the ETU, roads, and travel to the ETU. IBM SPSS version 25 was used to conduct the Odds Ratio based on the output of the findings from ArcGIS. ArcGIS was used to determine the distance from the village of symptom onset and the ETU where the case went for treatment and then SPSS was used to conduct the statistical analysis. In addition to the odds ratio, SPSS was used to create descriptive statistics of the Ebola case data. The target population of this study are people dwelling in Ebola affected areas of the African continent.

Univariate Analysis

The initial dataset of 8,440 records was used to conduct the first step in the analysis for study. This provided descriptive statistics for the entire dataset prior to excluding records that were missing key variables. The following analysis used the final dataset that contains all the required variables.

Descriptive Statistics

The initial dataset had 8,440 records and, of which, 48.6% (4,098) were female and 51.4% (4,342) were male. The ages ranged from 1 to 115 years of age. There were eight records showing the age between 100 to 115, which may have been an error in the data collection. Twenty-two percent of the cases were 18 years of age and below, 21.9% were 19 – 30 years of age, 40.2% were 31 – 64, and 11.3% were 65 and above, (see Figure 6). There are 15 counties in Liberia and the frequency of cases from the various counties ranged from 22 in Maryland County and 4,997 in Montserrado County, which is the capital (see Figure 7). Ebola Cases by County and Figure 8. Ebola Cases by District shows a map of the cases by district. Not all the cases had the District of symptom onset therefore there are only 8,107 cases illustrated in Figure 8 and Figure 9. Map of Case locations and ETUs shows the distribution of the cases by location and the location of the ETUs.

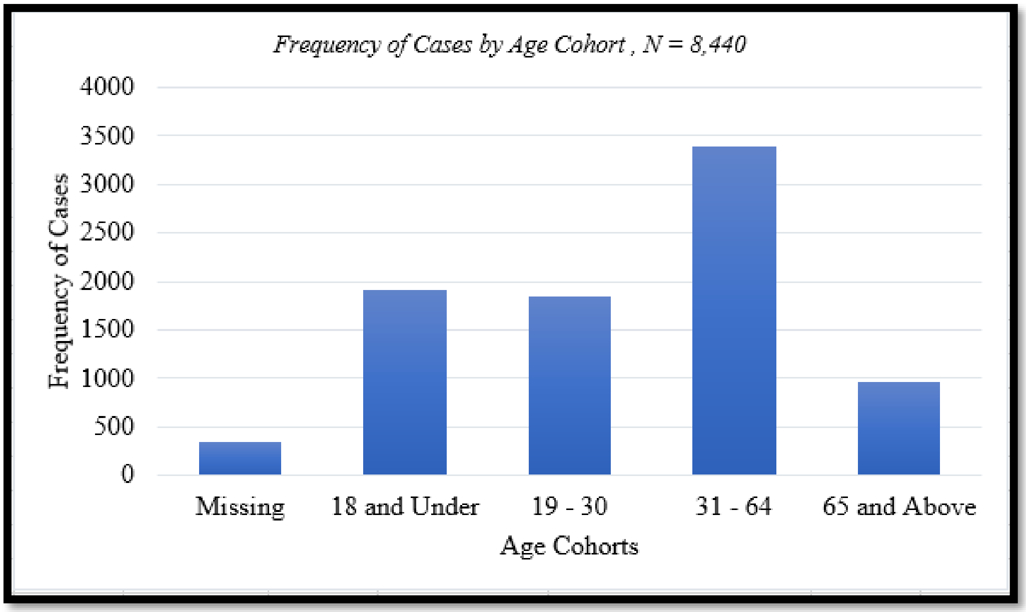

Figure 6. Cases by age cohort.

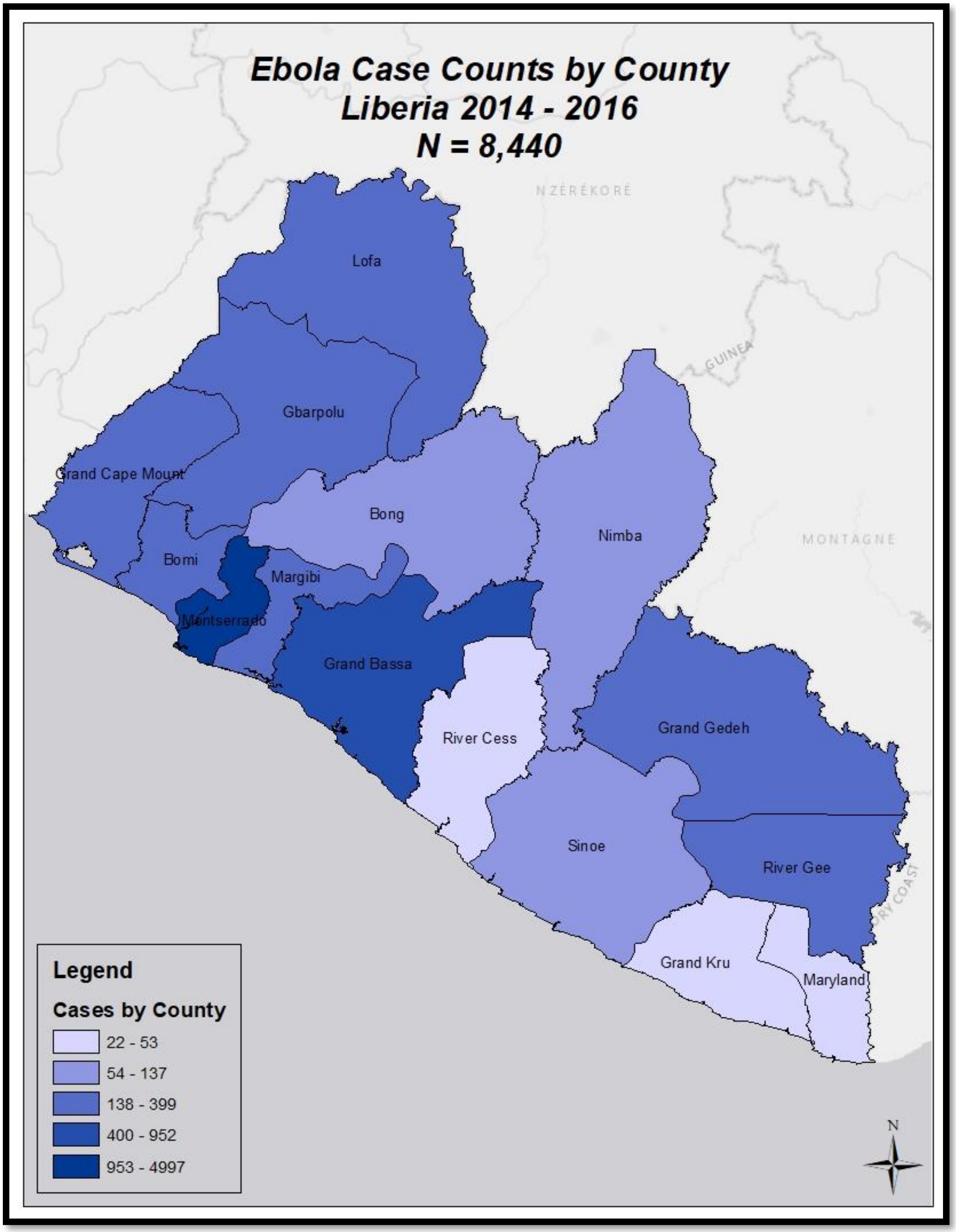

Figure 7. Ebola cases by county.

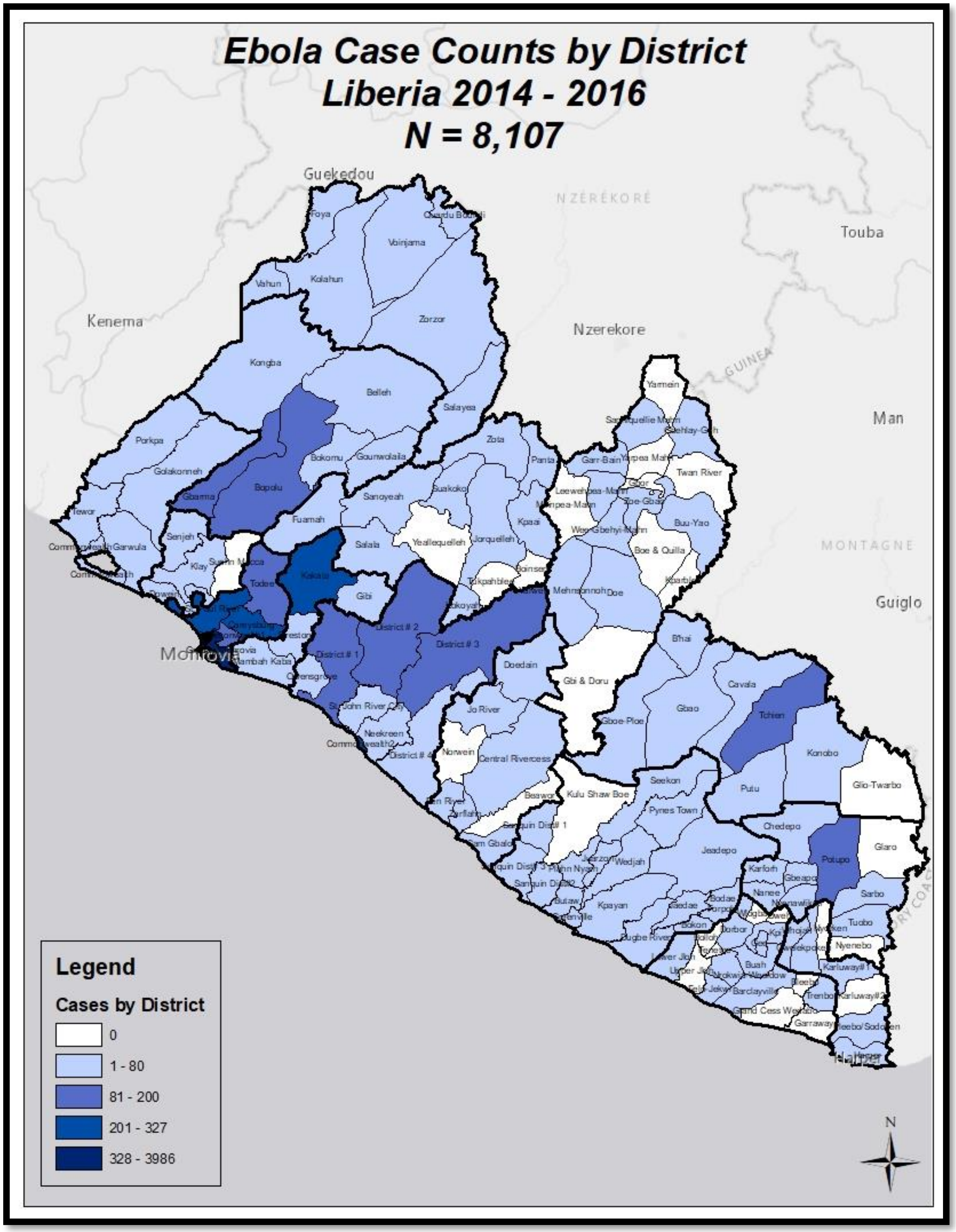

Figure 8. Ebola cases by district.

