Modelling the Climatic Drivers of Cholera Dynamics in Northern Nigeria Using Generalised Additive Models

Dr. Auwal F. Abdussalam
Department of Geography, Faculty of Science,
Kaduna State University, PMB 2339, Tafawa Balewa Way, Kaduna.
Nigeria
Email: abdussalamauwal@gmail.com

Abstract

Nigeria is one of the countries with high risk of cholera disease. This study modelled the disease dynamics in terms of its interannual variability using climatic variables and other confounding factors in proxy. Monthly counts of clinically diagnosed hospital-reported cases of cholera were obtained from three hospitals in northern Nigeria for 25-year period spanning 1990-2014. Generalized additive models were fitted to aggregated monthly cholera counts. Explanatory variables include monthly time series of maximum and minimum temperature, relative humidity, and rainfall from weather stations nearest to the hospitals, and the number of cases in the previous month. The effects of other unobserved seasonally-varying climatic and non-climatic risk factors that may be related to the disease were collectively accounted for as a flexible monthly-varying smooth function of time in the generalized additive models, s(t). Results reveal that the most important explanatory climatic variables are the monthly means of daily maximum temperature and monthly rainfall totals. Accounting for s(t) in the generalized additive models explains more of the monthly variability of cholera compared to those models that do not account for the unobserved factors that s(t) represents. And this emphasises the importance of including other factors. The skill score statistics of a model version with all explanatory variables lagged by 1-month suggest the potential to predict cholera cases in northern Nigeria up to a month in advance to aid decision makers.

Keywords: Cholera, Nigeria, Climate, GAMs

1.0 Introduction

A strictly human pathogen, *vibrio cholarea* is a gram-negative *bacillus* that mainly causes cholera (Marin, 2013). Despite the existence of several bacteria that can cause cholera, this bacterium is responsible for causing large epidemics (WHO, 2012). The disease is among the infectious diseases that are causing burden of morbidity and mortality in several countries around the world, most especially in Africa.

The natural habitat of the bacterium *Vibrio cholerae* is aquatic, with ideal conditions often reported as brackish waters (Munslow and O'Dempsey, 2010; Singleton et al., 1982; Lipp et al., 2002). Human illness may be caused if an individual ingests food or untreated water which contains sufficient levels of the bacterium (Constantin de Magny et al., 2007b; Hashizume et al., 2008; Talavera and Pérez, 2009). Symptoms include acute diarrhoea and vomiting (Mwasa and

Tchuenche, 2011), if cases are not treated a fatality rate of up to 50% may be expected as a result of kidney failure or dehydration with deaths occurring in as little as two hours after symptoms emerge (Talavera and Pérez, 2009; Kelvin, 2011; Penrose et al., 2010; Rabbani and Greenough, 1999; Munslow and O'Dempsey, 2010).

The survival and abundance of *Vibrio cholerae* in the aquatic environment has been shown to be related to several climate variables. For example, the bacterium is able to rapidly multiply if sufficiently high water temperatures are reached (Reyburn et al, 2007; Munslow and O'Dempsey, 2010; Bouma and Pascual et al, 2001). Further, it has been shown that sunlight and salinity are significantly associated with the abundance of the bacterium in its surroundings (Fernández et al., 2009; Paz, 2009; Hashizume et al., 2011; Rajendran et al., 2011; Cash et al., 2009; Lipp et al., 2002).

The discovery that the bacterium which causes cholera has a relationship with climate variability has led to a wealth of research detailing the exact relationship between climate and the number of cholera cases in areas where the disease is prevalent (Borroto and Martinez-Piedra, 2000). On a less regular basis, extreme weather such as floods, droughts and heat waves may reduce or accentuate the impact of the climate on the regular seasonal peak of a cholera outbreak (Hashizume et al., 2008; Constantin de Magny et al., 2007; Traerup et al., 2010; Emch et al., 2010; Borroto and Martinez-Piedra, 2000; Pascual et al., 2002).

Seasonal cycles in both climate and cholera have been associated around the world (Rajendran et al., 2011; Constantin de Magny et al., 2007b; Constantin de Magny et al., 2007a; Hashizume et al., 2008; Constantin de Magny et al., 2008; Matsuda et al., 2008; Emch et al., 2008; Reyburn et al., 2011; Fernandez et al., 2009; Akanda et al., 2009; Pascual et al., 2002). Relationships have also been found between cholera and climate variability such as the El Niño-Southern Oscillation (ENSO), or the Indian Ocean Dipole (IOD), and so there is a suggestion that both seasonal and interannual variability in cholera might be related with climate (Pascual et al., 2000; Pascual et al., 2002; Hashizume et al., 2011; Matsuda et al., 2008; Emch et al., 2010).

Other non-climatic factors may also play important roles in the risk of cholera transmission. Among these factors are: socioeconomic, cultural, behavioural practices, and migration, etc. Previous studies have established the association between non-climatic factors and cholera risk. For example, cholera has been entitled 'the disease of poverty' (Matsuda et al, 2008; Shahid, 2009; Snowden, 2008). Poverty is often associated with a lack of basic sanitation infrastructure and treated water infrastructure (Rabbani and Greenough, 1999; Pascual et al., 2002), both of which are commonly observed in developing countries (Mandall et al., 2011). If proper sanitation infrastructure does not exist then the bacterium could spread via faecal contamination is much more effective exposing far more people to the bacterium. Similarly if more individuals use untreated water it is logical to theorise that many more people will be exposed to the bacterium (Constantin de Magny et al., 2008; Reiner et al., 2012; Rajendran et al., 2011; Talavera and Perez, 2009; Matsuda et al., 2008; Emch et al., 2008).

