Published in final edited form as:

Appl Geogr. 2018 January ; 90: 272–281. doi:10.1016/j.apgeog.2017.10.003.

An Exploration of the Spatiotemporal and Demographic Patterns of Ebola Virus Disease Epidemic in West Africa Using Open Access Data Sources

Vasile A. Suchar^a, Noha Aziz^b, Amanda Bowe^c, Aran Burke^d, and Michelle M. Wiest^{e,*}

^aUniversity of Idaho, Department of Statistical Science, 875 Perimeter Drive MS1104, Moscow, Idaho 83844-1104, USA. vasiles@uidaho.edu

^bUniversity of Idaho, Department of Statistical Science, 875 Perimeter Drive MS1104, Moscow, Idaho 83844-1104, USA. noha.aziz@gmail.com

^cUniversity of Idaho, Department of Statistical Science, 875 Perimeter Drive MS1104, Moscow, Idaho 83844-1104, USA. abowe@uidaho.edu

^dUniversity of Idaho, Department of Statistical Science, 875 Perimeter Drive MS1104, Moscow, Idaho 83844-1104, USA.

^eUniversity of Idaho, Department of Statistical Science, 875 Perimeter Drive MS1104, Moscow, Idaho 83844-1104, USA. mwiest@uidaho.edu

Abstract

The purpose of this study was to investigate the utility of exploratory analytical techniques using publicly available data in informing interventions in case of infectious diseases outbreaks. More exactly spatiotemporal and multivariate methods were used to characterize the dynamics of the Ebola Virus Disease (EVD) epidemic in West Africa, and propose plausible relationships with demographic/social risk factors. The analysis showed that there was significant spatial, temporal, and spatiotemporal dependence in the evolution of the disease. For the first part of the epidemic, the cases were highly clustered in a few administrative units, in the proximity of the point of origin of the outbreak, possibly offering the opportunity to stop the spread of the disease. Later in the epidemic, high clusters were observed, but only in Liberia and Sierra Leone. Although not definitely factors of risk, in the setting in which the epidemic arose, our analysis suggests infrastructure, access to and use of health services, and connectivity possibly accelerated and magnified the spread of EVD. Also, the spatial, temporal, and spatiotemporal patterns of epidemic can be clearly shown – with evident application in the early stages of management of epidemics. In particular, we found that the spatial-temporal analytic tool SaTScan may be used effectively during the evolution of an epidemic to identify areas for targeted intervention. In the case of EVD epidemic in West Africa, better data and integration local knowledge and customs may have been more useful to recognize the proper response.

*Corresponding author.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Graphical Abstract

Keywords

Ebola; West Africa; spatiotemporal analysis; demography; cluster analysis

1. Introduction

The West African Ebola epidemic of 2014 - the largest in history - arose in a much different cultural setting than previous outbreaks. Previous outbreaks had occurred in isolated villages, whose people had experience with Ebola and were unlikely to travel great distances to seek medical care. In contrast, continuous movement of people from their villages (even while very sick), across borders from Guinea to either Sierra Leone or Liberia, and into urban centers, drove the rapid spread of Ebola to neighboring West African countries, into cities, in a matter of days (WHO, 2016a). Over three years, 28,616 confirmed, probable and suspected cases have been reported in West Africa, resulting in 11,310 deaths (WHO, 2016a). The magnitude of this epidemic and the difficulty containing it suggests the need for better understanding of dynamics of the EVD.

While it is well recognized that interventions such as isolation of patients and safe and sanitary funerals and burials played a vital role in controlling the epidemic, as did the people's own adaptation (G. Chowell & Nishiura, 2014; Richards, 2015; Rivers, Lofgren, Marathe, Eubank, & Lewis, 2014), the purpose of this study was to investigate if exploratory analytical techniques using publically available data can provide insights into epidemic dynamics and propose plausible relationships with demographic/social risk factors. More exactly, principal components, spatial, temporal, and spatiotemporal analysis were used to assess the patterns of the EVD epidemic, identify the areas and time intervals of high risk, and identify the associated risk factors which can influence the risk of infection.

The analysis of spatiotemporal-distributed disease data can be used to identify the presence of absence of areas with significant differences in risk (Sherman, et al., 2014), identify possible periodical patterns in the behavior of the disease (Marek, Tušek, & Pászto, 2015), and propose effective responses to outbreaks (Martins-Melo, Ramos, Alencar, Lange, & Heukelbach, 2012). However, these spatiotemporal analytic techniques cannot actually identify the factors leading to the observed patterns. Therefore, multivariate methods were used to explore if there are any patterns within the socio-demographic variables considered, and any relationship between these variables and the EVD case counts.

The paper is structured as follow: Section 2 describes the data processing and the statistical methodology. Section 3 presents the results of our analysis. Section 4 discuss the results, while section 5 contains the recommendations and conclusions.

2. Methods

2.1 Data processing and description

To calculate the daily and weekly number of cases in each administrative district, a dataset provided by OCHA ROWCA on the Humanitarian Data Exchange (HDX, 2015), that compiled the number of cases released by various sources including the WHO, national health ministries, and other sources was used. The dataset recorded daily cumulative total, confirmed, probable and suspected cases, as well as new cases and the number of deaths from March 24, 2014, up to March 28, 2015. The records cover six countries in West Africa: Guinea, Liberia, Mali, Nigeria, Senegal, and Sierra Leone, at various administrative units' levels.

For the study, the most severely affected countries by the EVD epidemic were selected: Guinea, Liberia, and Sierra Leone, with a total of 63 administrative units. Additional data was collected from published reports (Fink & Sheri, 2014; HumanitarianResponse, 2016; WHO, 2016b). As a result, the coverage of the case counts was extended, and some of the missing entries (June 2014 -August 2014), and errors (end of 2014 – beginning of 2015) were corrected. Figure 1 shows the weekly case counts based on the original and appended datasets. Supplementary material file S1 describes briefly the method used in calculating the case counts. The final datasets had daily and weekly Ebola virus counts and rates for the three West African countries (aggregated over 63 administrative units) from December 06, 2013 to March 28, 2015.

