Pure and Applied Mathematics Journal
2021; 10(6): 127-138
http://www.sciencepublishinggroup.com/j/pamj
doi: 10.11648/j.pamj.20211006.12
ISSN: 2326-9790 (Print); ISSN: 2326-9812 (Online)

Fitting Spatial Joint Model for U.S Regional Influenza-like Illness (ILINet) Data Set

Azizur Rahman, Arifa Tabassum, Mariam Akter

Department of Statistics, Jahangirnagar University, Savar, Dhaka, Bangladesh

Email address:
rahman.aziz83@gmail.com (A. Rahman)

To cite this article:
Azizur Rahman, Arifa Tabassum, Mariam Akter. Fitting Spatial Joint Model for U.S Regional Influenza-like Illness (ILINet) Data Set. *Pure and Applied Mathematics Journal*. Vol. 10, No. 6, 2021, pp. 127-138. doi: 10.11648/j.pamj.20211006.12

Received: October 25, 2021; **Accepted**: November 13, 2021; **Published**: December 29, 2021

Abstract: Background: Influenza is commonly known as the flu, which is a viral infectious disease that attacks our respiratory systems, such as the nose, throat, and lungs. Several studies have been performed on influenza determinants, concentrating on the role of biological and behavioral risk factors at the personal level to reduce the burden of the disease. However, few studies conducted to identify geographical patterns of infectious disease and its associated factors. Objective: This study aimed to provide a step-by-step process of finding the geographic patterns of influenza cases and the role that they can be determined by the racial factor. Method: In this study, first non-spatial and spatial models were estimated, and then a step-by-step procedure was used to fit a spatial joint model to the US Influenza Like Illness (ILINet) dataset using a single predictor: percentage of African American people in each state. Results: Findings revealed that for both non-spatial and spatial models, the racial variable was positively associated with standard morbidity ratio (SMR) and was highly statistically significant ($p<0.0001$). In addition, it showed that there was a large residual spatial dependency for the spatial joint model, which meant for our dataset, the spatial component explained much of the variability. Conclusion: Researchers that desire to create a joint special model from the ground up in the instance of infectious illness modelling can benefit from this research.

Keywords: Spatial Analysis, Influenza-Like Illness, Heterogeneity, Standardized Morbidity Rates (SMRs)

1. Introduction

Influenza is commonly known as the flu, which is a viral infectious disease that attacks our respiratory systems, such as the nose, throat, and lungs. For most individuals, it automatically resolves on its own but sometimes it shows deadly complications. Initially, flu shows symptoms like a common cold with a runny nose, sneezing, and sore throat. Common colds usually develop slowly, whereas flu tends to come suddenly. When someone with the infection coughs, sneezes or talks can spread flu viruses. Hence, it is a type of communicable disease. People at higher risk with this communicable disease include children with less than 5 years of age-especially under 12 months, older adults (age 65+), and pregnant women up to two weeks of postpartum, people with a weak immune system, and also people with a chronic disease like liver disease, heart disease, asthma and kidney disease (www.mayoclinic.org) [1].

Nearly 500,000 deaths occurred due to influenza disease around the world per year and about 5 to 10% of deaths occurred per year in the US [2]. Although the numbers vary widely from season to season, the estimated annual human cost of influenza is 610,660 life-years lost and 3.1 million hospitalization days in the US [3]. As an ongoing project to reduce mortality due to influenza, recently World Health Organization (WHO) has taken a global influenza reduction strategy for the period of 2019-2030 [4]. Yearly flu season in North America starts on the 40th week of each year (1st week of October). CDC started recording data from October, week 40, 1997. From then, CDC published routinely weekly unrevised unweighted Influenza-like-Illness (ILI) activity level data (gis.https://gis.cdc.gov/grasp/fluview/fluprotaldashboard.html) which measures the percentage of out-patients seeking medical attention with ILI symptoms. According to CDC a person having 100°F or greater temperature in his/her, body with a cough and/or a sore throat without a known cause other than influenza is consider symptoms for ILI [5]. Moreover, CDC reported that flu disparities occur among racial and ethnic minority groups (www. cdc.gov). Many studies have investigated the factors and spatial patterns of

influenza seasonality [6-11]. Some studies identified humidity and temperature were associated factors for seasonal flu onset [12, 13]. A demographic metric such as population density and school-aged children have identified two major influential sources of influenza [14, 15].

