SPATIO-TEMPORAL MODELLING OF ENDEMIC-EPIDEMIC CHOLERA IN NIGERIA

Abstract

A multivariate negative binomial model to capture spatio-temporal endemic-epidemic of infectious disease is used in this study to describe the spatial and temporal pattern of cholera outbreak in Nigeria. Weights are used in the epidemic component for the disease spread across the geographical neighbouring regions and temporal variation of disease incidence is modelled using endemic component. Weekly count data on cholera from the Nigeria Centre for Disease Control, Surveillance and Epidemiology Department (NCDC SED) between January 1st and November 19th, 2018 was used to illustrate the model. The study has shown that negative binomial model with both the seasonality and autoregressive components provided adequate fit for the cholera count data and also perform better than without seasonality and auto regression for capturing the spatio-temporal pattern of cholera disease.

Keywords: disease; surveillance; epidemic proportion; negative binomial model; spatio-temporal dependence; cholera

1.0 Background

Cholera is an acute watery diarrhoeal disease caused by the ingestion of food or water contaminated with the toxigenic strains of Vibrio cholerae serogroups O1 or O139 Clemens *et al.* (2017). Cholera is often characterised by watery diarrhoea, with or without vomiting, and severe dehydration, resulting in death if left untreated (Microbiology Society, 2016). The Case Fatality Rate (CFR) from untreated cholera can be as high as 30–50%, but prompt administration of rehydration therapy can reduce it to as low as 1% as reported in (Microbiology Society, 2016). The global estimates for cholera cases and deaths are about 2.9 million and 95,000 per year, respectively by Ali *et al*. (2015), disproportionately affecting sub-Saharan African countries especially since the onset of the seventh pandemic in 1961 Clemens *et al.* (2017). For instance, 17 African countries reported over 150,000 cholera cases from all the outbreaks in 2017.

In Nigeria, cholera is an endemic and seasonal disease, occurring annually mostly during the rainy season and more often in areas with poor sanitation. Cholera transmission is closely linked to inadequate access to clean water and sanitation facilities from regions with dense populations to other neighbouring regions. Historically, Nigeria has experienced several cholera outbreaks characterised by high CFRs, notable ones being the epidemic of 1991 which resulted in 59,478 cases and 7654 deaths, and the CFR of 12.9% reported for that outbreak remains the highest for the country to date. Furthermore, another major cholera outbreak occurred in Kano state in March, 1999, with cases spreading to Adamawa and Edo states by May of that year; and the outbreak resulted in 26,358 cases and 2085 deaths. From January to December 2010, Nigeria reported 41,787 cases and 1716 deaths (CFR 4.1%) across 18 states (Dalhat *et al.*, 2014). The last major cholera outbreak prior to 2018 was in 2014, during which the number of cases recorded surpassed over half of the number of cases recorded between 2012 and 2013 as well as between 2015 and 2017 (Dalhat *et al.,* 2014). In line with global evidence, however, it is likely that cholera burden in Nigeria is underestimated due to factors ranging from differences in case definitions and completeness to social, political, and economic disincentives for reporting cholera (Mengel *et al.,*2014). Nonetheless, in response to the increasing global cholera burden, the Global Task Force on Cholera Control (GTFCC) in (2017) launched the Global Roadmap Strategies which seek to reduce cholera-related deaths by 90% as well as eliminate cholera infections in at least 20 out of the 47 endemic countries by 2030 (Global Task Force on Cholera Control, 2017). Nigeria has taken fundamental steps toward attaining these goals by deploying Oral Cholera Vaccines (OCVs) in cholera hotspots. Since the first deployment in September 2017 to date, million doses of OCVs have been deployed, albeit in a reactive context, across several hotspot areas, predominantly in the northern states (example, Borno, Bauchi, Yobe and Adamawa states) of Nigeria. Also in line with the GTFCC recommendations, Nigeria is finalising its National

Strategic Plan of Action on Cholera Control. Despite the aforementioned efforts toward cholera prevention and control, the cholera outbreak of 2018 however, reaffirms the serious public health threat of cholera and, importantly, the need for the country to adopt holistic counter measures.

In recent years, there has been an increase in research activity regarding stochastic models for a complete understanding of disease transmission and persistence. Some novel approaches have been developed in (Becker, 1989 and Becker & Britton.1999), although these often involve making unrealistic modelling assumptions, which in turn affects the reliability of the conclusions. The probably best studied stochastic models for the spread of an infectious disease over time are mechanistic models such as the chain binomial model and related continuous time models such as the susceptible–infected–removed (SIR) model (Andersson & Britton, 2000; Daley & Gani, 1999) . These models tried to model the infection mechanism of a disease, based on data of a completely observed infection process. Morton and Finkenstädt (2005) proposed a stochastic discrete time version of the SIR model for infectious diseases. A disease is assumed to transmit within and between communities when infected and susceptible individuals interact. Markov chain Monte Carlo methods were used to make inference about these unobserved populations and the unknown parameters of interest. It is important to note that in a surveillance setting, data on infectious diseases are typically available as rates or counts which are aggregated on spatial and temporal basis. Recently, there has been much interest in the statistical analysis of multivariate time series of disease counts, where each component, for example, corresponds to the number of counts in a specific geographical region or in a certain age group. Kleinman *et al*. (2004) proposed a generalized linear mixed model approach for the spatio-temporal modelling of disease counts. The model assumes a binomial distribution for the counts and includes independent seasonal and spatial effects. However, it does not assess the problem of past outbreaks. Knorr-Held &