For demographic and health information about West Africa, datasets provided by USAID on the Demographic and Health Surveys (DHS) Program website (USAID, 2016) were used. DHS collects, analyzes, and disseminates population, health, HIV, and nutrition data in over 90 countries. More details about the data collected in West Africa can be found in the DHS reports specific to each country (INS, 2013; LISGIS, 2014; SSL, 2014). For this research, the focus was on surrogate variables for factors hypothesized to be possible risk factors in the transmission of diseases: having access to bicycles, motorcycles or cars, women having a hospital delivery, women having a doctor or medical professional present at birth, education level, literacy, and reading newspapers, listening to radio or watching TV, sharing a toilet with other households, number of children living at home, and the mean number of STD and

sexual partners. These factors represent access to information, transportation, healthcare, and behavior that might modify risk of exposure. The data was aggregated to the administrative unit, and variables were expressed as a percent of respondents having that attribute.

2.2 Spatial analysis

Spatial analysis methods were used to evaluate the geographical distribution of the weekly Ebola infection rates. The administrative units from which the case counts were recorded were considered the units for the analysis.

The presence of spatial dependence was assessed using Global and Local Moran's I indexes for each of the 70 weeks of the epidemic considered in our dataset. The rates of infection were used instead of case counts since, generally, the number of cases is correlated to the underlying population size, and sometimes spatial autocorrelation may be detected only as an artifact of the spatial distribution of the population (R. S. Bivand, Pebesma, & Gomez-Rubio, 2008). For the local Moran's I analysis, Holm p-value adjustments were used to assess the significance of each test (Brunsdon & Comber, 2015).

In the analysis, three distance measures were considered: first two are commonly used in spatial analysis: a *contiguity based neighbors matrix*, and a *centroid based distance matrix* (R. S. Bivand, et al., 2008). Since it was suggested that the dispersal of Ebola virus is supported by the proximity of infected people to main roads (Hui-Jun, et al., 2015), a *population-weighted road distance matrix* (Mitze, 2012) was also considered. To calculate the population-weighted road distance matrix, a list of major cities with their complementary population sizes for the three states considered was compiled from various internet sources (Brinkhoff, 2015; Wikipedia, 2015), but not limited to them. The web-based information available for these countries is scarce, inconsistent, and the town names were often different from source to source. In the end, a list of 83 cities from across all administrative units was compiled. The road-based distances (in km) between all them was calculated using the `ggmap` function in R (Kahle & Wickham, 2013) and Google Maps for the city pairs unrecognized by the R package. The population-weighted distances between administrative units were calculated as described by Mitze (Mitze, 2012).

The contiguity based weight matrix was row-standardized, and inverse-distance weight matrices were generated for the centroid and road distances. The analysis was conducted in Rlanguage (R Core Team, 2016), using the R packages `PBS mapping`, `spdep`, and `ape` (R. Bivand & Piras, 2015; Paradis, Claude, & Strimmer, 2004; Schnute, Boers, & Haigh, 2015).

2.3 Temporal analysis

To test for temporal dependence, a multivariate ARMAX model was considered (Shumway & Stoffer, 2011). The multivariate ARMAX model expressed the counts of new Ebola cases, in a given administrative unit, as a linear combination of the trend and past counts of Ebola cases in all the other administrative units.

The case counts at time t were expressed as:

$$y_{t,i} = \alpha_i + \beta_i t + \sum_{k=1}^K (\phi_{i,i} y_{t-k,i} + \sum_{j=1}^N \phi_{i,j} y_{t-k,j}) + w_{t,i} \quad (1)$$

for each of the $i = 1, 2, \dots, N$ administrative units.

Where: $k = 1, 2, \dots, K$ is the k -order time lag, $j = 1, 2, \dots, N$ indicates the administrative units j . $y_{t,i}$ represents the case counts at time t and location i . $y_{t-k,i}$ and $y_{t-k,j}$ are the case counts at time $t-k$ and locations i and j , respectively. And $w_{t,i}$ represents the residuals assumed correlated over the locations i , but independent over time.

The analysis was conducted in R-language (R Core Team, 2016), using the R package *vars* (Pfaff, 2008a, 2008b). Model residuals were checked to see if they fit the model assumptions: tests for the absence of serial correlation (Portmanteau test), heteroscedasticity (multivariate ARCH test) and normality (Jarque-Bera test). Non-normality and conditional heteroscedasticity are not often a concern for the validity of the models, especially in this case where the model was not considered final, but may help better the model deficiencies and the underlying properties of the data (Luetkepohl, 2011).

2.4 Spatial-temporal analysis

A third exploratory analytical approach looks at the retrospective spatiotemporal cluster analysis for the high and low incidence of the weekly Poisson-distributed count cases at each location. An analysis was conducted using the SaTScan software for the spatial and space-time scan statistics (Kulldorff, 2009).

For each location and time step, the scan analysis expects, under the null hypothesis, that the number of cases is proportional to its population size. The alternate hypothesis is that there is an elevated risk within the scanning window as compared to outside (Kulldorff, 1997, 2009; Kulldorff, Athas, Feuer, Miller, & Key, 1998). A maximum likelihood ratio test statistic and a p -value are calculated for a number of Monte Carlo replications (for this study set at 9999). Identified clusters are ordered based on their likelihood ratio test values (Kulldorff, 1997, 2009). The program scans for clusters of geographical size between zero and some user-defined upper limit, called population at risk. The authors recommend to use values of 50% for the upper limit of the population at risk especially when in doubt. It should be noted that population at risk is not referring to susceptible as defined in SIR models, but rather as a geographical susceptibility. In the current research, several upper limits of the percent of population at risk (10 to 50%) were tested, and the results were compared.

In addition, the usefulness of the spatial-temporal analysis for real-time prioritization of interventions was evaluated. To do so, prospective spatial-temporal analyses were conducted at monthly intervals, starting with week six of the epidemic. The prospective analysis identifies spatial-temporal clusters that are current, i.e. existent (or “live”) at the end date of the dataset analyzed.

2.5 Analysis of risk factors

In order to place the results of the spatial, temporal, and spatial-temporal analysis in the geographical context, the DHS data was used in our investigation. A principal components analysis (PCA) of the selected variables was conducted, and the dynamics of the EVD epidemic were visually assessed in the context of the relationships proposed by the PCA and knowledge of the social interactions between ethnic/religious groups in West Africa. PCA is used to identify patterns in multi-dimensional data and express it such that it highlights their similarities and differences (Johnson & Wichern, 2002). In this case, the methodology aimed to identify which districts were more similar along the socio-demographic characteristics and evaluate if these similarities correlate to similarities in EVD epidemic dynamics.

