The effect of poverty on mosquito-borne illness across the United States

1Kavita Kar, 2Russanne Low
1Wheeler High School, Marietta, Georgia
2Institute for Global Environmental Strategies (IGES), Arlington, Virginia

SUMMARY
Mosquito-borne diseases are a major issue across the world, and the objective for this project was to determine the characteristics that make some communities more susceptible to these diseases than others. We hypothesized that low-income communities are more susceptible to mosquito-borne diseases. To test this hypothesis, we identified and studied characteristics that make communities susceptible to mosquito-borne diseases, including water in square miles, average temperature, population, population density, and poverty rates per county. We used regression analysis to understand the relationship between the above variables and the total number of mosquito-borne disease cases. We found that the population of a county is the best indicator of the prevalence of mosquito-borne diseases. Rate of poverty in the county was not found to be a significant indicator of mosquito-borne diseases. This may be because poverty aggregated at county level does not describe income level variation within the county. In this research, both rural and urban counties were analyzed. Although rural counties have high rates of poverty, the population of rural counties are spread out, so they are at a lower risk of disease. Therefore, we conclude that population data needs to be studied in conjunction with poverty levels to truly understand the relationship between poverty and mosquito-borne diseases.

INTRODUCTION
Mosquitoes are deadly creatures because they are effective vectors of an array of diseases. Each year, 700 million people worldwide contract mosquito-borne diseases and 1 million die due to these diseases. Mosquitoes are an issue in both developed and developing countries (1). Mosquito-borne disease rates in the United States are on the rise and could spike if active prevention is not practiced. In 2016 alone, there were 47,461 reported cases of mosquito-borne diseases in the U.S. (2). There are several established factors that affect the presence of mosquitoes, including climate (humidity, temperature, rainfall, etc.), freshwater availability, and the presence of blood sources (3). Climate is a large predictor of mosquito abundance; mosquitoes are drawn to warmer places and places with more rainfall (4). Mosquitoes also breed in areas where there is stagnant freshwater, such as near lakes and ponds (3). These environmental concerns are more pressing for certain parts of the U.S., like the southeast region, which has hospitable environments for mosquitoes (4).

Aedes albopictus, a species of mosquito that can carry many diseases such as dengue, chikungunya, yellow fever, and zika, has been noted to travel only up to 300 feet in search of a blood meal, and other species have a maximum flight range of 1–3 miles. This short flight range of mosquitoes creates the situation where mosquitoes are only able to exist within a certain community (5). This also means that mosquitoes do not travel between communities, and mosquitoes born in a community will only stay in that community. It has even been noted that in Baltimore, there are three times as many mosquitoes in low-income neighborhoods than in high-income neighborhoods (6). An emerging field of research describes the effect of poverty on the mosquito population in a region and the characteristics of poverty that make communities most susceptible to increased mosquito populations.

The city of Baltimore, Maryland has been a model city to study poverty in conjunction with mosquito populations. In Baltimore, the largest contributor to the disparity in mosquito populations is the land cover that surrounds the neighborhoods studied. Land cover describes the surface of the earth and what potential habitats can exist in these neighborhoods. Since Baltimore itself has a homogenous climate and the sample sites were not near freshwater sources, it is likely that the neighborhood and the fauna were the biggest influences on the number of mosquitoes in the region (7). Characteristics of low-income communities like poor waste management, high building abandonment, many rodent blood sources, and poor pest control lead to an increased population of mosquitoes (7).

Historically, low-income communities suffer with poor waste management. Rainwater collects in small pieces of trash and debris and creates a breeding ground for mosquitoes (7). The median household income was also negatively associated with the quantity of garbage (7). Lower income communities are more likely to have large numbers
of abandoned buildings, which often collect rainwater in stagnant pools that offer another breeding ground for mosquitoes (6). While wealthy communities have on average one abandoned structure per 100 buildings, median income communities have five abandoned structures per 100, and low-income communities have 26 abandoned structures per 100 (7). Additionally, people in lower income communities may not have resources to eliminate mosquito habitats and are likely to have disused containers near them. A study found that *Ae. albopictus* were 83% more abundant in disused containers and could potentially contribute to the spread of diseases (8).