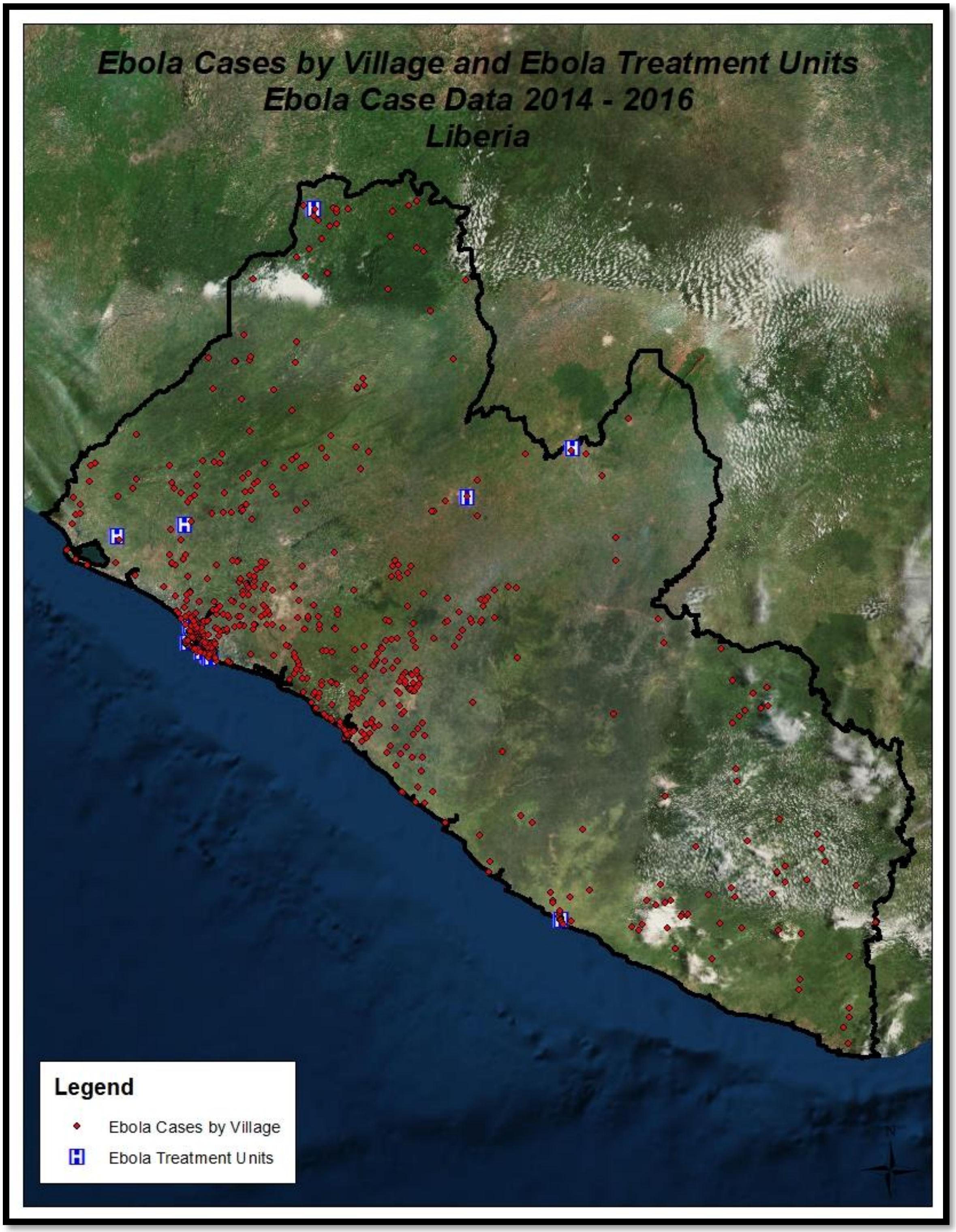

Figure 9. Map of case locations and ETUs.

To conduct the analysis to answer the research questions, the data required gender, age, ETU, Final Status, and distance. When all the records were excluded that did not contain these variables, the final data consisted of N = 916 records. These records where complete but univariate analysis was conducted to test for outliers. Using univariate analysis, outliers were identified for age and distance, see Figure 10. Histogram of Age shows a histogram of the data by age illustrating a slight positive skew in the data and potential outliers of higher ages.

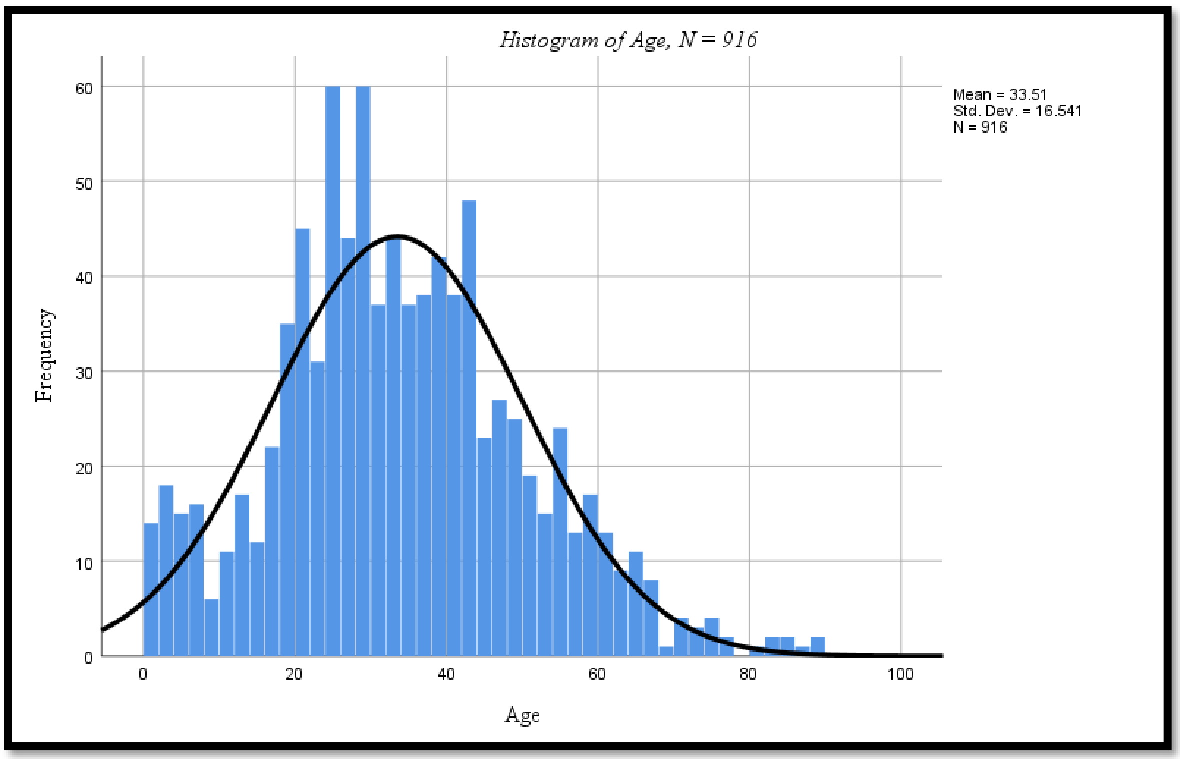

Figure 10. Histogram of age.

In addition to a histogram by age, a Boxplot was used to identify the outliers by age. The boxplot identified 13 cases as outliers which were removed from the data resulting in N = 903, Figure 11.

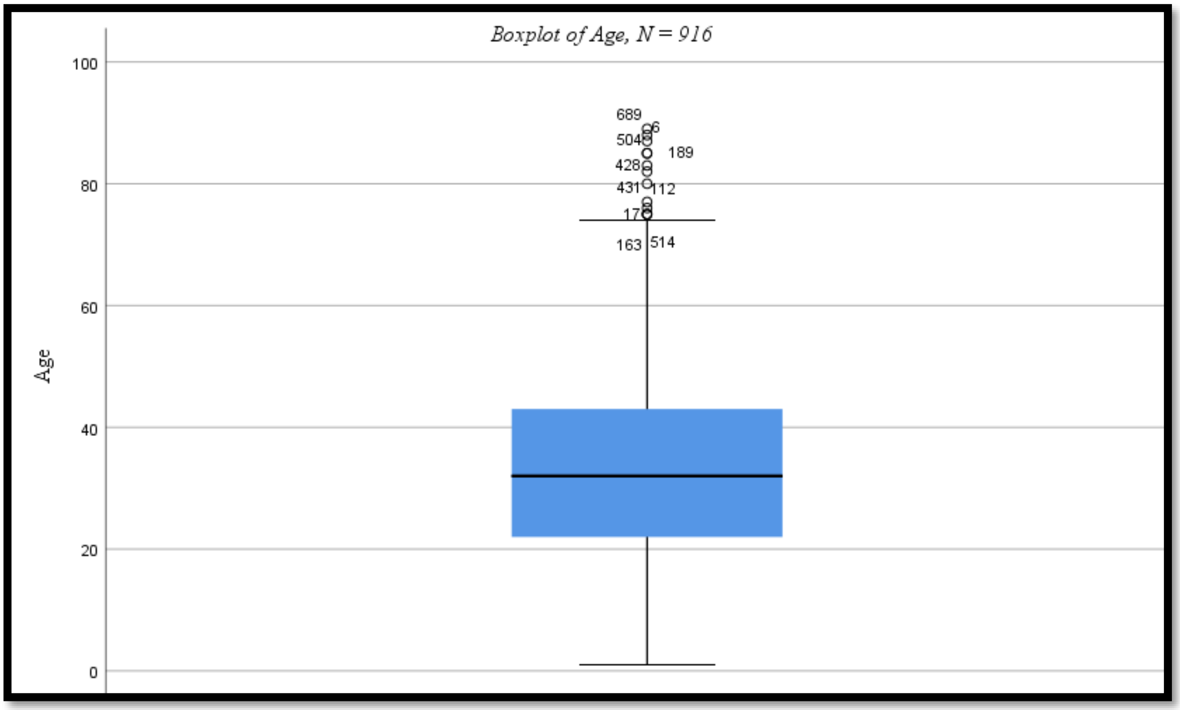

Figure 11. Boxplot of age.

After the removal of the outliers for age, the descriptive analysis was conducted again. The histogram still shows a slight positive skew in the data, but it is a more normal curve, Figure 12. The boxplot shows there are no more outliers for age, Figure 13.

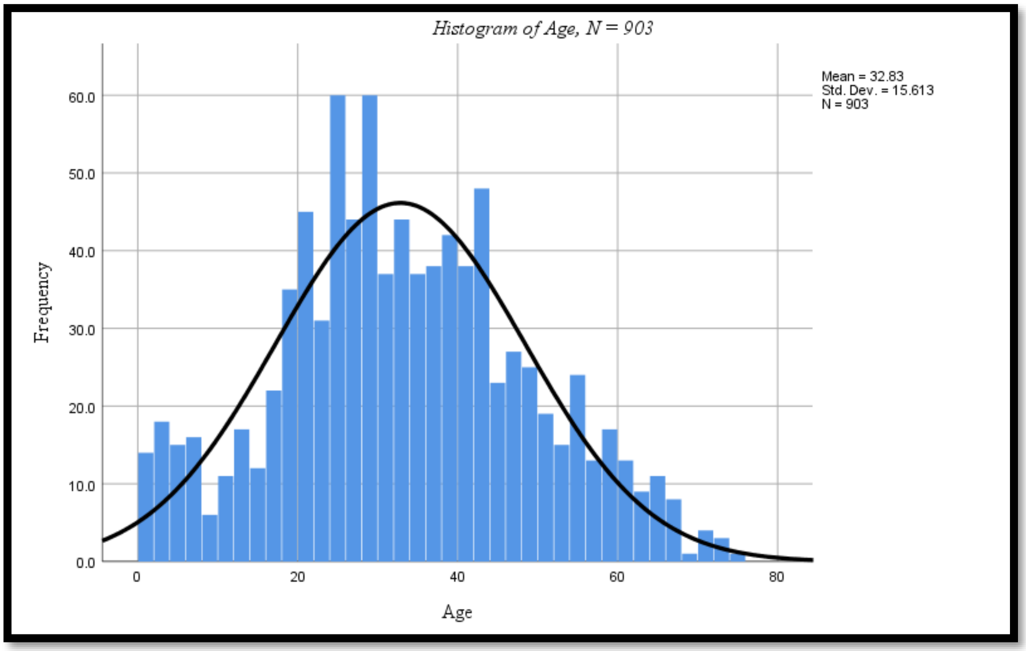

Figure 12. Histogram of age with outliers removed.