An analysis conducted in central Tennessee found that neighborhood socio-economic indicators such as race (*African American or not*), percent below poverty, and percent of female-headed household were also directly associated with influenza hospitalization rate [16]. Another study conducted in Ontario, Canada, revealed that the aboriginal population was positively associated with both non-spatial and spatial regression models of pneumonia and influenza hospitalization rate [17]. In this study, our aim was to develop a joint spatial model to examine the underlying spatial variation in reported influenza-like illness across 53 states in the United States and determined whether the race especially the percent of African American people (PAfA) in each state act as a driving factor in observed spatial heterogeneity. To do this, we set up the following questions. The associated Research Questions are:

1) Are there any differences between the raw and fitted Standard Morbidity Ratio's (SMR) of Influenza disease over U.S states?
2) Are there any counties with relatively high SMR's?
3) Does ethnic characteristics can influence the spatial variation of SMR's?

2. Dataset Description

The dataset of ILI visits is publicly available on the Center for Disease Control (CDC) website (https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html) and the covariate proportion of African American population size for each US states was obtained from Wikipedia website (https:worldpopulationreview.com). Our dataset consists of aggregated ILI cases (Y_i) for the 2018/2019 year reported by the providers on weekly basis and a covariate represents the percent of African American population size (x_i) for each state. We calculated corresponding expected ILI cases (E_i) standardizing with respect to area specific population size. Finally, we obtained a shapefile from the "*cdcfluview*" package in R, which consists of the longitude and latitude of each area. Our unit of spatial analysis was state-level, not county-level.

3. Methodology

The best approach for modeling health events may not always be a linear model, particularly when the interested outcome variables are counts or rates, or when we are interested in estimating health risk from binary data. In such cases, it is customary to use generalized linear models (GLMs), with Poisson and logistic regression models in special cases. In our dataset, the outcome variable is the yearly total observed ILI cases (Y_i) reported by public health service providers for each US state in 2018/2019 year. As an expected relevant predictor of influenza disease burden, we considered ethnic metric: the proportion of African American population size based on literature findings. We performed the following data processing steps for states level ILI cases: i) we assumed that our data follows Poisson distribution as $Y_i \sim Pois(E_i\lambda_i)\ i = 1, \cdots, m$, m represents the number of states or areas under consideration; ii) we assumed that the expected ILI cases (E_i) was calculated through direct standardization process as $E_i = N_i \left(\frac{\sum_{i=1}^{m} Y_i}{\sum_{i=1}^{m} N_i}\right)$, where, N_i is population at risk for *i* th area; iii) we calculated standardized morbidity ratio (SMR's) which is the ratio of observed to expected cases; iv) we plotted the histogram of SMRs for the US ILI data to have an idea of how SMRs are distributed; v) as an exploratory analysis, we plotted SMR values against the covariate and vi) as a starting point of modeling strategy, we then fitted a loess curve to SMR versus the rate of African American population (PAfA). This fitted curve will assist us to identify the appropriate functional form of the relationship between outcome and predictor. Our model-building strategy was as follows:

Step 1: We conducted the simple linear regression model considering SMRs as dependent variable and the proportion of African American in each state as a single covariate (see equation 1 in model structure section) assuming that all the observations are independent, and residuals follow normal distribution.

Step 2: We conducted the simple Poisson log-link linear in covariate model defined in equation 2 (see model structure section) assuming dispersion parameter equal to 1 (mean=variance). At this stage, we checked the statistical significance of the covariate and looked for relative risk of area-based association between incidence and covariate.