Cholera cases are currently being reported from six continents (Charles and Ryan, 2011; Colwell, 1996; Pascual et al., 2002). Areas with substantial problems with the disease include the Indian subcontinent, as well as many areas in South America and Africa (Hashizume et al., 2011; Emch et al., 2008). Recently, the World Health Organisation (WHO) has documented a growth in the number of reported cases worldwide over the last decade (WHO, 2011) with an increase of

24% in 2004-2008 with respect to 2000-2004 (Reyburn et al., 2011). As a result, cholera is now considered a major concern for public health globally by the WHO (Akanda et al., 2009). An outbreak may cause a direct economic burden to a country through increased hospital costs (Traerup et al., 2010; Talavera and Pérez, 2009), while indirect effects can include restricted and reduced food trade with other countries and lower levels of tourism if an outbreak causes sufficient international alarm. The complicated and widespread effects of an outbreak may actually impede development efforts within a country (Traerup et al., 2010). In 2005 and 2006, 95% and 87% (respectively) of the global number of cholera cases were reported from Africa (Paz, 2009; Fernández et al., 2009; Mandal et al., 2011). In 2010, Nigeria reported the second highest total number of cases internationally and the highest number within Africa (WHO, 2011). Historically, the greatest strain of the disease has been felt in Asia, which is the source of all of the seven pandemics (Reyburn et al., 2011; Colwell, 1996). Consequently, much of the current literature surrounding the subject is biased with respect to Asian countries, notably Bangladesh.

For the past thirty years cholera in Africa and Nigeria in particular, has been occurring in sporadic cases, outbreaks, or even in large epidemics mainly caused by two serogroup; *vibrio cholarea* O1 and O139 (Charles and Ryan, 2011; Harris et al., 2012; Marin, 2013). In recent times, newly-observed strains capable of causing more intense morbidity and mortality have been found in numerous parts of Africa and Asia (WHO, 2012).

The current study aims to statistically model the climatic drivers of monthly cholera incidence in northern Nigeria, while collectively accounting for the effects of all unobserved climatic and non-climatic factors that may be related to cholera, such as social and behavioural practices. This paper is among the few that studied the relationship between meteorological conditions (reflected by variables from station observations) and cholera in Nigeria. The model development and results are based on 25-years (1990 – 2014) of clinically diagnosed hospital-reported cases of cholera.

2.0 Materials and methods

The study employed the use of Generalised Additive Models (GAMs) to model the metrological drivers of cholera in northern Nigeria while accounting for other non-meteorological confounding factors. Data were obtained from the Nigeria Meteorological Agency (NIMET) and selected hospitals.

Table 1: Summary of station characteristics and meteorological data at four stations in Northern Nigeria

City	Station ID	Lat	Lon	Elev. (m)	Variables obtained from all stations
					Max. Temp. (^{o}C), Min. Temp. (^{o}C)
Kano	65 0460	12° 03'	08° 32'	476	Rain (mm), Relative Humidity (%)
Sokoto	65 0100	12° 55'	05° 12'	351	
Gusau	65 0150	12° 10'	06° 42'	463	

2.1 Study site and regional meteorological conditions

Northern Nigeria is located in the African Sahel savannah region, and currently has an estimated projected population of over 100 million people based on the 2006 census. The regional climate is characterized by two seasons, a short wet season from June to September, and a prolonged dry season for the remainder of the year. Daytime maximum temperatures remain consistently high throughout the year with maxima during March-May (up to 47°C), while relative humidity is low during the dry season, and increases during the wet season. These mean regional climate conditions are mainly a consequence of the West African Monsoon (WAM) system, which exhibits large spatiotemporal variability (e.g. Cornforth 2012), especially with respect to regional rainfall distributions.

2.2 Epidemiological data

Three hospitals were selected in northern Nigeria for cases collections based on their proximity to meteorological stations with long-term records of measurements, and consistency in reporting disease cases. In Nigeria, the FMoH classifies four categories of hospitals based on ownership status: federal, state and local public hospitals, and private hospitals. Personal communication with FMoH staff prior to the data collection indicated that the state-owned hospitals best suited the above criteria because most of the infectious disease cases are treated at these hospitals.

In addition to the monthly case data from these three hospitals (1990-2014), weekly records of suspected cholera cases at the district level (from all hospitals in a district) from the respective districts between 2007 and 2014 was obtained from the WHO for the purpose of evaluation. The proportion of cases between WHO the and hospital records was rather constant across the study period, providing evidence that the records are of sufficient quality and that the hospital records are regionally representative.

