BMJ Open

Geospatial clustering and correlates of deaths during the Ebola outbreak in Liberia: a Bayesian geoadditive semiparametric analysis of nationally representative cross-sectional survey data

Fiifi Amoako Johnson , Barbara Sakyi

To cite: Amoako Johnson F, Sakyi B. Geospatial clustering and correlates of deaths during the Ebola outbreak in Liberia: a Bayesian geoadditive semiparametric analysis of nationally representative cross-sectional survey data. *BMJ Open* 2022;**12**:e054095. doi:10.1136/bmjopen-2021-054095

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-054095).

Received 16 June 2021
Accepted 17 June 2022

Department of Population and Health, Faculty of Social Sciences, College of Humanities and Legal Studies, University of Cape Coast, Cape Coast, Ghana

Correspondence to
Professor Fiifi Amoako Johnson; famoakojohnson@ucc.edu.gh

ABSTRACT

Objective To investigate the extent of geospatial clustering of reported deaths during the Ebola outbreak in Liberia and the covariates associated with the observed clustering.

Design Cross-sectional study.

Participants Male and female respondents from the 2019–2020 Liberia Demographic and Health Survey. The analysis covered 11 928 (women=7854 and men=4074) respondents for whom complete data were available.

Outcome measures The outcome variable was the death of a household member or relative during the Ebola outbreak in Liberia, coded 1 if the respondent reported death and 0 otherwise.

Methods We applied the Bayesian geoadditive semiparametric regression to examine the extent of geospatial clustering of deaths at the district-level and community-level development and socioeconomic factors associated with the observed clustering.

Results Almost a quarter (24.8%) of all respondents reported the death of a household member or relative during the Ebola outbreak. The results show that deaths were clustered within districts in six (Grand Cape Mount, Bomi, Monsterrado, Margibi, Gbarpolu and Lofa) of the 15 counties in Liberia. Districts with high death clustering were all near or shared borders with Sierra Leone and Guinea. The community-level development indicators (global human footprint, gross cell production and population density) had a non-linear associative effect with the observed spatial clustering. Also, respondents' characteristics (respondent's age (non-linear effect), educational attainment and urban-rural place of residence) were associated with the observed clustering. The results show that death clustering during outbreaks was constrained to poor settings and impacts areas of moderate and high socioeconomic development.

Conclusion Reported deaths during the Ebola outbreak in Liberia were not randomly distributed at the district level but clustered. The findings highlight the need to identify at-risk populations during epidemics and respond with the needed interventions to save lives.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ In many low-income and middle-come countries, fatalities in times of epidemics are under-reported, and there are no registers to identify geographic areas where deaths are particularly clustered.
⇒ The present study has used cross-sectional data of reported deaths of household members and relatives during the Ebola outbreak in Liberia to identify geographical clusters (districts) where reported deaths at the time of the epidemic were statistically significantly high.
⇒ The identified clusters align with reported Ebola virus transmission rates at the district level.
⇒ The deaths are not directly linked to the Ebola virus disease and there are no records to verify their veracity; however, all things being equal, if Ebola deaths during the outbreak were not concentrated within some districts, then we will expect a random spatial scatter of deaths.

INTRODUCTION

Ebola, a viral infection known to be one of the deadliest pathogens to infect humans, has been a global public health concern since its outbreak in 1976, near the Ebola River in the Democratic Republic of Congo (DRC).[1 2] The transmission mode is primarily through human-to-human contact with infected body fluids, causing severe and acute systemic diseases.[2] Since the first outbreak in 1976, there have been over 20 reported outbreaks in sub-Saharan Africa, mainly in Sudan, Uganda, DRC and Gabon, with the largest and most complex outbreaks, between 2014 and 2016, predominantly in Guinea, Sierra Leone and Liberia.[1 3 4] More cases and fatalities were reported in this outbreak than all the others combined.[5]

Analysis of clinical data from Guinea, Liberia and Sierra Leone revealed an overall average case fatality rate of 62.9% (95% CI 61.9% to 64.0%).[6] There were more than 28 000 cases of the Ebola virus disease (EVD), of which over 11 000 were fatal.[2 5] However, there are high uncertainties about the precise number of cases and fatalities and their spatial distributions.[7–9] Nevertheless, understanding the geospatial distribution of cases and deaths at the local administrative level where health planning and interventions are implemented and monitored is essential for responding to epidemics. Thus, there is the need to identify geographic clusters where unfavourable epidemic outcomes are high and their associated covariates. Given the lack of a comprehensive register of reported deaths attributable to Ebola in Liberia, we used reported deaths of household members and relatives during the outbreak which occurred between 30 March 2014 and 1 June 2016[10 11] to investigate the extent of geospatial clustering (at the district level) of fatalities at the time of the epidemic and the covariates associated with the observed spatial clustering. All things being equal, if Ebola deaths during the outbreak were not clustered, then we will expect a random spatial distribution.

Previous outbreaks of Ebola were limited to remote, rural settings allowing containment efforts to be more effective. However, the widespread nature of the 2014–2016 outbreak was attributed to highly mobile populations and densely populated regions being affected at the early stages of the epidemic.[12] Liberia was one of the most affected countries. About 10 680 confirmed cases and 4810 deaths of EVD were reported in the country.[13 14] Liberia being one of the least developed countries in the world with a weak and underfunded health system, inadequate access to healthcare, further compromised by the outbreak,[13] under-reporting of cases and fatalities could be substantial.[8]

Before the outbreak of the Ebola epidemic, Liberia struggled with a weak health system that was distressed and weakened by a protracted civil war.[15] With just 50 doctors for a population of 4.3 million, there was limited capacity to respond to the epidemic given its magnitude.[13] The epidemic had severe impacts on the health system, including the loss of health workers. It was estimated that Ebola killed 0.11% of the general population, compared with 8.07% of healthcare workers.[16] In addition, more than 1.5 million of the population live in communities with extreme poverty, widespread job losses and food insecurity, weak surveillance systems, poor public health infrastructure and an adult literacy rate below 43%.[17] Over 70% of the urban population live in slums characterised by poor sanitation and overcrowding.[17] These potential super-spreaders typify potential spatial concentrations of cases and fatalities. Thus, to effectively and efficiently respond to epidemics, there is the need to identify hotspots of reported deaths during the outbreak. It is also essential that the associated covariates are examined to inform epidemic response.

METHODS

Sample

The data are derived from the 2019–2020 Liberia Demographic and Health Survey (LDHS). The LDHS adopted a two-stage stratified cluster (census enumeration areas) sampling design. A total of 325 clusters were selected, constituting 9745 households. Within the selected households, 8065 women aged 15–49 years and 4249 men aged 15–59 years were interviewed. The LDHS collected detailed information on respondents' experiences and practices during the Ebola outbreak, along with demographic, socioeconomic and other health data. The analysis covered 11 928 (women=7854 and men=4074) survey participants. A total of 2951 respondents representing 24.8% (weighted estimate) reported the death (95% CI 24.0 to 25.5) of a household member or relative.

