January 2014

Estimating The Incidence Of Japanese Encephalitis In Divisions Of Bangladesh

Craig Michael Rothenberg
Yale University, craig.rothenberg@yale.edu

Follow this and additional works at: http://elischolar.library.yale.edu/ysphtdl

Estimating the Incidence of Japanese Encephalitis in Divisions of Bangladesh

Craig Rothenberg
Yale School of Public Health
Division: Epidemiology of Microbial Disease
Class of 2014

Acknowledgements

   I would like to extend my deep gratitude to Dr. James Childs for his guidance and support throughout the project, as well as to Dr. Daniel Weinberger for his assistance in the development of the mathematical model.

Abstract

Japanese Encephalitis (JE) is a mosquito-borne viral disease responsible for 30,000 to 50,000 reported cases and up to 15,000 reported deaths every year. The disease is a significant public health concern in Bangladesh, but there is no regular surveillance or immunization for JE in the country. The objective of this study is to determine whether a practical model to estimate the incidence of JE in Bangladesh can be developed using data on population, environmental characteristics, and/or vector distribution for JE-endemic countries in Asia and the Western Pacific. Information on JE incidence, area, land cover, climate, population characteristics, land elevation, and distribution of mosquito vectors was collected for JE-endemic areas. Sources of the information included population and agricultural censuses, the Food and Agricultural Organization of the United Nations, the WHO, WorldClim, the CIA World Factbook, and mosquitomap.org. Generalized linear models examined the association between the variables and the outcome of Japanese Encephalitis. The best model was used to estimate the incidence of JE in each division of Bangladesh. The most statistically significant model used a negative binomial distribution and included variables for population density (p=.0052), mean annual temperature (p<.0001), annual range of temperatures (p<.0001), and mean temperature of the warmest quarter (p<.0001), with the population size as an offset variable. The estimated incidence in each division of Bangladesh ranged from 2.6 to 5.9 cases per 100,000 population. The division with the highest risk for JE was Rajshahi. A pilot vaccination program in that division may be more cost-effective than in other divisions. Limitations in data quality may have hindered the utility of several variables, so active case-finding may be useful for validating the model.

TABLE OF CONTENTS

Introduction

Japanese encephalitis (JE) is a mosquito-borne viral disease caused by the *Flavivirus* Japanese encephalitis virus (JEV) and is one of the most common form of epidemic encephalitis in the world (Weaver & Barrett, 2004). It was first formally identified as a disease during an outbreak in 1924 on the island of Okinawa, Japan, although the clinical characteristics had been recognized in Japan as early as 1871 (Lewis, Taylor, & et al., 1947). Initial symptoms generally appear a few days after infection, manifesting as nonspecific febrile illness that may include rhinitis, diarrhea, and chills (Solomon et al., 2000). These symptoms may be followed by headache, vomiting, reduced level of consciousness, and convulsions (Solomon et al., 2000).

Figure 1 shows the geographic distribution of JE risk. The disease is found in much of Asia and in the western pacific ("Centers for Disease Control and Prevention," 2012; Fischer, Lindsey, Staples, & Hills, 2010). Transmission of the virus is seasonal and peaks in the summer and fall in temperate areas, but can occur year-round in tropical and subtropical locations (Fischer et al., 2010). Approximately 2 billion people live in areas with risk of JEV transmission, and 700 million in areas with transmission risk are under the age of 15 (Keiser et al., 2005).

Approximately 30,000 to 50,000 cases of JE are reported every year, although additional cases may go unreported due to lack of high-quality surveillance and laboratory testing capabilities (Ghosh & Basu, 2009). There are up to 15,000 deaths reported annually, and half of all survivors may experience neurologic sequelae that can last for years (Weaver & Barrett, 2004). Children are most susceptible to the disease (Campbell et al., 2011) and are also particularly vulnerable to the more severe sequelae associated with infection: convulsions are reported in 85% of infected children but only in 10% of infected adults (Solomon et al., 2000).

The varying presentation of clinical features of the disease in children can result in misdiagnosis. Abnormal behavior, for example, may be the only presenting feature in certain children and may result in an initial misdiagnosis of mental illness (Solomon et al., 2000). Furthermore, some children may experience a single convulsion followed by rapid recovery of consciousness, leading physicians to provide an inaccurate diagnosis of febrile convulsion (Solomon et al., 2000). The difficulty associated with accurately diagnosing the disease can make the provision of appropriate treatment problematic.

JEV, the pathogen responsible for causing the disease, is a single-stranded RNA virus transmitted by arthropod vectors (Weaver & Barrett, 2004). *Culex tritaeniorhynchus* mosquitoes are the primary vectors of the pathogen, although there is some evidence that JEV can also be carried by other members of the *Culex, Aedes, Anopheles, Mansonia,* and *Armigeres* genera (Weaver & Barrett, 2004). Vectors are often found in flooded rice fields as well as in other stagnant collections of standing water such as swamps and marshes (Self et al., 1973). The habitats of the vectors are commonly located in rural areas, but more cases of JE tend to be reported in urban areas (Masuoka et al., 2010), possibly due to reporting biases as well as differences in the size of the population at risk in each area. The presence of *Culex tritaeniorhynchus* has also been positively correlated with temperature and precipitation, and JE incidence has been negatively correlated with elevation (Masuoka et al., 2010). The primary amplifying hosts of the pathogen are pigs and birds, both of which can develop high-titer viremia- optimal for infecting mosquitoes- without exhibiting clinical disease themselves (Weaver & Barrett, 2004). There is research indicating that a number of other species are capable of acquiring infections, but that they are generally unable to develop a level of viremia

sufficient to infect mosquitoes and therefore are largely dead-end hosts (Weaver & Barrett, 2004).

JE represents a serious public health concern in Bangladesh. The temperature and precipitation in a significant proportion of the country are favorable to mosquito proliferation (Erlanger, Weiss, Keiser, Utzinger, & Wiedenmayer, 2009). The incidence of the disease in the country is believed to have increased in recent decades due to population growth and an increase in environments suitable for mosquito reproduction (Erlanger et al., 2009). Additionally, rice paddy area and pork production increased by 30 and 45 percent respectively between 1990 and 2005 (Erlanger et al., 2009), indicating that the area suitable to disease vectors and the number of amplifying hosts for the pathogen have grown in recent years and may contribute to disease spread.

Effective vaccines against JE exist, but the implementation of vaccination programs for large populations can be expensive (Fischer et al., 2010). Several JE-endemic countries, including Bangladesh, do not currently have widespread vaccination (Fischer et al., 2010). However, there is significant evidence suggesting that vaccination programs, particularly those targeting children or infants, may offer a cost-effective approach to manage JE. A cost-benefit analysis of a JE vaccination program in Thailand found that the vaccination of children in high incidence areas produced savings in treatment cost, disability care, and future lifetime earnings (Siraprapasiri, Sawaddiwudhipong, & Rojanasuphot, 1997). Savings were even greater when the vaccination was performed on children at the age of 18 months (Siraprapasiri et al., 1997). Another study comparing costs and outcomes for children in Shanghai, China, indicated that immunization resulted in cost savings and might be economically justifiable in developing countries of Asia with endemic JE (Ding, Kilgore, Clemens, Wei, & Zhi-Yi, 2003). Similar

studies in Bali, Indonesia, and in Cambodia also found their respective immunization programs to be highly cost-effective (Liu, Clemens, Kari, & Xu, 2008; Touch et al., 2010). Researchers have suggested that Bangladesh should consider a pilot project to introduce vaccines to high-incidence areas (Paul et al., 2011) based on the threat that JE poses to the population. However, since a pilot program won't target the entire population of Bangladesh, it is important to identify which divisions in Bangladesh represent areas of high JE incidence in order to determine where a pilot program would be most appropriate.