Richardson (2003) recommended a model for space-time meningococcal disease data distinguishing an 'endemic' pattern for periods of no outbreaks and a 'hyper endemic' pattern that models possible outbreaks in the data. While the endemic pattern is built in the spirit of chronic disease models including structured time, space, and seasonal effects, the hyper endemic pattern allows for an auto regression on functions of counts of the same and neighbouring regions which can be switched on and off according to a two-stage hidden Markov model. Empirical Bayes techniques for space-time disease surveillance have been proposed by Böhning (2003). Mugglin *et al*. (2002) offered a log-linear Poisson model for space-time influenza data of Scotland assuming the logarithm of the mean to depend on a multivariate Gaussian autoregressive process, where the innovation can switch between three levels, an endemic level, an epidemic level, in case of an outbreak, and a third level for the decline of the counts after the outbreak. They treated disease counts as a realization from an underlying multivariate autoregressive process where the relative risk of infection incorporates a space–time dynamic. Sebastiani *et al*. (2006) used dynamic Bayesian networks to integrate four different data streams into a multivariate model for influenza surveillance.

Two of the studies making spatio-temporal forecasts of influenza mentioned in Chretien *et al*. (2014) used the HHH model, the multivariate time-series model framework for aggregated surveillance data first proposed by Held *et al*. (2005). Since first being proposed, this model has been further developed in Paul *et al.* (2008), Paul & Held (2011), and Meyer & Held (2014). Additionally, an analytical method for obtaining the first two moments of multivariate path forecasts produced within the HHH framework is presented in Held *et al*. (2017). Meyer & Held (2014) proposed modelling infection transmission in space using a power law for short-time human travel. Together, these studies allow for long-term predictions of infectious disease in both space and time.

A related problem in the current literature is the lack of research on the contributions of different model components to improved prediction (Johansson *et al*., 2016). One exception is the study of Yang *et al*. (2016), in which they examine whether the inclusion of spatial dynamics improves influenza forecast at different observation scales (borough or neighbourhood) in New York City. They find spatial network data to improve borough level predictions but degrade neighbourhood-level predictions, possibly due to poor signal to noise ratios at this finer granularity. Additionally, Meyer & Held (2014) find power models to produce better predictions of final counts, epidemic curves, and regional final counts. Held *et al*. (2017) also find power model predictions to perform better than predictions from models with simpler spatial assumptions at several aggregation levels.

The aim of this study is to adapt multivariate negative binomial model developed by Paul *et al.* (2008) for modelling the cholera in Nigeria and implement the model for the analysis of 2018 cholera count data. The objectives are to investigate spatio-temporal characteristics of cholera and to describe the epidemiology of cholera outbreak in Nigeria.

2.0 Materials and Methods

2.1 Data Used

Secondary surveillance data spanning on weekly numbers of cholera counts between January 1st and November 19th, 2018 were obtained from the NCDC SED (primarily mandated for the coordination of cholera outbreak surveillance and response activities in Nigeria). The 20 states affected by the outbreak and their corresponding geopolitical regions were: Anambra and Ebonyi (south-east); Adamawa, Borno, Bauchi, Gombe and Yobe (north-east); Abuja, Kogi, Kwara, Nasarawa, Niger and Plateau (north-central); and Jigawa, Kaduna, Kano, Katsina, Kebbi, Sokoto and Zamfara (north-west).

The study population comprised individuals classified as having suspected cholera (herein: cholera cases) during the outbreak period. In accordance with the NCDC guidelines for

preparedness and response to acute watery diarrhoea outbreak by Nigeria Centre for Disease Control (2017), a cholera case was defined as the detection of a cluster of persons aged two years or older with acute watery diarrhoea and severe dehydration or dying from acute watery diarrhoea from the same area within one week. In line with best practice in the context of a cholera outbreak (Dalhat *et al.*, 2010), however, children under the age of two years who met the case definition were included in the current study as cholera cases. A confirmed cholera was defined as a cholera case in which Vibrio cholera O1 or O139 was isolated in the stool by microbiological investigation, Nigeria Centre for Disease Control (2017).

The selection process of study records, January 1st to November 19th, 2018. Twenty out of 36 states (plus the federal capital city of Nigeria, Abuja) were affected by the cholera outbreak. The original records was 44,198 cases, while a total number of 194 cases were duplicated. After removal of duplicated ones, a final number of records was 44,044. Records without an epidemiological week was only 8 cases and a final record used in this study was 43,996 cases and 836 deaths.

2.2 Methodology

The study employed multivariate negative binomial model developed by Paul *et al.* (2008) to model the 2018 cholera outbreak in Nigeria. The additive model was decomposed into two components: an endemic component and an epidemic component. to account for exogenous factors and infectiousness, respectively. The epidemic components are driven by past counts in the region of interest and in neighbouring regions, while the endemic component captures exogenous factors such as seasonality and population demographics.