Outcome variable

The outcome variable, respondents who reported that a household member or relative died during the Ebola outbreak (30 March 2014 and 1 June 2016) in Liberia, was binary coded 1 if the respondent reported a death and 0 otherwise.

Explanatory variables

The explanatory variables were derived from the LDHS and the Liberia Geospatial Covariate Datasets (LGCD).[18] The LGCD data were extracted for 2014–2015, coinciding with the Ebola outbreak in Liberia. It provides geospatial community-level development and socioecological data for buffers surrounding 2 km of urban and 10 km of rural DHS survey clusters.[18] The LDHS and LGCD data sets were merged at the individual level. The LDHS provides an individual-level case identification code (CASEID) and a cluster-level identification code (V001). These two codes reflect an individual and the cluster in which he/she lives. The LGCD provides a cluster-level identification code (DHSCLUST), which matches V001 in the LDHS. Using V001 and DHSCLUST, we merged the LDHS and LGCD at the individual level. Therefore, all individuals within a cluster have the same geospatial covariate information. The geospatial covariates were contextual variables that describe the characteristics of the group in each cluster, rather than characteristics of an individual.

We grouped the explanatory variables into two categories: community development factors derived from the LGCD and socioeconomic characteristics of the respondents derived from the LDHS. The classification of the community development factors was based on the theory that geography plays a vital role in the development of communities,[19] and infectious disease outbreaks threaten poor and marginalised communities more than any others.[20 21] This has been attributed to the lack of access to healthcare services, resources and public health infrastructure to prevent, diagnose and treat infections.[22–26] For example, in Liberia, it is reported that the outbreak mainly affected the impoverished and remote communities with poor physical infrastructure,

including roads, proper sanitation and health facilities.[20] The community-level development indicators selected for the analysis were: built population, global human footprint, nightlights composite, proximity to national borders, travel time to the nearest settlement with 50 000 people or more and population density. The selected indicators reflect three primary influences of the development of an area: density (agglomeration, scale economies), distance (spatial mobility and access) and division (spatial integration of economies).[19] See table 1 for the definition of the selected community development indicators.

Table 1 Selected community development and socioeconomic factors

Community development factors (community level)		
Variable	**Definition**	**Type of variable**
Built population	An index ranging from 0 to 1, where 0 represents extremely rural and 1 extremely urban. The index reflects built-up presence, remoteness and access to resources. Growth in built-up presence is usually due to population and economic growth, urbanisation, growth of smaller settlements into larger ones and expansive land development, with accompanying challenges such as air pollution and uncontrolled and unplanned urban growth.[39]	Continuous
Global human footprint	An index ranging from 0 (low) to 100 (high) covering human population pressure (population density), human land use and infrastructure (built-up areas, night-time lights, land use/land cover) and human access (coastlines, roads, railroads, navigable rivers). The index measures human pressure on the environment and reflects the use of land resources and the growth of infrastructure and amenities.[40]	Continuous
Gross cell production	The average purchasing power parity (PPP) in 2005 US dollars. It considers economic, demographic and geophysical characteristics of an area, including climate (precipitation and temperature), terrain (elevation, roughness), location indicators, population and luminosity. As a result, the indicator reflects the spatial distribution of incomes and favourable economic environments within countries.[41]	Continuous
Nightlights composite	The average night-time luminosity of the area shows the differentiation of regions based on the density of population and the degree of electrification of dwellings, commercial and industrial premises and infrastructure. The higher the index, the higher the level of socioeconomic development of an area.[42]	Continuous
Proximity to national borders	The geodesic distance (in metres) to the nearest international borders.	Continuous
Travel times	The average time (minutes) required to get to a settlement of 50 000 or more people. An indicator of access to people, resources and markets.[19]	Continuous
Population density	The number of persons per square kilometre.	Continuous
Socioeconomic factor (individual level)		
Variable	**Coding**	**Type of variable**
Age of respondent		Continuous
Sex of respondent	0=male, 1=female	Categorical
Educational attainment	0=no education, 1=primary, 2=secondary, 4=higher	Categorical
Frequency of reading newspaper or magazine	0=not at all, 1=less than once a week, 2=at least once a week	Categorical
Frequency of listening to the radio	0=not at all, 1=less than once a week, 2=at least once a week	Categorical
Community development factors (community level)		
Variable	**Definition**	**Type of variable**
Frequency of watching television	0=not at all, 1=less than once a week, 2=at least once a week	Categorical
Frequency of using internet last month	0=not at all, 1=less than once a week, 2=at least once a week, 3=almost every day	Categorical
Household wealth status	0=poorest, 1=poor, 2=middle, 3=rich, 4=richest	Categorical
Place of residence	0=urban, 1=rural	Categorical
Region	0=North Western, 1=South Central, 2=South Eastern A, 3=South Eastern B, 4=North Central	Categorical

The socioeconomic factors we considered for the analysis were: the age and sex of the respondent, educational attainment, frequency of reading newspapers or magazines, listening to the radio and watching television, household wealth status, urban-rural and regional place of residence (table 1). We considered the socioeconomic indicators to examine if reported deaths were skewed to poor and marginalised households and investigate if respondents' background characteristics influenced their reporting of a death. Correlations among the selected predictors were assessed using the interval-by-interval Pearson's R and the ordinal-by-ordinal Spearman correlation for the categorical-by-categorical covariates, nominal-by-interval eta for the continuous-by-categorical covariates and the Pearson's R for the continuous-by-continuous variables (online supplemental file S2). The results showed very low correlations among the variables, except for built population and population density (R^2=0.883). The results, therefore, show low potential for multicollinearity.

Statistical analysis

We used one-way analysis of variance (ANOVA) technique to examine the mean distributions of the continuous covariates aggregated by respondents who reported and those who did not report a death. To check for the normality assumption associated with the one-way ANOVA test, we computed and plotted the standardised residuals of the outcome measure and the continuous covariates and assessed if they follow a normal distribution. Further, we used the Levene test to examine the assumption of homogeneity of variances. Examination of the data revealed platykurtosis distribution of the residuals and in some cases non-normal distribution. The Levene test revealed a violation of the homogeneity of variance assumption (p<0.05) for the three variables—gross cell production, nightlight composite and proximity to national borders. The results are presented in online supplemental file S1. In this regard, the Welch test was used to examine statistically significant differences in the means of the continuous covariates for respondents who reported and those who did not report a death. The Welch test was preferred over the Kruskal-Wallis H non-parametric test because studies have shown that it is unstable when data are non-normal.[27]

We also examined the percentage distribution of respondents who reported a death by the categorical covariates of the background characteristics of the respondents, using χ^2 test to assess statistically significant differences. To examine the extent of geospatial clustering of reported deaths at the district level during the Ebola outbreak and the covariates associated with the observed spatial patterns, we used the Bayesian geoadditive semiparametric (BGS) regression technique.[28] The BGS approach allows for simultaneous estimation of non-linear effects of the continuous covariates and the fixed effects of the categorical and continuous covariates in addition to the unobserved spatial effects, both spatially structured and unstructured.[28] The advantage of this technique is that it allows for the true underlining relationship between the outcome variable and continuous covariate variables to be explored. The analysis was conducted at the district level, where health programmes are implemented and monitored. The LDHS provides the geographic coordinates for the centroid of the geographic clusters (enumeration areas) which can be linked to the districts in which the respondents reside. Thus, area-level analysis was conducted with the district as the geographic unit of focus. It is worthwhile to note that, although deaths during the Ebola outbreak were reported by household members, the LDHS did not provide information on households' geographic location, for confidentiality and anonymity purposes. Thus, we are unable to provide information on households' location and their weights, as well as each case and its distance from other cases.