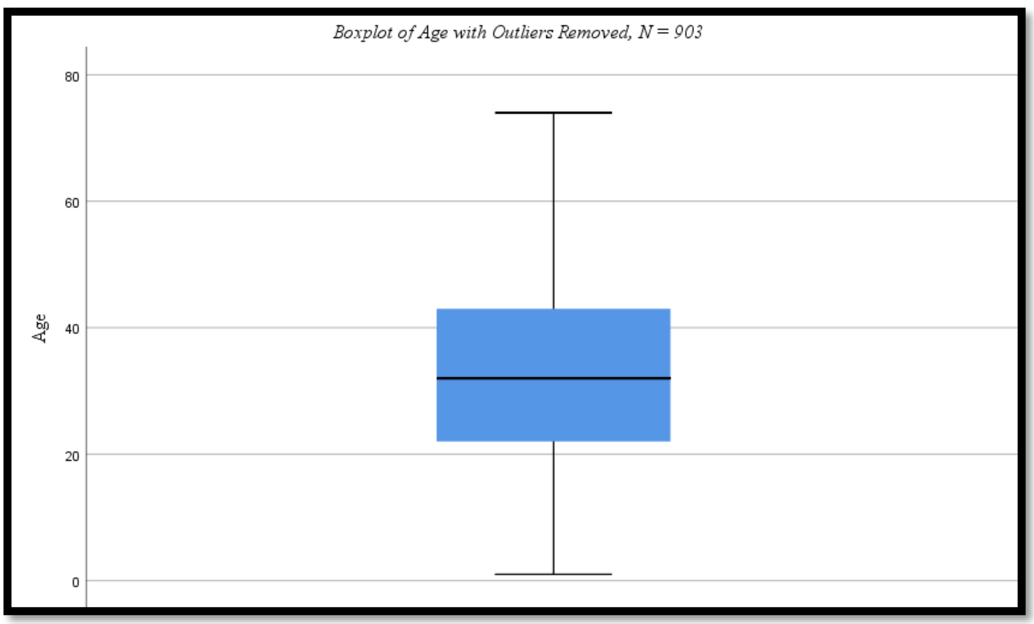

Figure 13. Boxplot of age with outliers removed.

With the age outliers removed for the overall dataset, the next step was to see if there were any outliers by Final Status. A boxplot was created for Alive and Dead by age which identified two outliers for the Alive category, Figure 14. Boxplot of Age by Final Status. These two outliers were removed, therefore, reducing the dataset to N = 901, Figure 15. Boxplot of Age by Final Status with Outliers Removed.

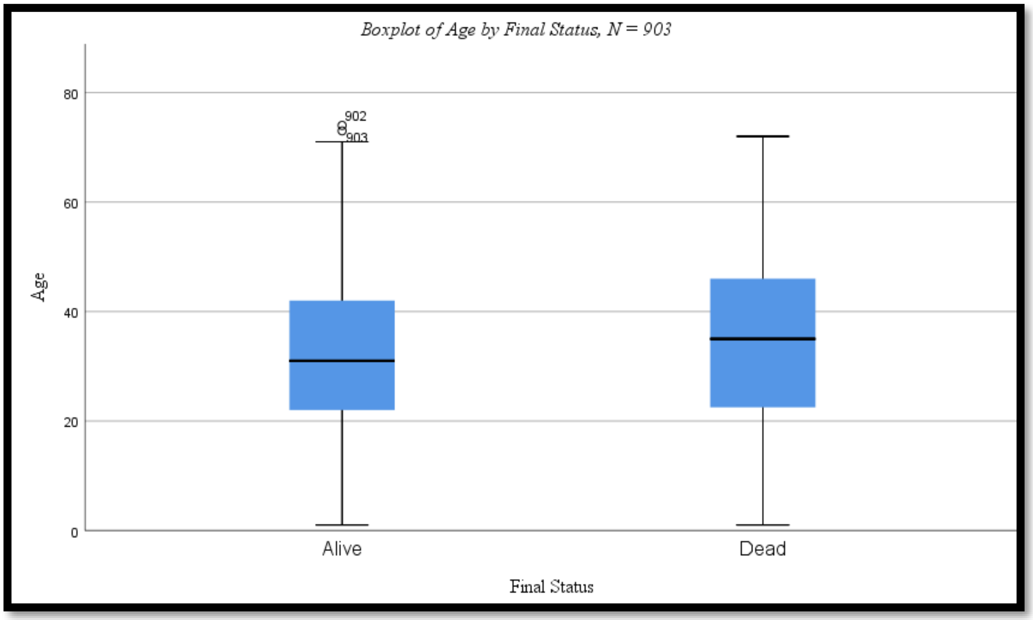

Figure 14. Boxplot of age by final status.

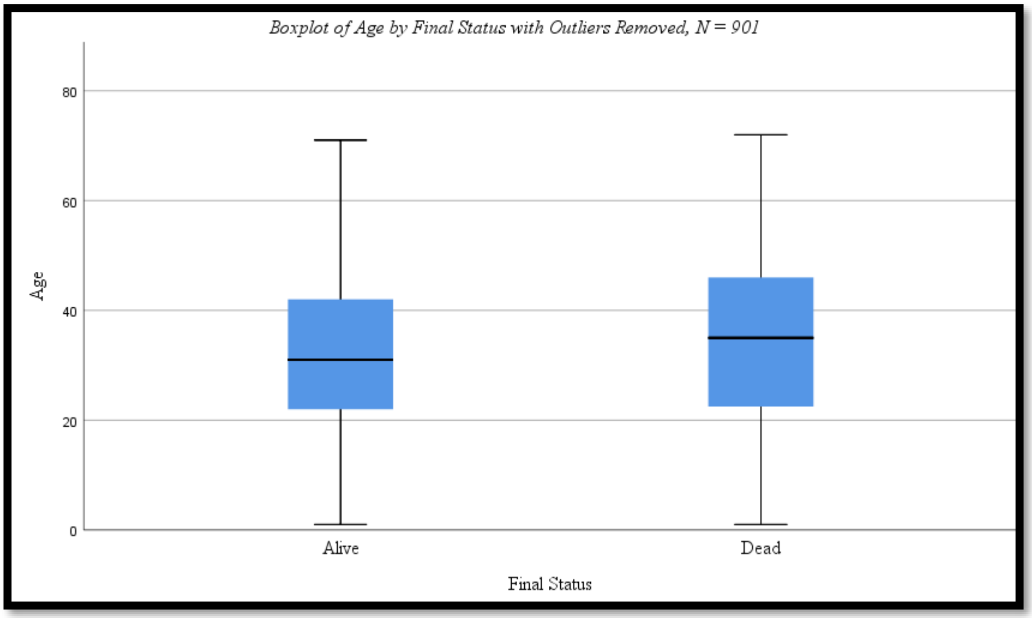

Figure 15. Boxplot of age by final status with outliers removed.

The next variable to examine for outliers was distance. A histogram was created for distance to check the distribution of the data which showed a positive skew in the data, Figure 16. Histogram of Distance from Village to ETU. A boxplot was then created to see if outliers were creating the positive skew, Figure 17. Boxplot of Distance from Village to ETU. The boxplot identified 16 outliers in the data due to distance.

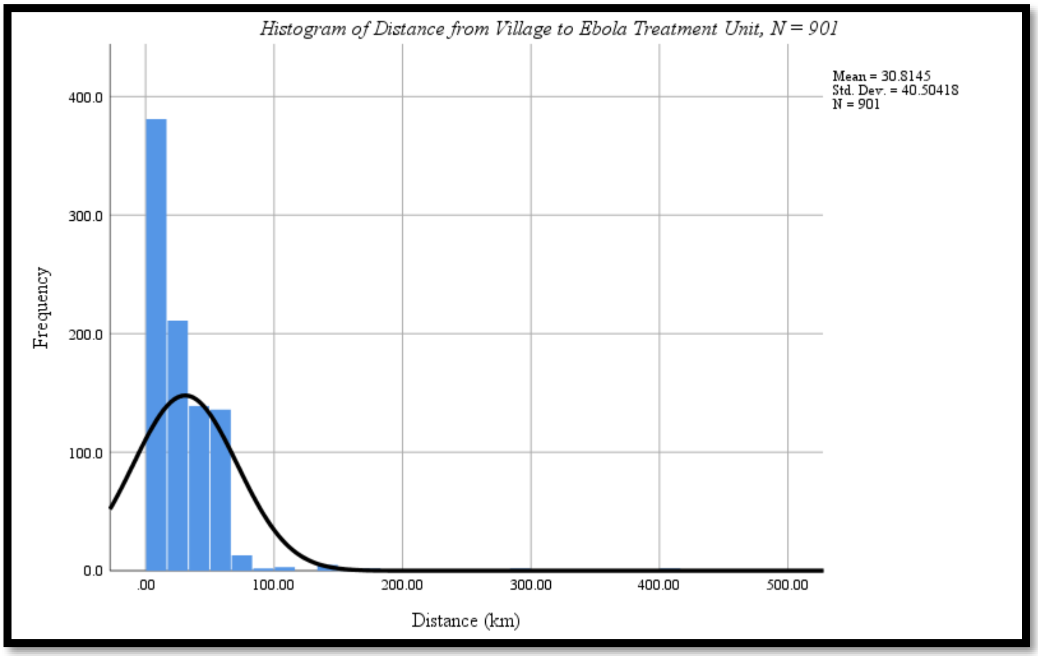

Figure 16. Histogram of distance from village to ETU.

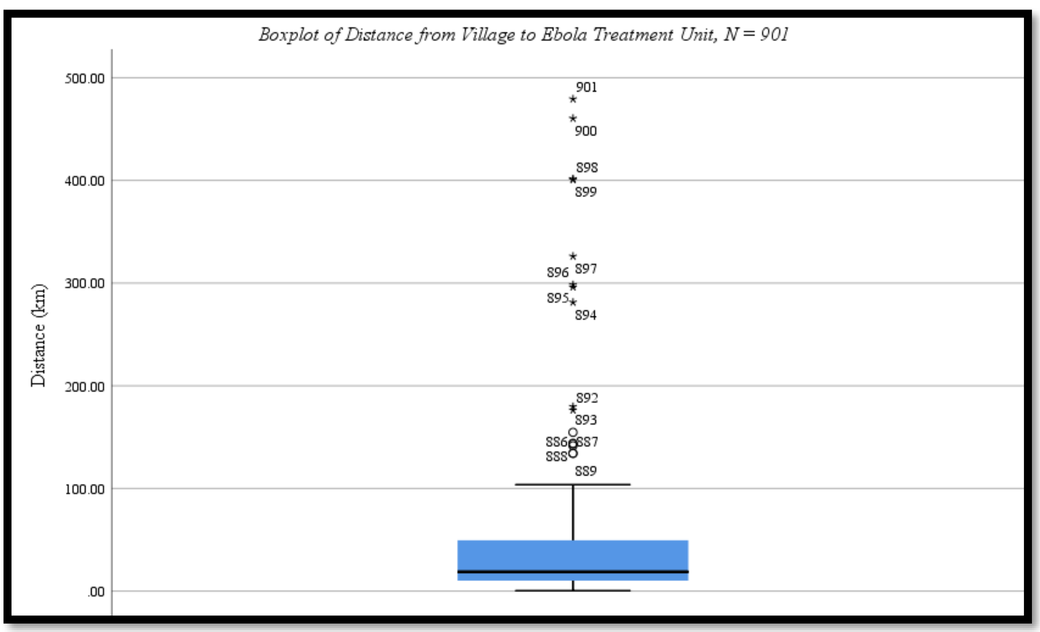

Figure 17. Boxplot of distance from village to ETU.

