Modeling of the Deaths Due to Ebola Virus Disease Outbreak in Western Africa

Robert J. Milletich[1], Norou Diawara[1,*] and Anna Jeng[2]

[1]*Mathematics and Statistics Department, Old Dominion University, 4700 Elkhorn Ave., Norfolk, VA 23529, USA*

[2]*School of Community and Environmental Health, College of Health Sciences, Old Dominion University, 4608 Hampton Blvd. Health Sciences Building 3140 Norfolk, VA 23529, USA*

Abstract: *Problem*: The recent 2014 Ebola virus outbreak in Western Africa is the worst in history. It is imperative that appropriate statistical and mathematical models are used to identify risk factors and to monitor the development and spread of the disease.

Method: Deaths data due to Ebola virus disease (EVD) in Guinea, Liberia, and Sierra Leone from October 10, 2014 to March 24, 2015 were collected *via* Situation Reports published by the World Health Organization [1]. Conditional autoregressive (CAR) models were applied to account for the spatial dependency in the countries along with the temporal dimension of the disease. Bayesian change-point models were used to identify key changes in growth and drop time points in the spatial distribution of deaths due to EVD within each country. Country-specific Poisson and negative binomial mixed models of covariate effects were applied to understand the between-country variability in deaths due to EVD.

Results: Both CAR models and generalized linear mixed models identified statistically significant covariate effects; however, the CAR models depended on the interval of data analyzed, whereas the mixed models depended on the underlying distribution assumed. Bayesian change-point models identified one significant change-point in the distribution of deaths due to EVD within each country.

Practical Application: CAR models, Bayesian change-point models, and generalized linear mixed models demonstrate useful techniques in modeling the incidence of deaths due to EVD.

Keywords: Ebola Virus Disease, Conditional Autoregressive Model, Bayesian Analysis, Change-Point Model.

1. INTRODUCTION

Ebola virus disease (EVD) is an often fatal human and primate disease that has an estimated fatality rate of up to 90% [1]. The recent 2014 Ebola epidemic is the largest outbreak in history and affected multiple countries in Africa, particularly Guinea, Liberia, and Sierra Leone. Although other countries outside of Africa were also affected, the countries in Western Africa experienced the most cases of infection and death due to EVD. Thus, given the severity of the situation, recent attention has been focused on utilizing mathematical and statistical models to ascertain the development and spread of EVD and help authorities focus their efforts on important risk factors and surveillance efforts to help eradicate EVD.

The inherent nature of Ebola disease calls for the need to use features that account for the spatial dependency between neighboring or communicating countries along with the temporal dimension of the disease. Replicate data are collected and the statistical procedure used to analyze them must account for spatial and temporal characteristics to be processed.

It is necessary before any action to determine its actual progression, at least as accurately as reasonably possible. As mentioned in Dallatomasina *et al*. [2], epidemiologic characteristics of the outbreak using different countriesand even in the rural areas is vital. McKinley *et al*. [3] usedapproximate likelihood-based inference for epidemic models and expressed the challengesdue to the evaluation of the likelihood. Those authors proposed a Markov Chain Monte Carlo and Sequential Monte Carlo algorithms for parameter estimates. Statistical inference in a stochastic discrete-time based (SEIR) epidemic model has been proposed by Lekone and Finkenstadt [4]. Extension of that model to the continuous time type has been developed by McKinley *et al*. Those authors suggested a time point at which the transmission parameter decays. Following the same idea, we have proposed a change point in the weekly count of deaths due to EVD, and build models at two levels where the counts are modelled without and with the presence of the change point, respectively. Our goal is to bridge the gap by presenting several probabilistic models from the statistical tools and the needs of EVD management centers.

*Address correspondence to this author at the Mathematics and Statistics Department, Old Dominion University, 4700 Elkhorn Ave., Norfolk, VA 23529, USA; Tel: (757) 683.3886; Fax: (757) 683.3885; E-mail: ndiawara@odu.edu

E-ISSN: 1929-6029/15

Dealing with heterogeneous and spatial populations, the hierarchical Bayesian model or the conditional autoregressive model (CAR) as proposed by Millar [5] or Hossain *et al*. [6] could serve as an alternate methodology. Millar [5] stated that the use of hierarchical Markov Chain Monte Carlo methods are computationally viable methods. We have constructed CAR models using the full data and then used data based on disjoint time intervals identified from the change point analyses.

Count data are often modeled using the Poisson distribution, which assumes that the mean and variance are equal. The existence of overdispersion is frequent in modeling count data, especially data measured over time. Ignoring overdispersion can lead to underestimating standard errors in regression coefficients and thus biased statistical inferences (e.g., Type I Error; Gardner *et al*. [7]). As recommended by Navarro *et al*. [8], the negative binomial distribution can be used to model the part of the variance, which the Poisson distribution is unable to identify. We construct generalized linear mixed models to model the deaths due to EVD using a Poisson distribution and with a negative binomial distribution.

The goals of this paper are to use data on the recent outbreak of EVD in Western Africa to build and compare statistical models to understand the development of disease. We identify starting and dropping time points of the disease as well as growth periods, and identify statistical covariates of the disease. The data on deaths due to EVD were collected in Guinea, Liberia, and Sierra Leone from October 10, 2014 to March 24, 2015. Models used in this study included Bayesian disease mapping models, change-point models, and generalized linear mixed models. One major role of the paper is to build a stochastic model in order to capture some of the disease characteristics. Doing so, we are providing pointers as to how to assist medical workers in their efforts to manage the epidemic as suggested in Dallatomasina *et al*. [2].

The paper is organized as follows: data description and preliminary analyses as shown in Section 2, and the description of the space time-mixture of the Poisson regression with the use of the CAR model along with the Bayesian change-point model as described in Section 3. The model formulations, estimation processes and prior distributions are presented. Next, we introduce the generalized linear mixed model with both a Poisson and negative binomial distribution. Section 4 discusses the results and comparisons from all the models. Lastly we end with a conclusion in Section 5.

2. THE DATASET

Data on counts of deaths due to EVD in Guinea, Liberia, and Sierra Leone from October 10, 2014 to March 24, 2015 were collected *via* Situation Reports from the World Health Organization's (WHO) website (http://apps.who.int/ebola/en/current-situation/ebola-situation-report). Due to the inconsistency in how frequently Situation Reports were published by the WHO for each country, deaths due to EVD data were aggregated into weekly counts, for a total of 25 weeks per country. Given the association between lower temperatures and EVD outbreak in Africa (Ng *et al*. [9]), data on mean temperatures (˚F) were collected as a covariate. In particular, mean temperatures for Guinea, Liberia, and Sierra Leone, respectively. To keep consistent with deaths due to EVD, temperature data were also aggregated into weeks, for a total of 25 weekly mean temperatures for each country.

To be used as an offset term in some of the statistical models, internally standardized expected death counts, E_{it}, were calculated based on recommendations by Carlin and Louis [10] as:

$$E_{it} = n_{it}\overline{y},$$

where $\overline{y} = \sum_{it} y_{it} / \sum_{it} n_{it}, i = 1,2,3 \text{ and } t = 1,2,\ldots,25$. Here, n_{it} represents the number of individuals at risk for country i at time *t*. Calculated as such, the expected death counts represent the average death rate over all countries for the entire observation period.

2.1. Preliminary Analyses

A preliminary study shows basic information about the disease. Table **1** presents a summary of these statistics by country and aggregated across country. The average number of deaths due to EVD over the 25 week period for Guinea is 60.96 (*SD* = 25.17) for Guinea, 89.28 for Liberia (*SD* = 61.90), and 111.68 for Sierra Leone (*SD* = 76.32). The number of deaths range from 28 to 112 for Guinea, 18 to 247 for Liberia, and 4 to 307 for Sierra Leone. Figure **1** presents the distribution of number of deaths due to EVD aggregated across country. It shows a peak and the study is to investigate the inference under each model and location.

The data was further divided into each country. Figure **2** shows the number of deaths due to EVD for

Table 1: Descriptive Statistics of Deaths Due to Ebola Virus Disease by Country

	Deaths Due to Ebola Virus Disease					
Area	**Mean**	***SD***	**Min**	**Median**	**(25%, 75%)**	**Max**
Guinea	60.96	25.17	28	54	(41, 84)	112
Liberia	89.28	61.90	18	72	(45, 121)	247
Sierra Leone	111.68	76.32	4	85	(60, 150)	307
Combined	87.31	61.43	4	70	(45, 112)	307

Note. SD = standard deviation; 25% = 25th percentile; 75% = 75th percentile. Estimates were obtained from weekly counts based on 25 weeks of observed data.