2.3 Meteorological data

Digital records of four meteorological variables from airport-based stations in each of the three cities were obtained from the Nigerian Meteorological Agency (Table 1) between 1990 and 2014. Quality control was carried out for all variables, specifically maximum and minimum temperatures and rainfall were checked using an R-based software tool *Rclimdex.r 1.0*. This has been developed and maintained by the Climate Research Division (CRD, 2008) of the Meteorological Service of Canada on behalf of the Expert Team on Climate Change Detection and Indices (ETCCDI). This tool is capable of identifying duplicate dates, out-of-range values based on a defined threshold, outliers, coherence between maximum and minimum temperatures (Tmax > Tmin), and consecutive days with equal values. A few values of maximum temperature that were below their daily minimum counterparts were detected; also a few outliers in rainfall values were identified and corrected. Corrections were made using information from days before or after the problematic value (e.g., Aguilar et al., 2005). While relative humidity was manually quality controlled by removing obvious spurious values based on knowledge of the regional climate; too few values were removed to affect the overall quality and continuity of the meteorological data. Generally, less than 7% of the data was affected in this process. Monthly averages, totals or percentages were then computed from the quality controlled daily values.

Both lag zero and one-month lagged meteorological data were included as explanatory variables in the development of the models.

2.4 Model Development

This study employed the use Generalised Additive Model (GAM) to investigate the relationships between meteorological drivers and cholera dynamics in northern Nigeria. For statistical model development, monthly cholera counts for the three hospitals in (Kano, Sokoto, and Gusau) were aggregated, and variables of the corresponding three meteorological stations averaged. Monthly cholera counts were aggregated in order to minimize the effect of bias in reporting to individual hospitals, and as well to have a regional perspective which is the intent of this study.

GAMs are a flexible extension of GLMs and are comprehensively described by Hastie and Tibshirani, (1999). Because of the additive smoothing function within the GAM (discussed below), it can collectively accounted (albeit not specifically) for the effects of all unobserved climatic and non-climatic factors that may be related to cholera.

GAM was subsequently applied to model the meteorological drivers of cholera on the monthly variability of clinically diagnosed hospital-reported cases of cholera. During model development, several predictor variables such as previous disease incidence; warm days and nights; monthly maximum values of daily maximum and minimum temperatures (above 98th percentile); and mean monthly values of meteorological variables were considered. Also different lags of meteorological variables ranging from 0 – to 2 month were tested in order to select the best combination of predictor variables that could better explain the variability of diseases. Lag zero and one month lagged meteorological variables appeared to be best and were selected for model input. Collinearity diagnostics and autocorrelation checks were performed, and explanatory variables were selected through a process of manually entering and removing variables from the model in a stepwise selection process, with a criterion of elimination being a p-value <=0.05 when testing the significance of the coefficient estimate. The explanatory variables include meteorological variables aggregated monthly between the three selected stations (Table 1), as well as the previous incidence of cholera in some models. Additionally, all other unobserved seasonally-varying climatic and non-climatic factors that may influence the disease were represented in GAMs by a smooth function of time, *s(t)*, which was modelled as a low-degree cubic spline that changes monthly over the course of the annual cycle (Dukic et al., 2012). The variable *s(t)* is the so-called "additive function" characteristic of GAMs, and captures the seasonality effect in a way that can be viewed as a smoothed analogue of the month-specific effects. It is assumed that *s(t)* is common to all years (i.e., that there is no intercept that can be applied to adjust the function for a specific year). The assumption is that the clinically diagnosed hospital cholera counts, y_{it} follow independent Poisson distributions (thus a log-link function was used (Cameron and Trivedi, 1998)), with mean μ_{it}, where i =1,..., 25 denotes the years, and t =1,...,12 denotes the month within each year. The GAM formulation is thus:

$$\log(\mu_{it}) = s(t) + X_{it}\beta$$

The expected cholera count in year i in month t therefore depends upon the vector of coefficients *β*, which contains the effects of climate variables collected in the covariate matrix X_{it}, and upon the effects of the unobserved seasonally-varying factors, *s(t)*. The GAM is fitted and the coefficients *β* and the parameters for the smooth function *s(t)* are estimated.

Three GAMs were fitted (denoted as models A, B, C), model A was fitted with non-lagged climatic variables and cases from the previous month; model B, with only the 1-month lagged explanatory climatic variables and cases from the previous month; and model C is the same as GAM A, except previous cases were excluded. All three GAMs were tested for a variety of degrees-of-freedom (DOF) for the fit of $s(t)$, but it was found that those models in which $s(t)$ has 4 DOF have the best fit. In summary, model A is intended to be optimal *explanatory* model of cholera cases, whereas model B by using only lagged variables is intended to be the optimal *predictive* model (with 1-month lead time). Model C, which does not include previous cases, is intended to be used for future climate change studies (since the number of cases in the previous month is unknown in the future). The best models were selected by minimising the Bayesian Information Criteria (BIC) (Dobson, 2010). Variable selections were made separately for each model, although the same variables were retained in all models. The retained variables include mean monthly maximum and minimum temperatures, precipitation totals, average relative humidity, and previous cases. All models were fit within R statistical software (R Core Team, 2015).