The 16 outliers due to distance were removed from the data reduce the dataset to N = 885. Another histogram was created to examine the distribution of the data which showed the data was still positively skewed but less than before the outliers were removed, Figure 18. Another boxplot was created to ensure no more outliers existed in the overall dataset due to distance, Figure 19.

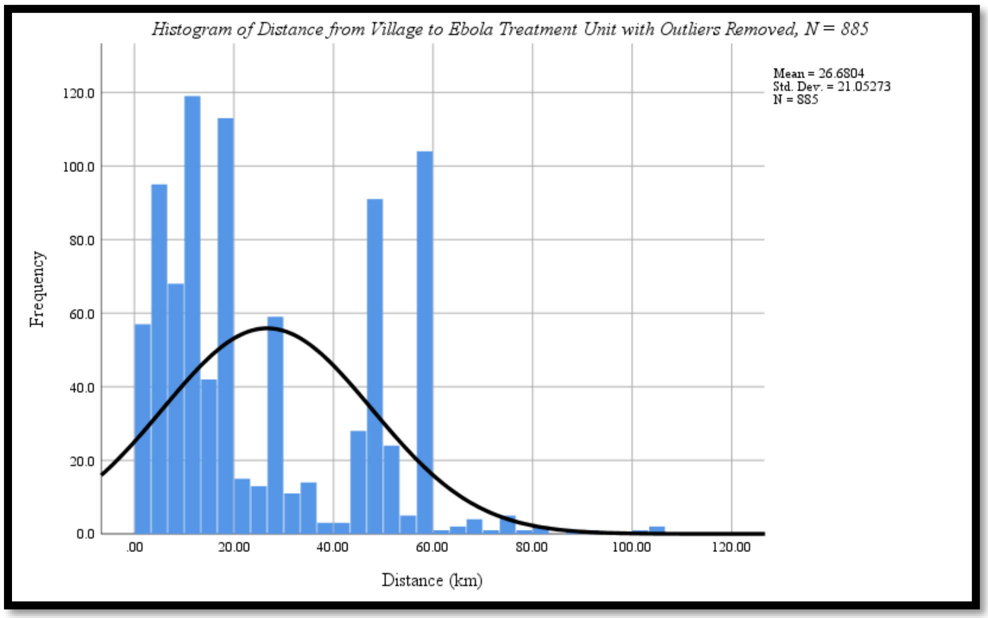

Figure 18. Histogram of distance from village to ETU with outliers removed.

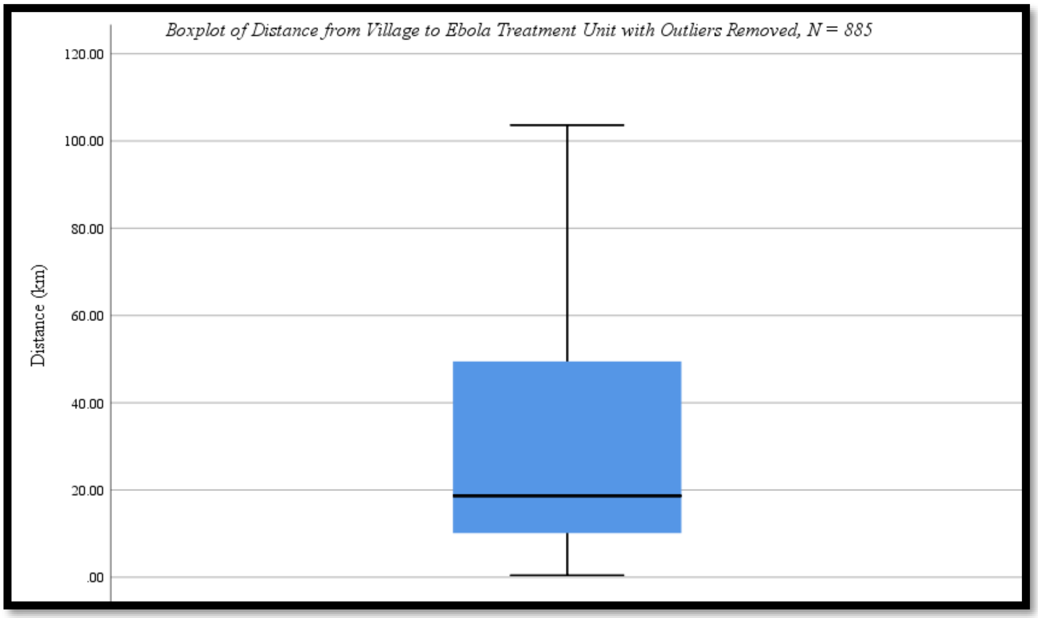

Figure 19. Boxplot of distance from village to ETU with outliers removed.

The last step to check for outliers was to analyze outliers by distance and Final Status. Using a boxplot again to check for outliers for distance by Alive or Dead, the data did not show any outliers, Figure 20. With all of the outliers removed, the final dataset is *N* = 885.

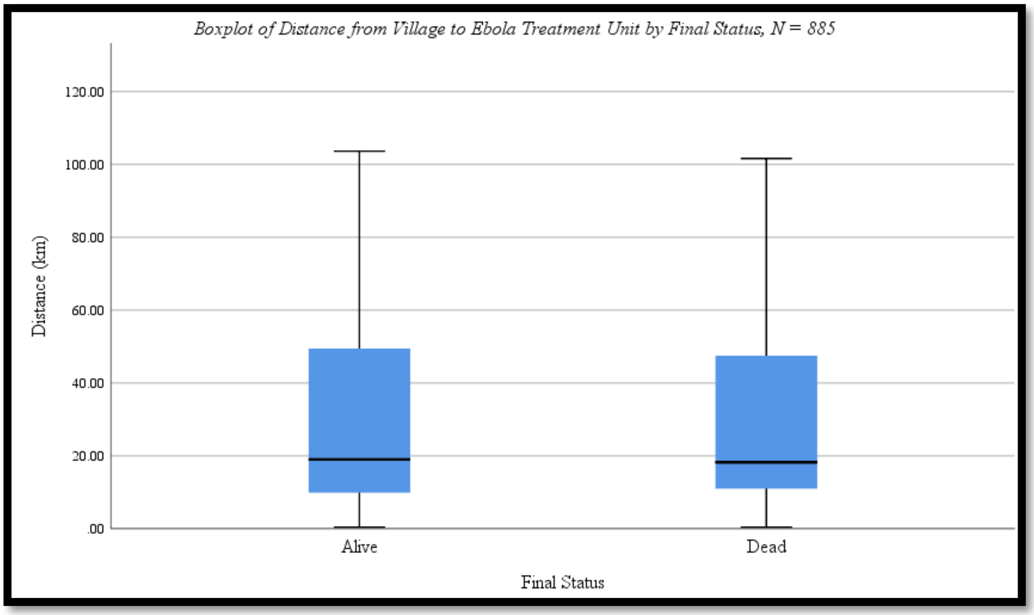

Figure 20. Boxplot of distance from village to ETU by final status.

With all the outliers removed, the descriptive analysis of the final dataset, *N* = 885, was conducted. The mean age for the final dataset was 32.66 and the mean distance from the village to the ETU was 26.68 km, Table 2. Descriptive Statistics of Final Dataset. Females represented 41.5% (367) of the cases while males represented 58.5% (518) of the cases and the Final Status, 81.4% (720) were alive and 18.6% (165) were dead, Table 3. Frequency and Percent by Gender and Final Status.

Table 2

Descriptive Statistics of Final Dataset

	Age	Distance
N	885	885
Mean	32.66	26.68
Median	32	18.66
Mode	28	49.47

Table 3

Frequency and Percent by Gender and Final Status

		Frequency	Percentage
Gender	Female	367	41.50%
	Male	518	58.50%
Final Status	Alive	720	81.40%
	Dead	165	18.60%

Table 4 shows the frequency and percentage of Ebola cases for each ETU that was used in analysis. They ranged from 1 case in Foya to 234 at ELWA 2. The largest percent of cases were at two ETUs; ELWA 2 and ELWA 3, which accounted for 50.6% of the cases, Figure 21 and Figure 22. Histogram of Distance from Village to ETU for Final Dataset, illustrates the distribution of distance from village to ETU for the final dataset.

Table 4

Ebola Cases by ETU

ETU	Frequency	Percentage
Foya	1	0.10%
Unity	2	0.20%
Suakoko	3	0.30%
Sinje	5	0.60%
Ganta	8	0.90%
Tubmansburg	16	1.80%
Island Clinic	18	2.00%
Greenville	27	3.10%
MMU	5	6.00%
China	82	9.30%
MoD	134	15.10%
German	136	15.40%
ELWA 3	214	24.20%
ELWA 2	234	26.40%
Total	885	100.00%

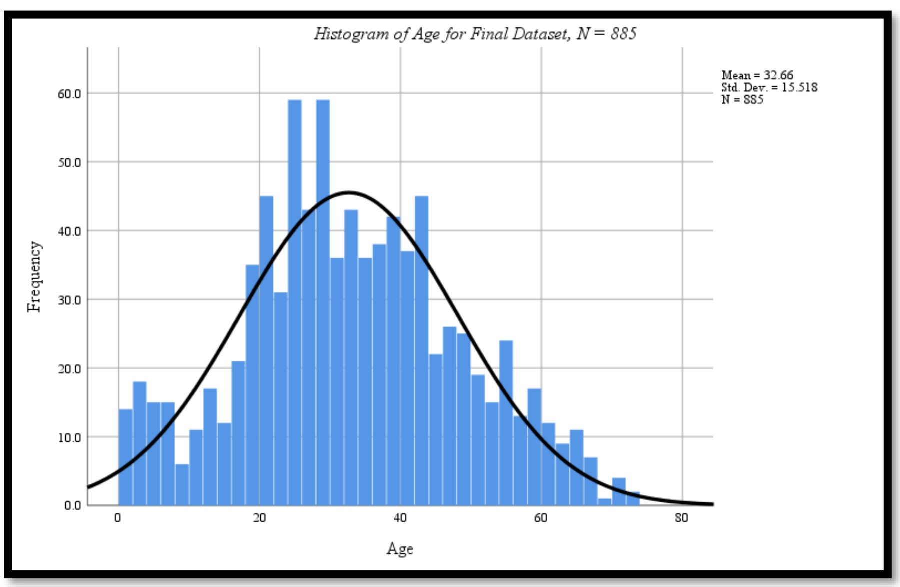

Figure 21. Histogram of age for final dataset.

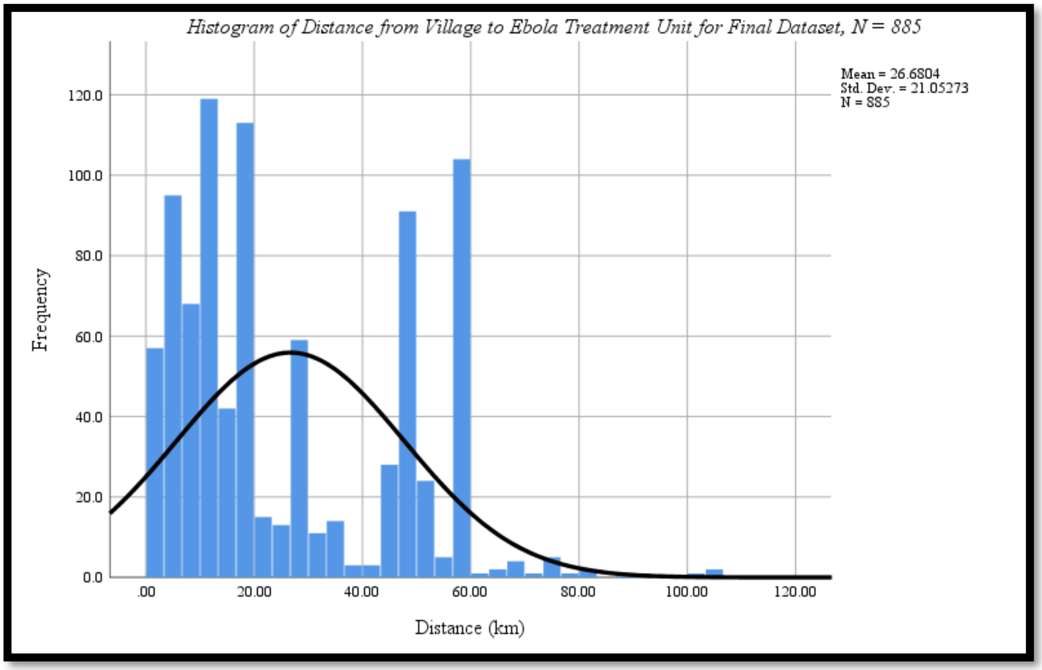

Figure 22. Histogram of distance from village to ETU for final dataset.