The outcome variable of interest y_{ij} was coded 1 if a respondent i in district j reported that a household member or relative died during the Ebola outbreak and 0 otherwise. In this regard, the outcome variable y_{ij} follows a binomial distribution with the expected probability π_{ij} of reporting a death. Thus, the logistic model linking the probability π_{ij} of reporting a death is of the form:

$$y_{ij}|\eta_{ij} \sim B(\pi_{ij}) \quad (1)$$

$$\pi_{ij} = \mathrm{P}\,(y_{ij} = 1|\eta_{ij}) = \frac{\exp\,(\eta_{ij})}{1+\exp\,(\eta_{ij})} \quad (2)$$

where η_{ij} is the covariate of interest. If we have a vector $x'_{ij} = (x_{ij1}, \ldots, x_{ijk})'$ of k continuous covariates and $\lambda'_{ij} = (\lambda_{ij1}, \ldots \lambda_{ijd})'$ a vector of d categorical covariates, then the predictor η_{ij} can be specified as:

$$\eta_{ij} = \alpha\lambda'_{ij} + \beta x'_{ij} \quad (3)$$

where α is a vector of unknown regression coefficients for the categorical covariates, λ'_{ij}, β is a vector of unknown regression coefficients for the continuous covariates x'_{ij}.

To account for non-linear effects of the continuous covariate and the spatial correlation of the proportion of respondents who reported death, the BGS framework, which replaces the strictly linear predictors with flexible semiparametric predictors, was adopted. The model is thus specified as:

$$\eta_{ij} = \alpha\lambda'_{ij} + f_k x'_{ijk} + f^{spat\,(S_i)} \quad (4)$$

where $f_k\,(x)$ are the non-linear smoothing function of the continuous variables x_{ijk}, and $f^{spat(S_i)}$ accounts for unobserved spatial heterogeneity at district j (j=1, …, S), some of which may be spatially structured (correlated) and others unstructured (uncorrelated). The spatially structured effects show the effect of location by assuming that geographically close areas are more similar than distant areas, while the unstructured spatial effect accounts for spatial randomness in the model. Equation 5 is thus specified as

$$\eta_{ij} = \alpha\lambda'_{ij} + f_k x'_{ijk} + f^{str(S_i)} + f^{unstr(S_i)} \quad 5$$

where f^{str} is the structured spatial effects, and f^{unstr} is the unstructured spatial effects and $f^{spat(S_i)} = f^{str} + f^{unstr}$. The spatially structured effects depict the extent of death clustering and the associative effects of unaccounted predictor covariates, which may be spatially clustered or random. The smooth effects of continuous factors are modelled with P-spline priors, while the spatial effects are modelled using Markov random field priors.

We used the posterior modes of the structured spatial effects and their corresponding probabilities at 95% nominal level to examine the spatial correlates of the outcome variable at the district level. The posterior probabilities at the 95% nominal level show districts where reported deaths were statistically significantly high (high positive estimates of the posterior mode), significantly low (high negative estimates of the posterior mode) and where the effects were not significant (estimated posterior mode not significantly different from zero). The estimated posterior mode of the spatial effects characterises unexplained spatially correlated covariate information. Thus, using a sequential modelling approach, we were able to identify districts where the community development and socioeconomic covariates were spatially correlated with the observed clustering of deaths during the Ebola outbreak.

To examine if the reported deaths were geospatially clustered at the district level and if the community-level development indicators and the respondents' socioeconomic characteristics were associated with the observed clustering, we fitted a series of models. Model 0 was a null (constant) model. Model 1 accounted for only the spatial effects. Model 2 included the community development indicators, and model 3 added the socioeconomic factors. Only covariates significant at p<0.05 were retained in the model. The statistical software R was used for the analysis.[29]

It is important to note that the outcome measure for this study is not the prevalence of death (death rate) but respondents who reported that a household member or relative died during the Ebola outbreak in Liberia. The LDHS did not collect information on the number of people who died during the outbreak and the number of household members or relatives who died during the outbreak was also not reported. In this regard, the total number of people exposed was not available, and thus relative risk could not be calculated and the OR was used as a measure of the strength of association.[30] To ensure that the ORs from the logistic regression are not overestimated we conducted a sensitivity analysis by fitting a Poisson model to the data, computed the prevalence ratios and compared them to the ORs.

To avoid model overfit, we added the variables in a progressive manner (sequentially) and check if there is any significant improvement to the model using the Akaike information criterion (AIC) and Bayesian information criterion as measures of improvement to the model fit. To assess the extent of spatial autocorrelation in the residuals, we computed the Moran's I statistics and their corresponding p values. The assumption of statistical independence and identical distribution of the residuals was deemed violated where p<0.05.[31]

RESULTS

Descriptive analysis

Table 2 shows the weighted percentage of respondents who reported the death of a household member or a relative during the Ebola outbreak by the socioeconomic background characteristics of the respondents. Nearly one-quarter (24.8%) of the respondents reported the death of a household member or relative. The rate was higher for urban residents (25.4%) when compared with rural residents (23.8%); however, the difference was not large enough to be statistically significant. Statistically, a significantly higher percentage of deaths were reported in the North Western region (32.0%) when compared with the South Central (26.4%), South Eastern A (21.2%), South Eastern B (21.9%) and North Central (21.5%) regions. Male (26.3%) respondents were significantly more likely to report a death (26.3%) compared with females (23.9%). Regarding educational attainment, respondents with secondary and higher education were more likely to report a death. Table 2 also shows that respondents with access to information (newspaper, radio and television) were statistically significantly more likely to report a death. A lower percentage of respondents from the poorest households (19.3%) reported the death of a household member or a relative during the Ebola outbreak when compared with those from poor (26.8%), middle (24.8%), rich (25.3%) and the richest (27.0%) households.