In order to assess which factors of risk may be associated with the observed trends, correlation tests were used for the total number of cases and selected principal components.

3. Results

3.1 Spatial analysis

Global Moran's I: Figure 2 presents a summary of the global Moran's I values over the 70 week period. Each pair of plots for the three weight matrices indicate the changes in the global Moran's I values, the p-value and the 0.05 level of significance line over time. For all distance matrices considered, we observe a pattern of alternating significant positive autocorrelation with non-significant autocorrelation. Moran's I values tend to increase over time, indicating an increase in the spatial autocorrelation as the disease evolved, and intervention measures take place. Moran's I values ranged from -0.03 to about 0.2 for the centroid and road distance matrices, and from -0.04 to 0.63 for the contiguity weight matrix. Week 30 in the epidemic period seem to be the first week with significant positive autocorrelation for all weight matrices.

Local Moran's I: The complete local Moran's I analysis can be found in Supplementary material files S2 to S4. Due to the low number of administrative units with non-zero cases, until week 27 there were only a few time periods when local Moran's I values could be calculated. Overall, the results for all weight matrices were similar, but the clusters for the population-weighted road distance were less significant. Figure 3 shows the local Moran's I plots (p values are not shown) for three representative time periods. Initially, there is a cluster of the disease around three administrative units, Gueckedou and Macenta (Guinea), then in Lofa (Liberia). From week 27 to week 32–33 a hot spot of the epidemic can be identified at the border area of the three countries Kailahun (Sierra Leone), Lofa (Liberia) and Gueckedou (Guinea). The disease seems still localized in that area (surrounding counties have very dissimilar values). Up to week 40, the epidemic continues to be mainly clustered in the tri-state area, but significant values can be observed in Liberia, in the Montserrado area which becomes the center of a cluster of the Liberia outbreak, on and off until week 63. From week 51 until the end of the covered epidemic period, another cluster of significant autocorrelation can be observed in the NW region of Sierra Leone (Port Loko, Bombali, and Kambia).

Overall, the local spatial analysis highlights the initial cluster of the Ebola epidemic in the tri-state area, followed by a second cluster in Liberia, and a third in the NW Guinea. The results indicate that for several weeks, the outbreak was fairly localized, but later as it spread in West Africa, affected more heavily the highly populated areas, and their neighbors. A cluster of similar low values can be seen in the NE of Guinea almost for the entire duration of the epidemic.

3.2 Temporal analysis

The analysis was conducted for the daily cases data with maximum time lag of five days. Bigger time lags could not be tested due to overfitting. The analysis for weekly cases data could not be fit due to overfitting, while the model for the weekly infection rates leads to computational errors.

Most models for the daily cases data with time lags of five days had multiple R-squared values above 85%, and only one administrative unit showed nonsignificant temporal dependence (Dingiraye). The models with higher time lags performed better than the ones with lower lags. The complete temporal analysis (424 pages) is available by request. The hypothesis of no serial correlation was rejected (the noise is not white), suggesting that the model does not fully capture the temporal dependence component. Also, the hypothesis of normality was rejected, while the heteroscedasticity test could not be performed due to overfitting.

3.3 Spatiotemporal analysis

The results of the 20% and 50% population at risk spatiotemporal clustering analysis are shown in Figure 4 and for 10% to 50% population at risk in Table 1. The choice of different percentages of the population at risk yielded a wide range of number of clusters, from eleven to one cluster(s) for 10% and 50% population at risk, respectively. But regardless of the values considered, we see clusters of significantly higher than expected case counts centered on Liberia and Sierra Leone from week 35 to 64, and clusters of significantly lower than expected case counts in the first 35 weeks centered in Guinea, and along the border of the affected area. The same pattern was observed in the local spatial analysis, with the outbreak moving from the point of origin (in the tri-state area) directionally towards parts of Liberia and Sierra Leone. Looking at the results of the 20% population at risk clustering analysis, two stages of the epidemic can be distinguished: First, two clusters of significantly lower case counts for weeks one to 36 (clusters 3 and 4), followed by two high case counts clusters in Liberia for weeks 36 to 53 (clusters 2 and 5), and by one high case counts cluster in Sierra Leone for weeks 43 to 65 (cluster 1). The same pattern was observed during the spatial analysis, with an initial period when the outbreak was highly localized, followed by an explosion of cases in Liberia, and then in Sierra Leone. If we examine the 50% population at risk results, we see one high case counts cluster over the Liberia and Sierra Leone.

The prospective clustering analysis at various time periods during the epidemic is presented in Supplementary material file S5. Table 2 summarizes the counties in high clusters in the prospective spatial-temporal analysis. Five of the six initial counties in high clusters are still

present in high clusters at the end of the study period. Also, at least 50% of counties in a “live” high cluster at the end of the month are still in a high cluster a month later.

3.4 Analysis of risk factors

Principal components analysis (PCA) was used to determine the sources of maximum variance across the counties from the DHS selected variables. The first four components of the PCA captured approximately 77% of the variation, with the first component capturing 40%. Figure 5 plots the first two principal components.

The first component had positive loadings on education, literacy, reading the paper, and health care facility use for births, while no education and number of children were negatively loaded. This component is highlighting the differences between the urban centers and suburban areas, with Monserrado, Conakry, Western Urban and Western Rural with much higher education and literacy rates. However, many outlying counties in Liberia and Sierra Leone also had fairly high scores for the first principle component, while, aside from Conakry, most counties in Guinea were scored very low.

The second principle component captured 18% of the variation and had positive loads on access to motor vehicles, higher education, watching television, and having a doctor present at births. The third component captured 12% of the variance and weighted access to bicycle, motorcycle, and car heavily. One might interpret these two components as capturing sources of variation as socioeconomic status.

The forth principle component captured 7% of the variation. This component had positive loadings on hospital delivery and doctors present at birth, and negative weights on prevalence of STDs and number of sex partners. This component seems to capture different health infrastructure across the region.

We correlated the first four principle component scores of each county with the rate of infection for that county, to determine if the variation captured by PCA was associated with the burden of disease. We found strong evidence the first principle component was positively correlated (0.55; (0.35, 0.70)). There is very weak evidence that the third principle component was positively correlated (0.27; (0.02, 0.49)) with the county infection rate. There was no evidence of association with the second and forth components ((-0.38, 0.10); (-0.38, 0.10)).