2.5 Model Validation and Relative Influence

The robustness and accuracy of models were assessed using the cross validation correlation (CVC) (Kohavi, 1995), the root mean square error (RMSE) (Geisser, 1993), and the skill score (Murphy, 1998) techniques. All three statistics were computed for observed versus predicted values for each model. To perform the cross validation, the data was partitioned into 3 consecutive subsets of equal length. One of these subsets is then successively excluded, fitted the model on the remaining data and computed the fitted values for the excluded subset. The fitted values that were obtained were then combined into one time series for ease of comparison with the "full model" (i.e., based on all 25 years of data). The skill score provides a measure of the prediction accuracy of the models by comparing the models' predicted RMSE, E_{pre}, with that of a reference model E_{ref}.

$$\text{Skill score} = 1 - (E_{pre} / E_{ref})$$

In this case, E_{pre} represents the RMSE of the monthly model-predicted cases compared to the observed cases, while, E_{ref} represents the RMSE of the long term monthly mean of the observed cholera cases, also compared to the observed cases for each month and year. The reference model is thus a persistence model: even if there is no model at all, they could predict cases for a given year and month by assuming that they will equal the long-term average of cases for that month, based on observations from other years. The skill score is the percentage of improvement or deterioration of a given model's RMSE with respect to the reference model.

To gain a perspective on the average comparative importance of each covariate for a given month, relative influence (RI) was computed by estimating the effect of each covariate with respect to all the covariates in the model, based on the long-term monthly means of each covariate over the 25 year period, as well as the monthly value of $s(t)$. The RI for a GAM is calculated as a percentage of all terms for a given month as follows:

$$RI = 100 * \frac{\overline{X}_{t,\vartheta}\beta_{\vartheta}}{\left[|s(t)| + \sum_{v=1}^{n}\overline{X}_{t,v}|\beta_v|\right]}$$

Where in the numerator $\overline{X}_{t,\vartheta}$ is the long term mean of X for a given month, t, and ϑ denotes a particular variable within the vector of X (e.g., Tmin), and β_{ϑ} is the coefficient that corresponds to that particular variable. In the denominator $s(t)$ is the month-specific additive function (which does not vary by year) and $\sum_{v=1}^{n}\overline{X}_{t,v}|\beta_v|$, is the sum of all other model terms $\overline{X}\beta$ over all variables from **v=1** to **n** (maximum and minimum temperature, relative humidity, dust, etc), for a given month t. The absolute values are taken for the coefficients, β and $s(t)$ in the denominator, in order to omit negative terms in the equation, otherwise, the RI of a given term may be inflated due to cancellation of negative and positive terms.

Table 2: Estimates for GAM A, B and C. Model A is fitted non-lagged climate variables, B is fitted with only 1-month lagged climate covariates, while C has the same composition with A, but previous cases are not included.

Predictors	Model A			Model B			Model C		
	coef.	*std. error*	*p-value*	*coef.*	*std. error*	*p-value*	*coef.*	*std. error*	*p-value*
Monthly mean Tmax (C°)	0.351	0.027	0.001	0.316	0.006	0.001	0.362	0.023	0.000
Monthly mean Tmin (C°)	0.203	0.017	0.101	0.157	0.008	0.007	0.256	0.039	0.028
Rain (mm)	0.256	0.141	0.021	0.295	0.021	0.101	0.286	0.173	0.031
Relative humidity (%)	0.056	0.017	0.001	0.142	0.103	0.001	0.142	0.102	0.041
Previous cases	0.107	0.005	0.201	0.132	0.012	0.104	na	na	na

Abbreviations: coef, coefficient; std error, standard error

3.0 Results

In recent years Nigeria has experienced increase in cholera cases and deaths, for example, in 2010 alone, between the month of January and December, the country reported about 41,784 cases and 1716 deaths (CFR 4.1%) from 222 districts in 18 states (WHO, 2012) in which most of the cases comes from the northern part. Generally, both the individual and aggregated counts of the monthly hospital-reported cholera cases exhibit a marked annual cycle, with yearly disease maxima occurring between the month of April and August.