The mean of the continuous socioeconomic and community-level development indicators aggregated by respondents who reported and those who did not report a death of a household member or other relatives during the Ebola outbreak in Liberia is shown in table 3. The mean age of respondents who reported a death (31.8 years) was significantly (p<0.05) higher compared with those who did not (29.3 years) report a death. The community development factors statistically significantly (p<0.05) associated with reporting a death were the gross cell production, nightlight composite and travel time to the nearest main settlement with a population of 50 000 or higher. The mean gross cell production for respondents who reported death was significantly lower (US$428.7) when compared with those who did not report a death (US$468.4). The nightlight composite shows that respondents in more developed communities (mean=0.82) were more likely to report a death when compared with those in less developed communities (mean=0.73). Also, the travel time showed that respondents with shorter travel time to main settlements were more likely to report a death. The built population, global human footprint,

Table 2 Weighted percentage distribution of respondents who reported death of a household member or other relatives during the Ebola outbreak in Liberia by socioeconomic factors

Socioeconomic factors	Percentage	95% CI	P value	Sample size
Overall	24.8	24.0 to 25.5		11 928
Type of place of residence			0.059	
Urban	25.4	24.4 to 26.4		4699
Rural	23.8	22.6 to 25.0		7229
Region			<0.001	
North Western	32.0	29.0 to 35.0		1669
South Central	26.4	25.2 to 27.5		3335
South Eastern A	21.2	18.2 to 24.2		1893
South Eastern B	21.9	18.7 to 25.0		2230
North Central	21.5	20.1 to 22.8		2801
Sex			0.004	
Female	23.9	23.0 to 24.9		7854
Male	26.3	25.0 to 27.7		4074
Highest educational level			<0.001	
No education	23.5	22.0 to 25.0		3586
Primary	22.9	21.3 to 24.5		3489
Secondary	25.7	24.6 to 26.9		4326
Higher	28.8	25.9 to 31.8		527
Frequency of reading newspaper or magazine			<0.001	
Not at all	23.6	22.8 to 24.5		10 258
Less than once a week	29.7	27.6 to 31.9		1263
At least once a week	28.6	25.0 to 32.1		407
Frequency of listening to the radio			<0.001	
Not at all	20.3	19.1 to 21.5		4770
Less than once a week	27.2	25.8 to 28.7		3834
At least once a week	27.6	26.1 to 29.0		3324
Frequency of watching television			0.003	
Not at all	23.7	22.7 to 24.7		8071
Less than once a week	25.8	24.2 to 27.5		2547
At least once a week	27.1	25.1 to 29.0		1310
Wealth index for urban/rural			<0.001	
Poorest	19.3	17.6 to 21.0		2696
Poorer	26.8	24.9 to 28.6		2496
Middle	24.6	22.9 to 26.3		2375
Richer	25.3	23.6 to 27.0		2209
Richest	27.0	25.3 to 28.7		2152

proximity to national borders and population density were not significantly associated with reporting a death.

BGS regression

The estimated posterior variance of the continuous covariates and the posterior ORs of the categorical covariates of reporting a death during the Ebola outbreak in Liberia and their corresponding 95% credible intervals are shown in table 4, along with their model summary statistics. Model 3A shows the posterior ORs from the final model fitted with the logit link function, while model 3B shows the posterior prevalence ratios from the model fitted with Poisson link function (table 4). The results show that the posterior ORs from the logistic model and posterior prevalence ratios from the Poisson model were in the same direction and there was no evidence that logistic regression markedly overestimated the ORs. We

Table 3 Mean of the continuous covariates by respondents who reported and those who did not report the death of a household member or other relatives during the Ebola outbreak in Liberia

Indicators	Mean	95% CI of mean	P value
Age of respondent (years)			<0.001
Reported a death	31.8	31.4 to 32.1	
Did not report a death	29.3	29.0 to 29.5	
Built population			0.226
Reported a death	0.33	0.32 to 0.35	
Did not report a death	0.32	0.31 to 0.33	
Global human footprint			0.103
Reported a death	41.1	40.9 to 42.0	
Did not report a death	41.0	40.7 to 41.3	
Gross cell production			<0.001
Reported a death	428.6	418.4 to 439.0	
Did not report a death	468.4	462.0 to 474.8	
Nightlight composite			0.003
Reported a death	0.82	0.77 to 0.878	
Did not report a death	0.73	0.70 to 0.76	
Proximity to national borders (m)			0.240
Reported a death	1717.5	1642.4 to 1792.5	
Did not report a death	1665.8	1623.1 to 1708.5	
Travel time			0.017
Reported a death	79.3	75.3 to 83.3	
Did not report a death	85.2	82.8 to 87.6	
Population density			0.596
Reported a death	2144.2	2042.4 to 2245.9	
Did not report a death	2112.6	2054.7 to 2170.4	

used a sequential model-building approach to analyse the associations of the community development and socioeconomic covariates on reporting a death. Interpretation of the model coefficients is based on the final model (model 3A).

The Moran's I statistics for model 1 was −0.142 with a corresponding p value of 0.0318, indicating spatial autocorrelation of the residuals and therefore violation of the assumption of statistical independence and identical distribution of the residuals, when only the spatial effects are accounted for in the model. When the community development indicators and the socioeconomic factors were added in models 2 and 3, the Moran's I statistics reduced to −0.0782 and −0.0721, with corresponding p values of 0.1768 and 0.2167, respectively. This indicates that the assumption of statistical independence and identical distribution of the residuals was not violated for the final model.

Geospatial clustering of reported deaths

The estimated deviance and AIC for model 0 (null model) were 13 020.0 and 13 022.0, respectively (table 4). When the spatial effects were included in the model (model 1), the deviance and AIC reduced by 456.8 and 326.1, respectively. The high reduction in the deviance and AIC after the spatial effects were included in the model indicates that deaths were not spatially randomly distributed but clustered. Figure 1A shows the districts where the posterior mode of the structured spatial effects was positive and statistically significantly high (clusters of high reported deaths) at the 95% nominal level. The figure shows clustering of high reported deaths in 21 districts in six of the 15 counties in Liberia: Grade Cape Mount, Bomi, Gbarpolu, Montserrado, Margibi and Lofa.

Community development factors associated with the observed spatial clustering

The community development indicators were included in model 2 (table 4). This led to a reduction of 88.7 and 53.0 in the deviance and AIC, respectively. The considerable decline in deviance and AIC suggests that the community development factors significantly correlate with the odds of reporting a death. The community development factors, global human footprint, gross cell production and population density were identified to have nonlinear associations with the observed spatial clustering of reported deaths (figure 2A–C).

Table 4 Posterior variance estimates of the spatial effects at the district level and the non-linear covariates, along with the posterior ORs and posterior prevalence ratios of the categorical covariates for reporting a death of a household member or relative during the Ebola outbreak in Liberia, their corresponding 95% credible intervals and model summary statistics

	Model 0	Model 1	Model 2	Model 3A POR (95% CI)	Model 3B PPR (95% CI)
District-level variance					
Structured spatial effects (SSE)		0.141	0.137	0.156	0.0417
Unstructured spatial effects (USE)		0.096	0.096	0.101	0.0875
% change in SSE					
Variance of the non-linear effects					
Community development factors					
Global human footprint			0.055	0.058	0.0218
Gross cell production			0.230	0.211	0.0494
Population density			0.020	0.020	0.0052
Socioeconomic factors					
Age of the respondent				0.014	0.0078
Covariates					
Socioeconomic factors					
Educational attainment					
No formal education				1.00	1.00
Primary				1.27 (1.13 to 1.44)**	1.19 (1.07 to 1.32)**
Secondary				1.39 (1.23 to 1.56)**	1.27 (1.15 to 1.40)
Higher				1.27 (1.02 to 1.59)*	1.19 (0.77 to 0.96)
Place of residence					
Urban				1.00	1.00
Rural				0.80 (0.70 to 0.93)**	0.86 (0.77 to 0.96)**
Model summary statistics					
Deviance	13 020.0	12 563.2	12 474.5	12 287.7	13 283.2
AIC	13 022.0	12 695.9	12 642.9	12 476.6	13 435.7
BIC	–	13 184.9	13 262.9	13 172.5	13 997.2
GVC	–	1.087	1.083	1.069	0.6547
Change in deviance	–	456.8	88.7	186.8	–
Change in AIC	–	326.1	53.0	166.3	–

**P<0.01; *p<0.05.
AIC, Akaike information criterion; BIC, Bayesian information criterion; CI, credible intervals; GVC, Generalized Cross Validation; POR, posterior OR; PPR, posterior prevalence ratio.