4. Discussion

4.1 Spatial analysis

Global Moran’s I: The selection of the “best” weight matrix was data-driven (Dray, Legendre, & Peres-Neto, 2006). The contiguity matrix lead to the best model performance or best Moran’s I (Getis & Aldstadt, 2004). The contiguity matrix also satisfied other proposed recommendations, such as: (1) under-specified matrices (fewer neighbors) should be preferred instead of overspecified weight matrices (extra neighbors), and (2) variables showing a good deal of local spatial heterogeneity should probably be modeled by fewer links in weight matrix (Getis & Aldstadt, 2004). As a warning, the selection of the

contiguity matrix as the “best” should be interpreted with caution, since Moran’s I test can incorrectly suggest the presence of spatial autocorrelation in the presence of other effects (Bivand et al., 2008; Viton, 2010).

Assuming that the contiguity matrix is capturing the spatial autocorrelation most accurately, the global Moran’s I analysis indicates high spatial heterogeneity (since each county is more alike its neighbors – with little influence from more distant districts). This may be an indicator of movement primarily occurring within neighboring counties, which lead to a spatial aggregation of cases. The pattern of alternating significant positive autocorrelation, where there was an EVD emergence shortly followed by a EVD diffusion (non-significant autocorrelation), can suggest several things: (1) people were leaving the high-risk areas, as soon as the disease reemerged, (2) localized medical response targeted preferentially, and effectively, the reemergence areas or (3) it is an artifact due to timing of case reporting. In time, there is an overall positive trend in the global Moran’s I values. This suggests that, as efforts were increasingly directed towards the treatment and prevention of the disease, the new outbreaks became more localized – a possible indicator that the intervention efforts were effective.

Since it was suggested that the dispersal of Ebola virus was supported by the proximity of infected people to main roads (Hui-Jun, et al., 2015), the population-weighted road distance weight matrix was expected to yield the highest Moran’s I values (i.e. had the best explanatory power). Instead, the contiguity matrix lead to higher Moran’s I values. The global spatial analysis did not support or negate this hypothesis, but it suggested that it is more plausible that there were other risk factors than main road proximity facilitating the spread of the EVD. For example, patterns of marriages and family ties may not adhere to the infrastructure in the region, and hence the transmission patterns in rural areas could have followed these family connections through rough and sometime treacherous routes (Richards, 2015).

Local Moran’s I: Overall, the results for all weight matrices were similar. But the clusters for the population-weighted road distance were less significant. It suggests the for this particular weight matrix, the underlying process was more stable (homogeneous) within the data, and the local values had about the same contribution to the global statistic. This suggests that the proximity to main roads probably was a factor of risk, but, as suggested by the global spatial analysis, was not the only factor (i.e., population size played also a role, as expected).

The local spatial analysis highlights the initial cluster of the Ebola epidemic in the tristate area, followed by a second cluster in Liberia, and a third in the NW Guinea. The results indicate that for several weeks, the outbreak was fairly localized, but later as it spread in West Africa, affected more heavily the highly populated areas, and their neighbors. A cluster of similar low values can be seen in the NE of Guinea almost for the entire duration of the epidemic.

A significant result proposed by the local spatial analysis was that for several weeks the disease was fairly localized in the tri-state area. This can be interpreted in two ways: (1) it is

possible that there was an opportunity to contain the outbreak for a fairly large period of time, or (2) despite the sustained efforts to contain the initial outbreak, the EVD broke and spread in all West Africa. Regardless of the interpretation, the results reiterate the need for strong, sustained containment efforts right from the beginning of any outbreak. Epidemic models usually propose exponential growth in the number of cases. And this was true, at least in Liberia and Sierra Leone. Before that exponential growth, however, an initial built-up period of several months (resembling an Allee-like effect) seemed to have been present in the EVD population.

4.2 Temporal analysis

The important results of the time series analysis were: (1) there was a strong temporal dependence in the changes in Ebola cases, and (2) the time series were not fully capturing the dynamics of the disease. A limitation of the method used consists in the ARMA assumption that the error term is white noise, approximately normally distributed with mean zero. But, by using count data, we violate this assumption since negative observations cannot occur. Moreover, the ARMA model approach ignores the fact that the data is discrete instead of continuous. Another limitation of the approach is that the administrative units with zero counts were eliminated from the analysis, and the missing counts had to be converted to zeros – which might alter the temporal dynamics of the epidemic. Therefore, in this context, the method can be used solely as a relative evaluation of the temporal dependence.

Overall the results discussed so far suggests that in the case of EVD epidemic there was both strong spatial and temporal dependence. Thus, it illustrates the need for spatiotemporal analysis.

4.3 Spatiotemporal analysis

The spatiotemporal clustering analysis indicated there was significant clustering of cases in time and space. There were significantly higher than expected case counts centered on Liberia and Sierra Leone from week 35 to 64. There were clusters of significantly lower than expected case counts in the first 35 weeks centered in Guinea and along the border of the affected area. The same pattern was observed during the spatial analysis, with an initial period when the outbreak was highly localized – followed by an explosion of cases in Liberia, and then in Sierra Leone. If we examine the 50% population at risk results, we see one high case counts cluster over the Liberia and Sierra Leone from week 37 to 64. A cluster of large size is indicative of areas of exceptionally low rates outside of the circle (Kulldorff, 2009). This confirms again the directionality of the spread of the Ebola virus disease: although it started in Guinea, it spread towards Liberia and Sierra Leone, with less than expected case counts in Guinea. While this result is hardly surprising in the context of the previous discussion, it still highlights the fact that Guinea had less than expected case counts for the entire epidemic.

The choice of percent population at risk value had a significant effect on the results. In this type of analysis, population at risk is not equivalent to the percent of population susceptible to the disease as defined in SIR models, but rather it refers to the spatial population at risk. Since the transmission chain was spatially limited (Ajelli, et al., 2015; Faye, et al., 2015;

Lau, et al., 2017), it is plausible that the results using smaller population at risk values are more reliable.