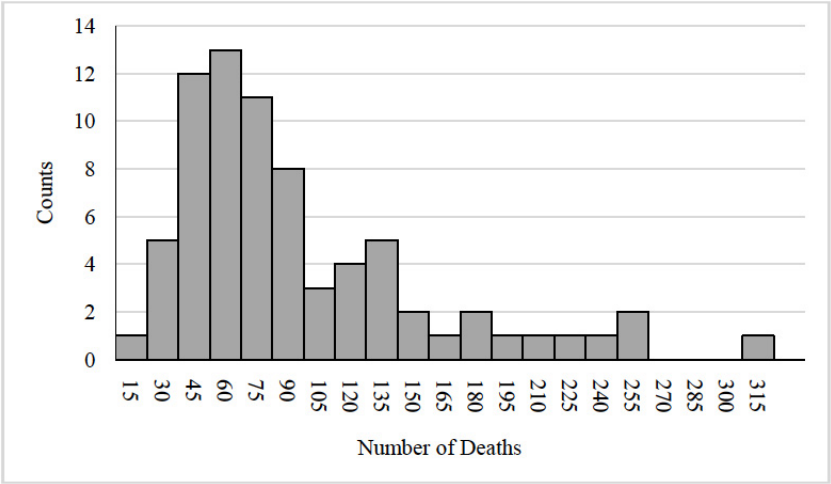


Figure 1: Histogram of the Number of Deaths Due to EVD Aggregated Across Country.

each country that appears to be decreasing over time. The pattern shown in Figure **1** can be seen in Figure **2** except for Sierra Leone. There is a change in the disease aggressiveness indicative of a change in the behavior of the disease growth. As reported by the USAID and its various offices, Liberia was one of the first countries to have received financial assistance and personal support from the US government. It had the most cases of suspected, probable and confirmed EVD cases as reported in http://www.usaid.gov/sites/default/files/documents/1864/09.17.14%20-%20USG%20West%20Africa%20Ebola%20Outbreak%20Fact%20Sheet%20%236.pdf. And it looks like the disease would be spread going northwards. We will test other basic characteristics associated with the data collected in those areas such as time and temperature.

3. STATISTICAL MODELS AND RESULTS

Given that the distribution of deaths due to EVD counts is skewed (see Figure **1**), the distribution of the counts is unlikely to follow a normal distribution. A distribution that is commonly used to model count data is the Poisson distribution. A discrete random variable, *Y*, is said to have a Poisson distribution, defined by a rate parameter $\lambda > 0$, if for $y = 0,1,2,\ldots$, the probability mass function of *Y* is given by

$$f(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!},$$

where *e* is Euler's number (2.71828…). The mean and variance of a Poisson-distributed random variable are both equal to λ, that is, $E(Y) = Var(Y) = \lambda$. In the case where data are overdispersed under the Poisson model, that is $Var(Y) > E(Y)$, the negative binomial model is more appropriate. The negative binomial distribution can be derived as a mixture distribution of a Poisson random variable, where the rate parameter itself is a random variable and follows a gamma distribution. A discrete random variable, Y, is said to have a negative binomial distribution, defined by $\mu, \kappa > 0$, if for $y = 0,1,2,\ldots$, the probability mass function of Y is given by:

$$f(Y = y) = \frac{\Gamma(\kappa^{-1} + y)}{\Gamma(\kappa^{-1})y!}\left(\frac{\kappa\mu}{1+\kappa\mu}\right)^y\left(\frac{1}{1+\kappa\mu}\right)^{1/\kappa},$$

where $\Gamma(\cdot)$ is the gamma function. The mean and variance of a negative binomial-distributed random

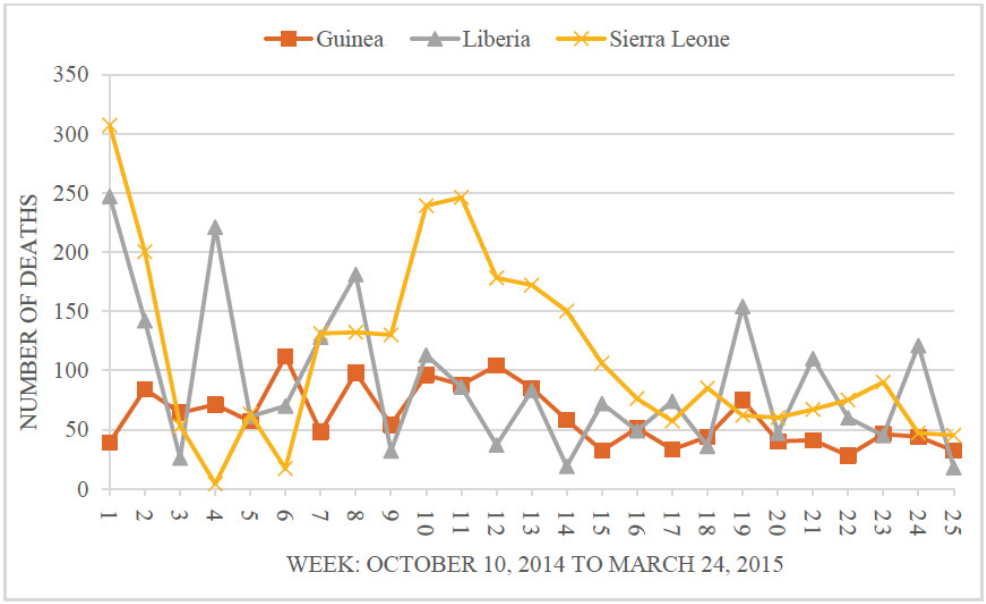


Figure 2: Number of Deaths Due to EVD by Week for Guinea, Liberia, and Sierra Leone.

variable are $E(Y)=\mu$ and $Var(Y)=\mu+\kappa\mu^2$, respectively. Importantly, as $\kappa\to 0$, the negative binomial approaches Poisson $(\lambda=\mu)$.

3.1. Conditional Autoregressive Bayesian Disease Mapping Models for Full Data

Disease mapping models are often applied in epidemiological settings to understand the incidence or prevalence of a specific disease. The data are counts of observed cases or deaths within multiple regions coupled with potentially relevant background information (Lawson [11]). One popular disease mapping model is the conditional autoregressive (CAR) model. This model gets its name from the type of random effect specified (i.e., conditional autoregressive) used to account for the spatial dependency among the observed data. Fitting hierarchical model using Markov Chain Monte Carlo (MCMC) methods is possible even for overdispersed count data.

Bayesian statistics have gained great attention due to the computing power available and the flexibility that the models offer. Applications can be found in many fields and in medical disease progression (Gelman *et al*. [12]). Computations are commonly carried out using R/WinBUGS and SAS. In the present study, for the EVD death count data, with the characteristics as time expressed in weeks and the temperature in degrees Celsius or Farenheit, the following CAR modelis specified as,

$$y_{it}\,|\,u_i, x_{it} \sim \text{Poisson}(\lambda E_{it}),$$

$$\log(\lambda)=\log(E_{it})+\beta_0+\beta_1\ \text{week}_i+\beta_2\ \text{temperature}_{it}+u_i,$$

for $i=1,2,3$ countries, $t=1,2,\ldots,25$ weeks, and $u_i\sim \text{CAR}(\sigma_u^2)$. In this case, E_{it} is treated as an offset so that the model can be rewritten as,

$$\log\left(\frac{\lambda}{E_{it}}\right)=\beta_0+\beta_1\ \text{week}_i+\beta_2\ \text{temperature}_{it}+u_i$$

so that the relative death rates are modeled rather than the observed death counts. The following independent diffuse prior distributions are specified for the model parameters,

$$\pi(\beta_0)\sim 1$$

$$\pi(\beta_1)\sim 1$$

$$\pi(\beta_2)\sim 1$$

$$\pi(u_i)\sim \text{Inverse Gamma}(1000,1000).$$

To facilitate interpretation of the intercept estimate, the covariate week is centered at 1 and the covariate temperature is centered around the mean of each group. Parameterized as such, the intercept represents the unadjusted relative death rate at week 1 (i.e., the beginning of the observed data) (Raudenbush & Bryk [13]). Descriptive statistics for the temperature variable are displayed in Table **2**.