Figure 2A shows that for a global human footprint of between 1 and 39, there is a trivial variation (OR between 0.90 and 1.11) in the associations with the observed spatial clustering of reported deaths. However, from a global human footprint of 40–55, the odds of reporting a death decreases from about 1.1 to 0.67 and then begins to increase to an OR of 1.11. The figure shows that the odds of reporting a death were lowest for those with a moderate (around 55) global human footprint. Regarding gross cell production, figure 2B shows higher odds of reported deaths for areas with average purchasing power parity (PPP) around US$1000.00, when compared with areas with lower PPP and those with higher PPP. The results also show a decline in the odds of reported deaths with increasing population density (figure 2C). Implying, the odds of reporting a death were lower in highly populated areas, indicating that reported deaths decreased with increasing urbanisation.

When community development indicators were included in the model, the posterior mode of the structured spatial effects became statistically insignificant (p>0.05) in 13 of the 21 districts with clustering of high reported deaths. These are the Gbarma district in the Gbarpolu county, the Commonwealth, Garwula, Golakonneh, Porkpa and Tewor districts in the Grand Cape Mount county, Quardu Boundi and Voinjama districts in the Lofa county, Firestone, Kakata and Mambah-Kaba districts in the Margibi county, and the Careysburg,

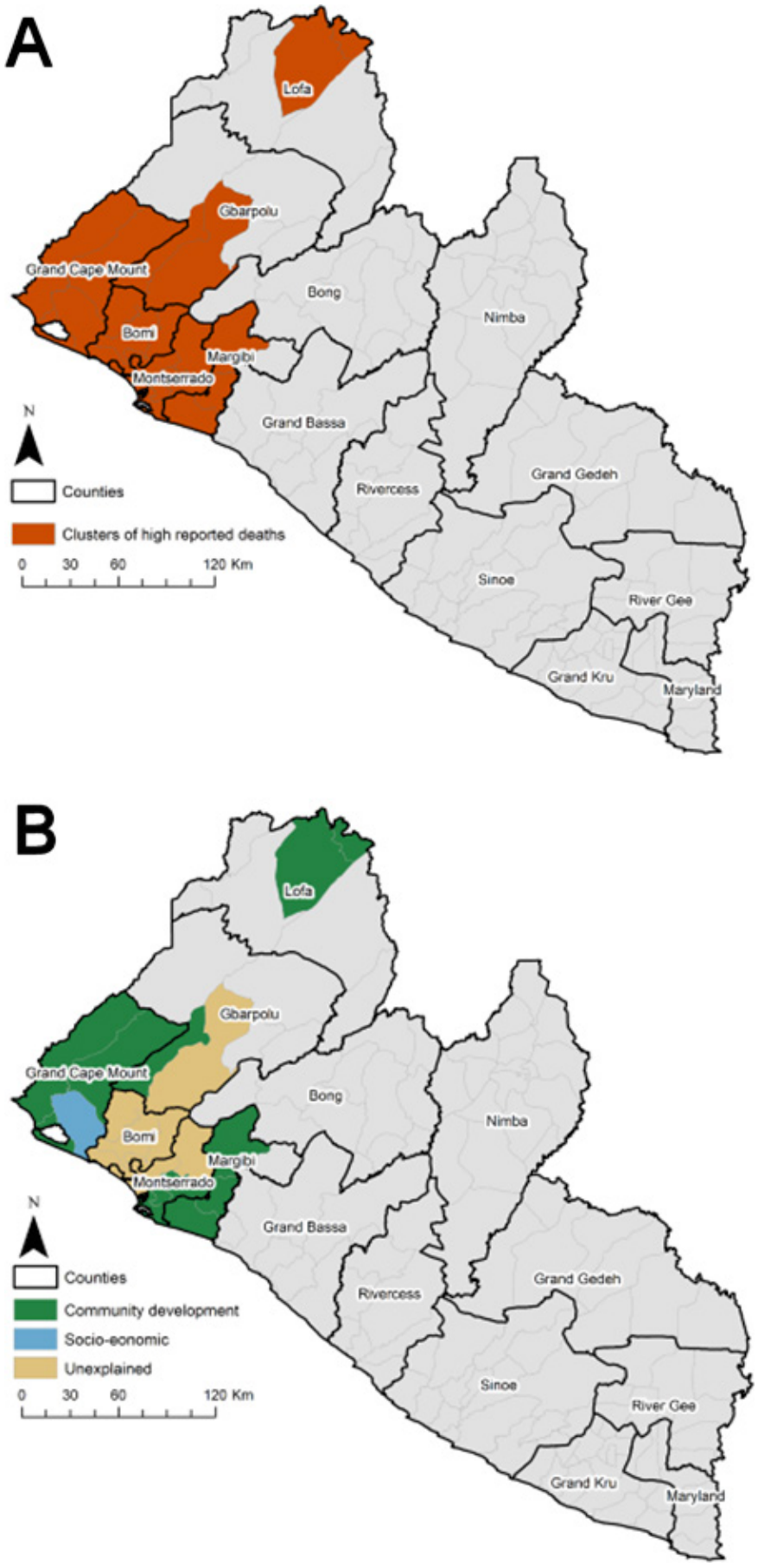

Figure 1 Geospatial (A) clustering of deaths during the Ebola outbreak in Liberia and (B) their geospatial correlates.