The true incidence of JE in Bangladesh is unknown since there are no specific diagnostic centers or surveillance systems in place to assess the threat (Erlanger et al., 2009). One study of the incidence of JE in the catchment areas of hospitals in three regions of Bangladesh found statistically significant differences in incidence of different areas (Paul et al., 2011). The estimated incidence in the division of Rajshahi was 2.7 per 100,000 population (95% confidence interval [CI]: 1.8-4.9), in Khulna was 1.4 (95% CI: 0.9 – 4.1) and in Chittagong was 0.6 (95% CI: 0.4 – 0.9) (Paul et al., 2011). The study did not estimate incidence in other parts of the country, and the estimates were derived using a novel low-cost approach that relied on social networks among community residents (Paul et al., 2011).   The researchers may have sacrificed some degree of accuracy in favor of frugality, since traditional house-to-house surveys are less likely to miss cases although they tend to be more expensive (Paul et al., 2011). Ultimately, there is still uncertainty regarding the disease incidence throughout the Bangladesh.

Without effective surveillance and reporting methods, it is difficult to develop public health interventions targeting the most at-risk populations. However, estimation of JE incidence based on information pertaining to risk factors for the disease may serve as a viable alternative to active surveillance, and may be valuable in planning driving public health policy and

interventions. To this end, an ecological niche model was developed by researchers using data on temperature, rainfall, elevation, and known locations of vector presence to identify habitats suitable for *Culex tritaeniorhynchus* survival in Asia and the western Pacific (Miller et al., 2012). The model could be useful in estimating JE risk, although it is not without its limitations. The model may not accurately reflect human cases of the disease since it only addressed the mosquito locations and did not account for human population density or distribution (Miller et al., 2012); however, used in conjunction with population data it may yield valuable information on disease risk in humans. The objective of this study is to determine whether a practical model to estimate the incidence of JE can be developed using data on population, environmental characteristics, or *Culex tritaeniorhynchus* distribution, and if said model could be used to estimate the incidence of JE in administrative divisions of Bangladesh.

Methods

Data Sources:

Countries reported by the World Health Organization (WHO) to be JE-endemic in the majority of the geographic area were included in the model development. The data on incidence of JE were collected from a systematic review by the WHO (Campbell et al., 2011). The WHO inferred these rates using: 1) case reports and laboratory confirmation (when available) by a variety of national, local, and hospital-based surveillance systems throughout Asia, and 2) data on population and age distribution within each country (Campbell et al., 2011). The WHO reported either a single incidence for an entire country or multiple differing rates for different administrative divisions (states, divisions, territories, or provinces) within the country.

Areas with high quality vaccination programs in place were excluded from analysis. Federal territories within Malaysia were excluded as observations due to the lack of available high-quality environmental data pertaining to their respective geographic areas. Singapore, although historically endemic to JEV, was excluded because the abolishment of pig farming in the country has eliminated the primary amplifying host of the disease; recent cases appear to have been imported and not to have originated within the country (Koh et al., 2006). Bangladesh was also not included from the development of the model as the objective of the study was to estimate the incidence in divisions of the country.

Country-level population estimates from 2011 were acquired from the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat (U. N. Secretariat, 2011). Country-level data on geographic area and age distribution were collected from the Central Intelligence Agency World Factbook. Population estimates and geographic area for administrative divisions within countries were acquired from each nation's respective population census in 2010 or 2011. The data on population and area were used to calculate population density per square kilometer. Information related to vaccination in each area was also collected: the WHO described the quality of the vaccination program for each observation as high-quality (high-quality vaccination programs were excluded from analysis), expanding, minimal, nascent, or nonexistent.

Several environmental variables were studied in relation to disease incidence. Data on reported area of rice sewn or harvested was acquired for each observation. The sources for data on rice are reported in Table 1. A rice measure for each area was calculated by dividing the reported area of rice sewn or harvested by the total area of the observation. The rice measure was included to serve as a proxy for breeding habitat of *Culex tritaeniorhynchus.* Other

environmental data was collected in the form of raster data. Using figures provided by

Worldclim (http://www.worldclim.org/bioclim), data on the following six environmental

variables were collected: mean altitude, annual mean temperature (Bio1), annual range of

temperatures (Bio7), mean temperature of the warmest quarter (Bio10), annual precipitation

(Bio12), and precipitation in the wettest quarter (Bio16). The data were collected in the format

of global raster datalayers at a spatial resolution of 30 arc-seconds, wherein the shading of each

cell represented the value associated for the corresponding variable. Additionally, a raster map

indicating probability of the presence of *Culex tritaeniorhynchus* based on the aforementioned

ecological niche model was attained from mosquitomap.org. Each of the raster datalayers are

depicted in figures 2-8.

Once the geographic data was collected, ArcMap 10.1 was employed to analyze zonal

statistics for each raster datalayer. The average value of all grid cells within the area of each

observation was calculated for the six environmental variables with spatial data. The result of

the calculation indicated mean values for the environmental variables of every observation. For

example, the mean temperature reported for grid cells representing the area of Dolpa, Nepal, was

averaged to produce a single mean temperature for that district. This process was repeated for

the raster layer indicating probability of the presence of *Culex tritaeniorhynchus.*

Mathematical Model:

The outcome of interest was the estimated count of cases, which was determined by

multiplying estimated incidence by population for each observation. Observations were

represented by the smallest administrative level for which an incidence was reported by the

WHO: when a single incidence was reported for an entire country, the entire country represented

a single observation; however, when different incidence was reported for different administrative divisions within a country, each administrative division counted as a single observation.

Generalized linear models were employed to analyze the association of independent variables with the outcome of JE cases. Models were tested using a Poisson distribution and a negative binomial distribution to determine which distribution would be more appropriate for the data. The log-population was used as an offset- a variable in the model to adjust for differences in population size. Three different types of models were tested, as shown in Table 2. Model A, the most basic model, estimated the count of cases for each observation based on the mean probability of the presence of *Culex tritaeniorhynchus*. Model B used the same variable from model A but included independent variables for population density and proportion of the population under age 15. Additionally, the presence or absence of an expanding vaccination program was represented by a categorical variable in model B in order to measure whether expanding vaccination was associated with an impact on incidence. Model C repeated model B but replaced the single variable for probability of vector presence with the other environmental variables as independent variables. Environmental variables in model C included rice measure (reported rice area sewn or harvested divided by total area), altitude, annual mean temperature, annual range of temperatures, mean temperature of the warmest quarter, annual precipitation, and precipitation in the wettest quarter (Table 2).

The geographic area of some observations displayed a wide level of variation for some of the environmental variables, which could not be represented by the single numerical value used in the model. For example, the value indicating mean elevation would not appropriately represent the geography of an observation if part of the land was covered by large mountains and another part of the same land was covered by low-lying areas. Therefore, three different forms

of Model C were tested. Model C1 tested each of the six environmental variables that were derived using spatial data with no adjustment to the data. Model C2 excluded 10% of the values with the highest standard deviation for each of these environmental variables. For example: for observations with standard deviations of mean altitude among the highest 10% of all observations studied, the corresponding mean altitude was not included in the model. Model C3 repeated model C2 but with an exclusion threshold of 50% rather than 10%. For each model, backwards elimination was then used to exclude the least significant variables until all remaining variables had a p-value of less than 0.1.

If any model produced a negative parameter estimate for the variable indicating absence of an expanding vaccination program, then the variable was eliminated from the model. For models C1, C2, and C3, any models that included multiple variables for similar environmental characteristics after backwards elimination were retested with interaction terms between those environmental variables. For example, if a model included variables for annual precipitation and precipitation in the wettest quarter, the model would be retested with an interaction term to test for statistically significant interaction between those two variables. Additionally, if any of models C1, C2, or C3 did not contain at least one variable pertaining to population characteristics and one variable pertaining to environmental characteristics after backwards elimination, it was not included in further analysis.

The completed models were used to estimate the case counts in divisions of Bangladesh. These counts were then divided by the population in order to estimate the incidence for each division. The resulting estimates for case counts and incidence were compared to information reported in the previously mentioned hospital-based study of disease caused by JEV in certain areas of Bangladesh (Paul et al., 2011) to examine the validity of the model.