Step 3: We then conducted quasi-Poisson regression as it provided us the general solution of over-dispersion assuming that variance is a function of mean. If we find large values for over-dispersion parameters and residual analysis shows any patterns, we may conclude that Poisson model is clearly inadequate. That means data may have spatial dependency.

Step 4: At this stage, we considered the spatial generalized linear mixed model (GLMM) to capture over-dispersion and the autocorrelation induced in the data by the latent spatial process with introducing the random effects in the model (see equation 3 in model structure section). To obtain estimates of the parameters of this model, we used Bayesian inference approach called empirical Bayes method and then full Bayesian method.

4. Model Structure

We present the most common version of our simple linear regression model with dependent variable SMR_i and the single covariate X_i, the rate of African American population for each area, as follows:

Model 1:

$$SMR_i = \beta_0 + \beta_1 X_i + \varepsilon_i, (i = 1, \cdots, m = 53) \qquad (1)$$

And we found the variable PAfA (X_i) had significant

(p=0.00825) effect on standardized morbidity rate for influenza disease (See Table 1). Moreover, the lowess smoother on the scatter plot showed a violation of the linear relationship between outcome and the predictor (see figure 1). Therefore, we conducted several possible polynomial orders of predictor to capture the non-linear functional relationship between predictor and outcomes. We found that polynomial order of 3, provide the better fit the functional relationship (see table 2). Since, our underlying dataset is a count data, therefore, the best approach for modelling this type of event may not be a usual linear regression model. Hence, it is customary to use generalized linear models (GLMs), with Poisson log-link linear in covariate model.

Now, we assume that our observed total ILI cases (Y_i) follows Poisson distribution with (μ_i). Then the log-linear model with single covariate (X_i) was defined as follows:

$$Y_i \sim_{ind} Poisson(\mu_i),\ (i = 1, \cdots, m = 53)$$

with $\mu_i = E[Y_i]$ and

Model 2:

$$\log \mu_i = \log E_i + \beta_0 + \beta_1 X_i \tag{2}$$

We carry out likelihood analyses using the *glm* function and the log-linear mean function defined in equation (2). Here, $\exp(\beta_1)$ represents the relative risk describing the area-based association between the incidence of influenza and the rate of African American population (PAfA) X_i. From the summary output of Poisson log-linear mean function without quasi-likelihood estimation approach (see table 3), we found that PAfA (X_i) has significant (p<0.0001) effect on standardized morbidity rate for influenza disease. Similar results were found for cubic model as well (see table 4). Since the Poisson model is restrictive in the sense that the variance is constrained to equal the mean, therefore, we conducted the estimation with quasi-likelihood approach to captures the over dispersion of estimates. Summary output of Poisson log-linear mean function (equation 2) with quasi-likelihood estimation approach is presented in table 5. Here, the over dispersion parameter was too high, which suggests that we need to go to step 4: build the spatial generalized linear mixed model (GLMM) approach to capture the over-dispersion problem by introducing the latent spatial process in the random effects term of the model. In this way, we could obtain smooth estimates for disease mapping as it reduced the instability inherent in SMRs based on small expected numbers.

The spatial GLMM is an example of a hierarchical model or a model defined in stages. At the first stage of the model, we define the distribution of the data given values of the random effects. At the second stage, we define the distribution of the random effects. By combining the first and second stages, we obtained inference about the data, considering the distribution of random effects, which leads us to use the Bayesian inference approach, an attractive alternative of likelihood approach in the estimation of the parameter of the model. The inference was made through the complete posterior distribution and we summarized it in terms of posterior moments, in most cases we used posterior median. We applied both empirical Bayes (MLE used for regression parameter and prior parameter) and full Bayes (prior distribution on regression and prior parameter) as of model-fitting approach.

The generic-model for state-wise observed total ILI cases y_i is:

$$y_i |\, \mu_i, \tau \sim f(y|\mu, \tau) \tag{3}$$

where $y = (y_1, \cdots, y_n)'$ denotes the vector of all observations. We modeled the mean (μ_i) of the observed disease cases, where $f(y|\mu, \tau)$ is the distribution of the likelihood of the observed total ILI cases, parameterized with mean $\mu = (\mu_1, \cdots, \mu_n)'$ and precision parameter τ as appropriate to the likelihood distribution.