In low-income communities, rats are common, and mosquitoes often feed on rats in addition to humans. This diversity of blood sources increases the total amount of sustenance for mosquitoes and can potentially support a larger mosquito population (6). There is also a disparity between the amount of money spent on mosquito control in low-income neighborhoods versus wealthy neighborhoods. The average mosquito treatment costs several hundred dollars, which may not be feasible for most low-income families to implement. Prices start around $325 and can go as high as $1,900 for seasonal control for typical pest control programs (3).

The effects of these issues include increased numbers of mosquitoes, which creates a higher number of potential vectors for mosquito-borne diseases to spread. We investigated the hypothesis that low-income communities experience mosquito-related illnesses at a higher rate than high income communities. The model created in this research includes explanatory variables aggregated at a county level: area in square miles, average temperature, population, population density, and the poverty rate per county. The dependent variable is the number of cases of neuroinvasive (most severe) mosquito-borne diseases in a county. Neuroinvasive cases of diseases are more deadly, and they constitute approximately one-third of all cases of West Nile Virus, the most common mosquito-borne disease in the US (9). This distinction was created as the reported cases of neuroinvasive disease are thought to be the most accurate indicator of arboviral activity in humans. In contrast, reported cases of non-neuroinvasive arboviral disease are more likely to be impacted by the disease awareness and healthcare seeking behavior in different communities, and by the availability and specificity of laboratory testing that is performed. The diseases included in the study are chikungunya, dengue, eastern equine encephalitis (EEE), Jamestown Canyon, La Crosse, St. Louis encephalitis, and West Nile virus. There are other mosquito-borne diseases present in the U.S., but data is consistent for only these diseases.

The scope of this research includes every county in the continental U.S. We chose Cobb and Dekalb counties to be studied in further detail because the similarities and differences between these counties may offer insights into why low-income communities experience mosquito-related illnesses at a higher rate than other communities. In addition to other variables, we investigated land cover elements in both counties. Any relationships identified in this research will add to our knowledge about social factors that influence rates of mosquito-borne diseases. Our findings could inform why and how governments and communities should allocate resources to minimize the spread of these diseases in vulnerable communities.

## RESULTS

To determine whether low-income communities experience mosquito related illnesses at a higher rate than high income communities, a multiple regression analysis was performed with the total cases of mosquito-borne diseases in the U.S. as the dependent variable. This data was collected from CDC ArboNet (9).

The model tested the following explanatory variables: poverty rate, population, population density, square miles of water, and average temperature. This data was collected from the 2010 U.S. Census at the county level. The multiple $R^2$ values of the data and estimates were analyzed to understand how well the independent variables explained the number of cases of mosquito-borne diseases.

Rate of poverty was the primary variable we investigated with this model. Population and population density were chosen to model the spread of diseases because densely populated areas provide many more hosts and hence mosquito-borne diseases are likely to spread quickly. Square miles of water, as well as average temperature, were included to understand if climatic differences across the U.S. contributed to the number of mosquito-borne diseases. Because the U.S. has several different climate zones, we conducted further regional analysis to account for the impact of climate on mosquito populations.

A multiple linear regression model was used to predict the number of cases of mosquito-borne diseases (DV) based on the variables mentioned above. A significant regression equation was found ($F(5, 1301) = 705.12, p < 0.00$), with an $R^2$ of 0.73000. This model predicts that cases of mosquito-borne diseases are equal to -0.00810 + 0.00055 (POPULATION) + 0.00680 (WATER) - 0.00310 (POPULATION DENSITY) - 0.33550 (POVERTY) - 0.13670 (AVG. TEMP), where cases are counted as number per county, population is counted as number of residents per county, water is counted as the square miles of water in a county, population density is counted as the population divided by the square miles of land in a county, poverty is counted as the percentage of people in a county that live below the poverty line, and the average temperature is counted as the average year-round temperature in a county in degrees Fahrenheit. The number of mosquito-borne diseases increased by 0.00055 for each additional person, increased 0.00680 for each square mile of water, decreased by 0.00310 for each person per square mile, increased 0.33550 for each percent increase in poverty rate, and decreased 0.13670 for each degree Fahrenheit. These results were analyzed beyond the national level.