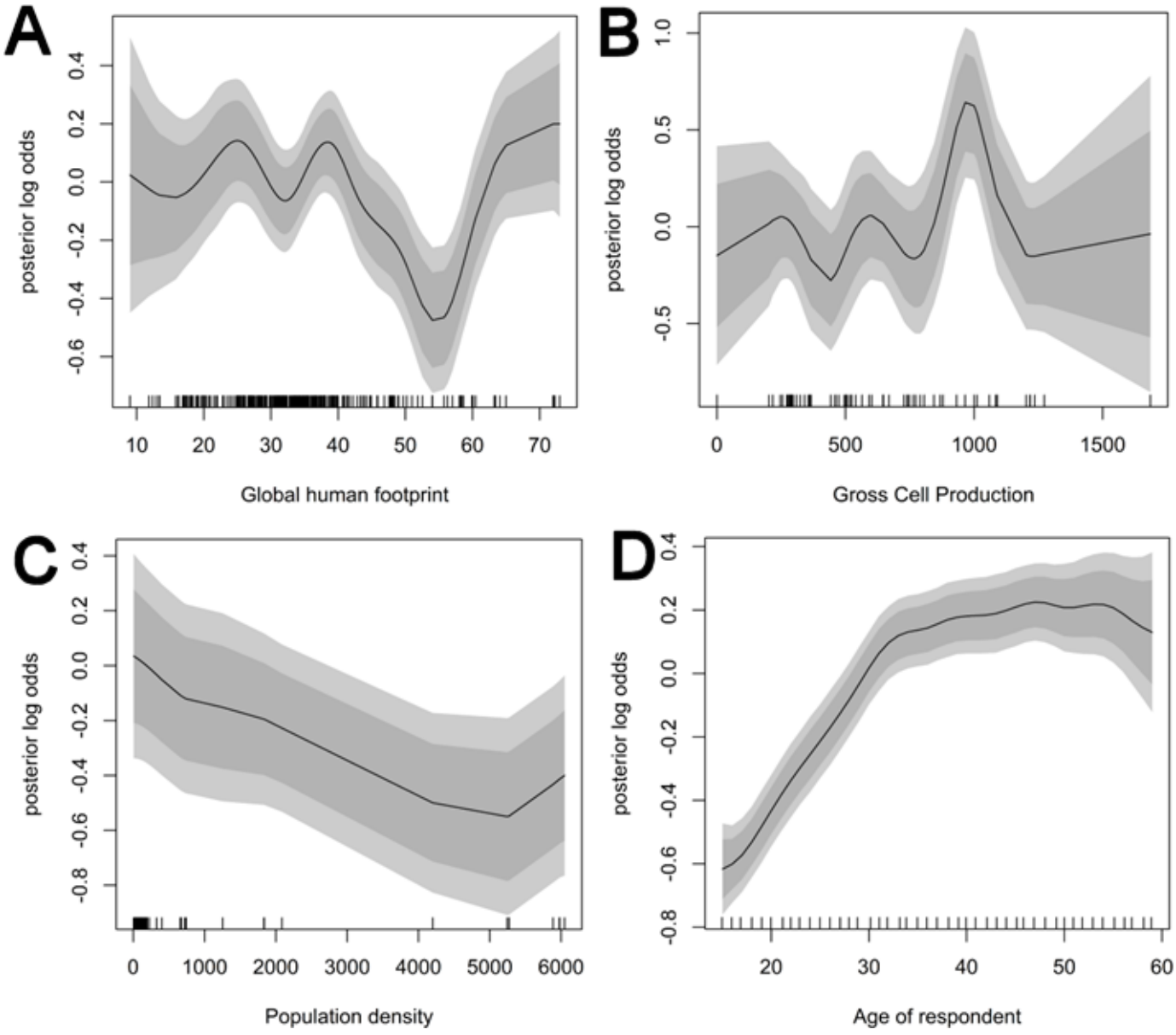

Figure 2 Non-linear associative effects of the continuous community development ((A) global human footprint, (B) gross cell production, (C) population density) and socioeconomic factors ((D) age of respondent) and the posterior log odds of reporting a death during the Ebola outbreak.

Commonwealth 1 and Greater Monrovia in the Montserrado county (figure 2B). This indicates that the community development indicators are associated with high reported deaths in these districts.

Socioeconomic factors associated with the observed spatial clustering

The respondents' socioeconomic factors were included in model 4 (table 3), leading to a decline of 186.8 and 166.3 in the deviance and AIC, respectively. The respondent's age showed a strong non-linear association with the odds of reporting a death (figure 2D). The odds of reporting a death is low from age 15 years (OR=0.61) and then increases linearly until about 30 years (OR=1.11) and remains constantly high between ages 30 and 55 years.

The categorical socioeconomic covariates identified to be statistically significantly ($p<0.05$) associated with the observed spatial clustering were educational attainment and urban-rural place of residence (table 4, model 3A). The estimated posterior ORs show that educated respondents had higher odds of reporting a death when compared with those with no formal education. Those with primary, secondary and higher level education were 27.0%, 39.0% and 27.0%, respectively, more likely to report a death when compared with those with no formal education. Further, respondents in rural areas were 20.0% less likely to report a death when compared with those residents in urban areas. The socioeconomic covariates, sex of respondent, frequency of reading newspaper or magazine, listening to the radio and watching television,

household wealth status and region of residence were not statistically significantly associated with the observed spatial pattern of reported deaths. The posterior mode of the structural spatial effects became insignificant in only one district (Garwula district in the Grand Cape Mount county) when the socioeconomic covariates were included in the model.

The observed spatial clustering of reported deaths in Dowein, Klay, Senjeh and Suehn Mecca districts in the Bomi county, Bopolu district in the Gbarpolu county and the St Paul River and Todee districts in the Montserrado county was neither associated with the community development factors nor the socioeconomic covariates. Thus, the observed spatial clustering of reported deaths in these districts was unexplained by the selected covariates. For policy relevance, we test for plausible interactions but none was statistically significant at $p<0.05$. The posterior mean of the unstructured spatial (random) effects is shown in figure 3. The figures show a random scatter of the posterior mean of the unstructured spatial effects for all the fitted models, further confirming low spatial autocorrelation of the residuals.

DISCUSSION

Our findings show that deaths during the Ebola outbreak in Liberia were not spatially randomly distributed at the district level but clustered. Deaths were significantly clustered in districts in the North Western, South Central and parts of the North Central regions. In the North Western region, we found clustering of high reported deaths in districts in the Grade Cape Mount, Bomi and parts of Gbarpolu counties, while in the South Central region, we found clustering in districts in Monrovia and its surrounding counties of Montserrado and Margibi. In the North Central region, clustering of high deaths was identified in districts in the Lofa county, one of the counties that reported the first case of Ebola in Liberia, and it shares borders with Guinea and Sierra Leone.[12 32]

Within Liberia, the Lofa county was reported as the third most important source of the Makona Ebola variant, the third worst affected county and a major epicentre early in the second wave of the outbreak.[33] However, Lofa's contribution to the spread of the Ebola virus to other counties was substantially minimal.[32 34] The minimal contribution of the Lofa county to the spread has been attributed to its remote location, isolation from populated settings of the country and poor connecting roads.[32] This narrative is consistent with the findings of our study. In the Lofa county, high death clustering was identified in only two adjoining districts: Quardu Boundi and Voinjama districts.

Documentary evidence shows that although the outbreak started in rural areas, it quickly spread to urban centres, including Montserrado and Margibi.[12] At the peak of the epidemic, the most affected counties included Montserrado and Margibi.[12 32] Contact tracing data also show that the neighbouring counties of Montserrado and Margibi were the primary sources for the spread of the Makona Ebola variant to other Liberian counties.[32 35] Our study also found significant clustering of deaths in these counties.

Regarding the country's North Western region, all districts in the Grand Cape Mount and Bomi counties and three districts in the Gbarpolu county were identified as clusters of significantly high reported deaths. Thus, indicatively, the Grand Cape Mount and the Bomi counties, in addition to the Montserrado and Margibi counties and the capital Monrovia experienced the highest levels of EVD infections in Liberia,[36] as reflected in our findings with regard to reported deaths during the Ebola outbreak in Liberia.