The estimated effect of other additional confounding factors represented in the GAM as a smooth function of time $s(t)$ is shown in Figure 1 for model A . The shape of the estimated function of both models is similar and follows the seasonality of the disease, with the months of April – August having the highest values of $s(t)$. The model estimates and standard errors for GAMs are presented in Table 2. All variables correlated positively with disease cases in both models, mean monthly maximum temperature and monthly rainfall totals appear to be the most important predictors in both the models. Model with meteorological variables lagged by one month appears to capture the monthly and interannual variability of the cholera cases more accurately.
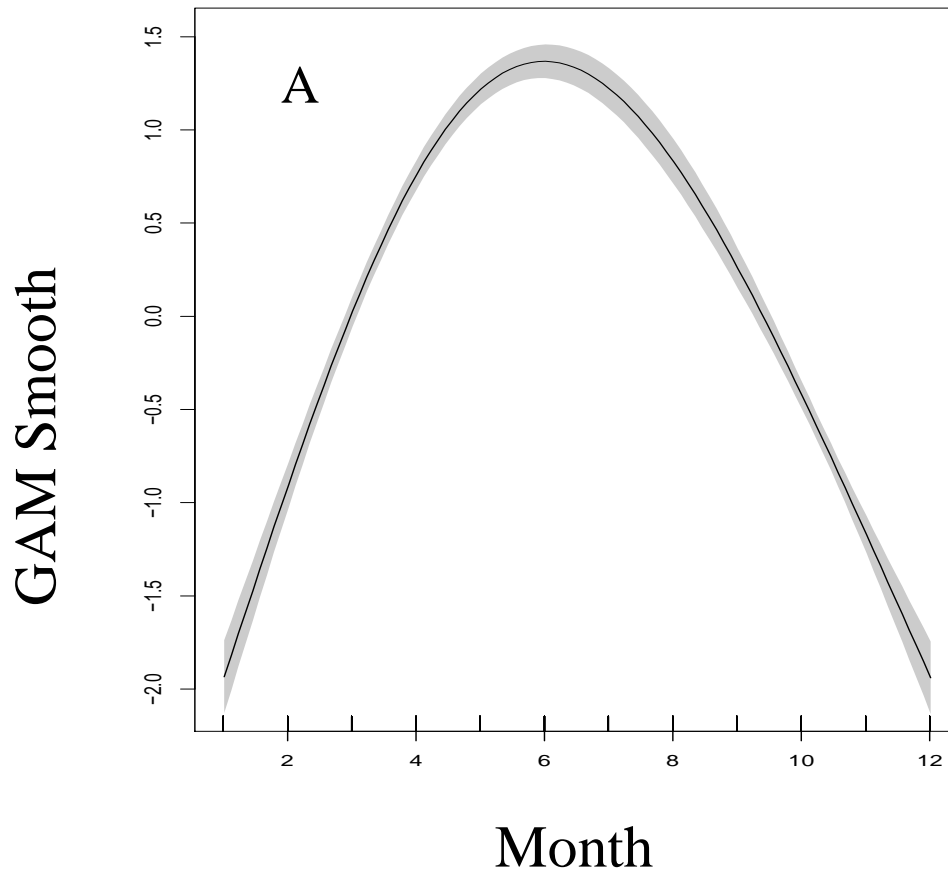
Figure 1: Estimated smooth function of time, $\hat{s}(t)$, across 12 months, for model **A**. The shaded area shows the 95% confidence interval about $\hat{s}(t)$.

Cross validation statistics are presented in Table 3, all three models show good skill. Lagged model (B) has improved statistic values if compared with non-lagged model A as measured by CVC and skill score (0.73 and 0.70). This demonstrates a good indication for the possibility of potentially predicting cholera cases with a month time lead in this region. Model C which is specifically designed for climate change studies also have a good skill. Predicted cases have a cross-validation correlation of 0.68 with 1990-2014 observed cases, and a skill score of 0.62, meaning the root-mean square error of the predicted cases yielded a 60% improvement over assuming the long term mean of cases is the value in each year (i.e., "persistence").

The relative influence for the four months in which cholera cases are generally in both models; mean monthly maximum temperature and monthly rainfall totals shows a comparatively important influence across the four months with the highest RI of about 35% and 28% across the months. The influence of mean monthly minimum temperature and that of average relative humidity remains almost the same across the months. The function $s(t)$ which accounts for unobserved explanatory variables, varies in influence from about 8-15%, while previous cases also remains relatively the same, with the highest influence in the month of July.

4.0 Discussion

In this study, GAM statistical techniques were employed to model the meteorological drivers and socioeconomic conditions on the interannual variability of cholera in northern Nigeria. GAMs were used to model the monthly aggregate counts of clinically diagnosed hospital-reported cholera cases from 1990 to 2014 in northern Nigeria, explanatory variables in this models includes mean monthly maximum and minimum temperatures, monthly rainfall totals, monthly average relative humidity, and 1-month previous incidence and *s(t).* The hospital case data exhibit a marked annual cycle, with yearly disease maxima occurring between the months of April and August, while the state level annual data indicates increase in cases of cholera with most of the cases being reported from the northern part of the country. Model performance was estimated by three-fold cross validation, RMSE, and skill score.

Table 3: Model validation results given CVC and Skill Score

Models	Kendal Correlation		Skill Score	
	Full[a]	**CV**[b]	**Full**[a]	**CV**[b]
A	0.703	0.685	0.659	0.641
B	0.734	0.713	0.704	0.672
C	0.681	0.652	0.623	0.592

[a]'Full' and [b]'CV' stands for full and 3-fold cross validated models respectively

The results indicated that both *explanatory* (fitted with non-lagged climatic variables) and *predictive* (fitted with only 1-month lagged climatic variables) models showed similar capabilities to fit values of cholera incidence. The best explanatory model (B) had a CVC of 0.73 with the observations, and a skill score of 0.70 out of 1.0. Additionally, the cross-validation version of model A (that was developed by systematically omitting the first, middle and last 8 years of case data) exhibited nearly identical statistics to the "full" model that was fit using all 25-years of case data, indicating that the model performance is not sensitive to the period chosen for development. The *predictive* model B was good in terms of fitting performance, as such; it has strong potential for short-term disease prediction.