Empirical Bayes Approach: Two stage Model (Poisson – Gamma model without covariate)

Now for our dataset, we assumed there were no covariates and assumed the first stage likelihood in equation (3) $f(y|\mu, \tau)$ is Poisson distribution, given by

Model 3:

$$Y_i|\theta_i, \beta \sim_{ind} Poisson(\mu E_i \theta_i) \tag{4}$$

where μ is the overall relative risk, act as intercept, and reflects differences between the reference rates and the rates in the study region.

At the second stage the random effects θ_i are assigned a prior distribution. We initially assumed that across the map the deviations of the relative risks from the mean, μ, are modelled by

$$\theta_i|\alpha \sim_{iid} Ga(\alpha, \alpha) \tag{5}$$

a gamma distribution with mean 1, and variance $\frac{1}{\alpha}$. The advantage of this Poisson-gamma formulation over naïve Poisson model was that it considered excess-Poisson variability resulting from unmeasured confounders and model misspecification. For this model, the relative risk was given by $RR_i = \mu\, \theta_i$ and has estimated mean $\widehat{RR_i} = \hat{\mu}\, E[\theta_i|\, \hat{\mu}, \hat{\alpha}] = E[RR_i] \times (1 - w_i) + SMR_i \times w_i$, with $\theta_i|\, \hat{\mu}, \hat{\alpha} \sim$ Ga $(\hat{\alpha} + y_i, \hat{\alpha} + E_i\hat{\mu})$ and $w_i = \frac{E_i\, \hat{\mu}}{\hat{\alpha} + E_i\, \hat{\mu}}$.

Empirical Bayes Approach: Two stage Model (Poisson – Gamma model with covariate)

Now for our dataset, we assumed that we have area level covariate, logarithm of population size and therefore, assumed that we have the mode

Model 4:

$$Y_i|\theta_i, \beta \sim_{ind} Poisson(\mu_i E_i \theta_i) \tag{6}$$

We assumed that across the map the deviations of the relative risks from the mean, μ_i, are modelled by equation (5), a gamma distribution with mean 1, and variance $\frac{1}{\alpha}$. Here, we obtained the estimates $\hat{\beta}, \hat{\alpha}$ by using maximum likelihood estimation over the marginal likelihood: a negative binomial distribution.

In the above two models, we applied an empirical Bayes

approach to estimates the parameters of the negative binomial model first (β, α) and then combined the gamma distribution with the data to obtain the empirical Bayes posterior distribution for the relative risks.

Full Bayesian Approach: Non-spatial Model (Poisson – lognormal model with covariate)

The Poisson-lognormal non-spatial random effect model is given by

Model 5:

$$Y_i|\beta, V_i \sim_{ind} Poisson(\mu_i E_i e^{V_i}), V_i \sim_{iid} N(0, \sigma_v^2) \quad (7)$$

Where, V_i are area-specific random effects that capture the unexpected log relative risk of disease in area *i*, $i = 1, \cdots, N = 53$. Here, $\theta_i = e^{V_i} \sim \log Normal\,(0, \sigma_v^2)$, whereas, for Poisson-gamma model θ_i has prior defined in equation (6). For this model we applied full Bayesian approach, hence, we need to specify priors both for i) the regression coefficient β and ii) the variance of the random effect σ_v^2. For regression parameters we choose non-informative priors, flat priors and we choose a gamma prior $dgamma(1,0.0260)$ for σ_v^{-2}, assumed that we had enough information for σ_v^2. This information was utilized in precision $\tau_v = \sigma_v^{-2}$ formula. In this non-spatial random effect model, the modeling of spatial dependence was much more difficult since spatial location was acting as a surrogate for unobserved covariates. Therefore, we moved to an appropriate spatial model which consists of both non-spatial and spatial structure random effects.