Results

173 observations from 16 countries were included in the development of the models. Case counts in observations ranged from 0 to 5595. Information on each variable is reported in table 3. Mean population density was 570.12 persons per square kilometer (95% CI: 0, 4323.40) although this value was skewed upward by two outliers, Jakarta (population density: 14,469 persons per square kilometer) and Macau (population density: 18,410 persons per square kilometer), both of which had population densities that were substantially higher than the population densities of other observations. The mean proportion of the population under age 15 was 0.28 (95% CI: 0.17, 0.38) and the mean value for the rice measure was 0.13 (95% CI: 0, 0.44). The mean probability of *Culex tritaeniorhynchus* presence was 0.17 (95% CI: 0, 0.58), and mean altitude above sea level was 917.38 meters (95% CI: 0, 3105.68). Average annual temperatures ranged from -1.08 degrees Celsius in the Manang district of Nepal to 28.07 degrees Celsius in the Indian Union Territory of Puducherry. The average annual range of temperatures was 21.69 degrees Celsius (95% CI: 2.86, 40.52), and the average temperature of the warmest quarter was 24.01 degrees Celsius (95% CI: 13.41, 34.60). Mean annual precipitation for each observation ranged from 298 millimeters to 3655 millimeters, and average precipitation in the wettest quarter was 903 millimeters (95% CI: 257, 1550).

All models developed using a Poisson distribution resulted in ratio of the deviance to degrees of freedom that was significantly different from 1, ranging from 9.5 to 128. Models developed with a negative binomial distribution resulted in a ratio of the deviance to degrees of freedom between 0.9 and 1.1. The negative binomial distribution was deemed more appropriate for modeling the data and was used to test the variables.

Table 4 displays the model parameters, parameter estimates, and p-values of the final models. Model A produced a parameter estimate of 0.65 for probability of *Culex tritaeniorhynchus* presence, although this result was not significant at a 95% confidence level (p=.10). The Akaike Information Criterion (AIC) for this model was 1624.46. When Model B was tested using all variables, two variables were eliminated for lack of adequate statistical significance: first eliminated was the variable for absence of an expanding vaccination program (p=.30), and second eliminated was the variable for proportion of population under age 15 (p=.26). The remaining two variables in the model, probability of *Culex tritaeniorhynchus* presence and population density, were again not significant at a 95% confidence level. The parameter estimates for the two variables were 0.78 (p=.05) and -0.0001 (p=.09), respectively. The AIC of model B was 1624.22.

After backwards elimination, model C1 contained three remaining variables that were statistically significant at a 95% confidence level. The variables and their associated parameter estimates were population density with a parameter estimate of -0.0001 (p=.01), absence of expanding vaccination program with an estimate of -0.80 (p<.01), and annual range of temperatures with an estimate of -0.005 (p<.01). The AIC for model C1 was 1611.88. The variable for absence of an expanding vaccination program was then removed from the model since it was represented by a negative parameter estimate, so the final version of model C1 had two variables: population density, with a parameter estimate of -0.0001 (p=0.02), and annual range of temperatures, with an estimate of -0.002 (p=0.02). The AIC according to this model was 1626.29.

Model C2 contained similar variables as model C1 after backwards elimination, but with slightly different parameters estimates: the parameter estimate for population density was again

-0.0001 (p=.01), the estimate for absence of expanding vaccination program was -0.83 (p<.01), and the estimate for annual range of temperatures was -0.005 (p<.01). The AIC for model C2 was 1393.60. When the variable for the vaccination program was removed, however, the parameter estimate for annual range of temperatures became statistically insignificant. The only statistically significant variable that remained in the final version of the model was population density. Since no statistically significant environmental variable remained in the model after elimination, model C2 was excluded from further analysis.

Model C3, the final model tested, contained four variables that were highly statistically significant at a 95% confidence level after backwards elimination. The variables included population density, with a parameter estimate of -0.0001 (p=.01), annual mean temperature, with a parameter estimate of 0.12 (p<.01), annual range of temperatures, with a parameter estimate of 0.03 (p<.01), and mean temperature of the warmest quarter, with a parameter estimate of -0.12 (p<.01). The AIC for this model was 390.14. Since three variables in the model pertained to temperature, the model was tested one more time with four interaction terms- one for each pair of variables pertaining to temperature, and one to test interaction between all three terms; however, there was no statistically significant interaction found between variables indicated in the model.

The estimated JE incidence reported by the hospital-based study in three divisions of Bangladesh multiplied by the population in each of these divisions produced the following case counts: 170 cases annually in Chittagong (95% CI: 114, 256), 220 cases in Khulna (95% CI: 141, 643), and 499 cases in Rajshahi (95% CI: 333, 906). The estimated case counts according to the models were compared to this data in order to examine their validity. Table 5 shows the estimated number of cases in each division of Bangladesh according to models A, B, C1, and C3.

Using data specific to divisions in Bangladesh, the estimated annual cases numbers in each of these divisions according to model A was 567 in Chittagong, 360 in Khulna, and 415 in Rajshahi. The division with the greatest estimated number of cases using this model was Dhaka with 1052. According to model B, the estimated number of annual cases was 539 in Chittagong, 359 in Khulna, and 399 in Rajshahi. Again, the expected number of cases in Dhaka was significantly greater at 958 than the expected number in any other division. Model C1 produced case estimates of 563 in Chittagong, 292 in Khulna, and 328 in Rajshahi. The highest estimate was 831 cases in Dhaka. Lastly, case estimates from model C3 were 993 cases in Chittagong, 693 cases in Khulna, and 1088 cases in Rajshahi. The division with the highest estimated number of cases was once again Dhaka with 2068 estimated cases.

Case numbers in each division were divided by their respective populations in order to determine estimated incidence. Table 6 displays the estimated incidence in each division of Bangladesh according to models A, B, C1, and C3. The estimated incidence per 100,000 population according to model A was 2.0 in Chittagong, 2.3 in Khulna, and 2.2 in Rajshahi. The division with the highest estimated incidence according to this model was Khulna. According to Model B, the estimated incidence was 1.9 in Chittagong, 2.3 in Khulna, and 2.2 in Rajshahi, with Khulna again representing the division with the highest estimated incidence. The estimated incidence according to model C1 was 2.0 in Chittagong, 1.8 in Khulna, and 1.8 in Rajshahi, with Sylhet reporting the highest incidence estimate. Finally, the estimated incidence according to Model C3 was 3.5 in Chittagong, 4.4 in Khulna, and 5.9 in Rajshahi, with Rajshahi representing the division with the highest incidence of any district.

Discussion

When the models were used to estimate the case counts in divisions of Bangladesh, all four successful models indicated that the Dhaka division had the greatest estimated number of cases annually in the country. Based on these results, a vaccination program in this division would likely reduce the burden of disease more than a vaccination program in any other division. However, it may be of interest to the government to vaccinate individuals in a division with the highest incidence rather than the highest number of cases since there are significant costs associated with vaccination of a large population and since the government may want to maximize the benefit for the associated cost in a pilot vaccination program. The division with the highest estimated incidence was not consistent across all models. According to model A and model B, the division with the highest estimated incidence was Khulna. The division with the highest estimated incidence according to model C1 is Sylhet, and according to model C3 is Rajshahi.

The negative binomial distribution appears to be more appropriate in model development than the Poisson distribution because the ratios of the deviance to degrees of freedom for models using the negative binomial distribution were close to one, while the ratios produced using the Poisson distribution were significantly different from one and in some cases exceeded 100. This suggests that the negative binomial distribution produces a better model fit and was justifiable for use in the model development.

Using the negative binomial distribution, four of the models (models A, B, C1, and C3) were successfully developed with p-values of less than 0.1 for all variables. With regards to model C2, the only remaining variable after backwards elimination of insignificant variables was population density. This model met the criteria for exclusion from further analysis because it did

not include at least one environmental variable and one population variable in its final version-population density in the absence of other variables is not an appropriate measure for the incidence of JE.