Although, many studies have explained for the role of meteorological conditions in the incidence of cholera (e.g., de Magnay et al., 2012; Fernendez et al., 2009; Trearup, 2010), our results emphasis the need for factoring socioeconomic conditions in modelling frameworks of this kind. Also, results from the study have supported the findings that temperature, rainfall and humidity may facilitate both the transmission and the development of invasive cholera. The influence of climate on the cholera dynamic has been well established in Asia (e.g., Bouma and Pascual, 2001; Pascual et al., 2000), South America (e.g., Colwell, 1996; Speelmon et al., 2000), and in Africa (e.g., de Magnay et al., 2012; Fernendez et al., 2009; Trearup, 2010). The link between temperature increase and the amplification of cholera incidence have been well reported (Colwell, 2002; de Magney et al., 2008; Lipp et al., 2002; Louis et al., 2003). Also, cholera outbreaks are characterised by strong seasonality corresponding with heavy rainfall and warm temperatures (Reyburn et al., 2011). Islam et al. (2009) reported significance of temperature and sunshine hours to cholera outbreaks both in summer and winter seasons in Matlab, Bangladesh, while rainfall and associated river levels were found to have influence on cholera patterns in Bangladesh (Akanda et al., 2009; Hashizume et al., 2007).

In northern Nigeria, extreme wet and dry conditions may be playing important roles in the disease prevalence. For example, extreme events like flooding and drought may increase the risk of cholera, although Carrel et al. (2010) found no significant correlation between high cholera incidence and households residing in flood control areas in Matlab, Bagledesh. Contamination of drinking water may be caused by heavy monsoon floods; on the other hand drought may help to facilitate the growth of bacteria in ponds and rivers. Natural disasters like floods, earthquakes, storms, and drought are usually related to an increasing risk of water-borne infectious diseases (Watson et al., 2007). For example, the Haitian earthquake on January 12th 2009 led to a cholera epidemic (Farmer et al., 2011).

Unobserved seasonally-varying non-climatic factors such as accessibility to safe drinking water, literacy, population density, poverty, societal and behavioural practices – which are represented in a very basic sense by the function $s(t)$ – are likely to enhance the transmission of the disease in this region. For example, in urban slums in norther cities like Kano, during wet season, cholera pathogens may be easily transported into domestic water sources like wells, which might subsequently enhance transmission. Additionally, since the disease is contagious, it could be easily transmitted via food due to overcrowding in such areas.

5.0 Conclusions

Results from this study indicates the role of specific meteorological conditions in explaining and predicting monthly cholera variability in northwest Nigeria, and also emphasize the importance of additional risk factors that may be linked to societal and behavioural practices. With respect to the latter, we are currently limited to only collectively accounting for these unobserved seasonally-varying climatic and non-climatic risk factors via functions such as $s(t)$. Identifying and quantifying such factors, and improving disease surveillance, will likely enhance our understanding and ability to predict and reduce cholera incidence.

The results emphasises the importance of including socioeconomic factors in studies of this nature, this is because socioeconomic variables help in explaining the disease. The new data provided by this study will serve as a basis for the potential prediction of cholera in Nigeria which could help authorities in controlling or even avoiding outbreaks. Also, with the growing concern of the potential impact of climate change on the dynamic of infectious diseases in the future, this study has provided a background for assessing the future impact, which is the next step of this study.

6.0 Acknowledgments

References

Aguilar, E., Barry, A. A., Brunet, M., Ekang, L., Fernandes, A., Massoukina, M., Mbah, J., Mhanda, A., Do Nascimento, D. J., Peterson, T. C., Umba, O. T., Tomou, M. & Zhang, X. (2009). Changes in temperature and precipitation extremes in western central Africa, Guinea Conakry, and Zimbabwe, 1955-2006. Journal of Geophysical Research-Atmospheres, 114.

Akanda, A.S., Jutla.A.S. and Islam.S. (2009) Dual peak cholera transmission in Bengal Delta: A hydro climatological explanation. Geophysical Research Letters, 36, L19401.

Borroto, R.J. and Martinez-Piedra, R. (2000) Geographical patterns of cholera in Mexico, 1991-1996. International Journal of Epidemiology, 29, 764-772.

Bouma, M.J. and Pascual, M. (2001) Seasonal and interannual cycles of endemic cholera in Bengal 1891-1940 in relation to climate and geography. Hydrobiologia, 460, 147-156.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*: Cambridge University Press.

Carrel, M., Voss, P., Streatfield, P. K., Yunus, M., & Emch, M. (2010). Protection from annual flooding is correlated with increased cholera prevalence in Bangladesh: a zero-inflated regression analysis. *Environmental Health, 9*. doi: 10.1186/1476-069x-9-13.

Cash, B.A., Rodo, X. and Kinter, J.L. (2009) Links between Tropical Pacific SST and Cholera Incidence in Bangladesh: Role of the Western Tropical and Central Extratropical Pacific. Journal of Climate, 22, 1641-1660.

Charles, R.C. and Ryan, E.T. (2011) Cholera in the 21st century. Current Opinion in Infectious Diseases, 24, 472-477.

Colwell, R.R. (1996) Global Climate and Infections Disease: The Cholera Paradigm. Science, 274, 2025-2031.

Colwell, R. R. (2002). A voyage of discovery: cholera, climate and complexity. *Environmental Microbiology, 4*(2), 67-69. doi: 10.1046/j.1462-2920.2002.00270.x.

CRD. (2012, April 14). Climate Research Division. RClimDex 1.0. Retreaved from http://cccma. seos.uvic.ca /ETCCDMI/software.shtml]

Cornforth, R. (2012). Overview of the West African Monsoon 2011. *Weather, 67*(3), 59-65. doi: 10.1002/wea.1896.