Table 2: Descriptive Statistics of Temperature by Country

	Temperature (˚F)					
Area	**Mean**	***SD***	**Min**	**Median**	**(25%, 75%)**	**Max**
Guinea	81.12	1.09	79	81	(80, 82)	83
Liberia	80.76	1.13	79	81	(80, 81)	83
Sierra Leone	80.52	1.42	77	81	(80, 82)	83
Combined	80.80	1.23	77	81	(80, 82)	83

Note. *SD* = standard deviation; 25% = 25th percentile; 75% = 75th percentile.

The CAR model is estimated in WinBUGS (Lunn *et al*. [14]) using a Metropolis-Hastings algorithm with 20,000 iterations, a burn-in period of 2,000 iterations, and thinning set to 5 to reduce the autocorrelation in parameter draws from the posterior distribution.

Results of the CAR model show that both covariates, week and temperature, are significantly associated with relative death rates due to EVD (see Table **3**). Specifically, the effects are negative indicating that an increase in time and temperature are associated with significantly *lower* death rates from EVD. Figure **3** reveals that the algorithm appears to converge for the intercept and slope parameters, but convergence is questionable for the variance component. Further examination reveals that the density plots (see Figure **4**) for the intercept and slope parameters are normally distributed, but the density plot for the variance component is severely positively skewed.

The model results displayed in Table **3** shows that both the temperature and week are significant and underlines the fact that there is need to investigate the convergence of the MCMC process. The negative association in the EVD outbreak and the mean temperatures is also parallel to the findings in Ng *et al*. [9]. Given the large standard deviation of the CAR random effect (*SD* = 742.4) for the model in Table **3**, coupled with the severely skewed density plot in Figure **4**, potential subsequences of data within each country over time are examined using Bayesian change-point models. As can be seen in Figure **2**, it appears that there is a change in the rate of deaths due to EVD for Guinea, Liberia, and Sierra Leone, with less variability towards the end of the observed data. Identifying important change-point(s) within each country will help separate different subsequences of data and may reduce the standard deviation of the CAR random effect.

3.2. Bayesian Change-Point Models

A Poisson process is used to model the frequency of occurrences. Because of interventions and aids from foreign countries, a change point analysis model is applied. That is, $Y_t \sim \text{Poisson}(\lambda)$ for $0 < t < k$ and $Y_t \sim \text{Poisson}(\mu)$ for $t > k$. As is common in change-point models, deaths due to EVD are assumed to follow a common distributional form (i.e., Poisson in this case), but have different parameters (i.e., rate or intensity) from one segment to another (Hawkins, [15]). A change-point model is estimated in each country separately to identify unique patterns of data that are independent of other countries. That is, for each

Table 3: Summary of Conditional Autoregressive Model

			95% Credible Interval	
Parameter	**Mean**	***SD***	***LL***	***UL***
Intercept	0.427	0.022	0.381	0.469
Week	-0.034	0.002	-0.037	-0.030
Temperature	-0.092	0.010	-0.113	-0.072
$\hat{\sigma}_u^2$	27.900	742.4000	0.482	66.930
Statistic	Estimate			
pD	90.789			
DIC	2552.970			

Note. *SD* = standard deviation, *LL* = lower limit, *UL* = upper limit. Week is centered at 1; temperature is centered around the mean of each country.

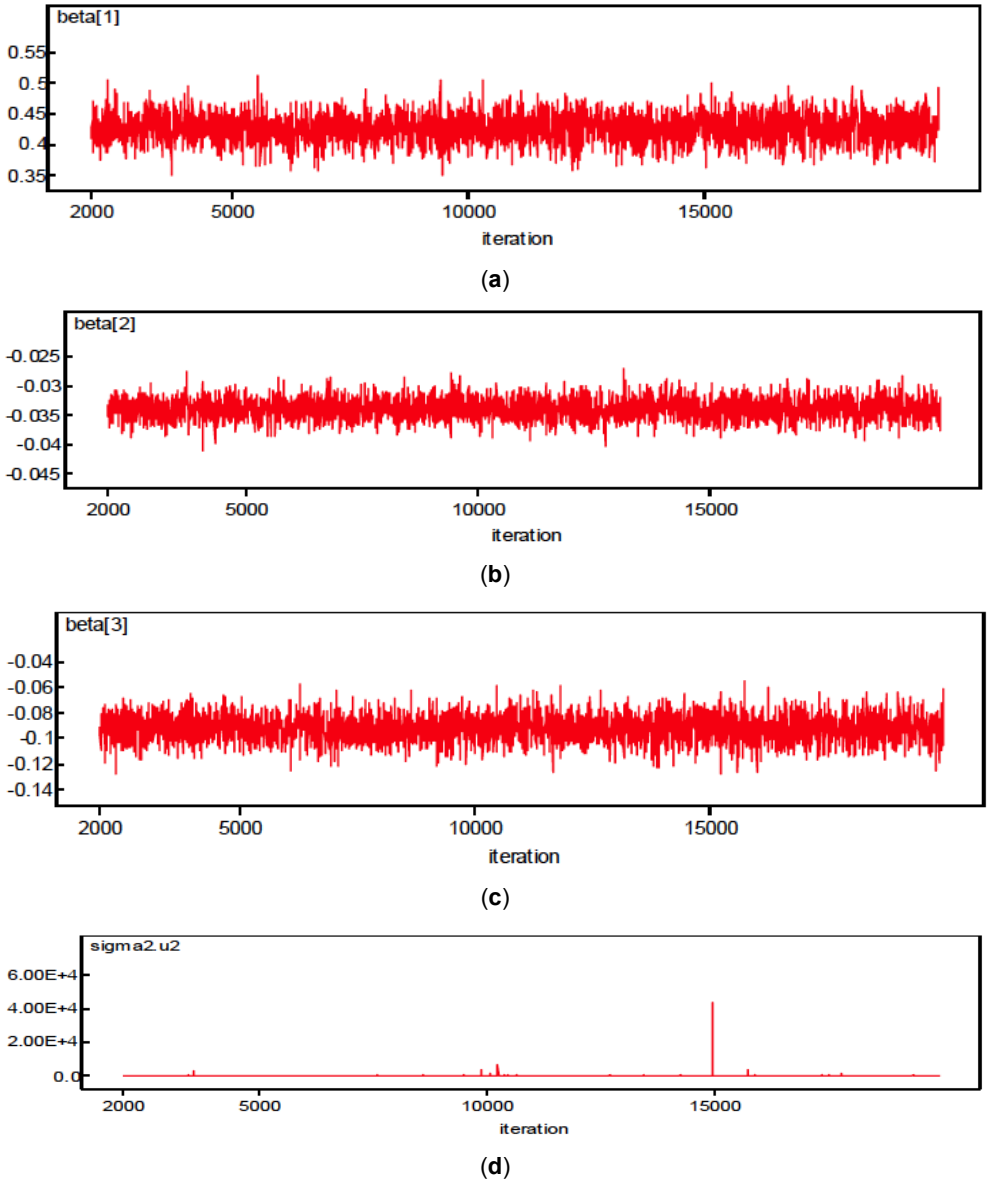


Figure 3: Time Series Plots for Parameters in CAR Model: (**a**) Intercept, (**b**) Slope for week, (**c**) Slope for temperature, and (**d**) CAR random effect.

country, let Y_t^c be the number of deaths due to EVD in week t for $t = 1,\ldots,25$ and country $c = 1,2,3$. Assuming that change point occurs at week k and that the number of deaths due to EVD in week t is a Poisson random variable, the change-point model is specified for each country as,

$$Y_t^c \sim \text{Poisson}(\lambda), \quad t = 1,\ldots,k; c = 1,2,3$$

$$Y_t^c \sim \text{Poisson}(\mu), \quad t = k+1,\ldots,n; c = 1,2,3$$

The following independent diffuse prior distributions were specified for k, λ and μ,

$$\pi(k) \sim \text{Uniform}(1,25)$$

$$\pi(\lambda) \sim \text{Normal}(0,1e6)$$

$$\pi(\mu) \sim \text{Normal}(0,1e6)$$

Parameters are estimated in SAS 9.3 ® (SAS Institute Inc. [16]) with PROC MCMC using a two-block random walk Metropolis algorithm with 20,000 iterations, a burn-in period of 2,000 and thinning set to 5 to reduce the autocorrelation in parameter draws from the posterior distributions. In the first block, the change point parameter is estimated; in the second block, the rate parameters of the Poisson distributions are estimated.