In addition to the descriptive statistics, the final dataset was mapped by district and county to see the geographic spread of the final data, Figure 23. Ebola Cases by County shows the final cases by county, see Figure 24. Ebola Cases by District shows the final cases by district.

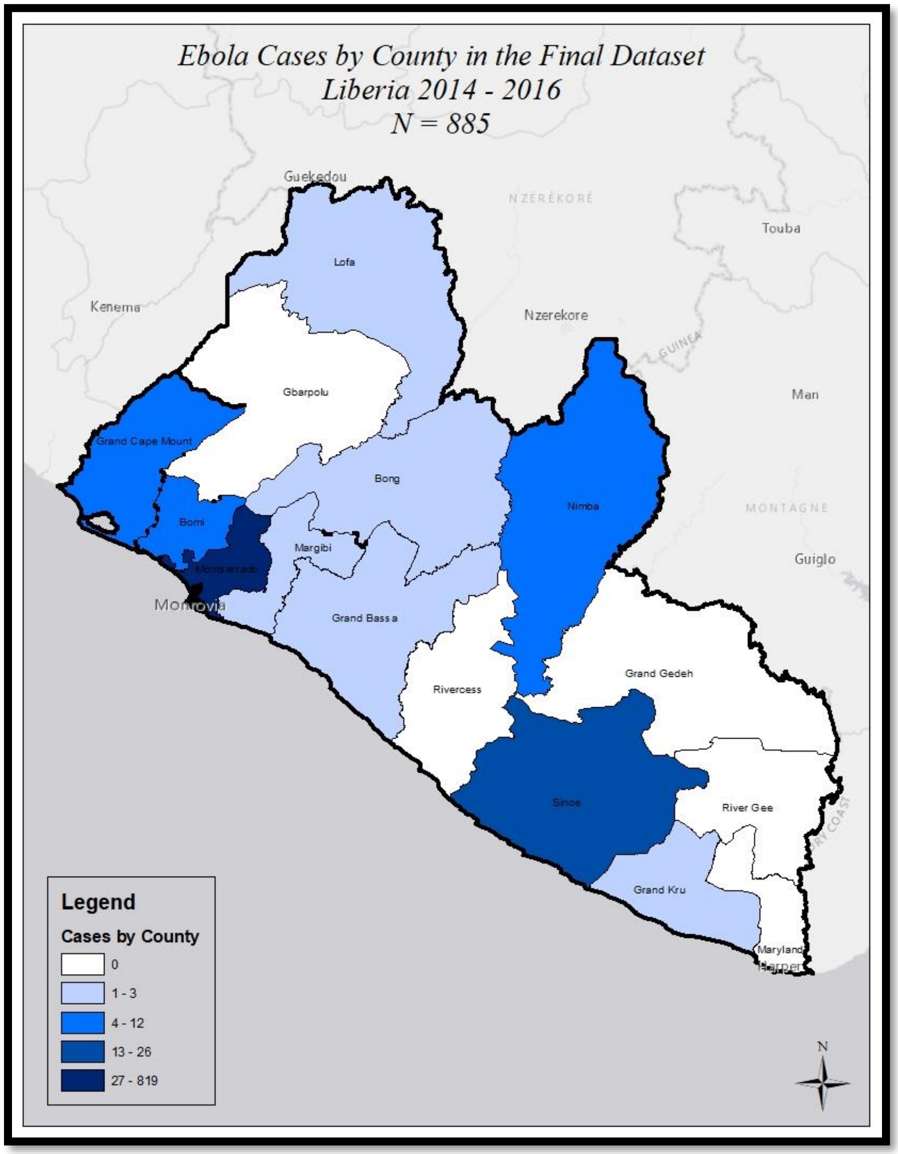

Figure 23. Ebola cases by county.

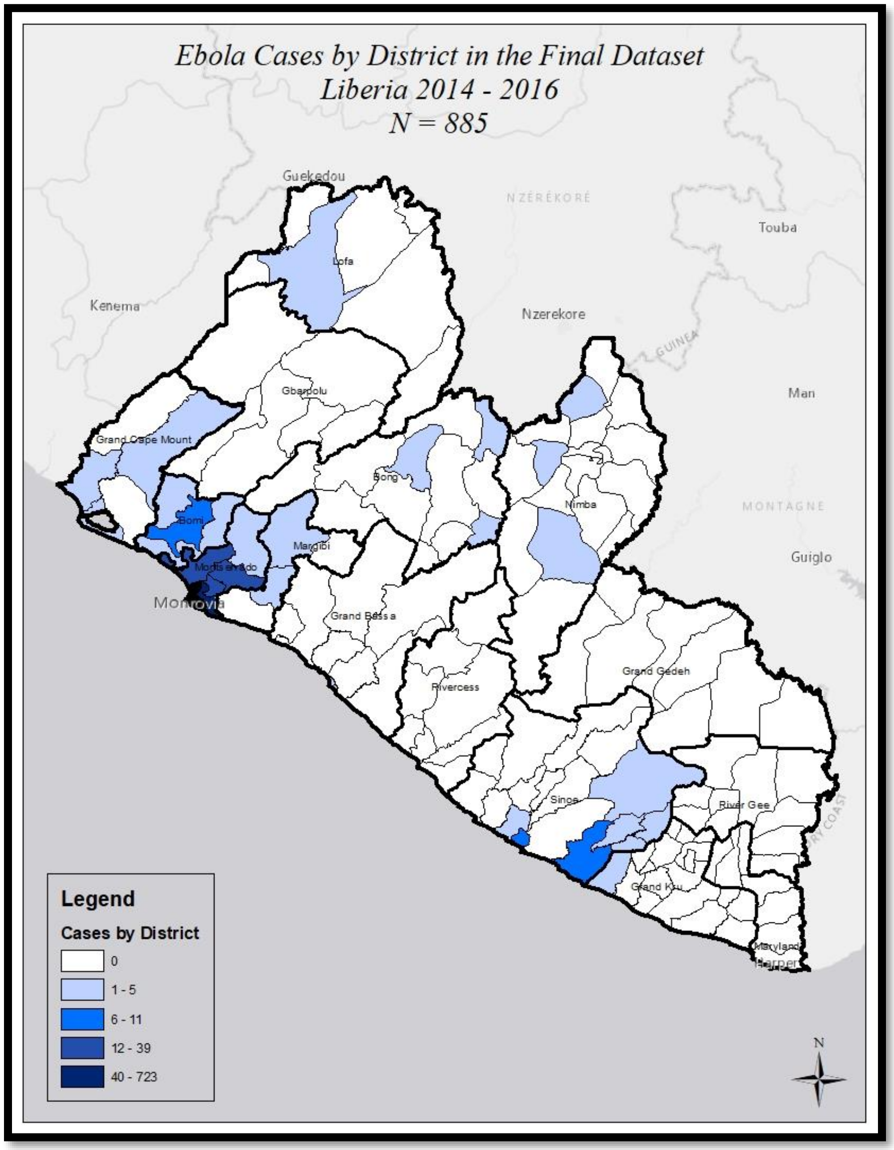

Figure 24. Ebola cases by district.

Analysis Results

Categorical Distance Variable

The association between distance and mortality can be analyzed with a continuous distance variable and/or a categorical distance variable. Since the dataset's limited size would present challenges to using a continuous variable for a logistic regression, a categorical distance variable was used. To test the association between categories of distance and mortality, SPSS was used to group the distance into three similar group sizes to ensure sample size was large enough to maintain statistical power. The first category was from 0.45 to 11.7 km, the second category was from 11.7 to 38, and the third was from 38.1 to 104, Table 5 *Frequency and Percentage of Cases by Distance Category*.

Table 5

Frequency and Percentage of Cases by Distance Category

Distance	Count	Percentage
0 km - 11.7 km	295	33.30%
11.71 - 38 km	314	35.50%
38.1 - 104 km	276	31.20%

Bivariate Analysis

To understand the data better, bivariate analysis was conducted on the final dataset that had an $N = 885$.

Cross-Tabulation and Chi-Square Analysis

A cross-tabulation was used to examine the relationship between gender and final status (alive or dead). Of the 885 cases, 367 (41.5%) were female and of those, 21.5% (79) died. The males had 518 (58.5%) of the cases and 16.6% (86) died. The cross-tabulation expected 68 female deaths and 97 male death but the observed numbers for both were less than expected (see Table 6). The results of the Pearson's r (Chi Square) indicate there is not an association between gender and final status since the Chi Square test was not statistically significant. The p-value is greater than 0.05 ($p > 0.05$) at the alpha ($\alpha = 0.05$) significance level, Table 7.

Table 6

Crosstabulation of Gender and Final Status

			Final Status		
			Alive	Dead	Total
Gender	Female	Count	288	79	367
		Expected Count	298.6	68.4	367
		Percent Female	78.50%	21.50%	100%
	Male	Count	432	86	518
		Expected Count	421.4	96.6	518
		Percent Male	83.40%	16.60%	100%
Total		Count	720	165	885
		Percent of Total	81.40%	18.60%	100%

Table 7

Chi-Square for Gender and Final Status

	Value	df	Significance
Pearson Chi-Square	3.433[a]	1	0.064
N of Valid Cases	885		

Note. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 68.42.

A cross-tabulation and chi-square were then used to test the association between distance and final status. The cross-tabulation provided the observed and expected deaths per distance category. Category 1 had 26.1% (43) of the deaths, category 2 had 44.2% (73), and category 3 had 29.7% (49), (see Table 8). The chi-square showed there is a statistically significant association between the distance categories and the final status (dead or alive), X^2 (2, N = 885) = 7.751, p < .05, therefore the null hypothesis in Research Question 1 is rejected. Even though the Chi-Square test showed an association between the distance categories and final status, the effect size was very weak, ϕ = .094, p < .05, (see Table 9).

Table 8

Crosstabulation of Distance Category and Final Status

			Final Status		
			Alive	Dead	Total
Distance Category	0 to 11.7 km	Count	252	43	295
		Expected Count	240	55	295
		Percent in 0 to 11.7 km	85.4%	14.6%	100.0%
		Percent in Final Status	35.0%	26.1%	33.3%
		Percent of Total	28.5%	4.9%	33.3%
	11.71 to 38 km	Count	241	73	314
		Expected Count	255	59	314
		Percent in 11.71 to 38 km	76.8%	23.2%	100.0%
		Percent in Final Status	33.5%	44.2%	35.5%
		Percent of Total	27.2%	8.2%	35.5%
	38.1 to 104 km	Count	227	49	276
		Expected Count	224	52	276
		Percent in 38.1 to 104 km	82.2%	17.8%	100.0%
		Percent in Final Status	31.5%	29.7%	31.2%
		Percent of Total	25.6%	5.5%	31.2%
Total		Count	720	165	885
		Expected Count	720	165	885
		Percent in Distance Category	81.4%	18.6%	100.0%
		Percent in Final Status	100.0%	100.0%	100.0%
		Percent of Total	81.4%	18.6%	100.0%

Table 9
Chi-Square and Phi for Distance Category and Final Status

	Value	*df*	Significance
Pearson Chi-Square	7.751[a]	2	0.021
N of Valid Cases	885		
Phi	0.094		0.021

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 51.46.