We also found that the observed geospatial clustering of deaths during the Ebola outbreak in Liberia was non-linearly associated with the community-level development indicators: the global human footprint (use of land resources and infrastructure development), gross cell production (estimate average PPP in 2005 US dollars) and population density. The results revealed that the community-level development indicators were associated with the observed clustering of deaths predominantly in the Grand Cape Mount, Margibi, Lofa and in some districts in Monsterrado.

With regard to the global human footprint, we found that the odds of high death clustering were lowest for areas with moderate (around 55) global human footprint and comparatively higher for areas with low and high global human footprint. Considering gross cell production, our findings show that the odds of high death clustering were low for areas with average PPP below US$500.00 and those above US$1200.00, but higher for those with PPP around US$1000.00. This indicates that deaths were more likely in areas with moderate to high community development. We also found that the odds of reported death clustering declines with increasing population density, showing that highly populated areas had lower odds of death clustering compared with less populated areas.

The 2016 Liberia Household Income and Expenditure Survey reported a substantially high absolute poverty rate (greater than 50.0%) across all counties, except the Montserrado county (20.3%), including Monrovia.[37] Although absolute poverty is high across most counties, the rates were comparatively modest for the Grand Cape Mount (53.7%), Margibi (52.2%) and the Gbarpolu (60.5%) counties when compared with counties such as River Gee (81.9%) and Maryland (84.0%). Similarly, the LDHS shows lower poverty levels and better access to amenities in the South Central region, particularly the Montserrado and Margibi counties, when compared with the Grand Cape Mount, Bomi and Gbarpolu counties in the North Western region.[38]

These developmental characteristics of the communities are reflective of the findings from our study. Death clustering was observed to be significantly high in the more developed counties of Montserrado (including Monrovia) and Margibi, and also the poor to moderate

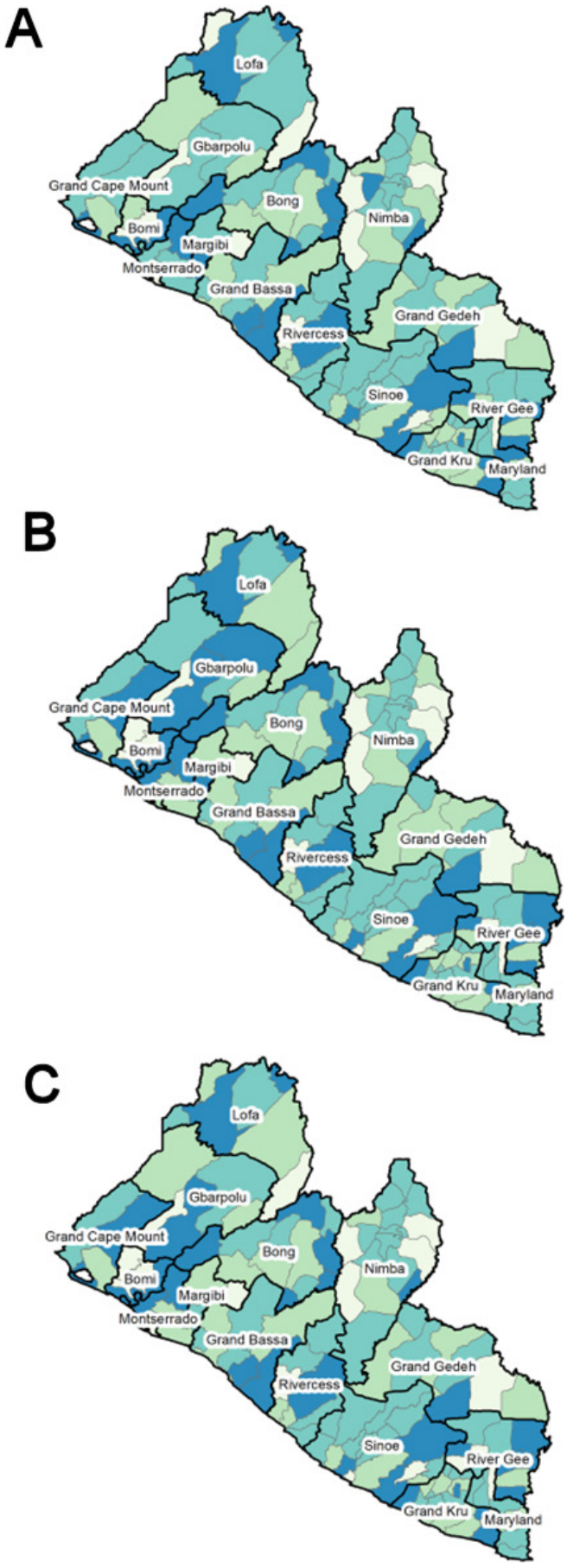

Figure 3 Posterior mean of the unstructured spatial effects for (A) model 1, (B) model 2, and (C) model 3A.

counties of Grand Cape Mount and Gbarpolu. The higher odds of reported death clustering in less populated areas reflect the remoteness and isolation of Lofa county.

The findings of our study show that during epidemics, deaths are likely to be clustered, and these may not always be in poor and marginalised communities. Therefore, interventions and policy responses should identify at-risk populations and respond accordingly to minimise fatalities.

Strengths and limitations of this study

Evidence shows that deaths during epidemics are underreported, and local-level registers of fatalities, particularly in the low-income and middle-come countries, are often non-existent. The present study has used cross-sectional data of reported deaths of household members and relatives during the Ebola outbreak to identify geospatial clusters (districts) where reported deaths at the time were statistically significantly high. The identified clusters are in line with reported Ebola virus transmission rates at the district level and their characteristics. Nonetheless, some limitations are worth noting. First, these are self-reported deaths not directly linked to Ebola. Also, there are no records to verify the veracity of these deaths. However, all things being equal, if Ebola deaths during the outbreak were not concentrated within some districts, then we will expect a random spatial scatter of deaths.

Contributors This paper is a joint research contributed by FAJ (FAJ, lead author/corresponding author) and BS (BS, coauthor). FAJ and BS conceptualised and designed the study. BS developed the literature review. FAJ extracted the data sets, conducted the statistical analysis and wrote the first draft. FAJ and BS provided the interpretation of the context-specific results, final drafting of the article and the final revision of the article. Both authors consent to the submission of the final version. FAJ is the guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. The data for the analysis come from The DHS Program Demographic and Health Surveys (https://dhsprogram.com/) which is available upon request.