Results demonstrate that the change points occur at approximately weeks 14, 9, and 16 for Guinea, Liberia,

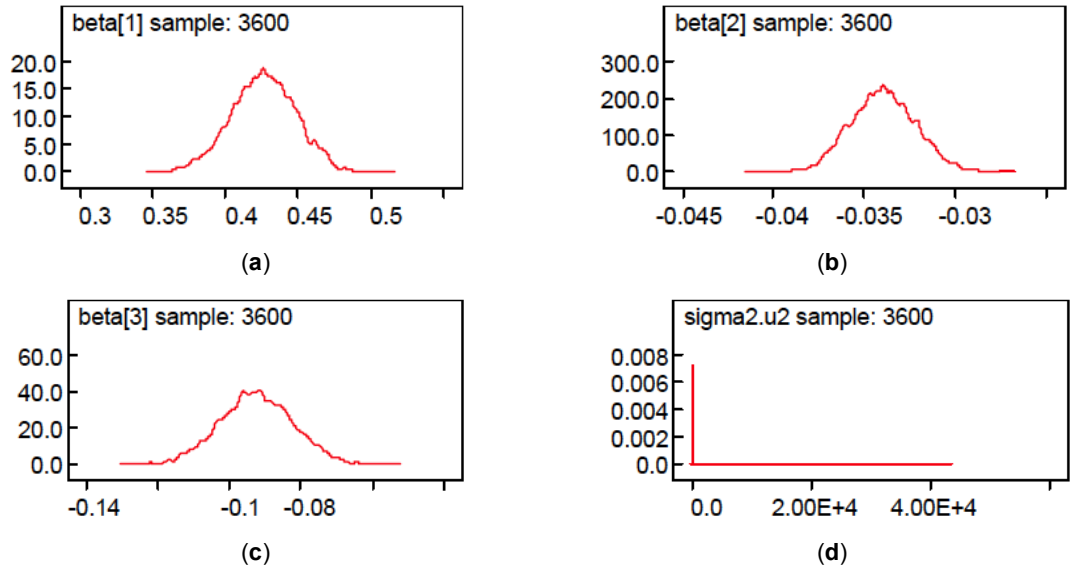


Figure 4: Density Plots for Parameters in CAR Model: (**a**) Intercept, (**b**) Slope for week, (**c**) Slope for temperature, and (**d**) CAR random effect.

Table 4: Summary of Change-Point Models by Country

	Country		
Statistic	**Guinea**	**Liberia**	**Sierra Leone**
$\hat{k}$			
Mean	13.991	8.502	15.475
SD	0.578	0.288	0.324
95% HPD interval	[13.000, 14.899]	[8.0512, 9.000]	[15.010, 15.981]
$\hat{\lambda}$			
Mean	76.286	134.500	142.000
SD	2.476	4.081	3.094
95% HPD interval	[71.643, 81.170]	[126.600, 142.700]	[135.800, 147.700]
$\hat{\mu}$			
Mean	43.045	68.106	66.665
SD	2.038	2.038	2.644
95% HPD interval	[39.092, 47.048]	[64.156, 72.022]	[61.706, 72.128]
pD	1.857	2.025	2.172
DIC	278.081	864.764	1055.747

Note. $\hat{k}$ = change point parameter estimate; $\hat{\lambda}$ = rate parameter estimate for first subsequence; $\hat{\mu}$ = rate parameter for second subsequence; *SD* = standard deviation; HPD = highest posterior density; pD = effective number of parameters; DIC = deviance information criteria.

and Sierra Leone, respectively (see Table **4**). Such an integrated estimation will be used later too. This suggests that Liberia was the first to have received support that has great impact, then followed by Guinea and then Sierra Leone. As expected, the rate parameters are smaller for each country after the change point, indicating a decrease in deaths due to EVD. Examination of time series plots suggest convergence of the algorithms for each parameter in the change-point models and the density plots suggest only one change-point is needed within the observed data for each country (see Figures **5-7**).

The analysis of the death for each country shows the mean and corresponding intervals are different for each country. In particular, it is suggestive of the change point described in Table **4**.

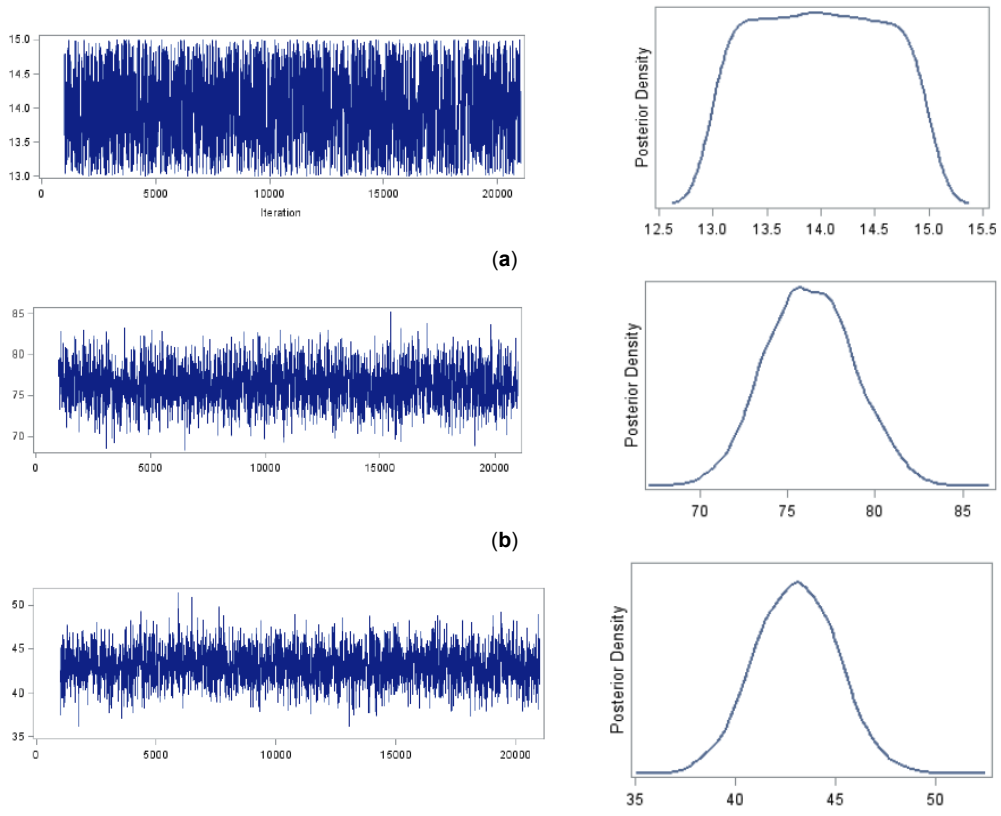


Figure 5: Time Series and Density Plots for Parameters in Change-Point Model for Guinea: (**a**) Change-point $\hat{k}$, (**b**) Rate parameter weeks 1-13 $\hat{\lambda}$, and (**c**) Rate parameter weeks 14-25 $\hat{\mu}$.