Models

Let y_{rt} denotes the number of cholera counts observed in unit r at time t, $r = 1, \dots, 4$ geographical regions over $t = 1, \dots, 47$ weekly time points. In this study, a unit is a single disease, cholera, observed in the four geographical regions. For example, $y_{1,1}$ denotes the number of cholera counts in a particular region in January 2018. The counts are assumed to be distributed as negative binomial, $y_{rt}|y_{i.t-1} \sim \text{NegBin}(\mu_{rt}, \varphi)$ with conditional mean

$$\mu_{rt} = \lambda_{rt} y_{r,t-1} + \exp(\eta_{rt}) \quad (1)$$

and the conditional variance of y_{rt} increases to; $\mu_{rt}(1 + \varphi\mu_{rt})$ with additional unknown overdispersion parameter $\varphi > 0$.

The negative binomial distribution implies some strong assumptions about the data. If the variance is considerably larger than the expected value, known as overdispersion, the negative binomial model is an appropriate which accounts for overdispersion by including an additional model parameter but if the mean and variance of random variable are assumed to be equal, then a Poisson distribution is an appropriate alternative model (Gardner *et al.* 1995). However, in this study, overdispersion was identical in every unit because this is a realistic assumption when the units are 'regions' Paul *et al.* (2008).

Paul *et al.* (2008) decomposed the disease incidence μ_{rt} additively into two parts; the first part is the 'epidemic' component;

$$\tau_{rt} = \lambda_{rt} y_{r,t-1} \quad (2)$$

where λ is an unknown autoregressive parameter. And the second part is the endemic' component;

$$\varsigma_{rt} = \exp(\eta_{rt}). \quad (3)$$

The epidemic component allows for capturing the occasional outbreaks, whereas the endemic component describes the endemic seasonal patterns. The model in Equation (1) is further adjusted to explain the spread of a disease across the neighbouring units. To incorporate this spatial spread, the sum of the previous number of counts $y_{j,t-1}$ in other

units $j, j \neq r$ is included as a potential explanatory variable for the disease incidence in unit r Held *et al.* (2005). An additional weight was introduced by Paul *et al.* (2008), hence the epidemic component of Equation (2) gives

$$\tau_{rt} = \lambda_{rt} y_{r,t-1} + \phi_r \sum_{j \neq r} \omega_{jr} \, y_{j,t-1} \qquad (4)$$

where $y_{j,t-1}$ denotes the number of counts observed in region j at time $t-q$ with lag $q \in \{1, 2, \ldots\}$, and ω_{jr} are the suitably chosen weights. The influence of $y_{j,t-1}, j \neq r$ on y_{rt} is quantified by the additional autoregressive parameters ϕ_r.

Hence, the weights ω_{jr} are chosen according to the relative population of each of the four geographical regions.

To account for temporal variation of cholera counts, the endemic log-linear predictor ς_{rt}incorporates an overall trend and a sinusoidal wave of frequency $w_s = 2\pi s/48$, for weekly data (since our data was a weekly count). As a basic district specific measure of cholera counts, the population fraction η_{rt} is included as a multiplicative offset. The endemic component of Equation (3) becomes:

$$log(\varsigma_{rt}) = \eta_{rt} = \alpha_{rt} + \sum_{s=1}^{S} (\gamma_s \sin(w_s t) + \delta_s \cos(w_s t)) \qquad (5)$$

where S is the number of harmonics to be included and s are Fourier frequencies. The parameter α_{rt} in Equation (5) allows for different incidence levels in each of the regions.

3.0 Results

Prevalence rates (number of cholera cases/total number of people (per 1000) for a particular region in the year 2018) of cholera for all the four regions were found to be as follows: 0.02316 (South-east), 0.05324 (North-central), 0.38476 (North-west), and 0.91920 (North-east). These estimates are evidence that cholera was highly prevalent in North-east, implying that the population of this region is at a higher risk than the populations of other regions

.

Table 1: Weight matrix of cholera cases in four affected geographical regions, Nigeria, 2018

Geographical	**South-east**	**North-central**	**North-west**	**North-east**

region				
South-east	1	2.8323	5.8277	2.7780
North-central	0.3531	1	2.0576	0.9809
North-west	0.1716	0.4856	1	0.4767
North-east	0.3599	1.0195	2.0977	1

Using the weights ω_{jr} introduced by Paul *et al.* (2008) given in Equation (4), a weight matrix of dimension 4 × 4 in Table 1 was calculated by considering the relative population of each of the affected regions. For example, when the disease transmits from South-east to North-central, the weight is calculated by dividing the population of North-central by that of South-east. However, two different negative binomial models that contained S = 1 seasonal term in the endemic component were fitted, depending on whether the seasonality term and the autoregressive component $\lambda_{rt}y_{r,t-1}$ were included in the linear predictor or not. The results of the fitted negative binomial model of Equations (4) and (5) are summarized in (Table 2) standard errors were given in parentheses.