ORCID iD

Fiifi Amoako Johnson http://orcid.org/0000-0003-0896-937X

REFERENCES

1 Kourtis AP, Appelgren K, Chevalier MS, *et al*. Ebola virus disease: focus on children. *Pediatr Infect Dis J* 2015;34:893–7.
2 Malvy D, McElroy AK, de Clerck H, *et al*. Ebola virus disease. *Lancet* 2019;393:936–48.
3 Briand S, Bertherat E, Cox P, *et al*. The international Ebola emergency. *N Engl J Med* 2014;371:1180–3.
4 Lo TQ, Marston BJ, Dahl BA, *et al*. Ebola: anatomy of an epidemic. *Annu Rev Med* 2017;68:359–70.
5 Kaner J, Schaack S. Understanding Ebola: the 2014 epidemic. *Global Health* 2016;12:53.
6 Garske T, Cori A, Ariyarajah A, *et al*. Heterogeneities in the case fatality ratio in the West African Ebola outbreak 2013-2016. *Philos Trans R Soc Lond B Biol Sci* 2017;372:1721:20160308.
7 Camacho A, Kucharski A, Aki-Sawyerr Y, *et al*. Temporal changes in Ebola transmission in Sierra Leone and implications for control requirements: a real-time modelling study. *PLoS Curr* 2015;7 doi:10.1371/currents.outbreaks.406ae55e83ec0b5193e30856b9235ed2
8 Dalziel BD, Lau MSY, Tiffany A, *et al*. Unreported cases in the 2014-2016 Ebola epidemic: Spatiotemporal variation, and implications for estimating transmission. *PLoS Negl Trop Dis* 2018;12:e0006161.
9 Towers S, Patterson-Lomba O, Castillo-Chavez C. Temporal variations in the effective reproduction number of the 2014 West Africa Ebola outbreak. *PLoS Curr* 2014;6 doi:10.1371/currents.outbreaks.9e4c4294ec8ce1adad283172b16bc908
10 Nyenswah TG, Kateh F, Bawo L, *et al*. Ebola and its control in Liberia, 2014-2015. *Emerg Infect Dis* 2016;22:169–77.
11 WHO. *Health worker Ebola infections in guinea, Liberia and Sierra Leone: a preliminary report 21 May 2015*. Geneva: Switzerland, 2015.
12 Coltart CEM, Lindsey B, Ghinai I. The Ebola outbreak, 2013-2016: old lessons for new epidemics. *Philos Trans R Soc Lond B Biol Sci* 20171721;372:297. doi:10.1098/rstb.2016.0297
13 Fall IS. Ebola virus disease outbreak in Liberia: application of lessons learnt to disease surveillance and control. *Pan Afr Med J* 2019;33:19074.
14 Wagenaar BH, Augusto O, Beste J, *et al*. The 2014-2015 Ebola virus disease outbreak and primary healthcare delivery in Liberia: time-series analyses for 2010-2016. *PLoS Med* 2018;15:e1002508.
15 Das D, Guerin PJ, Leroy S, *et al*. The largest Ebola outbreak– what have we learned so far. *J Med* 2015;16:1–4.
16 Evans DK, Goldstein M, Popova A. Health-care worker mortality and the legacy of the Ebola epidemic. *Lancet Glob Health* 2015;3:e439–40.
17 Fallah MP, Skrip LA, Gertler S, *et al*. Quantifying poverty as a driver of Ebola transmission. *PLoS Negl Trop Dis* 2015;9:e0004260.
18 Mayala B FT, Eitelberg D, Dontamsetti T. *The DHS program geospatial covariate datasets manual*. 2nd edn, 2018.
19 Linard C, Gilbert M, Snow RW, *et al*. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS One* 2012;7:e31743.
20 Kapiriri L, Ross A. The politics of disease epidemics: a comparative analysis of the SARS, Zika, and Ebola outbreaks. *Glob Soc Welf* 2020;7:33–45.
21 Quinn SC, Kumar S. Health inequalities and infectious disease epidemics: a challenge for global health security. *Biosecur Bioterror* 2014;12:263–73.
22 Arima Y, Shimada T. Epidemiological situation of Ebola virus disease in West Africa. *Uirusu* 2015;65:47–54.
23 Burkle FM, Burkle CM. Triage management, survival, and the law in the age of Ebola. *Disaster Med Public Health Prep* 2015;9:38–43.
24 McNamara LA, Schafer IJ, Nolen LD, *et al*. Ebola surveillance — guinea, Liberia, and Sierra Leone. *MMWR Suppl* 2016;65:35–43.
25 Plourde AR, Bloch EM. A literature review of Zika virus. *Emerg Infect Dis* 2016;22:1185–92.
26 Slavov SN, Otaguiri KK, Kashima S, *et al*. Overview of Zika virus (ZIKV) infection in regards to the Brazilian epidemic. *Braz J Med Biol Res* 2016;49:e5420.

27 Liu H. Comparing W comparing Welch's ANOVA, a Kruskal-W A, a Kruskal-Wallis test and tr allis test and traditional aditional ANOVA in case of Heterogeneity of Variance. Virginia Commonwealth University, Graduate School at VCU Scholars Compass 2015.
28 Kneib T, Tutz G. *Statistical modelling and regression structures*. Springer, 2010.
29 Umlauf N, Adler D, Kneib T, *et al*. Structured additive regression models: An R interface to BayesX. *J Stat Softw* 2015;63:1–46.
30 Ranganathan P, Aggarwal R, Pramesh CS. Common pitfalls in statistical analysis: odds versus risk. *Perspect Clin Res* 2015;6:222–4.
31 F. Dormann C, M. McPherson J, B. Araújo M, *et al*. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 2007;30:609–28.
32 Ladner JT, Wiley MR, Mate S, *et al*. Evolution and spread of Ebola virus in Liberia, 2014-2015. *Cell Host Microbe* 2015;18:659–69.
33 UNICEF. *UNICEF-Liberia Ebola virus outbreak SitReps 2014*. Monrovia, 2014.
34 Arwady MA, Bawo L, Hunter JC, *et al*. Evolution of Ebola virus disease from exotic infection to global health priority, Liberia, mid-2014. *Emerg Infect Dis* 2015;21:578–84.
35 Kuhn JH, Andersen KG, Baize S, *et al*. Nomenclature- and database-compatible names for the two Ebola virus variants that emerged in guinea and the Democratic Republic of the Congo in 2014. *Viruses* 2014;6:4760–99.
36 Stanturf JA, Goodrick SL, Warren ML, *et al*. Social vulnerability and Ebola virus disease in rural Liberia. *PLoS One* 2015;10:e0137208.
37 LISGIS. *Household income and expenditure survey 2016*. Monrovia: Liberia Institute of Statistics & Geo-Information Services, 2016.
38 LISGIS, MoH, and ICF. *Liberia demographic and health survey 2019-20*. Monrovia, Liberia and Rockville, Maryland, USA: Liberia Institute of Statistics and Geo-Information Services (LISGIS), Ministry of Health (Liberia), and ICF, 2021.
39 Ehrlich D, Balk D, Richard S. Measuring and understanding global human settlements patterns and processes: innovation, progress and application. *Int J Digi Earth* 2020;13:1–8.
40 Venter O, Sanderson EW, Magrach A, *et al*. Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nat Commun* 2016;7:12558.
41 Lawal O, Kalu IE. Measuring geographic distribution of economic activity in Nigeria using gross domestic product. *Ghana Journal of Geography* 2018;10:22–41.
42 Proville J, Zavala-Araiza D, Wagner G. Night-time lights: a global, long term look at links to socio-economic trends. *PLoS One* 2017;12:e0174610.