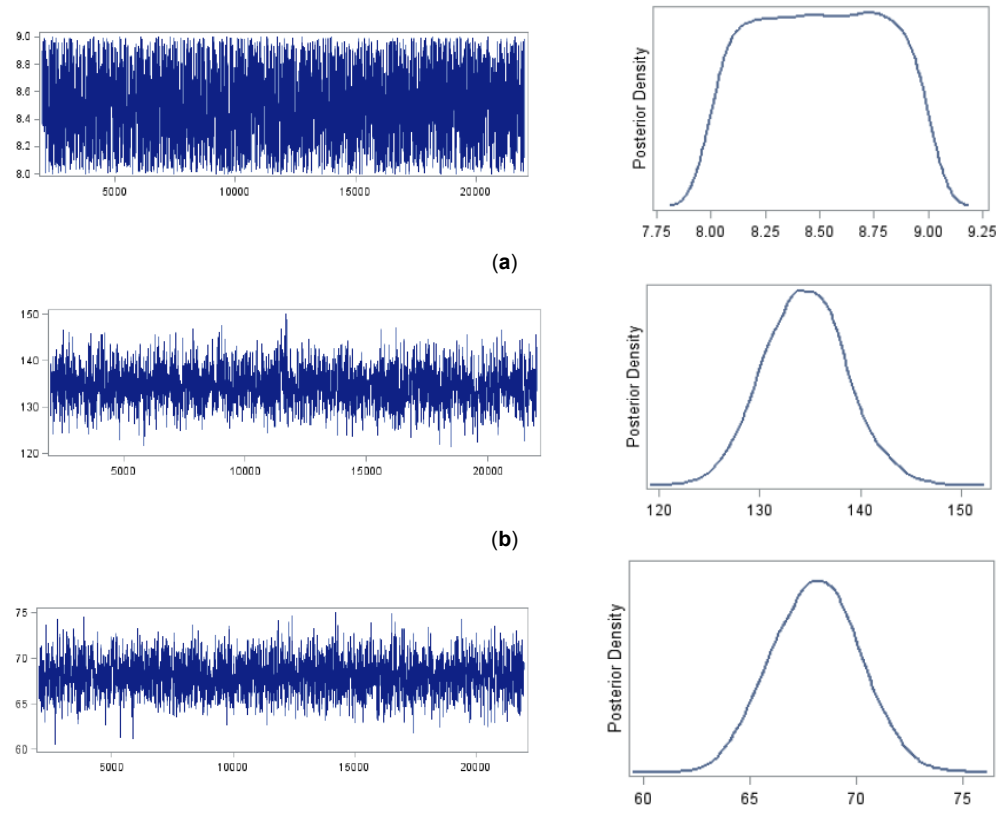


Figure 6: Time Series and Density Plots for Parameters in Change-Point Model for Liberia: (**a**) Change-point $\hat{k}$, (**b**) Rate parameter weeks 1-13 $\hat{\lambda}$, and (**c**) Rate parameter weeks 14-25 $\hat{\mu}$.

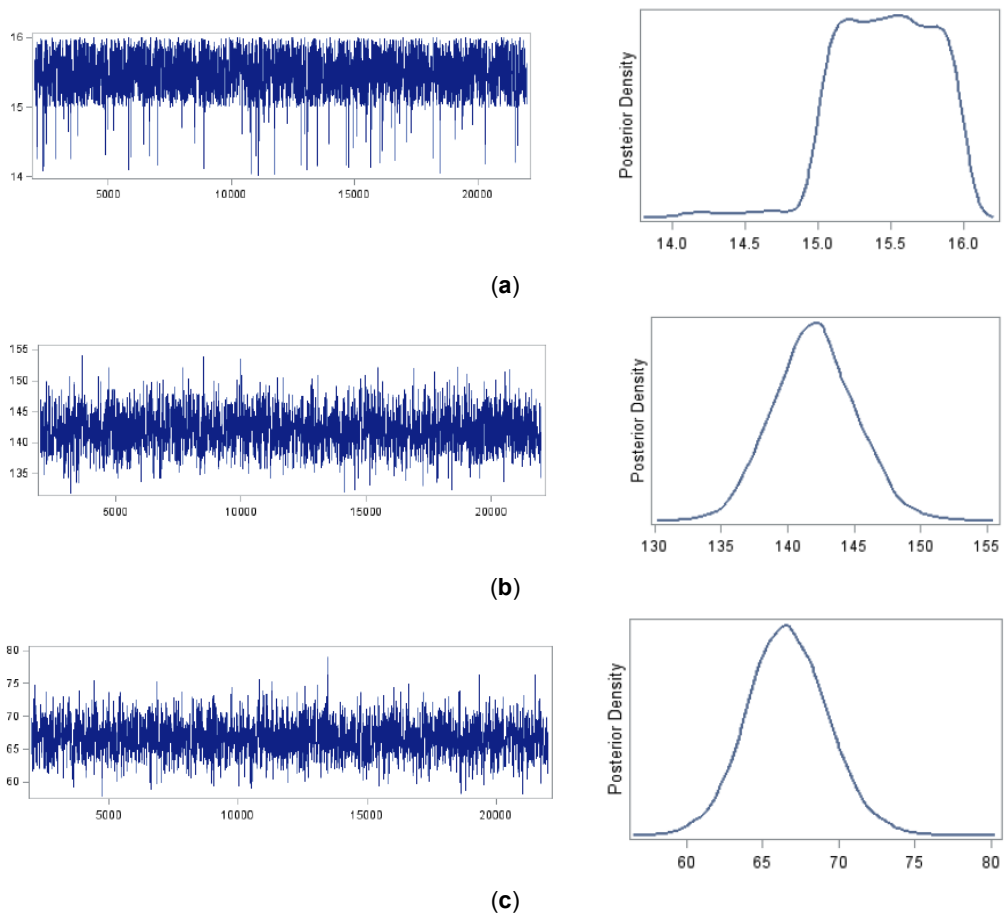


Figure 7: Time Series and Density Plots for Parameters in Change-Point Model for Sierra Leone: (**a**) Change-point $\hat{k}$, (**b**) Rate parameter weeks 1-13 $\hat{\lambda}$, and (**c**) Rate parameter weeks 14-25 $\hat{\mu}$.

The change point occurred earlier in Liberia, but also lasted the longest. Sierra Leone had a late occurrence in the change point, indicative of the fact that the disease has started there later on in time. But they have a shorter length before the change point was noticed. This is due to the fact that the awareness and support was much better targeted than in the other locations.

3.3. Conditional Autoregressive Bayesian Disease Mapping Models for Data Based on Change-Point Models

Given the occurrence of a change point in deaths due to EVD within each country, data are separated into two disjoint intervals before re-estimating the CAR models. Specifically, the deaths due to EVD data are separated into weeks 1-13 and weeks 14-25. In fact there is a belief that the epidemic will decrease after reaching a peak (http://www.cdc.gov/mmwr/pdf/other/su6303.pdf). Week 13 is used as the change point because it is the average approximate change point. The CAR model as described in Section 3.1, is specified for weeks 1-13,

$$y_{it} | u_i, x_{it} \sim \text{Poisson}(\lambda E_{it}),$$

$$\log(\lambda) = \log(E_{it}) + \beta_0 + \beta_1 \text{ week}_i + \beta_2 \text{ temperature}_{it} + u_i,$$

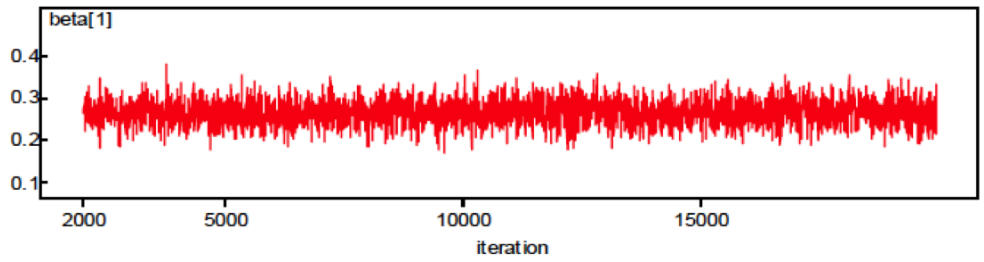
for $i = 1,2,3$ countries, $t = 1,2,\ldots,13$ weeks, and $u_i \sim \text{CAR}(\sigma_u^2)$. E_{it} is specified as an offset term. Similar to before, to facilitate interpretation of the intercept estimate, the covariate week is centered at 1 and the covariate temperature is centered around the mean of each group (this time using data only from weeks 1-13). Parameterized as such, the intercept represents the unadjusted relative death rate at week 1 (i.e., the beginning of the observed data). The same prior distributions and estimation procedure are followed in this CAR model as in the CAR model applied to the full data (see model at beginning of this section).