To ensure climate consistency, the data was further analyzed for each climatic region of the country: Southeast, Northeast, West, Southwest, and Midwest (Table 1). The $R^2$ values were consistent across all regions, except in the Southeast region where the $R^2$ was extremely low. A low $R^2$
value means that the model did not describe the cases of mosquito-borne diseases at all, and that there was either a data consistency issue, or another variable that was not analyzed. The data chosen for the number of cases was collected by CDC ArboNET, obtained as part of a voluntary program. One irregularity identified was the number of cases in Miami-Dade County. There were no reported cases recorded for the past 5 years in the ArboNET database for this county, when in fact there have been several confirmed cases (10).

**Cobb and Dekalb**

Further analysis was conducted on the poverty rates of Cobb and Dekalb counties. Cobb and Dekalb counties were studied in further detail because they have many common elements. They are adjacent counties in Georgia and have similar populations, population densities, average temperatures, and square miles of water; however, the poverty rate is different between these two counties. The poverty rate in Dekalb county is 14% whereas Cobb is at 9% (Figure 2), however the number of mosquito-borne disease cases in Dekalb is more than double that of Cobb. The land cover in both Cobb and Dekalb was analyzed to identify elements including canopy cover, buildings, and roads (Figures 1-2, 4). These elements create easily identifiable mosquito breeding grounds. The land cover distribution in both Cobb and Dekalb counties were found to be similar. Both Cobb and Dekalb counties had a high percentage of buildings because they are suburban areas. The map of poverty rates was superimposed on the county maps (Figure 3). There were a greater number of poor areas in Dekalb than in Cobb (11).

**DISCUSSION**

A multiple regression model was run with the total cases of mosquito-borne diseases as the dependent variable and poverty rate, the square miles of water, the population of a county, the population density of a county, and the average temperature of a county as explanatory variables. Each explanatory variable contributed to the model, and each had certain nuances.

**Population**

The population variable contributed to the greatest change in R². Urbanization creates suitable habitats for vector mosquitoes in which there are a reduced number
is small, mosquitoes only spread at a neighborhood level. Because the distance that mosquitoes travel be enough to understand the issues related to mosquito-borne illnesses. Because the distance that mosquitoes travel may be concentrated in a few urban areas. Density should also be studied on a granular level, potentially by Census Block Group. The most granular level of data available on ArboNET is at the county level, which may not be the best way to analyze population density because a high population density is conducive to the spread of predictors and human hosts are widely available. An alternate explanation to this is that there are simply more people for mosquitoes to infect in large cities, which leads to more mosquito-borne disease cases. Counties with high populations include metropolitan areas. Among the 10.6 million low-income working families in America, racial and ethnic minorities constitute 58%, despite only making up 40% of all working families nationwide. Ethnic minorities are also overrepresented in urban areas. Among urban residents, 56% are not white, compared with 32% in suburban and small metro counties and 21% in rural counties. In addition to this, minorities are less likely to receive adequate medical care, which puts them at a greater risk of complications from mosquito-borne diseases. Thus, incidence of mosquito-borne diseases across socioeconomic levels disproportionately affects ethnic minority communities. Understanding this issue is crucial when allocating city expenditure on public health measures. It will allow the government the ability to make public health decisions on a community-by-community basis.

**Population density**

Population density by county was not a significant contributor to the variance of R2. This is counterintuitive because a high population density is conducive to the spread in the mosquito-borne illnesses. However, county level may not be the best way to analyze population density because it relies heavily on total land area in a county, some of which may be uninhabited, and the majority of the population may be concentrated in a few urban areas. Density should therefore be studied on a granular level, potentially by Census Block Group. The most granular level of data available on ArboNET is at the county level, which may not be enough to understand the issues related to mosquito-borne illnesses. Because the distance that mosquitoes travel is small, mosquitoes only spread at a neighborhood level (5). Metropolitan areas are observed to be at a high risk of mosquito-borne diseases; thus, it is important to study the effects of mosquitoes in a more granular level here to implement preventative measures.