Only models C1 and C3 contained variables that were all statistically significant at a 95% confidence level, suggesting that they may be more reliable for estimating disease incidence. Both models included variables for population density and annual range of temperatures, but model C3 also included variables for annual mean temperature and mean temperature of the warmest quarter. The negative parameter estimate associated with population density in the models suggests that although cases are more often reported in urban areas, there may be a greater disease risk to those in rural areas. This negative correlation is in accordance with prior research as many environmental variables associated with JE, such as proximity to rice paddies, are also associated with rural areas and lower population density. In model C3, mean temperature was positively correlated with disease, as is expected since the mosquito vectors for the disease thrive in warmer climates. Greater annual range of temperatures is also associated with greater incidence of disease in model C3. This seems to reflect the fact that some areas with lower incidence, such as many districts in Nepal, are located in mountainous areas with consistently lower temperatures year-round which are not conducive to mosquito proliferation. Mountainous areas may also be less likely to have pig farms which could be more suited for flat areas. This means fewer reservoirs for the disease may be present in these areas. Interestingly, mean temperature of the warmest quarter was negatively correlated with disease in model C3. This may be because the deserts are found in many of the areas with the highest summer temperatures. Since rice paddies and other stagnant collections of standing water are generally not located in deserts, the vectors of disease may not thrive in these locations.

Since models C1 and C3 were the only models exclusively with highly statistically significant variables, they were considered more likely to be the best models. However, it was less clear which of these two models is superior. The AIC for model C3 is lower than the AIC for model C1, but model C3 excludes data that was included in model C1 so the AIC is not necessarily an appropriate measure for comparing the two models. Nonetheless, there is evidence that model C3 may be a more useful model than model C1. The incidence estimates of C1 were the lowest of the four successful models. Given that underestimation of JE incidence is a major concern in Bangladesh, the use of a model that is most likely to underestimate the disease incidence may be inappropriate. Additionally, model C1 only contains variables for population density and annual range of temperatures; this could be considered an insufficient amount of data with which to estimate disease rates. Model C3, however, includes multiple variables which may contribute to more accurate disease estimation. Importantly, it includes a variable for annual mean temperature which has previously been shown to be a strong predictor of vector presence. Based on model C3, the true incidence is likely highest in Rajshahi, and a vaccination program implemented there may be more cost effective than vaccination programs in other divisions of Bangladesh.

Each model demonstrated some weaknesses when these estimated case counts in Bangladesh were compared to the estimates of case counts inferred using the incidence data from the hospital-based study. The case count in Chittagong that was estimated using model A was 566, which was greater than the number of cases estimated using the hospital-based study (171, 95% CI: 114, 256). However, the model case counts for Khulna and Rajshahi both fell within the confidence interval of the validation data. Both model B and model C1 also overestimated the number of cases in Chittagong but predicted case numbers within the confidence intervals of

the validation data for Khulna and Rajshahi. Lastly, model C3 overestimated the case counts for all three divisions of Bangladesh when compared to the validation data of the hospital-based study.

The estimated incidence in each division also differed from the validation data. Again, models A, B, and C1 overestimated the incidence in Chittagong but accurately estimated the incidence in Khulna and Rajshahi while model C3 overestimated incidence in all three areas when compared to the hospital-based data. While this may bring into question the validity of model C3, there were some limitations in the quality and amount of validation data that may have compromised its usefulness. JE cases are consistently underreported since those who fall ill, particularly in developing nations, may not visit the hospital and receive a proper diagnosis (Campbell et al., 2011). The authors of the Bangladesh incidence study acknowledged that "hospital-based surveillance data underestimates the true burden of disease" and they made an effort to adjust for this by leveraging social networks in Bangladeshi rural communities to report cases (Paul et al., 2011). However, because they did not use a house-to-house survey due to the prohibitive cost associated with it, it is still possible that the true incidence was underestimated in the study. In addition, only 45% of admitted patients in the hospital-based study had a cerebrospinal fluid specimen or serum specimen collected within 7 or fewer days from the onset of illness. This means that some individuals with JE that presented to the hospitals may not have been diagnosed, which could further contribute to the underestimation. Furthermore, there was only data on JE in three divisions of Bangladesh available for validating the model, which may have compromised the utility of model validation.

A number of variables represented with spatial data that have previously been associated with JE risk were not significant in the models. Although a negative correlation between

elevation and disease incidence was identified in prior research, altitude was not found to be a significant variable in models C1, C2, or C3. Similarly, precipitation was reported to be positively correlated with the presence of *Culex tritaeniorhynchus* in prior studies, yet both measures of precipitation (annual precipitation and precipitation in the wettest quarter) were eliminated from the models due to failure to demonstrate statistical significance. The variables pertaining to climate, altitude, and mosquito presence may have failed to show strong statistical significance in the model because each variable was represented by a single value for a given area, which may have oversimplified the variability within that area. This is supported by the fact that model C3, which excluded values for environmental variables with a high standard deviation, had more variables with statistical significance than model C1, which did not exclude any values. Furthermore, there was substantial variation in the area of each observation. For example, some provinces of China were larger than the entire area of several countries. Many of the smallest observations used in the model were found in the north, far away from the equator. This could have influenced the significance of environmental variables in the model.

In addition, the single values representing environmental variables didn't take into consideration the more complex characteristics of population dispersion. If only a small proportion of an observation was suitable for JE vectors, it might suggest that that observation is not likely to be associated with a high disease incidence; however, if the population is largely distributed within the area with high probability of mosquito presence and is sparsely distributed in the other area within that observation, then the model would not produce reliable estimates for population density or probability of mosquito presence.

Some other variables expected to be associated with JE incidence also failed to show statistical significance in the models. The rice measure was also found to be insignificant in all

models when tested, even though past evidence has indicated that it is strongly associated with the presence of the pathogen vector. This could be due to the fact that rice area may only be associated with risk of disease to those living in close proximity to it. Because rice generally grows in rural areas with reduced population density, it may not noticeably contribute on a large scale to the risk in areas with more dense urban populations. Also, the method for reporting this variable differed among different nations. Some countries reported area of rice sewn, while others reported area of rice harvested; these may not necessarily represent similar measures. Additionally, proportion of the total population under age fifteen was not a significant variable in any model in spite of the fact that the incidence of JE is generally higher among younger individuals (Campbell et al., 2011). The difficulties associated with accurately diagnosing the disease in children and the inconsistency in diagnosing and reporting techniques in different areas may have influenced the validity of this variable.

Presence of an expanding vaccination program was included as a variable in the models to account for any effect vaccination had on reducing case numbers. Paradoxically, the models indicated that expanding vaccination was associated with *more* cases rather than fewer ones. This is likely due to the fact that vaccination programs are generally implemented in high-incidence areas and not in low incidence ones. Ultimately, this variable was therefore not useful in accounting for any reduction in disease incidence associated with vaccination. Reliable data reporting the proportion of each population that was vaccinated as well as the age distribution of the vaccinated population would be necessary to more accurately identify any vaccine-associated reduction of JE.

Failure to include some important variables may have limited the validity of the models. One major factor that may have limited the quality of the models was the lack of data on pig

farming. Pigs are important reservoirs for JEV, and their presence may contribute to disease spread. However, there was inadequate data on swine presence in several areas. Although pig farms exist in Bangladesh, there is insufficient data available on where they exist, how much land is devoted to them, or what proportion of the population lives or works in close proximity to them. Neither the Bangladesh Bureau of Statistics nor the Food and Agricultural Organization of the United Nations published reliable data on pig farming in the country.

Future efforts to model JE may be more successful if they employ spatial smoothing on raster data to give more weight to environmental variables in areas with greater population density. Consideration for additional interactions between environmental variables may also help improve future models. Models may also be more reliable if they use more specific data to represent proportion of population vaccinated within each observation. In addition, should better data on pig farming or rice production become available, it may contribute to improved efforts to estimate the incidence of JE in the future. Ultimately, the best way to produce a more accurate estimate of JE incidence and to validate the model would be to conduct active case finding. In the absence of high quality surveillance, however, model C3 may be a reasonable alternative to estimate disease incidence in divisions of Bangladesh.

References

Bangladesh Bureau of Statistics. (2011). Population and Housing Census 2011 Community
    Report, from
    http://www.bbs.gov.bd/WebTestApplication/userfiles/Image/Census2011/Barisal/Barisal/
    Barisal%20at%20a%20glance.pdf

Bangladesh, G. o. t. P. s. R. o. (2013). Estimates of Aman Rice. In B. B. o. Statistics (Ed.).
    Agargaon, Dhaka.