Table 2: Summary of the ML estimates and standard errors of the fitted models

Without seasonality and autoregression					
Distribution	$\hat{\lambda}$(SE)	$\hat{\varphi}$(SE)	$\widehat{\phi_r}$(SE)	Log-L	AIC
Negative binomial	- - - -	0.0664(0.3071)	0.9742(0.0429) 0.4444(0.0731) 0.8238(0.1060) 0.0453(0.0674)	-1384.71	1883.96
With both seasonality and autoregression					
Distribution	$\hat{\lambda}$(SE)	$\hat{\varphi}$(SE)	$\widehat{\phi_r}$(SE)	Log-L	AIC
Negative binomial	0.8903(0.0889) 0.1742(0.0186) 0.8903(0.0758) 0.8112(0.0857)	0.0816(0.3588)	0.0706(0.0100) 0.0351(0.0053) 0.1004(0.0665) 0.0305(0.0330)	-1353.04	1859.71

Table 2 shows that the maximum likelihood estimates of negative binomial model with both seasonal and autoregressive components λ were 0.8903(0.0889), 0.1742(0.0186), 0.8903(0.0758) and 0.8112(0.0857) for the four regions. These estimated values were the indication of the existing heterogeneity in the autoregressive effect and can be interpreted as the epidemic proportion of disease incidence (Held and Paul, 2012). This model maximizes

the log-likelihood function, log L, compared with the negative binomial model without the seasonal and autoregressive components, meaning that both the seasonality and the autoregressive components have to be included in the model. The calculated Akaike's Information Criterion (AIC) for the two models to check whether accounting for overdispersion is useful for these data, the negative binomial model with both seasonal and autoregressive also suggested to be a better one than the other with smaller AIC (1859.71).

3.1 Exploratory Analysis

Exploratory analysis of spatio-temporal characteristics of the study population is presented in (Table 3). Primary outcomes included attack rate (AR) and case fatality rate (CFR) were presented in (Table 4) and distribution of cholera cases by epidemiological week in (Table 5).

Table 3: Characteristics of the epidemic season by week from Jan. 1st to Nov., 19th 2018

Epidemic Week	Final count	Consecutive weeks (January)	Consecutive weeks (Season)
Week1	91	1st week	Dry
Week2	93	1st week	Dry
Week3	52	1st week	Dry
Week4	93	1st week	Dry
Week5	101	1st week	Dry
Week6	352	1st week	Dry
Week7	253	1st week	Dry
Week8	192	1st week	Dry
Week9	173	1st week	Dry
Week10	253	2nd week	Dry
Week11	292	2nd week	Dry
Week12	301	2nd week	Dry
Week13	521	2nd week	Rainy
Week14	542	2nd week	Rainy
Week15	901	2nd week	Rainy
Week16	1000	2nd week	Rainy
Week17	1203	2nd week	Rainy
Week18	1003	2nd week	Rainy
Week19	1201	2nd week	Rainy
Week20	1002	2nd week	Rainy
Week21	1351	2nd week	Rainy
Week22	1215	2nd week	Rainy
Week23	1308	2nd week	Rainy
Week24	1258	2nd week	Rainy
Week25	1108	2nd week	Rainy
Week26	1109	2nd week	Rainy
Week27	805	2nd week	Rainy

Week28	793	2nd week	Rainy
Week29	1000	3rd week	Rainy
Week30	1011	3rd week	Rainy
Week31	1364	3rd week	Rainy
Week32	1540	3rd week	Rainy
Week33	1030	3rd week	Rainy
Week34	1615	3rd week	Rainy
Week35	2614	4th week	Rainy
Week36	2764	4th week	Rainy
Week37	3064	4th week	Rainy
Week38	2300	4th week	Rainy
Week39	1968	4th week	Rainy
Week40	1558	4th week	Rainy
Week41	1302	4th week	Rainy
Week42	811	4th week	Rainy
Week43	532	4th week	Rainy
Week44	418	4th week	Rainy
Week45	301	4th week	Dry
Week46	102	4th week	Dry
Week47	126	4th week	Dry

The counts each week vary in several ways: epidemic week, final count, consecutive weeks (began on Sunday and ended on Saturday) and consecutive weeks (season) (Table 3). The data for week35, week36 and week37 (in red) represent the extreme nature of the cholera pandemic. Incidence peaks much higher, at 3064 counts, than at any other time in the data set. As expected, the majority (94.48%) of cholera cases were reported during the rainy season as reported (Table 5). The data for week5 represent the peak weak listed as 52 because the 2018 epidemic season is really in the first wave of the dry epidemic season.

Table 4: Distribution of cholera attack rates and case fatality rates by region, Nigeria, 2018

Geographical region	**Projected 2018 population**	**Cases**	**Deaths**	**Attack rate/100,000 population**	**CFR (%)**
South-east	8,852,569.68	205	8	2.32	3.90
North-east	24,593,161.29	22,606	251	91.92	1.11
North-central	25,072,869.86	1,335	62	5.33	4.64
North-west	51,590,496.43	19,850	515	38.48	2.59
Total	**110,109,097.26**	**43,996**	**836**	**127.43**	**1.90**

Table 4 shows distribution of cholera AR and CFR. AR for each reporting region was calculated using the estimated population of 2018, which was based on a 3.3% projected

growth rate from the 2006 national census results; the values were multiplied by 100,000. And CFR is the ratio of individuals classified as cholera cases who die to all those classified as cholera cases (alive and dead). CFR is expressed in percentage (%). The overall attack rate during the outbreak period was 127.43/100,000; specifically, North-east (91.92/100,000 population) and North-west (38.48/100,000 population) recorded higher ARs compared to other two geographical regions such as North-central and South-east with an ARs of 5.33/100,000 population and 2.32/100,000 population respectively. CFRs were generally high across all affected geographical regions, with about 75% of these regions recording higher CFRs above the national figure of 1.90%. Notably, north-central recorded the highest CFRs (4.64%); north-east recorded the second highest CFRs (3.90%) and north-west recorded CFRs (2.59%).