**Poverty**

Average poverty rate per county was not a significant contributor to variance. A possible reason for this is that county level aggregation is not indicative of large variation of incomes in different neighborhoods within a county. For example, in Cobb county, for the zip code 30068, the median household income is $117,475, and for the zip code 30168, the median household income is $44,911. Another example is in Dekalb county; in the zip code 30093, the median household income is $41,043, and in the zip code 30338, the median household income is $101,982 (16). These are large variations, and they affect the representation of poverty rate across a county. When this data is aggregated on a county level, these variations are lost. In prior studies about income and the effects of mosquitoes in Baltimore, data was collected and measured by street, and each part of the city was divided based on income block. This detailed separation allowed for better analysis of the effect of poverty on counts of mosquito-borne diseases. Rural communities need to be studied separately from urban communities. Rural counties have higher rates of poverty, but as the population is spread out, they may be at a lower risk of disease.

**Temperature**

The average annual temperature was not found to be a significant indicator of mosquito-borne disease in a county. The average annual temperature in the county may not be a good predictor because much of the United States has a temperate climate, where there are hot summers and cold winters. This may restrict the spread of mosquitoes to thrive in certain regions and certain periods of time. Mosquitoes thrive year-round in tropical climates because temperatures are extremely consistent and warm. In the U.S. where there is a large band of temperate areas, averaging the temperature over a year hides the unsuitable conditions in many areas. For example, in Chicago, there are cold winters and hot summers, whereas in a coastal city, the temperature is consistent year-round. These cities could both average to the same temperature, but their climates and the opportunity that they afford mosquitoes are different. Therefore, an additional analysis was conducted by region.

**Square Miles of Water**

The square mileage of water in a county was not found to be a significant contributor to the number of cases of mosquito-borne diseases. This may have to do with the distribution of water sources in a county, as mosquitoes lay eggs in shallow parts of water bodies. This impacts the number of mosquitoes in a county because there may be counties with some marshy areas with multiple shallow water
bodies and in other counties there may be one large lake. For example, a county with a high square mileage of water may include a large lake, which mosquitoes would not breed in.

Cobb and Dekalb
Based on the information provided by the Baltimore studies, it is possible that many of the buildings in the low-income areas in Dekalb may be ripe for mosquito-borne diseases. Dekalb as a county has many urban, poor areas. This observation needs to be further substantiated with mosquito surveillance programs to better inform disease models. This may provide an avenue for future research to find the connections between land cover and poverty rates.

Future overall program implementations
Overall, we created a model that described mosquito-borne disease cases across the United States and have found that that it is extremely likely that counties with a high population, such as counties with large cities are further at risk of mosquito-borne diseases. Additionally, while the data is not granular enough to determine whether poverty is correlated with mosquito-borne diseases, this model sets a framework from which future research can be conducted.

The current model can be developed further by adding more descriptive data such as humidity and rainfall, and by building it on a more granular level. This will help predict the occurrence of mosquito-borne diseases more accurately. The current impediment is that the ArboNET data is limited to a county level aggregation. It is not reported at a more granular level. Also, the data may not be accurately reported because it is a voluntary program; in impoverished areas, many cases go undiagnosed and unreported (5).

Disease data only reflects past outbreaks, which are not an effective predictor for future outbreaks. For an outbreak to occur, the population must have at least one infected person. This may mean an area with a high population of mosquitoes has not yet had an infected individual yet, but if an infected individual is introduced, the disease will spread quickly. This stresses the importance of mosquito surveillance data to track the mosquitoes rather than using past disease data.

As the Baltimore studies show, understanding the effects of poverty on mosquito-borne illness requires a more local level of analysis than the county data used in this study. The next step for this research would be to expand the surveillance of mosquitoes in other cities using the GLOBE Observer program. The GLOBE Observer program is a NASA citizen science initiative to track mosquito habitats and movement patterns. This program could be used to map mosquito habitat data to the average income in the census tract the collection point lies in. One recommendation is to immediately promote this program in highly populated cities to begin analysis and prevention.