A more important result came from the prospective spatial-temporal analysis (Supplementary file S6 and Table 2). The analysis was intended to assess the potential of the method to identify “live” spatial-temporal clusters of the EVD and their evolution throughout the extent of the epidemic. Most of the counties (five out of six) present in the early high clusters were still present at the end of the study period. An important question was: can the current “live” high clusters be used to predict the “live” clusters a month from now? The proportion of counties present in “live” clusters for two consecutive months ranged from 50 to 100%. Therefore, it is a strong possibility that concentration of resources in a current “live” cluster may be the *necessary* strategy to reduce the severity of an epidemic, but by no means should it be expected to be the *sufficient* strategy. The method was already used in identifying outbreaks of shigellosis in Chicago (Jones, Liberatore, Fernandez, & Gerber, 2006), and is currently used in the daily automated spatiotemporal analysis in New York for early outbreak detection of 35 reportable diseases (Greene, Peterson, Kapell, Fine, & Kulldorff, 2016). The current research indicated the tool can potentially be useful in effective early response to other diseases, such as Ebola, as long as the data can be collected, recorded, geocoded and analyzed in real-time.

4.4 Analysis of demographic and social metrics

Overall, the spatial analysis indicated the presence of spatial and temporal dependence of the EVD epidemic, identifying spatiotemporal locations of hot spots. But can we use existing demographic data to identify attributes of the population at risk so that interventions and messaging can be appropriately tailored?

In other analyses, a wide variety of factors were found to have a significant effect on the EVD outbreak dynamics. They include case isolation, safe burials, behavioral changes, differences in intervention control, population size, time to travel to large population centers, proximity to roads and hospitals, precipitation, and temperature seasonality (Gerardo Chowell & Nishiura, 2015; D'Silva & Eisenberg, 2017; Dudas, et al., 2017; Fang, et al., 2016; Zinszer, Morrison, Verma, & Brownstein, 2017). Also, Ebola spread most readily between members of the extended family or community (Ajelli, et al., 2015; Faye, et al., 2015), and/or over a short distance (Lau, et al., 2017).

The spatial-temporal analysis supports some of these proposed risk factors. For example, the analysis confirmed that population size and density might have increased the risk of exposure in the highly populated areas of Liberia and Sierra Leone, but not in Guinea. The high-density areas in Guinea had actually less than expected infection rates.

Also, the results of the PCA highlight the range of conditions under which Ebola spread across the three highly affected countries. If nothing else, a quick analysis of DHS survey data can help identify resources and important demographics to aid in intervention design and messaging. Further, the first principle component, which captures the diverse access to education, birth control, and hospitals, is moderately correlated with the average case rates

over the epidemic. The areas with higher infection risk (i.e., access to education, and health care) are also the areas with higher population density and connectivity.

Finally, our analysis suggests the combination of less infrastructure, slowing the rate of contact – and therefore the spread of EVD – along with the international response to curb the epidemic, resulted in lower case counts in Guinea than in more connected, more developed, Sierra Leone and Liberia. When we examine the epidemic curves in each county, most counties in Guinea were infected later, and the local outbreaks were more quickly resolved. We also see from Figure 6 that the usage of hospitals was associated with this pattern of earlier infection, and longer duration of the local outbreak in counties in Sierra Leone and Liberia. Given that most villages in Guinea would have heard about the epidemic, and received information from aid organizations and their government, it is also likely villages leaders had the opportunity to consider how they would implement control strategies before they experienced their first case.

While these type of results can inform interventions, we have to remember that the social variables used in the analyses can be considered, at most, baseline characteristics of the populations affected by the EVD outbreak. Moreover, without taking into account the direct effect of interventions, it is not possible to evaluate whether these are predictive.

4.5 Methodology limitations

There are a few limitations inherent to the methods themselves. For example, the spatial autocorrelation analysis is sensitive to spatial scale effects, the different polygons' shapes and sizes, and border effects. Further, the spatial autocorrelation test can incorrectly suggest the presence of spatial autocorrelation in the presence of other effects, the temporal analysis is clearly suggesting other effects beside the autoregressive terms, and the correlation analysis may be spurious. Even under the assumption that the previously mentioned limitations are not an issue in this case, the range of conclusions based on this analysis is rather limited. Spatial, temporal, and spatiotemporal patterns can be shown - with clear application in the management of the epidemic – but the drivers of the epidemics and the effects of interventions cannot be accounted for with the data we used. Finally, our dataset did not contain data on individual cases, and is based on aggregate data at the county level. Therefore, we recognize that some associations we are observing could suffer from ecological fallacy.

4.6 Data sources limitations

It is well-recognized that reliable and accurate information is essential to evaluate and improve the delivery of health services. In the case of epidemics of emerging infectious diseases and crises, data collection is often difficult if not impossible. Therefore, considerable efforts were placed to collect standardized, high quality data in West Africa prior and during the EVD disease. Still, it is still possible that the quality of the dataset may raise questions about the reliability of the analytic results:

First, the original dataset of case counts had gaps, incorrect counts, and unexplained drops in the cumulative case counts in several administrative units. Second, the early symptoms of Ebola are, for the most part, indistinguishable from malaria, and posed a major challenge

when identifying probable cases. While measures were taken to amend these problems, there is a certain degree of uncertainty about the calculated daily/weekly new cases. Third, it has been documented that there was under-reporting of the number of Ebola cases (Westcott, 2014; Zavis & Healy, 2016). Fourth, although the DHS data provide an immense source of information about certain aspects of West African socio-demographic developments, these data also had limitations: Only certain ethnic groups are defined for each country, with minority groups receiving the generalized designation of “other.” This limits the ability to investigate social similarities in border counties which are known to share people of the same ethnicities of neighboring countries.

5. Recommendations

This research investigated if exploratory analytical techniques using publically available data can inform interventions in case of infectious diseases outbreaks. More specifically, the methods were used to evaluate the dynamics and causes of the EVD epidemic in West Africa. The results showed that there was significant spatial, temporal, and spatiotemporal dependence in the evolution of the disease. For the first part of the epidemic, the cases were highly clustered in a few administrative units, in the proximity of the point of origin of the epidemic, offering the opportunity to stop the spread of the disease, pending a robust, directional intervention. Later in the epidemic, high clusters were observed only in Liberia and Sierra Leone. The spatial-temporal analytic tool SaTScan may be used effectively during the evolution of an epidemic. We also confirmed that DHS surveys can be used to characterize the demography, existence of resources, and typical usage of clinics and hospitals. These data can help guide resources and perhaps help predict spread, although we cannot claim to have definitely identified them as risk factors.