Results of the CAR model using data on weeks 1-13 now reveal that only temperature is significantly associated with relative death rates due to EVD (see Table **5**). Specifically, after controlling for week, increases in temperature are associated with significantly *lower* death rates from EVD. As expected,

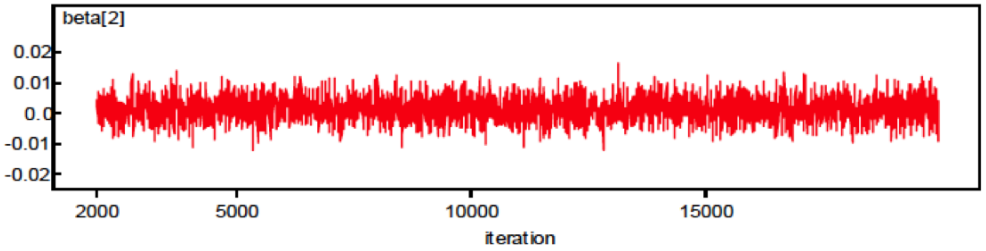
Table 5: Summary of Conditional Autoregressive Models Weeks 1-13

			95% Credible Interval	
Parameter	**Mean**	***SD***	***LL***	***UL***
Intercept	0.266	0.029	0.207	0.324
Week	0.001	0.004	-0.006	0.009
Temperature	-0.154	0.013	-0.179	-0.129
$\hat{\sigma}_u^2$	18.650	237.800	0.460	64.4500
Statistic	Estimate			
pD	17.667			
DIC	1632.760			

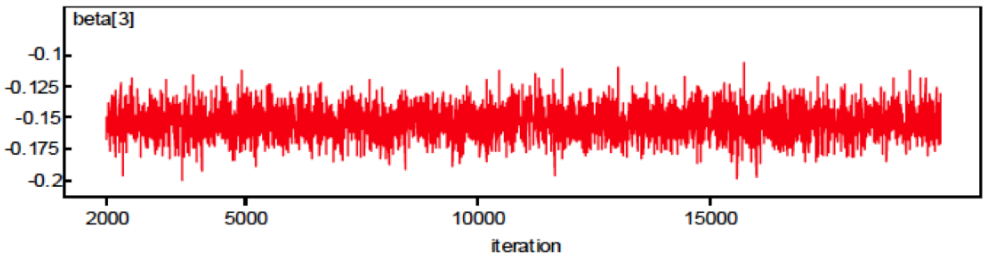
Note. SD = standard deviation, *LL* = lower limit, *UL* = upper limit. Week is centered at 1; temperature is centered around the mean of each country.



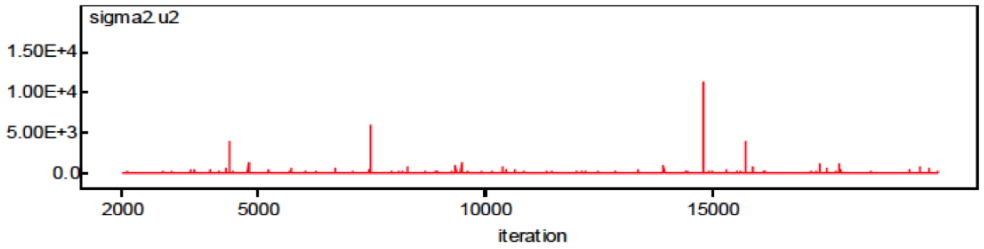
(**a**)



(**b**)



(**c**)



(**d**)

Figure 8: Time Series Plots for Parameters in CAR Model for Weeks 1-13: (**a**) Intercept, (**b**) Slope for week, (**c**) Slope for temperature, and (**d**) CAR random effect.

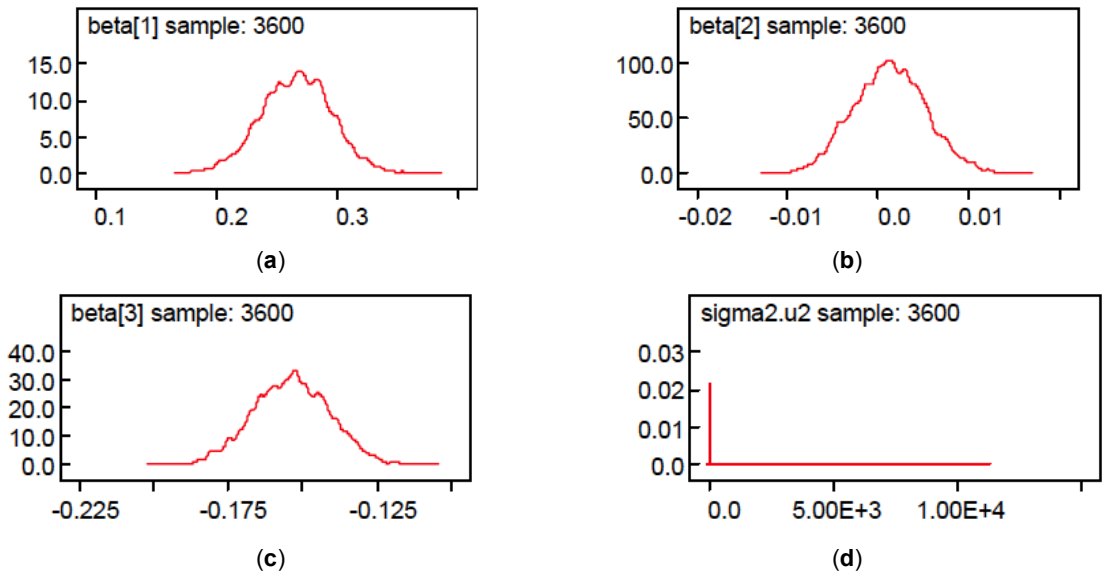


Figure 9: Density Plots for Parameters in CAR Model for Weeks 1-13: (**a**) Intercept, (**b**) Slope for week, (**c**) Slope for temperature, and (**d**) CAR random effect.

splitting the data into disjoint subsequences does reduce the standard deviation of the variance component (i.e., *SD* using weeks 1-25 = 742.4 vs. *SD* using weeks 1-13 = 237.8). Moreover, the model fit also improves (DIC = 1632.76 vs. DIC = 2552.97).

Figure **8** reveals a similar pattern to the CAR model using the full data, that is, that the algorithm appears to converge for the intercept and slope parameters, but convergence is questionable for the variance component. Again, the density plots (see Figure **9**) for the intercept and slope parameters appear normally distributed but severely positively skewed for the variance component.

Similarly, the CAR model isbuilt for weeks 14-25.

Results of the CAR model using data on weeks 14-25 now reveal that only week is significantly associated with relative death rates due to EVD (see Table **6**). Specifically, after controlling for temperature, increases in time are associated with significantly *lower* death rates from EVD. Similar to the CAR model applied to data on weeks 1-13, the smaller subset of data does reduce the standard deviation of the variance component (i.e., *SD* using weeks 1-25 = 724.4 vs. *SD* using weeks 14-25 = 81.53). The model fit also improves (DIC = 690.845 vs. DIC = 2552.97). However, the question is then to identify the effect of the foreign aids.

Figure **10** reveals a similar pattern to the previous CAR models, that is, that the algorithm appears to converge for the intercept and slope parameters but convergence is questionable for the variance component. The density plots (see Figure **11**) for the intercept and slope parameters appear normally

Table 6: Summary of Conditional Autoregressive Models Weeks 14-25

			95% Credible Interval	
Parameter	**Mean**	***SD***	***LL***	***UL***
Intercept	0.044	0.143	-0.248	0.315
Week	-0.017	0.008	-0.032	-0.002
Temperature	-0.041	0.022	-0.084	0.000
$\hat{\sigma}_u^2$	12.780	81.530	0.491	64.480
Statistic	Estimate			
pD	35.925			
DIC	690.845			

Note. *SD* = standard deviation, *LL* = lower limit, *UL* = upper limit. Week is centered at 1; temperature is centered around the mean of each country.