Methods
Datasets Used
Data analysis was performed at a county level in the U.S. and by different regions, as there are several climate zones that span the United States. The analysis was conducted by comparing the neuroinvasive cases of mosquito-borne illnesses to other factors including poverty rate, square miles of water area in the county, population, population density, and average annual temperature.

The data source for arboviral diseases was ArboNET, a voluntary national arboviral surveillance system (9). This is a CDC program and data are reported by healthcare providers. Since 2003, ArboNET has collected data for all notifiable arboviral diseases. The viruses studied are Chikungunya, Dengue, Eastern equine encephalitis (EEE), Jamestown Canyon, La Crosse, St. Louis encephalitis, and West Nile. This data includes only the neuroinvasive cases of all these diseases. This measure was taken to ensure that there was not a severe disparity between wealthy and poor communities in diagnosis, given that non-neuroinvasive cases are harder to diagnose and these individuals are less likely to seek treatment. The total cases of mosquito-borne disease were aggregated for the years 2014–2019, to create a set period where yearly climate variations would not affect the total number of cases.

Poverty rate per county was collected from the U.S. Census Bureau which describes the poverty rates across all counties in the country. The data represents a five-year average of poverty rates in each county. The measure of poverty adopted by this study is the percentage of individuals in a county that earn below the poverty line. County level poverty data allowed the same level of specificity as the disease data. The water data measures the square miles of water in a county and is sourced from USA.com created from the 2010–2014 U.S. Census. The source for population, temperature, and population density per county were also sourced from USA.com, which uses U.S. Census data (17–19).

The in-depth analysis of Cobb and Dekalb counties was done using data from Collect Earth Online (20). Collect Earth characterizes and classifies land cover data from satellite imagery. Collect Earth works with GeoDash, which uses information from Google Earth Engine. The data that Google Earth uses is from Landsat imagery.

Data Collection Procedure
The disease data was requested from the CDC. The Census data was first downloaded from USA.com in csv files. The data elements that were analyzed were not in a consistent format, so Microsoft Excel functions were used to clean and merge the datasets together.

Statistical modeling
Multiple regression analysis was used to understand the relationship between the explanatory variables and the total number of mosquito-borne disease cases, and then to understand how each explanatory variable was related to the
response variable. This step was crucial in testing the guiding hypothesis because it provided insight into which explanatory variable was the most connected to the total cases of mosquito-borne diseases. This allowed us to understand the variable poverty level in a county in comparison to other factors. The multiple regression analysis was conducted in Microsoft Excel. The total number of cases of mosquito-borne disease was set as the dependent variable. Poverty rate by county, square miles of water in a county, population of a county, population density of a county, and the average temperature of a county were all used as explanatory variables. The multiple R² values of the data and estimates were analyzed to understand how well the data fits the regression line. This analysis was conducted at a national level first. Because this analysis was performed over a broad region with several different climate zones, we identified that climate may impact each region's interpretation. For further analysis, the data was divided by the five regions (Northeast, Southeast, Southwest, West, and the Midwest).

Cobb and Dekalb
To conduct a deeper analysis for Cobb and Dekalb counties in Georgia, satellite images of the land encompassed by the counties were selected, and the outline of each county was superimposed onto the satellite data to identify land cover characteristics: canopy cover, buildings, and roads. This is significant in determining what a county’s physical features are, and the potential habitats it provides. A grid was placed over each county to simulate random sampling points for land cover data. At the location of each grid point, the land cover element was identified. After this initial land cover analysis, the land cover elements were compared, and the numbers of elements and their distributions were graphed to identify whether Cobb and Dekalb have the same land features. The land cover analysis was mapped to the regions of Cobb and Dekalb counties that had high rates of poverty. This was done to understand whether there are particular land features that arise in areas with high poverty rates.

Received: August 31, 2020
Accepted: February 15, 2021
Published: February 25, 2021

REFERENCES

Copyright: © 2021 Kar and Low. All JEI articles are distributed under the attribution non-commercial, no derivative license (http://creativecommons.org/licenses/by-nc-nd/3.0/). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.