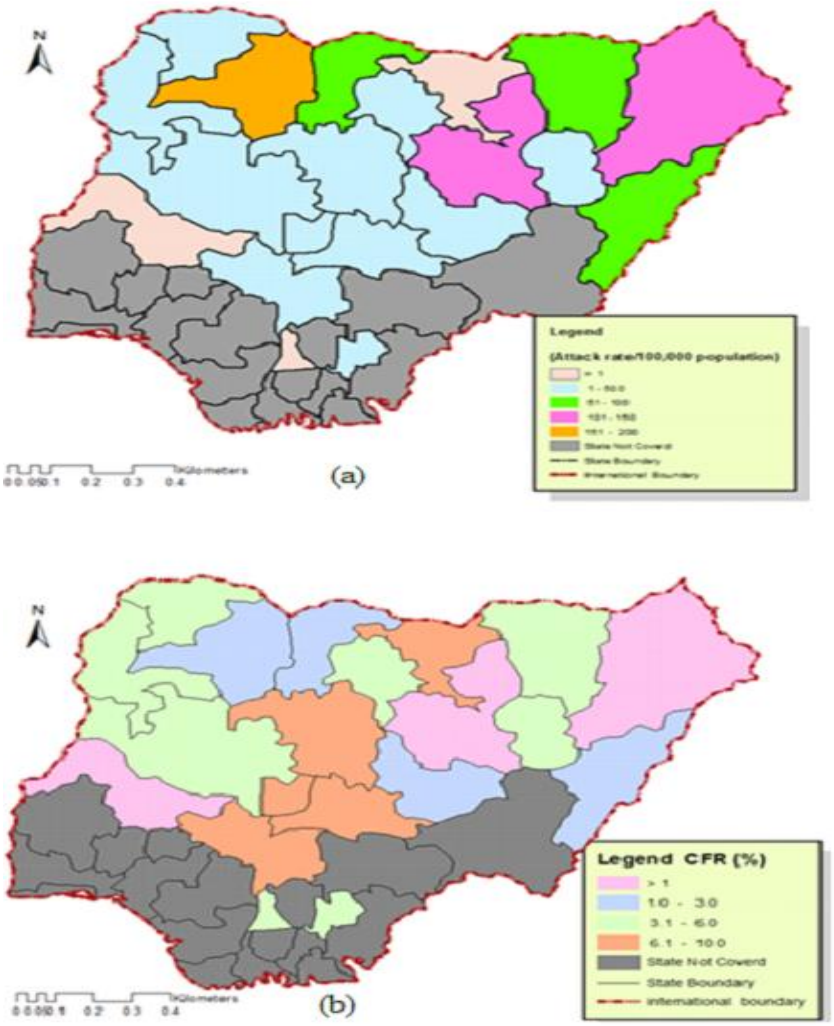

Figure 1: Spatial distribution of cholera in terms of: (a) attack rates and (b) case fatality rates.

The spatial distribution of cholera cases in terms of attack rates and case fatality rates across the 20 affected states is shown in Figure. 1. The modes of attack rates was found in category 6: 60 to 100(per 100,000) and category 5: 30 to 49.9(per 100,000) which were notably found in north-east and north-west regions. The spatial distribution of case fatality rates was more often in category 5: with 5.0% and above than categories 3 and 4: with (2.0 to 3.5%) and (3.5 to 5.0%) which were notably dispersed in north-central, and north-west regions.