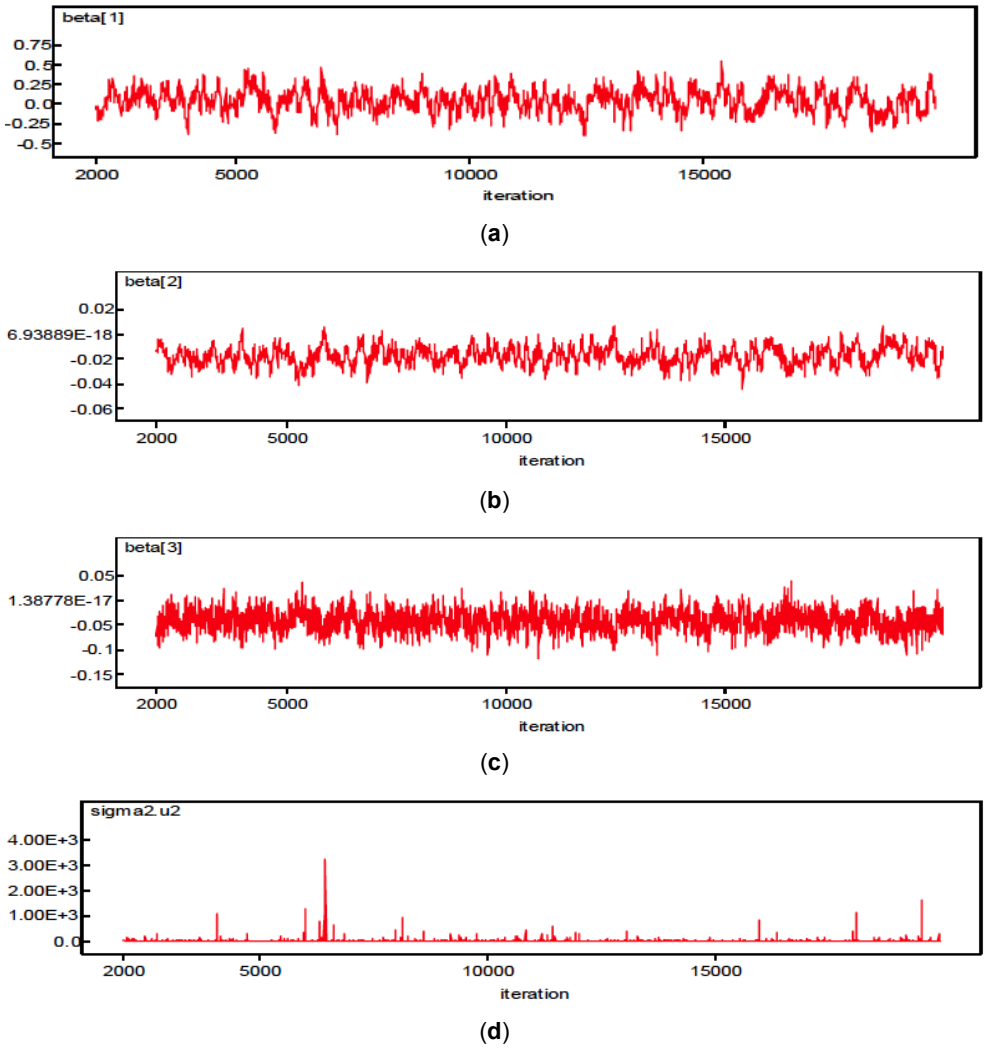


Figure 10: Time Series Plots for Parameters in CAR Model for Weeks 14- 25: (**a**) Intercept, (**b**) Slope for week, (**c**) Slope for temperature, and (d) CAR random effect.

distributed but severely non-normal for the variance component.

3.4. Generalized Liner Mixed Model

A popular set of models that are used to model non-normal repeated measures outcomes are generalized linear mixed models (GLMMs). These models share similar characteristics to linear mixed models, except the modeling can be extended to non-normal outcomes. For the EVD death count data, the following country-specific random intercept model is specified,

$$y_{it} | b_i, x_{it} \sim \text{Poisson}(\lambda E_{it}),$$

$$\log(\lambda) = \log(E_{it}) + \beta_0 + \beta_1 \text{ week}_i + \beta_2 \text{ temperature}_{it} + b_i,$$

for $i = 1,2,3$ countries, $t = 1,2,\ldots,25$ time points, and $b_i \sim \text{i.i.d.} N(0,\theta)$ represents the random intercept describing between-country variation. Similar to the CAR models, E_{it} is treated as an offset term.

The model is estimated in SAS 9.3 ® (SAS Institute Inc. [16]) using PROC GLIMMIX with a Poisson distribution specified for the outcome and a log link. Given that there is only one variance component term in the model (i.e., for the intercept), the variance components of the model refreely estimated with no constraints (i.e., TYPE = VC in SAS).

Results of the model show that both week and temperature are significantly related to relative death rates due to EVD. Specifically, increases in time and temperature are associated with significantly *lower* relative death rates due to EVD (see Table **7**). After controlling for other parameters such as week and temperature in the model, there does not appear to be significant between-country variation (VC = 0.585, *SE* =

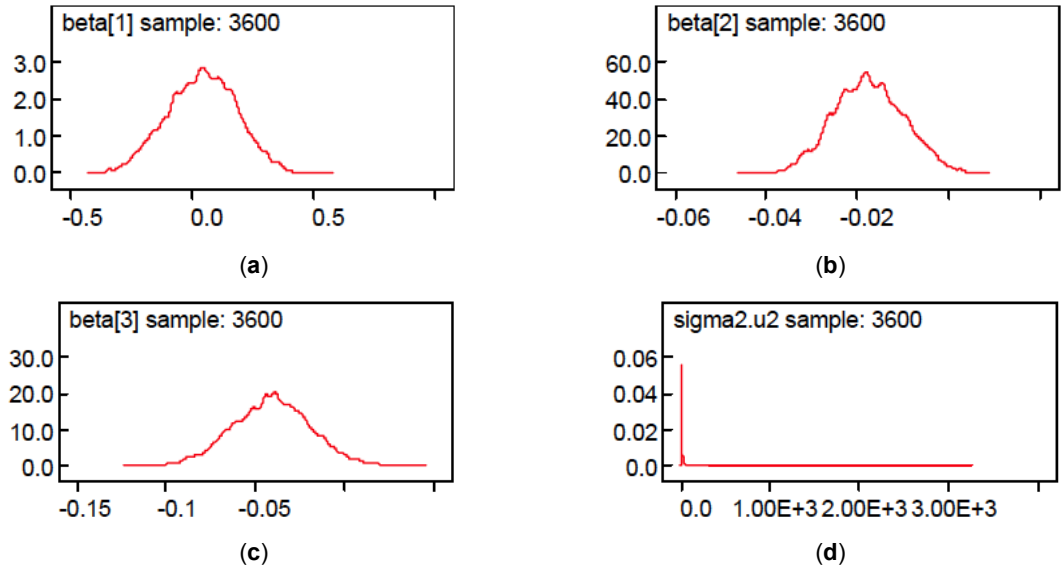


Figure 11: Density Plots for Parameters in CAR Model for Weeks 14- 25: (**a**) Intercept, (**b**) Slope for week, (**c**) Slope for temperature, and (**d**) CAR random effect.

Table 7: Summary of Generalized Linear Mixed Model: Poisson Distribution

			95% Confidence Interval	
Type	**Estimate**	***SE***	***LL***	***UL***
Fixed Effects				
Intercept	0.375	0.442	-1.527	2.277
Week	-0.030	0.002	-0.033	-0.026
Temperature	-0.110	0.010	-0.131	-0.089
Random Effects				
Overall VC	0.585	0.585		
Intercept (G)	-0.879	0.442	-1.760	0.002
Intercept (L)	0.507	0.442	-0.375	1.388
Intercept (SL)	0.372	0.442	-0.509	1.253
Statistic				
-2 Res LPL	1665.620			
Gen. χ^2	1818.360			
Gen. χ^2 / df	25.260			

Note. *SE* = standard error; *LL* = lower limit; *UL* = upper limit; VC = variance component; Res = residual; LPL = log pseudo-likelihood; Gen = generalized; G = Guinea; L = Liberia; SL = Sierra Leone. Week is centered at 1; temperature is centered around the mean of each country.

0.585). Moreover, the individual random effects for each country are not significantly different from zero.

The large value of generalized χ^2 / df indicates significant overdispersion in the data (Gardner *et al*. [7]). To model the overdisperson, the same country-specific random intercept is re-fit but using a negative binomial distribution instead of Poisson. Results show that the negative binomial model fit better than the Poisson model (-2 residual log pseudo-likelihood = 136.6 vs. 1665.62) and the data no longer exhibit a concern for overdispersion (generalized χ^2 / df = 1.00 vs. 25.26). However, in the new model, the only significant fixed effect is week, such that increases in time are associated with significantly lower death rates due to EVD. Similar to the Poisson model, after controlling for other parameters in the model, there does not appear to be significant between-country variation (VC = 0.552, *SE* = 0.563).