Full Bayesian Approach: Spatial Joint Model (Poisson – lognormal model with covariate)

We first consider the model

Model 6:

$$Y_i|\beta, \gamma, U_i, V_i \sim_{ind} Poisson(\mu_i E_i e^{V_i + U_i}) \quad (8)$$

with

$$\log(\mu_i) = f(x_i, \beta) + g(S_i, \gamma)$$

Where $f(x_i, \beta)$ is a regression model; and $g(S_i, \gamma)$ is an expression that may include to capture large scale spatial trend and S_i is the centroid of area *i*. The random effect V_i represents non-spatial over-dispersion defined in equation (7) and U_i are random effects with spatial structure.

Now, we assumed that $U = (U_1, \cdots, U_N)$ arise from a zero mean multivariate normal distribution with variance σ_u^2 and common correlation in all spatial directions $\tau_u^{-1} exp(-\varphi d_{ij})$, where $\tau_u^{-1} = \sigma_u^2$ and $\varphi > 0$. This model is called spatial joint model.

5. Results

As a descriptive analysis of covariate, we plotted the distribution of proportion of African American in each U.S states in figure 1 as below:

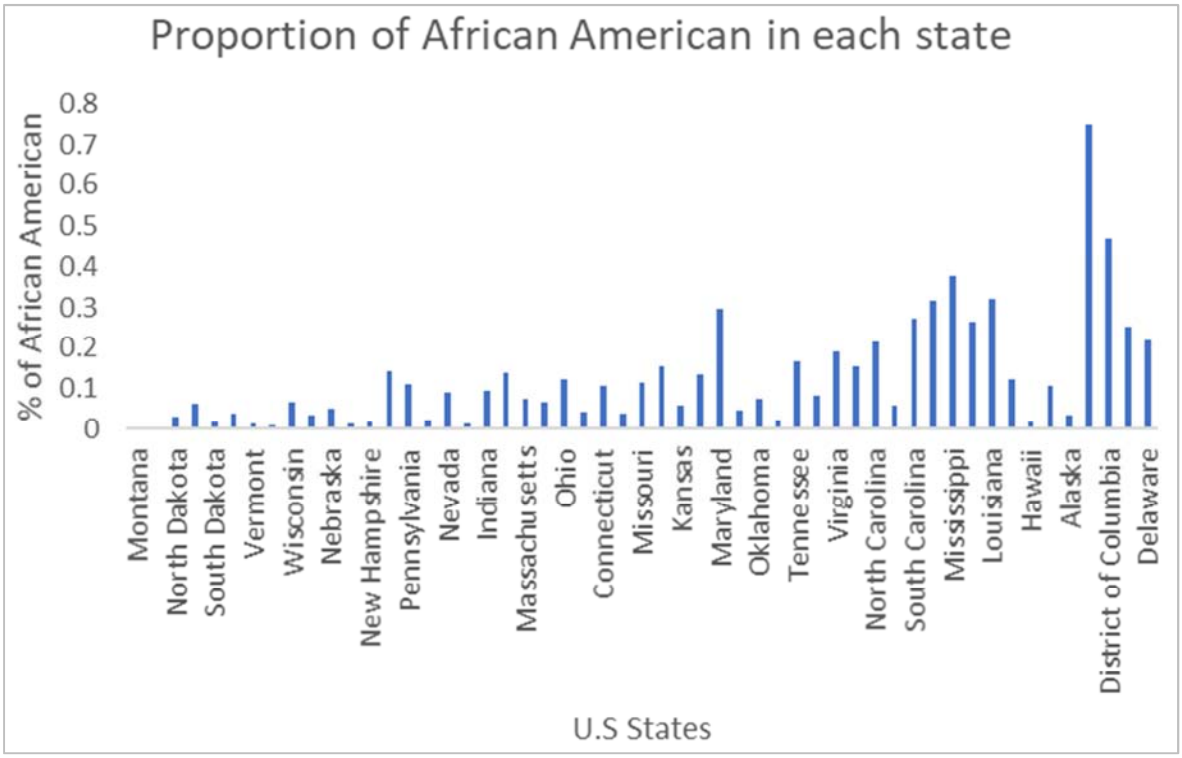

Figure 1. *Proportion of African American in each U.S states in 2019.*

From the figure 1, we observed that in U.S virgin lsland has highest rate of African American people and the second largest rate of this group of people was found in District of Columbia.