Based on these results, we believe that these exploratory techniques can be useful for monitoring purposes, as tools for early detection of potential outbreaks. Early in an outbreak, data is usually sparse, the potential risk factors are largely unknown, and resources are limited. The presented methods have the advantage of being fairly straight forward, require rather low resources, while the results are quite reliable. The presented analysis can indicate if and where disease clusters are. Later in the epidemic, as interventions and behavioral changes are shaping the dynamics of the outbreak, they have to be taken in consideration in the analytic efforts. This requires more sophisticated modeling approaches (which has been extensively addressed in the literature), better data, integration of local knowledge and customs, and substantially higher resources.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This publication was supported by the EPSCoR Program, National Science Foundation #IIA-1301792, the Mountain West Clinical and Translational Research - Infrastructure Network, NIH, National Institute of General Medical Sciences (NIGMS) #1U54GM104944-01A1, and NIGMS #P20GM104420. Special thanks go to Steven Radil for his suggestions, and the anonymous referees, whose comments have greatly improved the paper.

Appendix A:: Supplementary material

S1: Case counts calculation methodology

S2: Local Moran's I analysis for contiguity matrix.

S3: Local Moran's I analysis for centroid matrix.

S4: Local Moran's I analysis for road matrix.

S5: Prospective analysis.

References

- Ajelli M, Parlamento S, Bome D, Kebbi A, Atzori A, Frasson C, Putoto G, Carraro D, & Merler S (2015). The 2014 Ebola virus disease outbreak in Pujehun, Sierra Leone: epidemiology and impact of interventions. *Bmc Medicine*, 13 (1), 281. [PubMed: 26607790]
- Bivand R, & Piras G (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, 63, 1–36.
- Bivand RS, Pebesma EJ, & Gomez-Rubio V (2008). *Applied spatial data analysis with R*. Springer, New York, NY.
- Brinkhoff T (2015). City Population. In (Vol. 2015). <http://www.citypopulation.de/>.
- Brunsdon C, & Comber L (2015). *An introduction to R for spatial analysis and mapping*. Sage Publications, UK.
- Chowell G, & Nishiura H (2014). Transmission dynamics and control of Ebola virus disease (EVD): a review. *Bmc Medicine*, 12. [PubMed: 24460622]
- Chowell G, & Nishiura H (2015). Characterizing the Transmission Dynamics and Control of Ebola Virus Disease. *Plos Biology*, 13 (1), e1002057. [PubMed: 25607595]
- D'Silva JP, & Eisenberg MC (2017). Modeling spatial invasion of Ebola in West Africa. *Journal of Theoretical Biology*, 428, 65–75. [PubMed: 28551366]
- Dray S, Legendre P, & Peres-Neto PR (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196 (3–4), 483–493.
- Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, Park DJ, Ladner JT, Arias A, Asogun D, Bielejec F, Caddy SL, Cotten M, D'Ambrozio J, Dellicour S, Caro AD, Diclaro JW, Duraffour S, Elmore MJ, Fakoli LS, Faye O, Gilbert ML, Gevaio SM, Gire S, Gladden-Young A, Gnirke A, Goba A, Grant DS, Haagmans BL, Hiscox JA, Jah U, Kugelman JR, Liu D, Lu J, Malboeuf CM, Mate S, Matthews DA, Matranga CB, Meredith LW, Qu J, Quick J, Pas SD, Phan MVT, Pollakis G, Reusken CB, Sanchez-Lockhart M, Schaffner SF, Schieffelin JS, Sealfon RS, Simon-Loriere E, Smits SL, Stoecker K, Thorne L, Tobin EA, Vandi MA, Watson SJ, West K, Whitmer S, Wiley MR, Winnicki SM, Wohl S, Wölfel R, Yozwiak NL, Andersen KG, Blyden SO, Bolay F, Carroll MW, Dahn B, Diallo B, Formenty P, Fraser C, Gao GF, Garry RF, Goodfellow I, Günther S, Happi CT, Holmes EC, Kargbo B, Keita S, Kellam P, Koopmans MPG, Kuhn JH, Loman NJ, Magassouba NF, Naidoo D, Nichol ST, Nyenswah T, Palacios G, Pybus OG, Sabeti PC, Sall A, Ströher U, Wurie I, Suchard MA, Lemey P, & Rambaut A (2017). Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature*, *advance online publication*.
- Fang LQ, Yang Y, Jiang JF, Yao HW, Kargbo D, Li XL, Jiang BG, Kargbo B, Tong YG, Wang YW, Liu K, Kamara A, Daffae F, Kanu A, Jiang RR, Sun Y, Sun RX, Chen WJ, Ma MJ, Dean NE, Thomas H, Longini IM, Halloran ME, & Cao WC (2016). Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone. *Proceedings of the National Academy of Sciences of the United States of America*, 113 (16), 4488–4493. [PubMed: 27035948]
- Faye O, Boëlle P-Y, Heleze E, Faye O, Loucoubar C, Magassouba NF, Soropogui B, Keita S, Gakou T, Bah EHI, Koivogui L, Sall AA, & Cauchemez S (2015). Chains of transmission and control of

- Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *The Lancet Infectious Diseases*, 15 (3), 320–326. [PubMed: 25619149]
- Fink DG, & Sheri. (2014). Tracing Ebola's Breakout to an African 2-Year-Old.
- Getis A, & Aldstadt J (2004). Constructing the spatial weights matrix using a local statistic. *Geographical Analysis*, 36 (2), 90–104.
- Greene SK, Peterson ER, Kapell D, Fine AD, & Kulldorff M (2016). Daily Reportable Disease Spatiotemporal Cluster Detection, New York City, New York, USA, 2014–2015. *Emerging Infectious Diseases*, 22 (10), 1808–1812. [PubMed: 27648777]
- HDX. (2015). Sub-national time series data on Ebola cases and deaths in Guinea, Liberia, Sierra Leone, Nigeria, Senegal and Mali since March 2014. Humanitarian Data Exchange Retrieved November 7, 2015 from <https://data.hdx.rwllabs.org/dataset/rowca-ebola-cases> In.
- Hui-Jun L, Jun Q, David K, Xiao-Guang Z, Fan Y, Yi H, Yang S, Yu-Xi C, Yong-Qiang D, Hao-Xiang S, Foday D, Yu S, Cheng-Yu W, Wei-Min N, Chang-Qing B, Zhi-Ping X, Kun L, Brima K, George FG, & Jia-Fu J (2015). Ebola Virus Outbreak Investigation, Sierra Leone, September 28–November 11, 2014 *Emerging Infectious Disease journal*, 21 (11).
- HumanitarianResponse (2016). Sierra Leone Situation Reports. Government of Sierra Leone <https://www.humanitarianresponse.info/>.
- INS. (2013). Enquete demographique et de sante et a indicateurs multiples (EDS-MICS 2012) In. Institut National de la Statistique (INS) and ICF International, Conakry, Guinée and Calverton, Maryland, USA.
- Johnson RA, & Wichern DW (2002). *Applied multivariate statistical analysis* (5th ed.). Prentice Hall, Upper Saddle River, NJ.
- Jones RC, Liberatore M, Fernandez JR, & Gerber SI (2006). Use of a Prospective Space-Time Scan Statistic to Prioritize Shigellosis Case Investigations in an Urban Jurisdiction. *Public Health Reports*, 121 (2), 133–139.
- Kahle D, & Wickham H (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal*, 5 (1), 144–161.
- Kulldorff M (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26, 1481–1496.
- Kulldorff M (2009). SaTScanTM v9.0: Software for the spatial and space-time scan statistics. <http://www.satscan.org>. In. <http://www.satscan.org>.
- Kulldorff M, Athas WF, Feuer EJ, Miller BA, & Key CR (1998). Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health*, 88 (9), 1377–1380. [PubMed: 9736881]
- Lau MSY, Dalziel BD, Funk S, McClelland A, Tiffany A, Riley S, Metcalf CJE, & Grenfell BT (2017). Spatial and temporal dynamics of superspreading events in the 2014–2015 West Africa Ebola epidemic. *Proceedings of the National Academy of Sciences*, 114 (9), 2337–2342.
- LISGIS. (2014). Liberia Demographic and Health Survey 2013 In. Liberia Institute of Statistics and GeoInformation Services (LISGIS), Ministry of Health and Social Welfare, National AIDS Control Program, and ICF International, Monrovia, Liberia.
- Luetkepohl H (2011). Vector autoregressive models. *EUI Working Papers (ECO 2011/30)*, 33.
- Marek L, Tu ek P, & Pászto V (2015). Using geovisual analytics in Google Earth to understand disease distribution: a case study of campylobacteriosis in the Czech Republic (2008–2012). *International Journal of Health Geographics*, 14 (1), 1–13. [PubMed: 25563056]
- Martins-Melo FR, Ramos AN, Alencar CH, Lange W, & Heukelbach J (2012). Mortality of Chagas' disease in Brazil: spatial patterns and definition of high-risk areas. *Tropical Medicine & International Health*, 17 (9), 1066–1075. [PubMed: 22809055]
- Mitze T (2012). *Empirical modeling in regional science: towards a global time-space-structural analysis* (Vol. 657). Springer.
- Paradis E, Claude J, & Strimmer K (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290. [PubMed: 14734327]
- Pfaff B (2008a). *Analysis of Integrated and Cointegrated Time Series with R* (2nd ed.). Springer, New York, NY.

- Pfaff B (2008b). VAR, SVAR and SVEC Models: Implementation Within R Package vars. *Journal of Statistical Software*, 27 (4), 1–32.
- R Core Team. (2016). R: A language and environment for statistical computing. In. R Foundation for Statistical Computing, Vienna, Austria.
- Richards P (2015). *How a people's science helped end an epidemic*. Zed Books.
- Rivers CM, Lofgren ET, Marathe M, Eubank S, & Lewis BL (2014). Modeling the impact of interventions on an epidemic of ebola in sierra leone and liberia. *PLoS Curr*, 6.
- Schnute JT, Boers N, & Haigh R (2015). PBSmapping: Mapping Fisheries Data and Spatial Analysis Tools. In. <http://CRAN.R-project.org/package=PBSmapping>.
- Sherman RL, Henry KA, Tannenbaum SL, Feaster DJ, Kobetz E, & Lee DJ (2014). Applying Spatial Analysis Tools in Public Health: An Example Using SaTScan to Detect Geographic Targets for Colorectal Cancer Screening Interventions. *Preventing Chronic Disease*, 11, E41. [PubMed: 24650619]
- Shumway RH, & Stoffer DS (2011). *Time series analysis and its applications. With R examples* (3rd ed. Vol. Springer), New York, NY.
- SSL. (2014). Sierra Leone Demographic and Health Survey 2013. In. Statistics Sierra Leone (SSL) and ICF International, Freetown, Sierra Leone and Rockville, Maryland, USA.
- USAID. (2016). Demographic and Health Surveys (DHS) Program. <http://dhsprogram.com/>. In.
- Westcott L (2014). Sierra Leone Grapples With Spike in Ebola Numbers. In. @newsweek.
- WHO. (2016a). Situation Report. Ebola virus disease. 10 June 2016. In (Vol. 2016). World Health Organization <http://www.who.int/csr/disease/ebola/en/>.
- WHO. (2016b). WHO | Disease Outbreak News (DONs) <http://www.who.int/csr/don/en/>. WHO.
- Wikipedia. (2015). Wikipedia: The Free Encyclopedia. In (Vol. 2015). Wikimedia Foundation Inc. <http://www.wikipedia.org>.
- Zavis A, & Healy M (2016). Ebola cases in West Africa may be vastly underreported, WHO says. In. @latimes.
- Zinszer K, Morrison K, Verma A, & Brownstein JS (2017). Spatial Determinants of Ebola Virus Disease Risk for the West African Epidemic. *PLoS Curr*, 9, ecurrents.outbreaks.b494f492c496a396c472ec424cb4142765bb4142795.

Highlights:

- Data-driven spatiotemporal methods were used for EVD outbreak dynamics & risk factors
- Successive high clusters were localized in tristate area, Liberia and Sierra Leone
- Infrastructure and familial connectivity may have facilitated the dispersal of EVD
- Real time analysis can be useful in early response to epidemics
- Effective data collection is essential for real-time intervention strategies

Figure 1: Weekly case counts calculated from the original dataset (left), and calculated from the appended dataset (right).