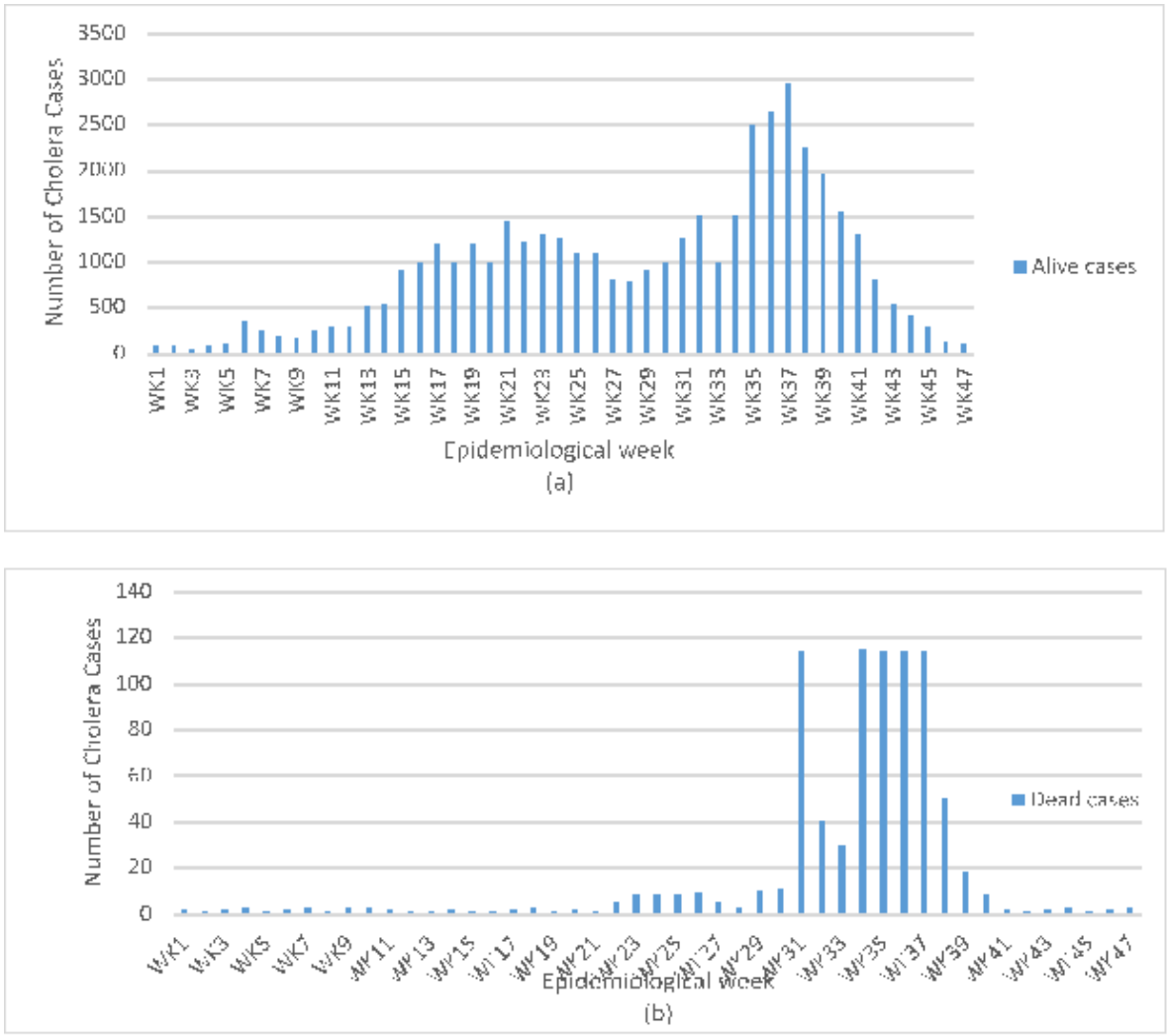

Figure 2: Time series plot by epidemic week, 2018.for (a) Alive and (b) Deaths

Figure 2 shows the epidemiological curve for cholera cases alive and deaths by epidemic week. The outbreak was characterised by four distinct epidemic waves and mirrored a propagated epidemic pattern, suggesting a person-to-person transmission. Notably, the majority of cholera cases occurred in the second and fourth waves, with a peak at week 37; there was however a preponderance of cholera deaths towards the end of the third wave and beginning of the fourth wave, with sporadic cases of death in between the two waves.

Table 5: Distribution of cholera cases by epidemiological week, Nigeria, 2018

Characteristic	Epidemiological week				
	Week 1–9 Cases (%)	Week 10–28 Cases (%)	Week 29–34 Cases (%)	Week 35–47 Cases (%)	**Total Cases (%)**
Age (years)					
< 5	386 (34.50)	4551 (25.96)	1422 (19.08)	3899 (21.79)	10258(23.32)
≥ 5	731 (65.33)	12,955 (73.91)	5744 (77.06)	12,947(72.35)	32377(75,59)
Missing	2(0.18)	22(0.13)	288(3.86)	1049(5.86)	1361 (3.09)
Sex					
Female	567 (50.67)	8782 (50.10)	3773 (50.62)	9200 (51.41)	22322(50.74)
Male	552 (49.33)	8746 (49.90)	3681 (49.38)	8695 (48.59)	21674(49.26)
Geographical region					
South-east	3 (0.27)	188 (1.07)	14 (0.19)	0 (0.00)	205(0.47)
North-east	874 (78.11)	12,372 (70.58)	807 (10.83)	8553 (47.80)	1335(3.03)
North-central	3 (0.27)	1083 (6.18)	158 (2.12)	91 (0.51)	19850(45.12)
North-west	239 (21.36)	3885 (22.16)	6475 (86.87)	9251 (51.70)	22606(51.38)
Season					
Dry	0 (0.00)	997 (5.69)	7454 (100)	311 (1.74)	2427(5.52)
Rainy	1119 (100)	16,531 (94.31)	0 (0.00)	17,584(98.26)	41569(94.48)
Outbreak setting					
Rural	887 (79.27)	3073 (17.53)	4093 (54.91)	7448 (41.62)	15501(35.23)
Urban	199(17.79)	14382(82.05)	3318(44.51)	10119(56.54)	28018(63.68)
Missing	33(2.95)	73(0.42)	43(0.58)	328(1.83)	477(1.08)

Cholera cases by epidemiological week, Table 5, indicates that higher number of cholera cases was recorded consistently in individuals aged 5 years or older throughout the outbreak period, particularly in week 29 to 34 during which they accounted for 77.06% of recorded cases. However, the distribution of cholera cases was about the same between males and females across all weeks. With respect to geographical regions, states from the north-east accounted for a higher number of cases in week 1 to 9 (78.11%) and week 10 to 28 (70.58%), whereas those from the northwest accounted for a higher number of cases in week 29 to 34 (86.87%) and week 35 to 47 (51.70%). As expected, the majority of cholera cases (94.48%) were reported during the rainy season. Epidemiological weeks during the cholera outbreak appeared too had been significantly influenced by seasonality.