Binary Logistic Regression

Binary Logistic Regression between Distance and Final Status

To understand the association between distance and final status better, a binary logistic regression was used to determine the Odds Ratio associated with distance. In this regression, a Binary Logistic Regression was used between distance and the categorical variable representing distance in three categories (0 to 11.7 km, 11.71 to 38 km, and 38.1 to 104 km). Final Status was the dependent variable and distance from the village to the ETU was the independent variable. Because a categorical variable for distanced was used, the categories needed to be set to show the reference category and the first category was used. This means that each category would reference the shortest distance category to determine the odds ratio. The logistic regression was statistically significant, X^2 (2, N = 885) = 7.711, $p < .05$, Table 10, Omnibus Tests of Model Coefficients, however the model only explained 1.4% (Nagelkerke R^2, Table 11) of the variance in death but correctly classified 81.4% of the cases. The cases in the second distance category, 11.71 to 38 kilometers from an ETU, are 1.775 times more likely to die than those in the first category, (see Table 12).

Table 10

Omnibus Tests of Model Coefficients

		Chi-Square	df	Significance
Step 1	Step	7.711	2	0.021
	Block	7.711	2	0.021
	Model	7.711	2	0.021

Table 11

Nagelkerke R Square Test

Step	- 2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	843.695[a]	0.009	0.014

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Table 12

Logistic Regression of Distance Categories and Final Status

				95% C.I.	
	Distance Category	Sig.	Odds Ratio	Lower	Upper
Step 1a	0 to 11.7 km				
	11.71 to 38 km	0.007	1.775	1.171	2.691
	38.1 to 104 km	0.303	1.265	0.809	1.978
	Constant	0.000	0.171		

a. Variable entered in step 1: Distance Category

Logistic Regression between Distance and Final Status while Controlling for Gender and Age

Research Question 1 examined the association between distance and mortality without controlling for potential confounders, but Research Question 2 examines the association while controlling for gender and age. To control for gender and age, a logistic regression was used but adding both as additional independent variables. Gender was set as a categorical variable with male as the reference. In addition, the distance categories were set as a categorical variable with the Category 1, the shortest distance category, as the reference category. The logistic regression was statistically significant, X^2 (4, N = 885) = 14.394, p < .05, (see Table 13, Omnibus Tests of Model Coefficients), however the model only explained 2.6% (Negelkerke R^2, see Table 14, Nagelkerke R Square Test) of the variance in death, which is higher than only using distance as the predictor variable, and correctly classified 81.4% of the cases, Table 17. The Hosmer and Lemeshow Test suggests that the model is a good fit for the data, p = .304 (> .05), (see Table 15, Hosmer and Lemeshow Test). The cases in the second distance category, 11.7 to 38 kilometers from an ETU, are 1.778 times more likely to die than those in the first category, (see Table 16, Logistic Regression for Distance Categories and Final Status Controlling for Age and Gender). Males had a 1.4 times lower risk of death due to EVD, (see Table 16, Logistic Regression for Distance Categories and Final Status Controlling for Age and Gender).

Table 13

Omnibus Tests of Model Coefficients

		Chi-Square	df	Significance
Step 1	Step	14.394	4	0.006
	Block	14.394	4	0.006
	Model	14.394	4	0.006

Table 14

Nagelkerke R Square Test

Step	- 2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	837.013[a]	0.016	0.026

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than.001.

Table 15

Hosmer and Lemeshow Test

Step	Chi-Square	df	Significance
1	9.476	8	0.304

Table 16

Logistic Regression for Distance Categories and Final Status Controlling for Age and Gender

		Sig.	Odds Ratio	95% C.I. Lower	Upper
Step 1a	0 to 11.7 km				
	11.71 to 38 km (1)	**0.007**	**1.778**	1.171	2.700
	38.1 to 104 km (2)	0.348	1.240	0.791	1.942
	Age	0.077	1.010	0.999	1.021
	Gender (1)	**0.044**	**1.423**	1.009	2.007
	Constant	0.000	0.105		

a. Variable entered in step 1: Distance Category, Age, Gender

Summary

Two research questions were hypothesized; is there an association between distance, from the village of symptom onset for Ebola and the Ebola Treatment Unit where they went for treatment, and mortality, and the second research questions is there the same association while controlling for gender and age. To test these two hypotheses, the Ebola case data for Liberia from the 2014 – 2016 Ebola outbreak was used.

The Ebola case dataset was provided by the Liberia Ministry of Health and consisted of 8,440 case records. However, due to missing key variables including, the ETU where they went for treatment, Final Status (alive or dead), gender, age, and location of symptom onset, the dataset was reduced to 916 records. This final dataset was then mapped and analyzed for network distance from the location of symptom onset to the ETU where they were treated. That distance was stored as an additional variable in the case dataset.

Through descriptive and univariate analysis, outliers were identified in both the age and distance variables and removed, therefore, reducing the final dataset to 885 case records. These 885 records were then used to answer the research questions. A Chi-Square Test of Association was used to test the hypothesis for Research Question 1. The Chi-Square showed there is a statistically significant association between the distance categories and the final status (dead or alive), X^2 (2, N = 885) = 7.751, $p < .05$, therefore the null hypothesis in Research Question 1 is rejected. Even though the Chi-Square test showed an association between the distance categories and final status, the effect size was very weak, $\phi = .094$, $p < .05$.

To test the hypothesis of Research Question 2, I used a Logistic Regression to test the association between distance and mortality while controlling for gender and age. The logistic regression, using the categorical variable for distance was statistically significant, X^2 (4, N = 885) = 14.394, $p < .05$, however the model only explained 2.6% (Negelkerke R^2) of the variance in death, which is higher than only using distance as the predictor variable, and correctly classified 81.4% of the cases. The Hosmer and Lemeshow Test suggests that the model is a good fit for the data, $p = 9.476$ (> .05). The cases in the second distance category, 11.7 to 38 kilometers from an ETU, are 1.778 times more likely to die than those in the first category. Males had a 1.4 times lower risk of death due to EVD.

Section 3 provided the methods, processes, and analysis for the study. It went through the data cleaning process, geospatial data creation, data organization, and analysis steps. The results indicated that the null hypothesis for both Research Question

1 and Research Question 2 were rejected based on statistically significant tests of association. Section 4 will provide discuss the "so what" aspect of the study. It will provide an interpretation of the findings, limitations, recommendations, and implications for both professional practice and social change.

Section 4: Application to Professional Practice and Implications for Social Change

Introduction

The purpose of this retroactive quantitative study was to examine the association between distance to care and mortality during a large outbreak. Numerous studies have been conducted to examine distance to care and maternal mortality or similar related health outcomes, but none focus on emergency situations where treatment units are constructed specifically for an outbreak or humanitarian event. To conduct this study, the Ebola case data from the 2014 to 2016 outbreak in Liberia was used. Research Question 1 examined the association between distance (calculated from the village of symptom onset and the ETU where they were treated) and mortality. Research Question 2 examined the same association but controlled for gender and age. The null hypothesis for both research questions was rejected due to statistically significant tests for association therefore there is an association between distance and mortality both when gender and age are controlled or not.

Interpretation of the Findings

The findings showed statistically significant results to reject the null hypothesis between distance and mortality for the Ebola outbreak from 2014 to 2016 both when controlling for age and gender or not. In addition, the results showed that males had a 1.4 times lower risk of death during the EVD outbreak. This finding shows that the location of treatment units, during large outbreaks, can impact the population served and potentially increase mortality if those distances are too far from the population. In addition, the findings show that women had a higher risk of dying during the outbreak.

Populations were 1.8 times less at risk of death during the Ebola outbreak if they were within 12 kilometers of an ETU. This is an important finding when planning treatment unit locations during large outbreaks or when planning health clinic locations for ongoing care for the general population.

This study aligns to previous studies that have focused on access to care challenges and maternal mortality or health care for women and children. Kenny et al, (2015) conducted a similar study in Liberia but focused on how distance impacted health care for maternal and child health care. Their findings determined that living closer to a facility increased the odds of getting antenatal care during pregnancy, delivering in a health facility, and receiving postnatal care. Hanson et al. (2015) conducted a similar study in Tanzania examining the association of distance and maternal mortality. A similar association was found. In this study, there was a strong association between distance and maternal mortality. When comparing mortality, the maternal mortality was 111 deaths per 100,000 live births for women who lived less than 5 km from the health facility compared to 422 deaths per 100,000 live births for women who lived more than 35 km (Hanson et al., 2015). An additional study in Tanzania examined the association between distance and childhood mortality and found that children who lived more than 5 km from the health care facility had a 17% increase risk of mortality compared to those who less than 5 km (Kadobero et al., 2012).

This presents significant challenges to countries, like Liberia, where the road network is minimal and, because many roads are dirt, they are not passable during the rainy season which is from July to November. This will impact travel time even if the distance is the same, therefore impacting health outcomes. Because the association between distance from health care has repeatedly shown negative health outcomes, it needs be planned for at all levels of the health system and during emergency-related events.

This study was grounded on a social ecological model within a GIS framework. To understand the impact of distance to care on mortality, it was important to understand the spatial relationships between location (individual) of symptom onset, district (community) of symptom onset, and county (policy) of symptom onset. Since the SEM is a multi-level model that frames society at these levels, it is an important guide for planning and response (UNICEF, 2016). These connections of how policy is derived at a county or national level will impact how health care is delivered at the district or community level. Those decisions will impact the mortality during large scale events, such as the Ebola Outbreak. The simple decision at a national level to place X number of treatment units in certain locations will negatively impact part of the affected population.

The GIS Framework pulled the data together in a spatial context, which is a key concept for decision making. Planning and decision making cannot occur without an understanding of the spatial connections of society and the barriers that increase risks to effective health care. By using a GIS framework to geolocate the cases by village or specific location, this allowed a spatial understanding of the distribution of the cases but also allowed additional analysis to be conducted to understand associations between geography and mortality. Creating a network distance, as opposed to straight line distance, provides a more accurate travel distance and allows that distance to be examined more closely for associations. Those associations are connected to the policy level decisions that are made at the national or county levels that ultimately impact the individuals.

The GIS Framework has matured more in countries with more resources but is more challenging in countries with less. In countries with more resources, conducting spatial analysis is relatively straightforward. Addresses can be geocoded to conduct spatial and network analysis which can then guide the next level of analysis. But in resource poor settings, this isn't possible. To conduct the geospatial analysis for this study, many aspects had to be created from scratch using imagery. Manually, digitizing the road or pathway networks, village locations, and ETU locations. The GIS Framework is flexible enough to allow this but the researcher, who wants to study in resource poor settings, must be willing to take this extra step. For emergency response GIS, there should be an expectation that these layers will not be available and the researcher, responder, or administer should expect and plan for imagery acquisition and analysis.

Limitations of the Study

There are several limitations to this study. The first is the sample size. Even though it is well beyond the requirement for statistical power, the 885 case records do not provide the same analysis possibilities as if the full 8,440 case records were used. This is the nature of a database during an outbreak though. The database evolved and changed during the outbreak as it grew and spread. In addition, the database was created using paper case investigation forms that were sent from the investigation teams in the field. Those forms were quickly filled out, some more than others, and forwarded sometimes hand to hand from across the country to the Ministry of Health in Monrovia. The data was then entered into EpiInfo and then DHIS2 but there are always potential errors with data entry. So, this is a sample from an imperfect dataset that had case data with many missing values. The decision to only use the case records with all of the key variables was a decision to use a smaller dataset but a complete one.