Figure 2:
Global Moran's I values (top) and p-values (bottom) for contiguity (black), centroid (dark grey) and road (light grey) distance matrices.

Figure 3: Local Moran's I values for the contiguity, centroid and road distances (left to right columns), for weeks 27, 43, and 66 (top to bottom rows). Yellow circles indicate districts with Local Moran's I p-values (Holm's method) < 0.05.

Figure 4:
Retrospective space-time clusters for 20% (top) and – 50% (bottom) population at risk.

Figure 5:
Principal component 1 vs. Principal component 2

Figure 6:
Hospital visits (%) kriging interpolation

Table 1:

Spatiotemporal clusters

Cluster	Administrative units	Weeks	Observed / expected cases ratio	Log Likelihood ratio	p-value
<i>10% population at risk</i>					
1.	Montserrado, Bomi, Margibi	39 to 53	14.01	9688.80	<0.001
2.	Western Rural, Western Urban, Port Loko	41 to 68	8.83	8750.47	<0.001
3.	Tonkolili, Bombali, Bo, Moyamba	43 to 56	5.71	2016.15	<0.001
4.	Dalaba, Pita, Labe, Mamou, Lelouma, Tougue, Koubia, Kindia	1 to 35	0	1210.20	<0.001
5.	Lofa, Macenta	36 to 51	6.71	1187.39	<0.001
6.	Kankan, Kerouane, Mandiana, Kouroussa, Kissidougou, Beyla	1 to 35	0.0059	1182.84	<0.001
7.	Nimba, Yamou, Bong, Grand Gedeh, River Cess, Nzerekore, Grand Bassa	1 to 34	0.0073	1066.95	<0.001
8.	Boke, Boffa, Fria, Telimele, Gaoual, Dubreka	1 to 35	0.041	828.11	<0.001
9.	Conakry	1 to 35	0.059	824.36	<0.001
10.	Kono, Gueckedou, Kailahun, Kenema	30 to 64	2.21	533.95	<0.001
11.	Forecariah, Kambia, Coyah	1 to 35	0.0020	488.15	<0.001
<i>20% population at risk</i>					
1.	Port Loko, Kambia, Western Rural, Western Urban, Moyamba, Forecariah, Bombali, Tonkolili	43 to 65	6.78	9921.75	<0.001
2.	Montserrado, Bomi, Margibi	39 to 53	14.01	9688.80	<0.001
3.	Siguiro, Dingiraye, Mandiana, Kouroussa, Kankan, Dabola, Tougue, Faranah, Koubia, Kerouane, Kissidougou, Mamou, Mali	1 to 35	0.0042	2426.32	<0.001
4.	Boffa, Fria, Dubreka, Boke, Telimele, Conakry, Coyah, Kindia	2 to 36	0.045	2027.60	<0.001
5.	Macenta, Lofa	36 to 51	6.71	1187.39	<0.001
<i>30% population at risk</i>					
1.	Pujehun, Grand Cape Mount, Bonthe, Bo, Kenema, Bomi, Moyamba, Kailahun, Montserrado, Gbarpolu, Tonkolili, Margibi, Kono, Western Rural, Port Loko, Lofa	35 to 64	4.93	14407.70	<0.001
2.	Dingiraye, Tougue, Dabola, Koubia, Kouroussa, Siguiro, Mali, Mamou, Dalaba, Labe, Faranah, Lelouma, Pita, Kankan, Mandiana, Koinadugu, Kissidougou, Kindia, Koundara, Gaoual	1 to 35	0.0029	3647.86	<0.001
<i>40% population at risk</i>					
1.	Pujehun, Grand Cape Mount, Bonthe, Bo, Kenema, Bomi, Moyamba, Kailahun, Montserrado, Gbarpolu, Tonkolili, Margibi, Kono, Western Rural, Port Loko, Lofa, Gueckedou, Bong, Western Urban, Grand Bassa, Bombali	35 to 64	4.78	21808.09	<0.001
<i>50% population at risk</i>					
1.	Kenema, Bo, Kailahun, Pujehun, Kono, Grand Cape Mount, Tonkolili, Gbarpolu, Bonthe, Gueckedou, Moyamba, Bomi, Lofa, Koinadugu, Montserrado, Bombali, Margibi, Port Loko, Kissidougou, Bong, Western Rural, Macenta, Kambia, Western Urban	37 to 64	4.72	22969.37	<0.001

Table 2:

Counties in high observed/expected clusters (black) at the end of the time period

County/week	14	18	22	26	30	34	38	42	46	50	54	58	62
Beyla			v	v				v					
BoCounty					v	v	v	v	v	v	v	v	v
Boffa					v	v							
Bombali									v	v	v	v	v
Bomi						v	v	v	v	v	v	v	v
Bong							v	v		v	v	v	v
Bonthe							v						
Conakry			v	v	v		v						
Coyah									v		v		
Dubreka					v	v							
Fria					v	v							
Gbarpolu		v			v	v	v	v	v	v			
Grand Bassa							v		v	v	v	v	
GrandCape								v				v	v
GrandGedeh										v	v	v	
Gueckedou	v	v	v	v	v	v	v		v	v	v	v	
Kailahun					v	v	v	v	v	v	v	v	
Kambia									v				
Kankan													
Kenema					v	v	v	v	v	v	v	v	
Kerouane	v												
Koinadugu												v	
Kono							v				v	v	v
Lofa	v	v	v	v	v	v	v	v	v	v	v	v	v
Lola			v	v				v					
Macenta	v	v	v	v	v	v		v	v	v	v	v	v
Margibi							v	v	v	v	v	v	v
Montserrado						v						v	v
Moyamba							v	v		v	v	v	v
Nimba							v			v	v	v	v
Nzerekore	v		v	v				v		v	v	v	
Port Loko							v	v	v	v	v	v	v
RiverCess										v	v		
Telimele					v	v							
Tonkolili									v	v	v	v	v
Western area rural							v	v		v	v	v	v
Western area urban								v		v	v	v	v
Yomou	v		v	v			v	v		v	v	v	v
Percent*	0.50	0.75	1	0.5	0.92	0.61	0.63	0.5	0.87	0.96	0.96	0.68	0.82

*Percent of counties that remains in a high cluster at the end of the next prospective analysis