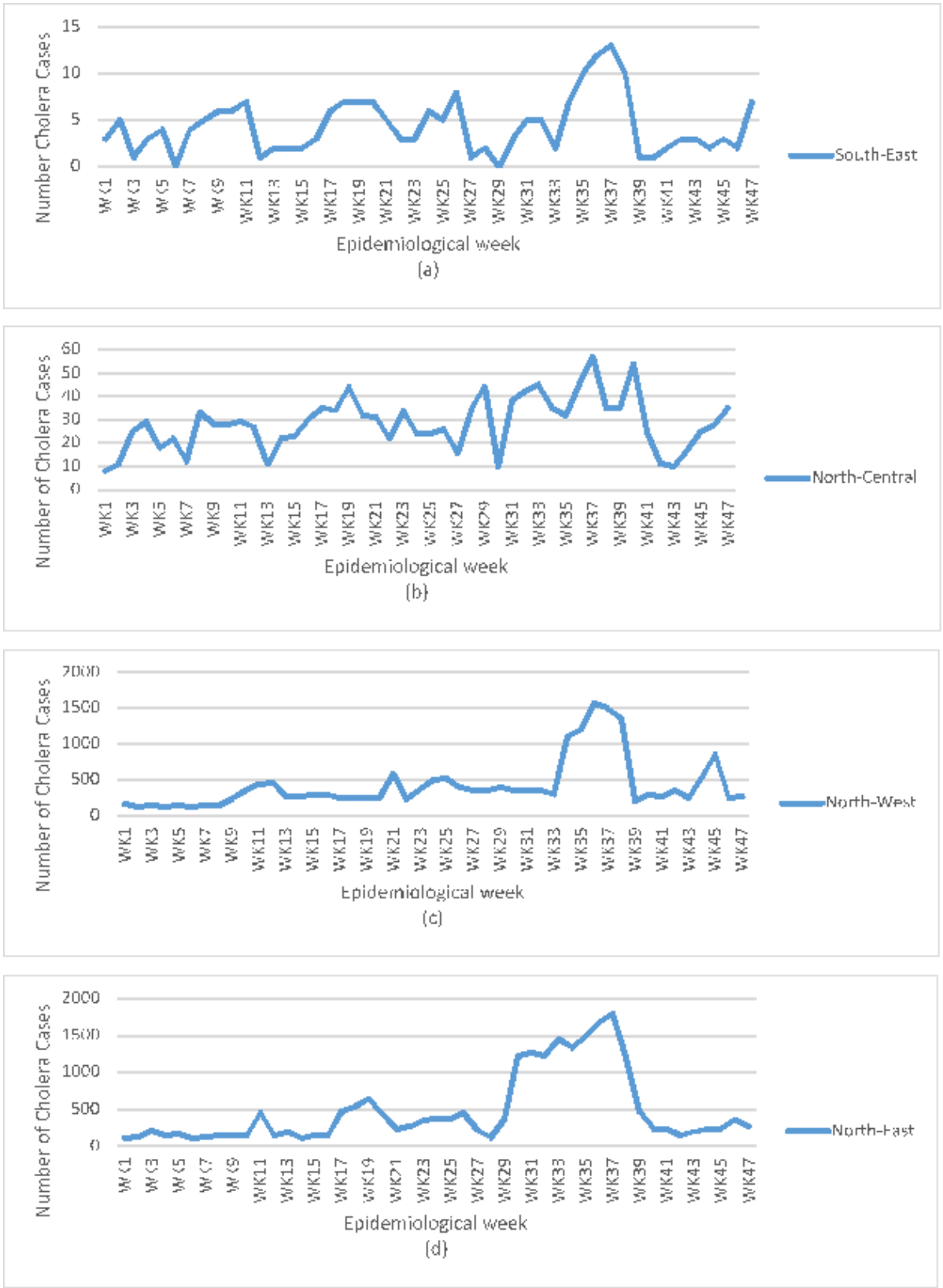

Figure 3: Time series plot by epidemic week, 2018 for (a) South-East (b) North-Central (c) North-West amd (d) North-East.

The plotted time series (Figure 3) for week1 through week47 (Figure 1) reveals a general seasonality to the data, with incidence peaking toward week 35 to week 37 the end of the second quarter/beginning of the third quarter, that is, in rainy season, as expected.

Conclusion

In this study, the spatio-temporal model of infectious disease counts developed by Paul *et al.* (2008) is used to describe the spatial and temporal pattern of cholera outbreak in Nigeria. The negative binomial model with both the seasonality and autoregressive components provided adequate fit for the cholera count data than without seasonality and auto regression. It also found to perform better for capturing the spatio-temporal dependence of a disease. This result coincides with the report of Paul *et al.* (2008) and Sifat & Israt (2011). This model is more efficient than a purely parameter-driven model by considering seasonality, over dispersion, and localized epidemics, which are common features of infectious disease data

For the exploratory analysis of spatio-temporal characteristics, the analyses shown that there three problems. First, there were tremendous increase in the cholera cases and deaths between weeks 35 to 47 which coincided with when the intensity of rainy season had begun to dwindle across the country and increased likelihood for water sources to be contaminated by floods around these periods.,

Second, the majority of cases in week 1 to 9, week 29 to 34 and week 35 to 47 were recorded in rural areas as many persons tend to rely more on unsafe water sources when water levels are decreasing towards the end of rainy season.

Lastly, the CFR of 1.9% in the current study is comparable with that for Africa at approximately 2% but almost twice as low as the value recorded by Dalhat and colleagues during the 2010 cholera outbreak in Nigeria, Dalhat *et al.* (2014). A CFR higher than the WHO recommended benchmark of < 1% **(**WHO, 2011; and WHO, 2017**)** is generally

considered high and Clemen *et al.*, (2017) reported that it was indication of inadequate clinical case management or quality of care.