Campbell, G. L., Hills, S. L., Fischer, M., Jacobson, J. A., Hoke, C. H., Hombach, J. M., . . .
    Ginsburg, A. S. (2011). Estimated global incidence of Japanese encephalitis: a systematic
    review. [Research Support, Non-U.S. Gov't

Review]. *Bull World Health Organ, 89*(10), 766-774, 774A-774E. doi: 10.2471/BLT.10.085233

. Centers for Disease Control and Prevention. (2012). *Geographic Distribution of Japanese
    Encephalitis Virus*, 2014, from
    http://www.cdc.gov/japaneseencephalitis/maps/index.html

China, N. B. o. S. o. t. P. s. R. o. (2010). *Sixth National Population Census of the People's
    Republic of China*.

Ding, D., Kilgore, P. E., Clemens, J. D., Wei, L., & Zhi-Yi, X. (2003). Cost-effectiveness of
    routine immunization to control Japanese encephalitis in Shanghai, China. [Research
    Support, Non-U.S. Gov't]. *Bull World Health Organ, 81*(5), 334-342.

Erlanger, T. E., Weiss, S., Keiser, J., Utzinger, J., & Wiedenmayer, K. (2009). Past, present, and
    future of Japanese encephalitis. [Historical Article

Research Support, Non-U.S. Gov't]. *Emerg Infect Dis, 15*(1), 1-7. doi: 10.3201/eid1501.080311

FAOSTAT. (2014). FAOSTAT. Faostat, from Food and Agriculture Organization of the United
    Nations http://faostat3.fao.org/faostat-gateway/go/to/download/Q/QC/E

Fischer, M., Lindsey, N., Staples, J. E., & Hills, S. (2010). Japanese encephalitis vaccines:
    recommendations of the Advisory Committee on Immunization Practices (ACIP).
    [Guideline]. *MMWR Recomm Rep, 59*(RR-1), 1-27.

Ghosh, D., & Basu, A. (2009). Japanese encephalitis-a pathological and clinical perspective.
    PLoS Negl Trop Dis, 3(9), e437. doi: 10.1371/journal.pntd.0000437

India, C. O. o. (2011). India Population Census 2011.

Keiser, J., Maltese, M. F., Erlanger, T. E., Bos, R., Tanner, M., Singer, B. H., & Utzinger, J.
    (2005). Effect of irrigated rice agriculture on Japanese encephalitis, including challenges
    and opportunities for integrated vector management. [Research Support, Non-U.S. Gov't

Review]. *Acta Trop, 95*(1), 40-57. doi: 10.1016/j.actatropica.2005.04.012

Koh, Y. L., Tan, B. H., Loh, J. J., Ooi, E. E., Su, S. Y., & Hsu, L. Y. (2006). Japanese encephalitis, Singapore. [Case Reports

Letter]. *Emerg Infect Dis, 12*(3), 525-526. doi: 10.3201/eid1205.051251

Lewis, L., Taylor, H. G., & et al. (1947). Japanese B encephalitis; clinical observations in an outbreak on Okinawa Shima. *Arch Neurol Psychiatry, 57*(4), 430-463.

Liu, W., Clemens, J. D., Kari, K., & Xu, Z. Y. (2008). Cost-effectiveness of Japanese encephalitis (JE) immunization in Bali, Indonesia. [Research Support, Non-U.S. Gov't]. *Vaccine, 26*(35), 4456-4460. doi: 10.1016/j.vaccine.2008.06.050

Masuoka, P., Klein, T. A., Kim, H. C., Claborn, D. M., Achee, N., Andre, R., . . . Grieco, J. (2010). Modeling the distribution of Culex tritaeniorhynchus to predict Japanese encephalitis distribution in the Republic of Korea. [Comparative Study

Research Support, Non-U.S. Gov't]. *Geospat Health, 5*(1), 45-57.

Miller, R. H., Masuoka, P., Klein, T. A., Kim, H. C., Somer, T., & Grieco, J. (2012). Ecological niche modeling to estimate the distribution of Japanese encephalitis virus in Asia. [Research Support, U.S. Gov't, Non-P.H.S.]. *PLoS Negl Trop Dis, 6*(6), e1678. doi: 10.1371/journal.pntd.0001678

. Paddy Statistics of Malaysia. (2012). Peninsular Malaysia: Department of Agriculture.

Paul, R. C., Rahman, M., Gurley, E. S., Hossain, M. J., Diorditsa, S., Hasan, A. M., . . . Luby, S. P. (2011). A novel low-cost approach to estimate the incidence of Japanese encephalitis in the catchment area of three hospitals in Bangladesh. [Research Support, U.S. Gov't, P.H.S.]. *Am J Trop Med Hyg, 85*(2), 379-385. doi: 10.4269/ajtmh.2011.10-0706

Republik Indonesia, B. P. S. (2010). Population of Indonesia by Province 1971, 1980, 1990, 1995 , 2000 and 2010, from http://www.bps.go.id/eng/tab_sub/view.php?kat=1&tabel=1&daftar=1&id_subyek=12&notab=1

Republik Indonesia, B. P. S. (2012). Food Crops, 2014, from http://www.bps.go.id/eng/tnmn_pgn.php?kat=3

Secretariat, N. P. C. (2012). National Population and Housing Census 2011. Kathmandu, Nepal: Central Bureau of Statistics

Secretariat, U. N. (2011). World Population Prospects: The 2010 Revision and World Urbanization Prospects: The 2011 Revision: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat.

Self, L. S., Shin, H. K., Kim, K. H., Lee, K. W., Chow, C. Y., & Hong, H. K. (1973). Ecological studies on Culex tritaeniorhynchus as a vector of Japanese encephalitis. *Bull World Health Organ, 49*(1), 41-47.

Siraprapasiri, T., Sawaddiwudhipong, W., & Rojanasuphot, S. (1997). Cost benefit analysis of Japanese encephalitis vaccination program in Thailand. [Research Support, Non-U.S. Gov't]. *Southeast Asian J Trop Med Public Health, 28*(1), 143-148.

Solomon, T., Dung, N. M., Kneen, R., Gainsborough, M., Vaughn, D. W., & Khanh, V. T. (2000). Japanese encephalitis. [Research Support, Non-U.S. Gov't]. *J Neurol Neurosurg Psychiatry, 68*(4), 405-415.

. Statewise Statistics. (2013, 2 February, 2013), 2014, from http://www.rkmp.co.in/rice-stats/statewise-statistics/

Statistics, C. B. o. National Sample Census of Agriculture (N. P. C. Secreteriat, Trans.). Nepal.

Statistics, M. D. o. (2010). Population and Housing Census of Malaysia.

Touch, S., Suraratdecha, C., Samnang, C., Heng, S., Gazley, L., Huch, C., . . . Soeung, S. C. (2010). A cost-effectiveness analysis of Japanese encephalitis vaccine in Cambodia. [Research Support, Non-U.S. Gov't]. *Vaccine, 28*(29), 4593-4599. doi: 10.1016/j.vaccine.2010.04.086

USDA Economic Research Service. (2012). China Agricultural and Economic Data.