Constantin de Magny, G., Guégan, J., Petit, M. and Cazelles, B. (2007a) Regional-scale climate-variability synchrony of cholera epidemics in West Africa. BMC Infectious Diseases, 7, 20.

Constantin de Magny, G., Cazelles, B. and Guégan, J. (2007b) Cholera Threat to Humans in Ghana is Influenced by Both Global and Regional Climate Variability. EcoHealth, 3, 223-231.

Constantin de Magny, G., Murtugudde, R., Sapiano, M.R.P., Nizam, A., Brown, C.W., Busalacchi, A.J., Yunus, M., Balakrish Nair, G., Gil, A.I., Lanata, C.F., Calkins, J., Manna, B., Rajendran, K., Bhattacharya, M.K., Huq, A., Bradley Sack, R. and Colwell, R.R. (2008) Environmental signatures associated with cholera epidemics. Proceedings of the National Academy of Sciences, 105(46), 17676-17681.

Constantin de Magny, G. C., & Colwell, R. R. (2009). Cholera and climate: a demonstrated relationship. Transactions of the American Clinical and Climatological Association, 120, 119-128.

Constantin de Magny, G. C., Thiaw, W., Kumar, V., Manga, N. M., Diop, B. M., Gueye, L., Colwell, R. R. (2012). Cholera outbreak in senegal in 2005: was climate a factor? PLoS One, 7(8), e44577.

Dobson, A. J. (2010). *An introduction to generalized linear models*: CRC press.

Dukic, V., Hayden, M., Forgor, A. A., Hopson, T., Akweongo, P., Hodgson, A., Monaghan, A., Wiedinmyer, C., Yoksas, T., Thomson, M. C., Trzaska, S. & Pandya, R. (2012). The Role of Weather in Meningitis Outbreaks in Navrongo, Ghana: A Generalized Additive

Modeling Approach. *Journal of Agricultural Biological and Environmental Statistics,* 17**,** 442-460.

Emch, M., Yunus, M., Escamilla, V., Feldacker, C. and Ali, M. (2010) Local population and regional environmental drivers of cholera in Bangladesh. Environmental Health, 9(2), 1-8.

Emch, M., Feldacker, C., Islam, M.S. and Ali, M. (2008) Seasonality of cholera from 1974 to 2005: a review of global patterns. International Journal of Health Geographics, 7(31).

Farmer, P., Almazor, C.P., Bahnsen, E.T., Barry, D., Bazile, J., Bloom, B.R., Bose, N., et al. (2011) Meeting Cholera's Challenge to Haiti and the World: A Joint Statement on Cholera Prevention and Care. PLOS Neglected Tropical Diseases, 5(5), e1145.

Fernandez, M.A.L., Bauernfeind, A., Jimenez, J.D., Gil, C.L., El Omeiri, N. and Gilbert, D.H. (2009) Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003-2006: analysis of a time series. Transactions of the Royal Society of Tropical Medicine and Hygiene, 103, 137-143.

Geisser, S. (1993). *Predictive inference: an introduction* (Vol. 55): CRC Press.

Harris, J. B., LaRocque, R. C., Qadri, F., Ryan, E. T., & Calderwood, S. B. (2012). Cholera. Lancet, 379(9835), 2466-2476.

Hashizume, M., Faruque, A.S.G., Terao, T., Yunus, M., Streatfield, K., Yamamoto, T. and Moji, K. (2011) The Indian Ocean Dipole and Cholera Incidence in Bangladesh: A Time-Series Analysis. Environmental Health Perspectives, 119(2), 239-244.

Hashizume, M., Armstrong, B., Hajat, S., Wagatsuma, Y., Faruque, A.S.G., Hayashi, T. and Sack, D.A. (2008) The Effect of Rainfall on the Incidence of Cholera in Bangladesh. Epidemiology, 19(1), 103-110.

Hashizume, M., Armstrong, B., Hajat, S., Wagatsuma, Y., Faruque, A. S., Hayashi, T. and Sack, D. A., (2007). Association between climate variability and hospital visits for non-cholera diarrhoea in Bangladesh: effects and vulnerable groups. International Journal of Epidemiology, 36, 1030-1037.

Islam, M. S., Sharker, M. A. Y., Rheman, S., Hossain, S., Mahmud, Z. H., Uddin, A. M. K., Ernst, R. (2009). Effects of local climate variability on transmission dynamics of cholera in Matlab, Bangladesh. Transactions of the Royal Society of Tropical Medicine and Hygiene, 103(11), 1165-1170. doi: 10.1016/j.trstmh.2009.04.016.

Kelvin, A.A. (2011) Cholera outbreak in the Republic of Congo, the Democratic Republic of Congo, and cholera worldwide. Journal of Infection in Developing countries, 5(10), 688-691.

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection.* Paper presented at the International joint Conference on artificial intelligence.

Lipp, E.K., Huq, A. and Colwell, R.R. (2002) Effects of Global Climate on Infectious Disease: the Cholera Model. Clinical Microbiology Reviews, 15(4), 757-770.

Louis, V. R., Russek-Cohen, E., Choopun, N., Rivera, I. N. G., Gangle, B., Jiang, S. C., Colwell, R. R. (2003). Predictability of Vibrio cholerae in Chesapeake Bay. *Applied and Environmental Microbiology, 69*(5), 2773-2785. doi: 10.1128/aem.69.5.2773-2785.2003.