Recommendations

1. There should be a continuous monitoring by technical working group of each affected state to know the current epidemiology of cholera in the country.
2. Provision of health facilities to Cholera Treatment Centres (CTCs)) in affected Local Government Areas (LGAs) for rapid testing of suspected cholera cases.
3. The study recommend a further research on the spatio-temporal cholera model that will detect routes of transmission.

References

Andersson, H. & Britton, T. *Stochastic Epidemic Models and their Statistical Analysis*, Lecture Notes in Statistics, Vol. 151, Springer-Verlag, NewYork (2000).

Ali M, Nelson A. R, Lopez A' L, *et al*. Updated global burden of cholera in endemic countries. PLoS Negl Trop Dis. 9:e0003832 (2015).

Becker, N. *Analysis of Infectious Disease Data*, Chapman & Hall, London (1989).

Becker, N.G. & Britton, T. *Statistical studies of infectious disease incidence*, J. R. Stat. Soc. Ser. B Stat. Methodol. 61, pp. 287–307 (1999).

Böhning, D. *Empirical Bayes estimators and non-parametric mixture models for space and time-space disease mapping and surveillance*, Environmetrics 14, pp.431-51 (2003).

Chretien, J.-P., George, D., Shaman, J., Chitale, R. A. & McKenzie, E. Inuenza forecasting in human populations: a scoping review. PLoS ONE 9 e94130 (2014).

Clemens J. D, Nair G. B, Ahmed T, *et al*. Cholera. Lancet; 390: 1539–49 (2017).

Daley D.J. & Gani, J. *Epidemic Modelling: An Introduction*, Cambridge University Press, Cambridge (1999).

Dalhat M. M, Isa A. N, Nguku P, *et al*. Descriptive characterization of the 2010 cholera outbreak in Nigeria. BMC Public Health. 14:1167 (2014).

Finkenstädt, B.F., & Morton, A.M. *Discrete-time modelling of disease incidence time series by using Markov chain Monte Carlo methods*, Appl. Stat. 54 (3), pp. 575–594 (2005).

Gardner, W., Mulvey, E. P., and Shaw, E. C. Regression analyses of counts and rates Poisson, overdispersed Poisson, and negative binomial models. Psychological Bulletin, 118 (3): 392–404.

Global Task Force on Cholera Control Ending Cholera: A Global Roadmap to 2030. Annecy (2017).

Held, L., Höhle, M. & Hofmann, M. *A statistical framework for the analysis of multivariate infectious disease surveillance data*, Statistical Modelling. 5, pp. 187–199 (2005).

Held, L., Meyer, S. & Bracher, J. Probabilistic forecasting in infectious disease epidemiology: The thirteenth Armitage lecture. Statistics in Medicine (2017).

Held, L & Paul, M. Modeling seasonality in space-time infectious disease surveillance data. Biometrical Journal, 54(6):824–843 (2012).

Johansson, M. A., Reich, N. G., Hota, A., Brownstein, J. S. & Santillana, M. Evaluating the performance of infectious disease forecasts: A comparison of climate-driven and season dengue forecasts for Mexico. Scientific Reports (2016).

Kleinman, K., Lazarus, R. & Platt, R. *A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism*, Am. J. Epidemiol. 159, pp. 217–224 (2004).

Knorr-Held, L. & Richardson, S. *A hierarchical model for space-time surveillance data on meningococcal disease incidence*, Appl. Stat. 52, pp. 169–183 (2003).

Mengel M, Delrieu I, Heyerdahl L, *et al*. Cholera outbreaks in Africa. Curr Top Microbiol Immunol; 379:117–44 (2014).

Meyer, S. & Held, L. Power-law models for infectious disease spread. The Annals of Applied Statistics 8, 1612-1639 (2014).

Microbiology Society Fact file: cholera: death by Diarrhoea (2016).

Mugglin, A.S., Cressie, N. & Gemmell, I. *Hierarchical modeling of influenza epidemic dynamics in space and time*, Stat. Med. 21, pp. 2703–2721 (2002).

Nigeria Centre for Disease Control Preparedness and response to acute watery Diarrhoea outbreaks: a guide for health workers and authorities in Nigeria (2017).

Paul, M., Held, L & Toschke, A. M. *Multivariate modeling of infectious disease surveillance data*, Stat. Med. 27, pp. 6250–6267 (2008).

Paul, M. & Held, L. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. Statistics in Medicine 30, 1118-1136 (2011).

World health organization. Cholera case fatality rate: situations and trends. Geneva: WHO; (2017).

World Health Organization. Cholera case fatality ratio (%). Geneva: WHO; (2011).

Sebastiani, P., Mandl, K. D., Szolovits, P., Kohane, I. S. & Ramoni, M. F. *A Bayesian dynamic model for influenza surveillance*, Stat. Med. 25 (2006), pp. 1803–1816 (2006).

Sifat S., &. Israt R. Md. Spatio-temporal modelling of infectious disease dynamics Journal of Applied Statistics, (2011).

Yang, W., Olson, D. R. and Shaman, J. Fofo recasts influenza outbreaks in boroughs and neighbourhoods of New York City. PLoS Computational Biology 12 e1005201 (2016).