Weaver, S. C., & Barrett, A. D. (2004). Transmission cycles, host range, evolution and emergence of arboviral disease. [Research Support, Non-U.S. Gov't

Research Support, U.S. Gov't, P.H.S. Review]. *Nat Rev Microbiol, 2*(10), 789-801. doi: 10.1038/nrmicro1006

Tables and Figures

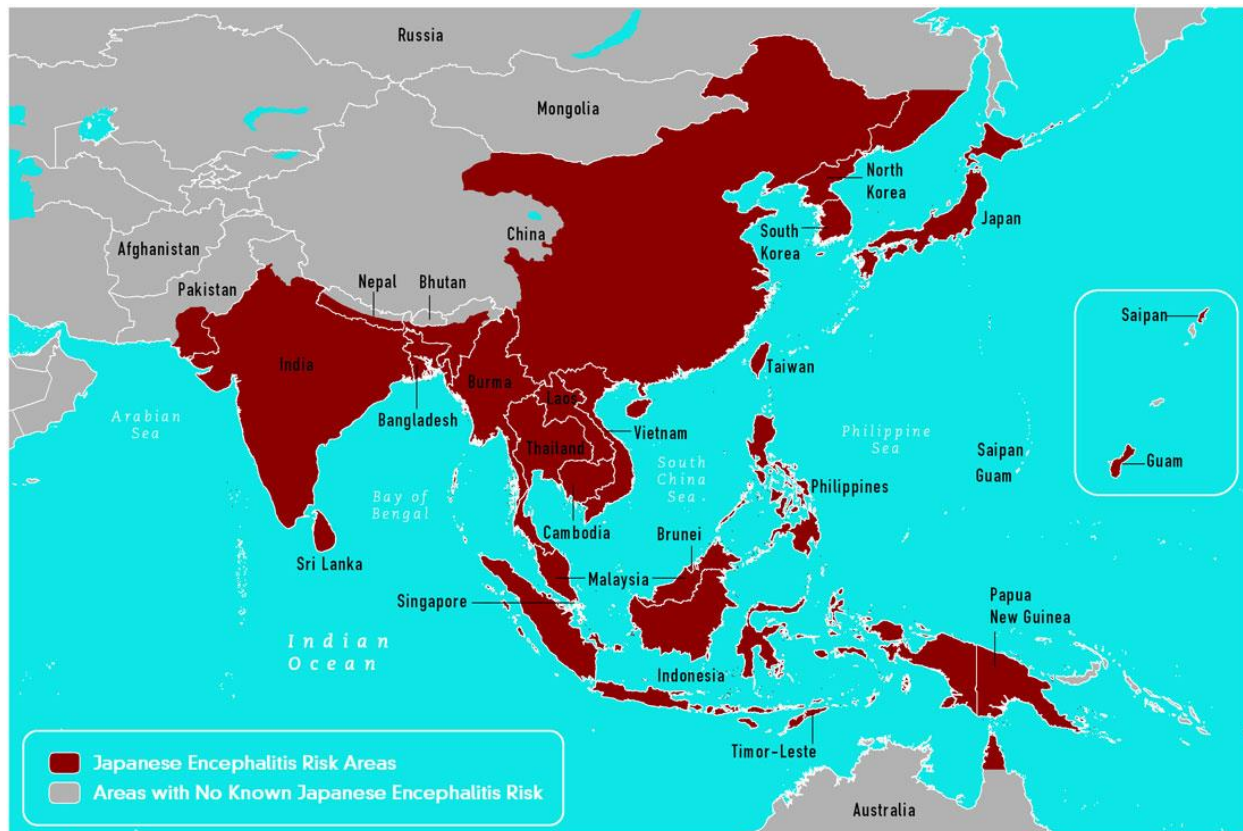Figure 1: Geographic Distribution of Japanese Encephalitis ("Centers for Disease Control and Prevention," 2012)

Figure 2: Estimated Probability of *Culex Tritaeniorhynchus* Presence

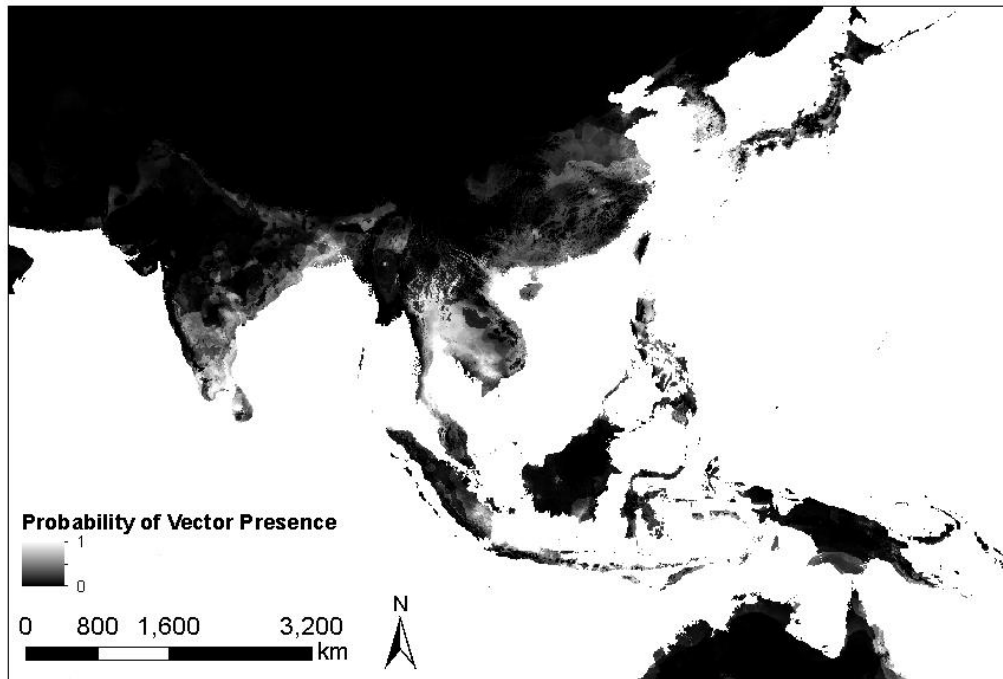**Estimated Probability of Culex Tritaeniorhynchus Presence**