The next decision was to use only the records with the ETU where they received treatment. This decision reduced the case records from 5,365 to 916. The 5,365 records had all of the variables except most were missing the ETU. One discussion was to use the 916 case records, since they have the ETU, and then use GIS to identify the closest ETU for the rest and use it. This would allow a sample size of 5,365. The issue with this method is that there is no guarantee that the case went to that ETU. Many people didn't go to any ETU. Also, ETUs were built throughout the response so just because 14 were used in this analysis, early in the response, there were only 2; one in Lofa and one in Monrovia (Nyenswah et al., 2016). Therefore, the process would have needed to break the cases down by symptom onset date, identify the ETUs that were open in that time frame, and then determine the closet one but this assumes that they went to one at all. These were too many assumptions; therefore, the decision was made to stay with the 916 since the ETU of treatment was part of the record and the rest of the variables were complete. This decision to limit the dataset probably affected the third distance category which was the farthest from the ETUs. In the overall dataset, there were many cases with a final status of death but, just by chance, the reduced dataset had more alive than dead cases and therefore probably does not represent the overall dataset very well. This miss representation probably led to the lack of significance in the farthest distance category.

Another limitation is the network distance. There are two limitations to the network distance; creating the network and actual travel. When creating the road network, the Liberia road geospatial layer was not complete but to calculate network distance every case needed to be on a road. Over 300 cases were not on roads, therefore, satellite imagery was used to zoom into the case, identify the pathway, and then conduct onscreen digitizing of the pathway to connect it to the road in the Liberia road layer. This was a best guess process which assumed that they traveled the pathway that was digitized, which was no guarantee. Since Liberia's rainy season is one of the heaviest in the world, many of these pathways were washed out (Sieh, 2017). Just because the pathway was visible on the date the satellite imagery was taken, doesn't mean it was usable on the day the case had symptoms and traveled to the ETU. So, there were two potential issues, creating the correct pathway that was used and assuming they took the shortest route.

Network distance provided the distance from the location of symptom onset to the ETU where they received treatment. This distance was then used to examine the association between distance and mortality. A better approach would have been travel time versus travel distance. In developed countries, these are relatively easily calculated since the road layer contains impedances such as speed limits, stop lights, one way, etc., but the Liberian road layer did not have any of these variables. To add them would have been very challenging. When people travel in Liberia, it may be a mixture of walking, riding a motorcycle, and then riding in a car. They may walk on a pathway to a larger pathway where they can wait for a motorcycle to take them to a primary road where they will wait for a car. This process has many variables involved which include weather, time of day, cars or motorcycles traveling, and during the Ebola outbreak many people did not drive. Also, if the person was very sick, someone may be required to carry them to the nearest road. Because of all the caveats and potential errors in creating a travel time, the decision to use travel distance instead, with the understanding of the limitations, was made early on.

Ideally, a continuous distance variable would have been used to provide an odds ratio per kilometer of distance traveled to an ETU but due to the dataset size, the statistical significance would have been challenging. Because of this, SPSS was used to calculate three distance categories to be used for the logistic analysis. The categorical distance logistic regression showed there is a higher risk of death due to distance and provides guidance for planning treatment units during an outbreak response but a per kilometer distance odds ratio would have been better.

Lastly, the logistic model used may not completely describe the data. The Cox & Snell R Square and Nagelkerke R Square were 0.016 and 0.026, respectively. This indicates that the model only explains between 1.6% and 2.6% of the variation in survival. In addition, there were additional variables in the dataset set, such as symptom onset date and when the person arrived at the ETU which would lend itself to a more robust analysis if used. The problem is that many of the records were missing this data so it would have reduced the dataset even further to only use the records that had all of these variables.

Recommendations

The findings of this study showed a statistically significant association between distance and mortality during the Ebola outbreak in Liberia. However, the findings were based on a limited subset of Ebola case dataset. The first recommendation would be for partners in Liberia to work with the Ministry of Health to go through all the past datasets and paper forms to create a more complete dataset from the Ebola outbreak. This was the largest Ebola outbreak in history and this data can provide additional guidance to the larger public health community.

Based on the findings in this study and other similar studies, there is an association between distance and negative health outcomes. This knowledge should be part of the planning process at all levels of public health. At the national level, mapping out the population and understanding that health facilities need to be close to those populations, in all seasons. Understanding that in a population that has very limited motor vehicle ownership, getting to care may involve walking and when someone is very sick, they may not go or when they do, it is too late (LISGIS, 2013).

It is very important for governments to have complete geospatial data layers for key aspects of their countries. This includes administrative boundaries, health facilities locations, villages, pathway, roads, rivers, etc. Using these layers for geospatial analysis and decision making is critical. Understanding how someone gets to care is as important as if they get to care. The maternal mortality rate in Liberia is one of the highest in the world (1,072 per 100,000 live births) and they often find that most deaths occur in the health facility (UNWomen, 2017). Women who live in remote areas may choose to deliver at home, due to distance, but when complications arise, it is very challenging to get to care therefore they may have died in the health facility but the real issue was that they waited at home until there were complications and then had to travel a long distance to the health facility.

In addition to the broader access to care issue, the findings from this study indicate that during large outbreaks and when treatment units are being constructed, an understanding of the population distribution and using geospatial tools to guide the location is critical. It is common to geolocate outbreak cases by home or village and use this information for spot maps or cluster analysis, but the next step is to systematically use these tools to geolocate treatment units based on case locations versus available property or where people “think” is a good location.

Implications for Professional Practice and Social Change

The implications for professional practice and social change are intertwined. Using the knowledge that there is an association between distance and negative health outcomes, public health practitioners and leadership should always consider it when creating health policies, determining funding levels, examining access to care issues, transportation decisions, infrastructure decisions, and analyzing health patterns. To fully understand when there are different morbidity or mortality rates across a population; geography, distance, and travel need to be considered. In countries like Liberia, weather also becomes a factor to plan around. If there is a health facility within 5 km of each populated area but, during rainy season, the creek that separates them becomes an impassable river, that facility is not usable for that population. The focus is often on having adequate health care professionals in the facilities but, if the people cannot get there or cannot get there in a timely manner, they will not be used as much as they should. This knowledge that should arm public health professionals to convince political leadership about the importance of connecting communities to health facilities, using community workers to advocate the need for access care, and expectations from citizens that the government will provide those services. Thinking spatially and strategically will change public health practice and, with that, create social change around expectations of the health system.

Conclusion

The Ebola outbreak in West Africa was first detected in Guinea in December 2013 and then it spread to Sierra Leone and Liberia ultimately resulting in over 28,000 suspected, probable, and confirmed cases of the Ebola virus (WHO, 2014; CDC, 2016). The first case was detected in the northern county of Lofa, which is along the border of Guinea and then spread to Monrovia (WHO, 2015). After entering Liberia, the virus soon spread throughout the country resulting in over 10,000 suspected, probable, and confirmed cases and almost 5,000 deaths (CDC, 2016). The virus spread slowly so there were, initially, only two ETUs constructed; one in Lofa and one in Monrovia (Nyenswah et al., 2016). If the outbreak were contained in those two areas, those two ETUs may have been enough, but it soon spread to other parts of the country as people traveled back to their home villages who were infected. Because the virus spread across the country, eventually in all 15 counties, people were required to travel long distances to an ETU. As the outbreak continued, more ETUs were built but cases started leveling off when many ETUs became available, so their construction was halted (Beaubien, 2014).

Distance to care is a well-documented physical barrier to positive health outcomes (Kenny et al., 2015; Hanson et al., 2015; Kadobera et al., 2012). These studies have focused on on-going health care challenges such as maternal and child mortality. What hasn't been studied is the association of distance and mortality during a large outbreak, such as the Ebola outbreak in Liberia. The intent of this study was to fill the knowledge gap about the association of distance and morality during a large event, such as an outbreak or humanitarian event. This knowledge could then be used to guide public health practitioners on the appropriate distances to treatment units to reduce the risk of mortality.

The two research questions centered around the association of distance from the village to the ETU and mortality with one controlling for gender and age and the other not. Both research questions were found to have a statistically significant association between distance and mortality. The study provided another example of how distance to care can increase the risk of mortality during an emergency.

This information can be used by public health professionals to improve public health responses, guide planning during a response, guide treatment unit location decisions, and illustrate the need for public health professionals to think spatially when planning for public health services. In addition to guiding public health professionals, this leads to social change on the way the public health and the larger public think about resource allocation, planning, expectations for services, and the importance of integrating geospatial science into the public health workflow.

References

Abramowitz, S. A., McLean, K. E., McKune, S. L., Bardosh, K. L., Fallah, M., Monger, J., … Omidian, P. A. (2015). Community-centered responses to Ebola in urban Liberia: The view from below. *PLoS Neglected Tropical Diseases*, 9(4), e0003706. doi:10.1371/journal.pntd.0003706

Aschengrau, A. & Seage III, G. (2008). *Essentials of epidemiology in public health*. Sudbury, Massachusetts: Jones and Bartlett Publishers

Beaubien, J. (2014). U.S. military response to Ebola gains momentum in Liberia. *National Public Radio*. Retrieved from https://www.npr.org/sections/goatsandsoda/2014/11/05/361796044/u-s-military-response-to-ebola-gains-momentum-in-liberia

Bell, D. M., Damon, I., Bedrosian, S. R., Johnson, V. R., McQuiston, J. H., & O'Connor, J. (2016). CDC's response to the 2014–2016 Ebola epidemic—West Africa and United States. *MMWR Suppl*, 65, 1-106. Retrieved from https://www.cdc.gov/mmwr/volumes/65/su/pdfs/su6503.pdf

Bowers, A. S. (2015). Sewer overflows and the vector mosquito proximity to human West Nile Virus infections (Doctoral dissertation, *Walden University*). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3736887)

Carter, S. E., O'Reilly, M., Walden, V., Frith-Powell, J., Umar Kargbo, A., & Niederberger, E. (2017). Barriers and enablers to treatment-seeking behavior and causes of high-risk practices in Ebola: A case study from Sierra Leone. *Journal of Health Communication*, 22(sup1), 31-38.

doi:10.1080/10810730.2016.1222034

Centers for Disease Control and Prevention [CDC]. (2014). *Ebola (Ebola Virus Disease) signs and symptoms.* Retrieved from https://www.cdc.gov/vhf/ebola/symptoms/index.html

Centers for Disease Control and Prevention [CDC]. (2014). *Preparing healthcare workers to work in Ebola treatment units (ETUs) in Africa: Ebola treatment unit.* Retrieved from https://www.cdc.gov/vhf/ebola/hcp/safety-training-course/assets/E-Lecture_ETU.pdf

Centers for Disease Control and Prevention [CDC]. (2015). *Ebola (Ebola Virus Disease) transmission.* Retrieved from https://www.cdc.gov/vhf/ebola/transmission/index.html

Centers for Disease Control and Prevention [CDC]. (2016). *Outbreaks Chronology: Ebola Virus Disease.* Retrieved from https://www.cdc.gov/vhf/ebola/outbreaks/history/chronology.html#modalIdString_outbreaks

Centers for Disease Control and Prevention [CDC]. (2017). *2014 Ebola outbreak in West Africa - reported cases graphs.* Retrieved from https://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/cumulative-cases-graphs.html

Cretney, R. (2014). Resilience for Whom? Emerging Critical Geographies of Socio-ecological Resilience. *Geography Compass*, 8(9), 627-640. doi:10.1111/gec3.12154

Environmental Systems Research Institute [Esri]. (2017). *The power of mapping*. Retrieved from http://www.esri.com/what-is-gis

Eros, E. (2016). Field mapping at scale in West Africa. *Missing Maps*. Retrieved from http://www.missingmaps.org/blog/2016/04/25/west-africa-mapping-hub-launch/

Fang, L. Q., Yang, Y., Jiang, J. F., Yao, H. W., Kargbo, D., Li, X. L., ... & Liu, K. (2016). Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone. *Proceedings of the National Academy of Sciences*, 113(16), 4488-4493. doi:10.1073/pnas.1518587113

Figueroa, M. E. (2017). A theory-based socioecological model of communication and behavior for the containment of the Ebola epidemic in Liberia. *Journal of Health Communication*, 22(sup1), 5-9. doi:10.1080/10810730.2016.1231725

Gartland, M. G., Taryor, V. D., Norman, A. M., & Vermund, S. H. (2012). Access to facility delivery and caesarean section in north-central Liberia: a cross-sectional community-based study. *BMJ open*, 2(6), e001602.