Mandal, S., Mandal, M.D. and Pal, N.K. (2011) Cholera: a great global concern. Asian Pacific Journal of Tropical Medicene, 4(7), 573-580

Marin, M. A., Thompson, C. C., Freitas, F. S., Fonseca, E. L., Aboderin, A. O., Zailani, S. B.,Vicente, A. C. P. (2013). Cholera Outbreaks in Nigeria Are Associated with Multidrug

Resistant Atypical El Tor and Non-O1/Non-O139 Vibrio cholerae. Plos Neglected Tropical Diseases, 7(2). doi: 10.1371/journal.pntd.0002049.

Matsuda, F., Ishimura, S., Wagatsuma, Y., Higashi, T., Hayashi, T., Faruque, A.S.G., Sack, D.A. and Nishibuchi,M. (2008) Prediction of epidemic cholera due to *Vibrio cholera* O1 in children younger than 10 years using climate data in Bangladesh. Epidemiology and Infection, 136, 73-79.

Munslow, B. and O'Dempsey, T. (2010) Globalisation and Climate Change in Asia: the urban health impact. Third World Quarterly, 31(8), 1339-1356.

Murphy, A. H. (1988). Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. *Monthly Weather Review, 116*(12), 2417-2424. doi: 10.1175/1520-0493(1988)116<2417:ssbotm>2.0.co;2.

Mwasa, A. and Tchuenche, J.M. (2011) Mathematical analysis of a cholera model with public health interventions. Biosystems, 105, 190-200.

Pascual, M., Bouma, M.J. and Dobson, A.P. (2002) Cholera and climate: revisiting the quantitative evidence. Microbes and Infection, 4, 237-245.

Pascual, M., Rodó, X., Ellner, S.P., Colwell, R. and Bouma, M.J. (2000) Cholera Dynamics and El-Niño-Southern Oscillation. Science, 289, 1766-1769.

Paz, S. (2009) Impact of Temperature Variability on Cholera Incidence in Southeastern Africa, 1971-2006. EcoHealth, 6, 340-345.

Penrose, K., Caldas de Castro, M., Werema, J. and Ryan, E.T. (2010) Informal Urban Settlements and Cholera Risk in Dar es Salaam, Tanzania. PLOS Neglected Tropical Diseases, 4(3), e631.

Rabbani, G.H. and Greenough, W.B. (1999) Food as a Vehicle of Transmission of Cholera. Journal of Diarrhoeal Diseases Research, 17(1), 1-9.

R Core Team. (2015, May 15). R language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Retrieved from http://www.R-project.org

Rajendran, K., Sumi, A., Bhattachariya, M.K., Manna, B., Sur, D., Kobaashi, N. and Ramamurthy, T. (2011) Influence of relative humidity in *Vibrio cholera* infection: A time series model. Indian Journal of Medical Research, 133, 138-145.

Reiner Jr, R.C., King, A.A., Emch, M., Yunus, M., Faruque, A.S.G. and Pascual, M. (2012) Highly localized sensitivity to climate forcing drives endemic cholera in a megacity. Proceedings of the National Academy of Sciences, 109(6), 2033-2036.

Reyburn, R., Kim, D.R., Emch, M., Khatib, A., Von Seidlein, L. and Ali, M. (2011) Climate Variability and the Outbreaks of Cholera in Zanzibar, East Africa: A Time Series Analysis. The American Journal of Tropical Medicine and Hygiene, 84(6), 862-869.

Shahid, S. (2009) Probable Impacts of Climate Change on Public Health in Bangladesh. Asia-Pacific Journal of Public Health, 22(3), 310-319.

Singleton, F.L., Attwell, R., Jangi, S. and Colwell, R.R. (1982) Effects of temperature and salinity on Vibrio cholera growth. Applied and Environmental Microbiology, 44(5), 1047-1058.

Snowden, F.M. (2008) Emerging and reemerging diseases: a historical perspective. Immunological Reviews, 225, 9-26.

Speelmon, E. C., Checkley, W., Gilman, R. H., Patz, J., Calderon, M., & Manga, S. (2000). Cholera incidence and El Nino-related higher ambient temperature. Jama-Journal of the American Medical Association, 283(23), 3072-3074. doi: 10.1001/jama.283.23.3072-a.

Talavera, A. and Pérez, E. (2009) Is cholera disease associated with poverty? Journal of Infection in Developing Countries, 3(6), 408-411.

Traerup, S.L.M., Ortiz, R.A. and Markandya, A. (2010) The Health Impacts of Climate Change: A Study of Cholera in Tanzania. BC3 Working Paper Series 2010-01. Basque Centre for Climate Change (BC3), Bilbao, Spain.

Watson, J.T., Gayer, M. and Connolly, M.A. (2007) Epidemics after Natural Disasters. Emerging Infectious Diseases, 13(1), 1-5.

WHO. (2011, August 11). Prevention and control of cholera outbreaks: World Health Organisation policy and reccomendations. Retreived from http://www.who.int/cholera/technical/prevention/control/en/

WHO. (2012, June 17). Cholera fact sheet N107, 2012. Retreaved from http://www.who.int/mediacentre/factsheets/fs107/en/]