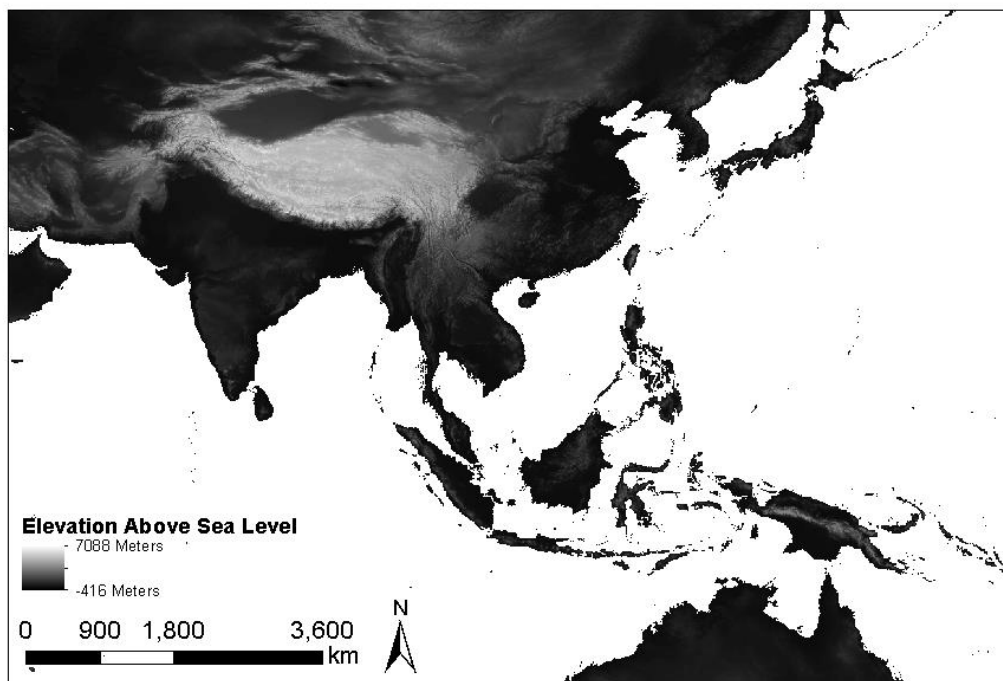Figure 3: Mean Altitude Above Sea Level

Mean Altitude

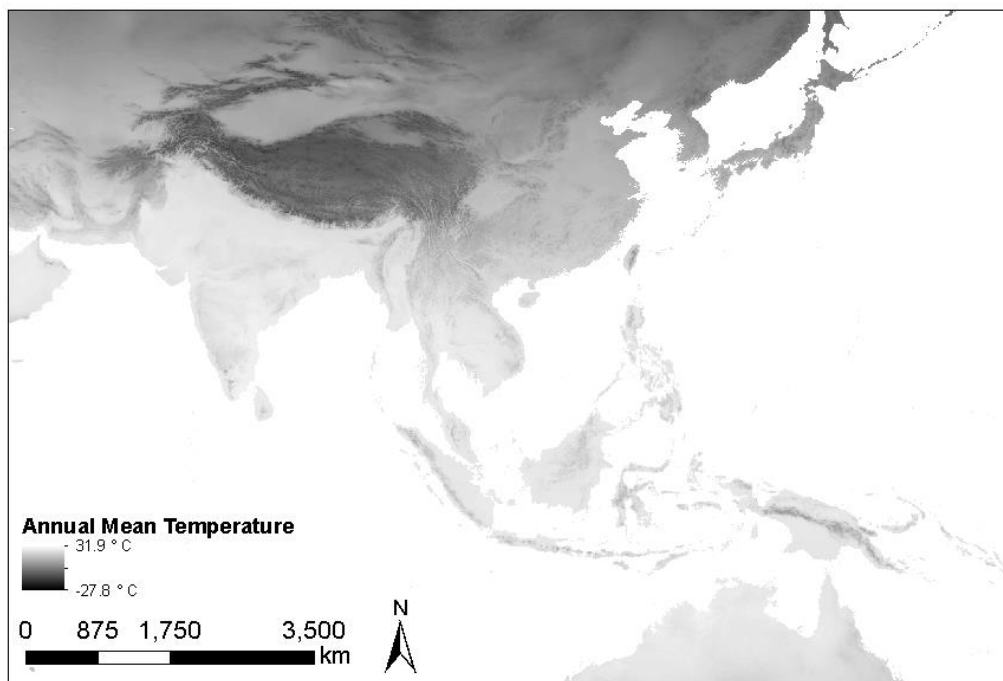Figure 4: Annual Mean Temperature

Annual Mean Temperature

Figure 5: Annual Range of Temperatures

Annual Range of Temperatures
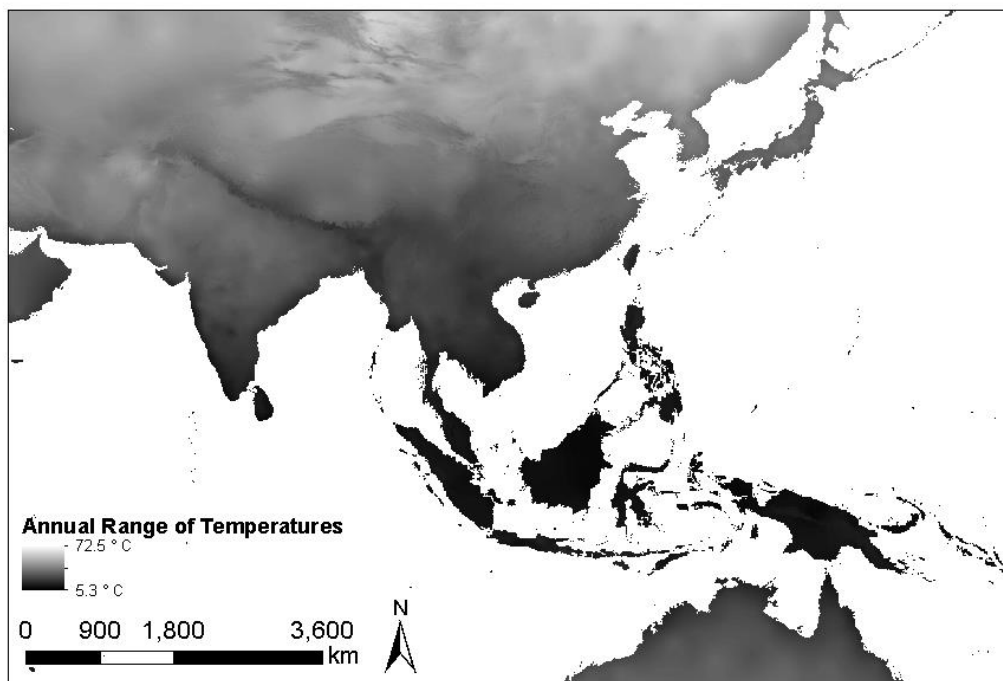
Figure 6: Mean Temperature of the Warmest Quarter
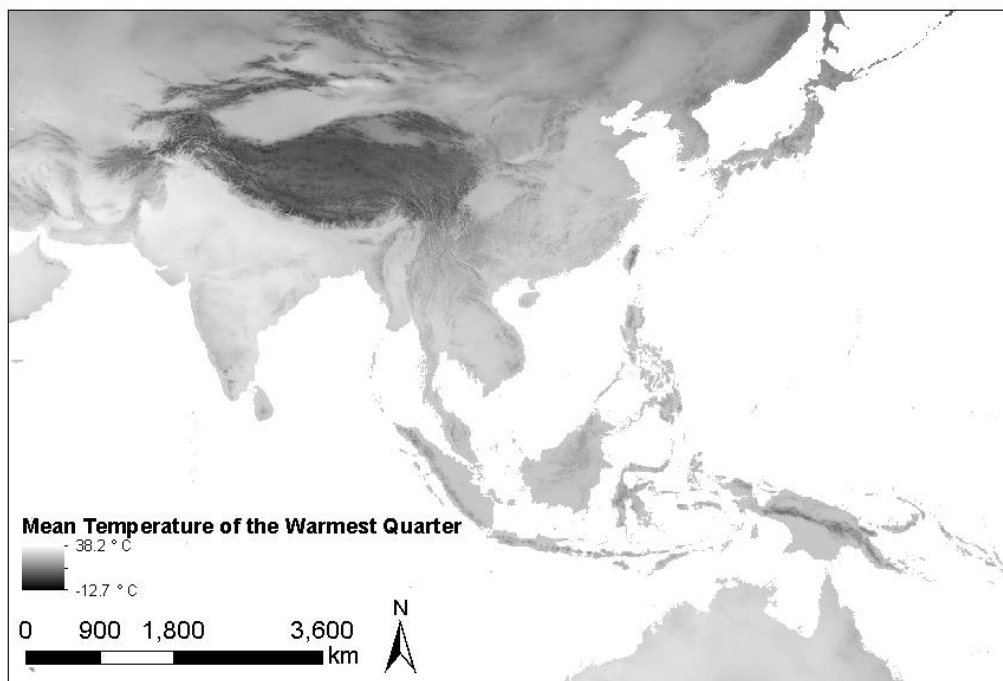

Mean Temperature of the Warmest Quarter
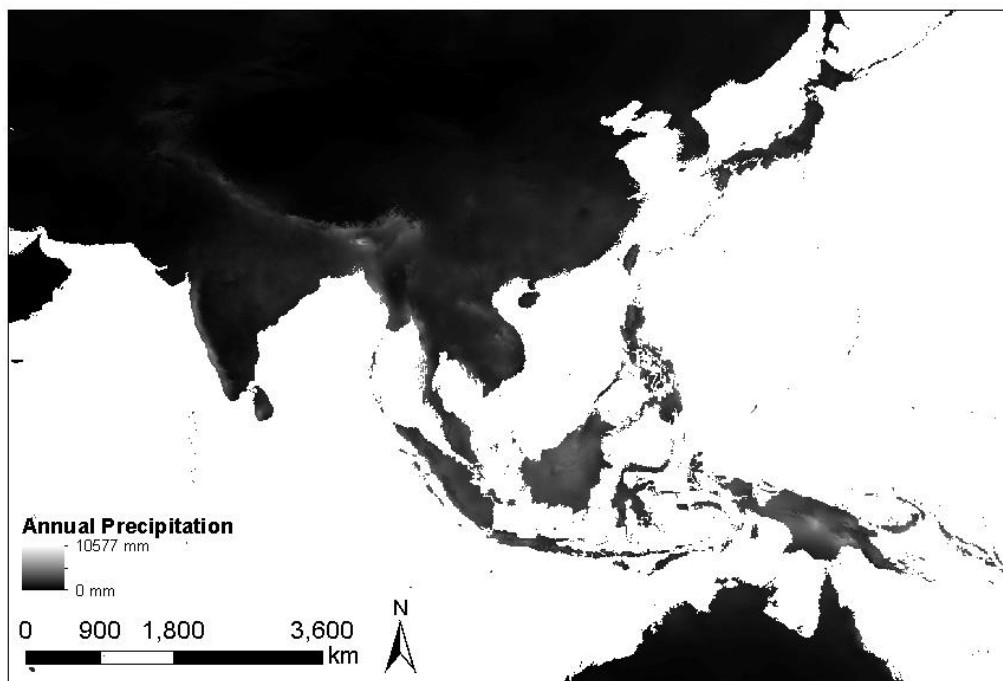
Figure 7: Annual Precipitation