Gleason, B. L., Foster, S., Wilt, G. E., Miles, B., Lewis, B., Cauthen, K., ... & Kamara, S. I. (2017). Geospatial analysis of household spread of Ebola virus in a quarantined village–Sierra Leone, 2014. *Epidemiology & Infection*, 145(14), 2921-2929. doi:10.1017/S0950268817001856

Hanson, C., Cox, J., Mbaruku, G., Manzi, F., Gabrysch, S., Schellenberg, D., ... & Schellenberg, J. (2015). Maternal mortality and distance to facility-based obstetric care in rural southern Tanzania: a secondary analysis of cross-sectional census data in 226 000 households. *The Lancet Global Health*, *3*(7), e387-e395. doi:10.1016/S2214-109X(15)00048-0

Healey, P. (2001). Planning theory: Interaction with institutional contexts. *Science Direct*. Retrieved from http://www.sciencedirect.com/topics/neuroscience/rational-planning-model

Hu, W., Dong, Q., & Huang, B. (2017). Correlations between road crash mortality rate and distance to trauma centers. *Journal of Transport & Health*, 6, 50-59. doi:10.1016/j.jth.2017.05.005

Kadobera, D., Sartorius, B., Masanja, H., Mathew, A., & Waiswa, P. (2012). The effect of distance to formal health facility on childhood mortality in rural Tanzania, 2005-2007. *Global Health Action*, *5*. doi:10.3402/gha.v5i0.19099

Kenny, A., Basu, G., Ballard, M., Griffiths, T., Kentoffio, K., Niyonzima, J. B., … Kraemer, J. D. (2015). Remoteness and maternal and child health service utilization in rural Liberia: A population–based survey. *Journal of Global Health*, 5(2), 020401. doi:10.7189/jogh.05.020401

Kentoffio, K., Kraemer, J. D., Griffiths, T., Kenny, A., Panjabi, R., Sechler, G. A., ... & Siedner, M. J. (2016). Charting health system reconstruction in post-war Liberia: a comparison of rural vs. remote healthcare utilization. *BMC health services research*, 16(1), 478. doi:10.1186/s12913-016-1709-7

Kramer, A. M., Pulliam, J. T., Alexander, L. W., Park, A. W., Rohani, P., & Drake, J. M. (2016). Spatial spread of the West Africa Ebola epidemic. *Royal Society Open Science*, 3(8), 160294. doi:10.1098/rsos.160294

Lafreniere, D., & Gilliland, J. (2015). “All the world's a stage”: A GIS framework for recreating personal time-space from qualitative and quantitative sources. *Transactions in GIS*, 19(2), 225-246. doi:10.1111/tgis.12089

Leligdowicz, A., Fischer, W. A., Uyeki, T. M., Fletcher, T. E., Adhikari, N. K., Portella, G., ... & Vanderschuren, A. (2016). Ebola virus disease and critical illness. *Critical Care*, 20(1), 217. doi:10.1186/s13054-016-1325-2

Liberia Institute of Statistics and Geo-Information Services [LISGIS]. (2013). *Liberia demographic and health survey – 2013*. Retrieved from https://dhsprogram.com/pubs/pdf/FR291/FR291.pdf

Lori, J. R., Munro-Kramer, M. L., Shifman, J., Amarah, P. N., & Williams, G. (2017). Patient satisfaction with maternity waiting homes in Liberia: a case study during the Ebola outbreak. *Journal of Midwifery & Women’s Health*, 62(2), 163-171. doi:10.1111/jmwh.12600

Luckow, P. W., Kenny, A., White, E., Ballard, M., Dorr, L., Erlandson, K., ... & Ly, E. J. (2017). Implementation research on community health workers’ provision of maternal and child health services in rural Liberia. *Bulletin of the World Health Organization*, 95(2), 113. doi:10.2471/BLT.16.175513

Ly, J., Sathananthan, V., Griffiths, T., Kanjee, Z., Kenny, A., Gordon, N., ... & Thomson, D. R. (2016). Facility-based delivery during the Ebola virus disease epidemic in rural Liberia: analysis from a cross-sectional, population-based household survey. *PLoS medicine*, 13(8), e1002096. doi:10.1371/journal.pmed.1002096

Lyseen, A. K., Nøhr, C., Sørensen, E. M., Gudes, O., Geraghty, E. M., Shaw, N. T., & on behalf of the IMIA Health GIS Working Group, C. (2014). A review and framework for categorizing current research and development in health-related geographical information systems (GIS) studies. *Yearbook of Medical Informatics*, 9(1), 110–124. doi:10.15265/IY-2014-0008

Macharia, P. M., Ouma, P. O., Gogo, E. G., Snow, R. W., & Noor, A. M. (2017). Spatial accessibility to basic public health services in South Sudan. *Geospatial Health*, 12(1). doi:10.4081/gh.2017.510

Makanga, P. T., Schuurman, N., Sacoor, C., Boene, H., Von Dadelszen, P., & Firoz, T. (2016). Guidelines for creating framework data for GIS analysis in low-and middle-income countries. *The Canadian Geographer/Le Géographe canadien*, 60(3), 320-332. doi:10.1111/cag.12295

McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica*, 23(2), 143–149. doi:10.11613/BM.2013.018

McNamara, L., Schafer, I, Nolen, L, Gorina, Y., Redd, J., Lo, T., Ervin, E., Henao, O., Dahl, B., Morgan, O., Hersey, S., & Knust, B. (2016). Ebola surveillance – Guinea, Liberia, and Sierra Leone. *Morbidity and Mortality Weekly Report (MMWR)*. Suppl 2016;65(Suppl-3):35–43. doi:10.15585/mmwr.su6503a6

Moisi, J. C. (2009). *Mortality, morbidity, immunization coverage and distance to health facilities in Kilifi District, Kenya* (Order No. 3339950). Available from ProQuest Dissertations & Theses Global. (304910685). Retrieved from https://search-proquest-com.ezp.waldenulibrary.org/docview/304910685?accountid=14872

Nyenswah, T. G., Kateh, F., Bawo, L., Massaquoi, M., Gbanyan, M., Fallah, M., … De Cock, K. M. (2016). Ebola and its control in Liberia, 2014–2015. *Emerging Infectious Diseases*, 22(2), 169–177. doi:10.3201/eid2202.151456

Ordaz-Németh, I., Arandjelovic, M., Boesch, L., Gatiso, T., Grimes, T., Kuehl, H. S., ... & Junker, J. (2017). The socio-economic drivers of bushmeat consumption during the West African Ebola crisis. *PLoS neglected tropical diseases*, 11(3), e0005450. doi:10.1371/journal.pntd.0005450

Preston, R. (1994). The hot zone. New York: Random House

Root, E. D., Lucero, M., Nohynek, H., Stubbs, R., Tallo, V., Lupisan, S. P., ... & Simões, E. A. (2016). Distance to health services modifies the effect of an 11-valent pneumococcal vaccine on pneumonia risk among children less than 2 years of age in Bohol, Philippines. *International Journal of Epidemiology*. doi:10.1093/ije/dyw217.

Shahid, R., Bertazzon, S., Knudtson, M. L., & Ghali, W. A. (2009). Comparison of distance measures in spatial analytical modeling for health service planning. *BMC Health Services Research*, 9, 200. doi:10.1186/1472-6963-9-200

Shoman, H., Karafillakis, E., & Rawaf, S. (2017). The link between the West African Ebola outbreak and health systems in Guinea, Liberia and Sierra Leone: a systematic review. *Globalization and health*, 13(1), 1. doi:10.1186/s12992-016-0224-2

Sieh, R. (2017). Bad roads, rainfall - major headache for candidates in Liberia’s upcoming elections. *Front Page Africa*. Retrieved from https://www.frontpageafricaonline.com/index.php/news/4996-bad-roads-rainfall-major-headache-for-candidates-in-liberia-s-upcoming-elections

Sullivan, L. (2016). *Power and sample size determination*. Retrieved from http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_power/BS704_Power_print.html

The World Bank. (2015). *Making progress: Liberia increases connections to remote areas, in spite of adversity*. Retrieved from http://www.worldbank.org/en/news/feature/2015/09/10/making-progress-liberia-increases-connections-to-remote-areas-in-spite-of-adversity

Ueberschar, N. (2010). Modelling of catchment areas for health facilities in Africa. *University of Applied Sciences*. Retrieved from http://www.irdes.fr/english/conferences-and-workshops/enrghi-15th-emerging-new-researches-in-the-geography-of-health-and-impairment-conference-10-june-2010/presentation-ueberschar.pdf

United Nations Children's Fund [UNICEF]. (2014). *Ebola outbreak response in West Africa*. Retrieved from https://www.unicef.org/appeals/files/20140915_HAC_Ebola_Response.pdf

United Nations Children's Fund [UNICEF]. (2016). *Communication for development (C4D)*. Retrieved from https://www.unicef.org/cbsc/index_65736.html

United Nations Women [UNWomen]. (2017). *Maternal Health gets a boost in Liberia*. Retrieved from http://www.unwomen.org/en/news/stories/2017/7/feature-maternal-health-gets-a-new-boost-in-liberia

Wagner, C. (2015). Q&A: How changing behaviors is helping stop Ebola's spread in West Africa. *United States Agency for International Development*. Retrieved from https://blog.usaid.gov/2015/06/qa-how-changing-behaviors-is-helping-stop-ebolas-spread-in-west-africa/

Wang, H., Shen, Q., & Tang, B. S. (2014). GIS-based framework for supporting land use planning in urban renewal: Case study in Hong Kong. *Journal of Urban Planning and Development*, 141(3), 05014015. doi:10.1061/(ASCE)UP.1943-5444.0000216

WHO Ebola Response Team. (2014). Ebola Virus Disease in West Africa — The first 9 months of the epidemic and forward projections. *The New England Journal of Medicine*, 371(16), 1481–1495. doi:10.1056/NEJMoa1411100

WHO Ebola Response Team. (2016). After Ebola in West Africa—unpredictable risks, preventable epidemics. *New England Journal of Medicine*, 375(6), 587-596. doi:10.1056/NEJMsr1513109

World Health Organization [WHO]. (2015). *Factors that contributed to undetected spread of the Ebola virus and impeded rapid containment*. Retrieved from http://www.who.int/csr/disease/ebola/one-year-report/factors/en/

World Health Organization [WHO]. (2015). *Health worker Ebola infections in Guinea, Liberia and Sierra Leone: a preliminary report 21 May 2015*. Retrieved from http://www.who.int/hrh/documents/21may2015_web_final.pdf

World Health Organization [WHO]. (2017). *Frequently asked questions on Ebola virus disease*. Retrieved from http://www.who.int/csr/disease/ebola/faq-ebola/en/

Zinszer, K., Charland, K., Kigozi, R., Dorsey, G., Kamya, M. R., & Buckeridge, D. L. (2014). Determining health-care facility catchment areas in Uganda using data on malaria-related visits. *Bulletin of the World Health Organization*, 92(3), 178-186. doi:10.2471/BLT.13.125260

Zinszer, K., Morrison, K., Verma, A., & Brownstein, J. S. (2017). Spatial Determinants of Ebola Virus Disease risk for the West African epidemic. *PLoS Currents*. doi:10.1371/currents.outbreaks.b494f2c6a396c72ec24cb4142765bb95