Table 8: Summary of Generalized Linear Mixed Model: Negative Binomial Distribution

			95% Confidence Interval	
Type	**Estimate**	***SE***	***LL***	***UL***
Fixed Effects				
Intercept	0.395	0.446	-1.526	2.315
Week	-0.031	0.009	-0.049	-0.013
Temperature	-0.087	0.054	-0.195	0.021
Scale	0.264	0.047		
Random Effects				
Overall VC	0.552	0.563		
Intercept (G)	-0.844	0.437	-1.760	0.002
Intercept (L)	0.500	0.437	-0.372	1.372
Intercept (SL)	0.344	0.437	-0.528	1.216
Statistic				
-2 Res LPL	136.600			
Gen. χ^2	71.680			
Gen. χ^2 / df	1.000			

Note. *SE* = standard error; *LL* = lower limit; *UL* = upper limit; VC = variance component; Res = residual; LPL = log pseudo-likelihood; Gen = generalized; G = Guinea; L = Liberia; SL = Sierra Leone. Week is centered at 1; temperature is centered around the mean of each country.

4. DISCUSSION

This study sought to explore the impacts of the disease in Western Africa. Several (four) statistical models were applied to data on deaths due to EVD in Guinea, Liberia, and Sierra Leone over a 25 week period. The factors influencing the disease based on the data available from the WHO website have been used to fit longitudinal structure models. When all the counts were aggregated across country, results of a CAR model demonstrated that both week and temperature had statistically significant negative effects on relative death rates due to EVD. However, the variance component representing the spatial dependency in the data was approximately zero. The results also show that Liberia that had the earliest aid was able to recover faster and that Sierra Leone seems to have been able to control the disease spread.

Bayesian change-point models revealed a significant change-point in the distribution of deaths due to EVD counts within Guinea, Liberia, and Sierra Leone. CAR models applied to the two disjoint intervals of data based on the change-point models showed that in weeks 1-13, only temperature had a statistically significant negative effect on relative death rates due to EVD, whereas in week 14-25, only week had a statistically significant negative effect on relative death rates due to EVD. Compared to using the full data, the CAR models applied to the disjoint data resulted in better model fit. In fact the CAR model is useful when the EVD incidence is increasing or decreasing phases. Overall, however, the variance components in these CAR models were approximately zero.

The findings match with the presumption that because of foreign aids, the disease has been under control. The changes in attitudes were very detrimental in slowing the disease. Although an average drop was noticed, approaching the behavior from each country behavior and resources is recommended. The country specific factors can be further explored.

When variance components are approximately zero, there are two primary interpretations from a statistical perspective: (1) after controlling for all other parameters in the model, there is not enough variation in the response to attribute any variation to the random effect, and (2) despite the near zero random effect, the random effect should be retained because it is essential to the dependent structure in the data (Kiernan, Tao, & Gibbs, [17]). In a statistical sense, it appears that little is gained by modeling the spatial dependency among data observed from contiguous countries.

Similar to the CAR model using combined counts from all countries, results of a country-specific, random intercept generalized linear mixed model revealed that both week and temperature had statistically significant negative effects on relative death rates due to EVD. However, model fit statistics indicated that deaths due to EVD under the Poisson model were overdispersed. The results under the negative binomial model were similar to those of the CAR model applied to data from weeks 14-25, that is, only week had a statistically significant negative effect on relative death rates due to EVD. Compared to the Poisson model, the negative binomial model resulted in a better fit to the data, however; both models had small variance components indicating negligible between-country variation after controlling for other model parameters.

5. CONCLUSION

EVD is a substantive public health concern in Western Africa that warrants appropriate statistical and mathematical models to understand the development and spread of the disease. Statistical investigations will increasingly shed light on the steps needed in minimizing the spread and the effort to control the disease. Results of the current study demonstrate that CAR models are a viable modelling framework for the spatially dependent nature of Ebola data. Furthermore, Bayesian change-point models are useful in identifying critical changes over time in the distribution of deaths due to EVD. Moreover, a country-specific negative binomial mixed model is useful in modeling the serial dependency in Ebola data while also accounting for additional variability in the data that the Poisson model is unable to capture.

As more data become available at the city-level, rather than the country-level, CAR models and other disease mapping models can be used to identify high risk clusters or areas where disease incidence is highest. Moreover, generalized linear mixed models can be applied to model the nested structure of the data with time nested within city, which is nested within country. These models will help identify key covariates at different levels (e.g., cost of delayed aid, socio-economic measures, city and country) that can ultimately help reduce the development of current and future EVD outbreaks. The research results have implication for predicting the spread of Ebola in a city, within a country and between countries. Such information is critical for health care professionals and management to take appropriate action to allocate medical resources and take preventive action to limit or stop the spread of this disease.

ACKNOWLEDGEMENTS

The authors are very grateful to the comments and suggestions from the editor. The reviews have greatly helped refine the manuscript.

REFERENCES

[1] World Health Organization [homepage on the Internet]. Ebola virus disease 2015. Available from: http://www.who.int/mediacentre/factsheets/fs103/en/.

[2] Dallatomasina S, Crestani R, Squire SS, Sylvester SJ, *et al.* Ebola outbreak in rural West Africa: epidemiology, clinical features and outcomes. Trop Med Int Health 2015; 20: 448-54. http://dx.doi.org/10.1111/tmi.12454

[3] McKinley T, Cook AR, Deardon R. Inference in epidemic models without likelihoods. The International Journal of Biostatistics 2009; 5(1): 1-40. http://dx.doi.org/10.2202/1557-4679.1171

[4] Lekone PE, Finkenstadt BF. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. Biometrics 2006; 62: 1170-77. http://dx.doi.org/10.1111/j.1541-0420.2006.00609.x

[5] Millar RB. Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes's factors. Biometrics 2009; 65: 962-69. http://dx.doi.org/10.1111/j.1541-0420.2008.01162.x

[6] Hossain MM, Lawson AB, Cai B, Choi J, Liu J, Kirby RS. Space-time areal mixture modeling: Relabeling algorithm and model selection issues. Environmetrics 2014; 25: 84-96. http://dx.doi.org/10.1002/env.2265

[7] Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psych Bulletin 1995; 118: 392-404. http://dx.doi.org/10.1037/0033-2909.118.3.392

[8] Navarro A, Utzet F, Puig P, Caminal J, Martin M. Negative binomial distribution versus Poisson in the analysis of recurrent phenomena. Gaceta Sanitaria 2001; 15: 447-52. http://dx.doi.org/10.1016/S0213-9111(01)71599-3

[9] Ng S, Basta NE, Cowling BJ. Association between temperature, humidity and ebolavirus disease outbreaks in Africa, 1976 to 2014. Euro Surveill 2014; 19: 1-11. http://dx.doi.org/10.2807/1560-7917.es2014.19.35.20892

[10] Carlin BP, Louis TA. Bayesian methods for data analysis. 3rd ed. Boca Raton, FL: CRC Press 2009.

[11] Lawson AB. Bayesian disease mapping: Hierarchical modeling in spatial epidemiology. 2nd ed. Boca Raton, FL: CRC Press 2013.

[12] Gelman A, Carlin J, Stern H, Rubin DB. Bayesian data analysis.3rd ed. London: Chapman & Hall 2013.

[13] Raudenbush SW, Bryk AS. Hierarchical linear models. 2nd ed. Thousand Oaks: Sage Publications 2002.

[14] Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Stats Comp 2000; 10: 325-37. http://dx.doi.org/10.1023/A:1008929526011

[15] Hawkins DM. Fitting multiple change-point models to data. Comp Stats Data Analysis 2001; 37: 323-41. http://dx.doi.org/10.1016/S0167-9473(00)00068-2

[16] SAS Institute Inc. Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc., 2011.

[17] Kiernan K, Tao J, Gibbs P. Tips and strategies for mixed modelling with SAS/STAT procedures. SAS 2012 Global Forum: Orlando, FL 2012.

Received on 19-09-2015 Accepted on 12-10-2015 Published on 03-11-2015

http://dx.doi.org/10.6000/1929-6029.2015.04.04.1