Annual Precipitation

Figure 8: Precipitation in the Wettest Quarter
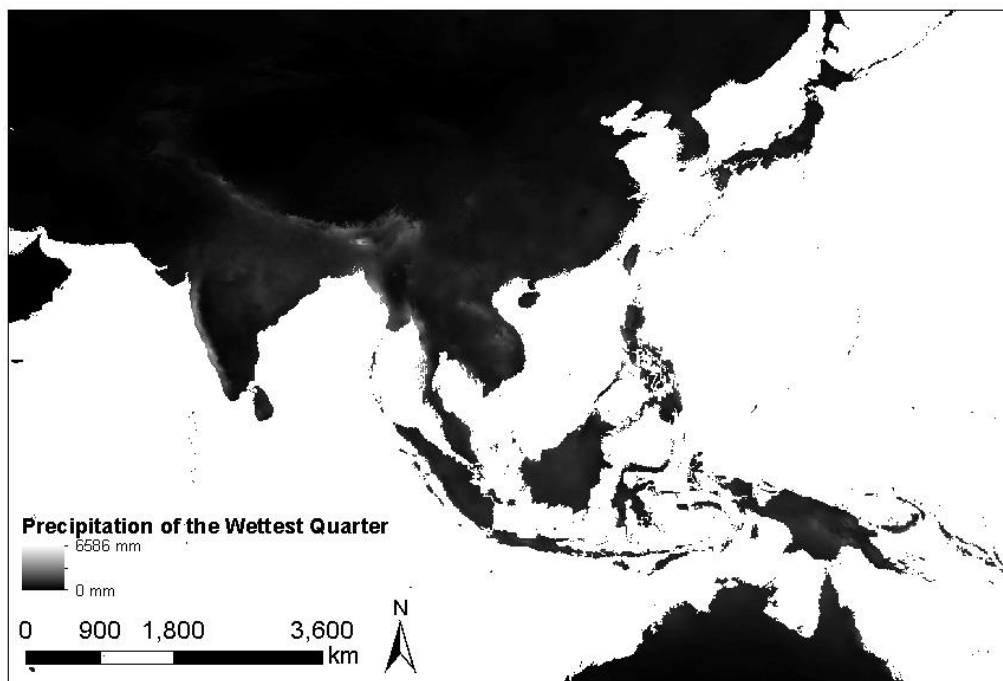

Precipitation of the Wettest Quarter

Table 1: Data on Incidence, Population, and Rice Area for each Country

Country	Administrative Level in which Incidence Data was Reported	Source of Data on Population	Source of Data on Rice Area
Brunei	Country	(U. N. Secretariat, 2011)	(FAOSTAT, 2014)
Bangladesh	Division	(Bangladesh Bureau of Statistics, 2011)	(Bangladesh, 2013)
Cambodia	Country	(U. N. Secretariat, 2011)	(FAOSTAT, 2014)
China	Province	(China, 2010)	(USDA Economic Research Service, 2012)
North Korea	Country	(Secretariat, 2011)	(FAOSTAT, 2014)
India	States/ Federal Territory	(India, 2011)	("Statewise Statistics," 2013)
Indonesia	Province	(Republik Indonesia, 2010)	(Republik Indonesia, 2012)
Laos	Country	(Secretariat, 2011)	(FAOSTAT, 2014)
Malaysia	States/ Federal Territory	(M. D. o. Statistics, 2010)	("Paddy Statistics of Malaysia," 2012)
Myanmar	Country	(Secretariat, 2011)	(FAOSTAT, 2014)
Nepal	District	(N. P. C. Secretariat, 2012)	(C. B. o. Statistics)
Papua New Guinea	Country	(Secretariat, 2011)	(FAOSTAT, 2014)
Philippines	Country	(Secretariat, 2011)	(FAOSTAT, 2014)
Sri Lanka	Country	(Secretariat, 2011)	(FAOSTAT, 2014)
Thailand	Country	(Secretariat, 2011)	(FAOSTAT, 2014)
Timor-Leste	Country	(Secretariat, 2011)	(FAOSTAT, 2014)
Vietnam	Country	(Secretariat, 2011)	(FAOSTAT, 2014)

Table 2: Variables tested in each model

Model	Parameters Tested
Model A	Mean probability of the presence of *Culex tritaeniorhynchus*
Model B	Mean probability of the presence of *Culex tritaeniorhynchus* Population density Percent of population under age 15 Expanding vaccination program
Model C	Population density Percent of population under age 15 Expanding vaccination program Rice Measure Mean altitude Annual mean temperature Temperature annual range Mean temperature of the warmest quarter Annual precipitation Precipitation in the wettest quarter

Table 3:

Variable	Mean	95% Confidence Interval
Population Density (per sq. km)	570.12	(0,4323.30)
Proportion of Population Under Age 15	0.28	(0.17,0.38)
Rice Measure	0.13	(0,0.44)
Altitude (m)	917.38	(0,3105.68)
Probability of Presence of *Culex tritaeniorhynchus*	0.17	(0,0.58)
Annual Temperature (°C)	19.57	(52.26,339.12)
Annual Temperature Range (°C)	21.69	(28.61,405.24)
Temperature of the Warmest Quarter (°C)	24.01	(134.15,346.01)
Annual Precipitation (mm)	1727.24	(318.52,3135.95)
Precipitation in the Wettest Quarter (mm)	903.99	(257.59,1550.39)

Table 4: Results of Valid Models (A, B, C1, and C3)

Model	Model Parameters	Parameter Estimate	Standard Error	Wald 95% Confidence Limits	Wald 95% Confidence Limits	Wald Chi-Square	P Value
MODEL A	Intercept	-10.9456	0.1014	-22.6574	-22.2597	49014.6	<.0001
	Probability of Vector Presence	0.6421	0.3894	-0.1211	1.4054	2.72	0.0992
MODEL B, AFTER ELIMINATION	Intercept	-10.9364	0.1008	-11.134	-10.7389	11773.1	<.0001
	Probability of Vector Presence	0.7797	0.404	-0.0121	1.5715	3.72	0.0536
	Population Density	-0.0001	0	-0.0002	0	2.95	0.086
MODEL C1 FINAL VERSION	Intercept	-10.4022	0.1888	-10.7723	-10.0321	3034.03	<.0001
	Population Density	-0.0001	0	-0.0002	0	5.5	0.019
	Bio7	-0.0018	0.0008	-0.0034	-0.0003	5.26	0.0218
MODEL C3 AFTER ELIMINATION	Intercept	-12.5877	1.1495	-14.8407	-10.3346	119.91	<.0001
	Population Density	-0.0001	0	-0.0001	0	7.79	0.0052
	Bio1	0.1163	0.0118	0.0931	0.1395	96.35	<.0001
	Bio7	0.0325	0.0037	0.0252	0.0398	76.33	<.0001
	Bio10	-0.1196	0.0145	-0.1481	-0.0911	67.58	<.0001

Table 5: Estimated Number of Cases in Divisions of Bangladesh According to the Four Final Models

	Model A	Model B	Model C1	Model C3
Barisal	188	188	162	312
Chittagong	567	540	563	993
Dhaka	1052	958	831	2068
Khulna	360	359	292	693
Rajshahi	416	399	328	1088
Rangpur	298	277	287	488
Sylhet	190	181	190	258

Table 6: Estimated Incidence in Divisions of Bangladesh According to the Four Final Models

	Model A	Model B	Model C1	Model C3
Barisal	2.26	2.26	1.94	3.75
Chittagong	1.99	1.90	1.98	3.49
Dhaka	2.22	2.02	1.75	4.36
Khulna	2.30	2.29	1.86	4.42
Rajshahi	2.25	2.16	1.78	5.89
Rangpur	1.89	1.76	1.82	3.09
Sylhet	1.92	1.83	1.92	2.61