We then drew histogram of SMR and scatter plot of the SMR and PAfA variable (with fitted loess and linear in X model) given as below:

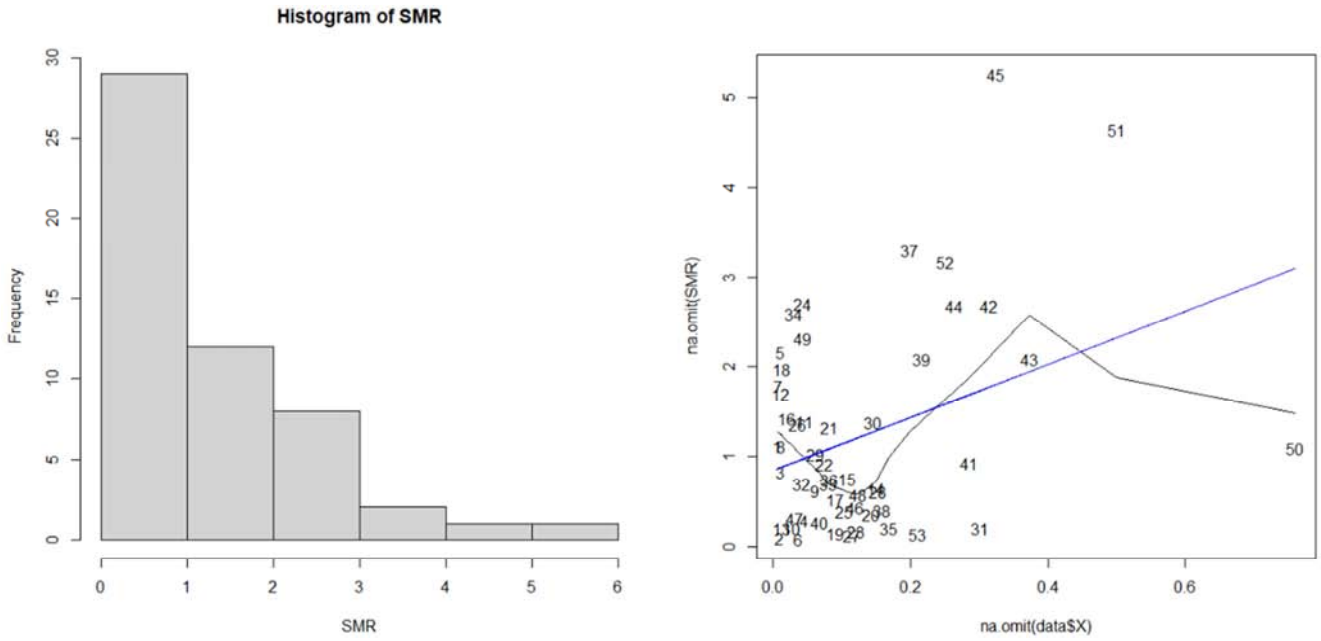

Figure 2. *Plot of (Y/E) versus proportion of African American people (PAfA). Solid line represents fitted line for linear in x model and other line represents loess fit.*

From the above figure, we observed that the linear in X model was not a good fit whereas lowess fit indicates a polynomial order of *x* variable may be a good choice of functional relationship between SMRs and racial variable (PAfA) X.

From the following tables, we observed that the cubic linear in X (PAfA) model was a good fit (provide highest adjusted R-square value compared to other models (see table 2)) and a good choice of functional relationship between SMRs and proportion of ethnic group of people (PAfA), whereas other fitted values were not good fit for our dataset. The output of linear in X and cubic in X model are given in the following table 1 and table 2.

Table 1. *Analysis of Maximum Likelihood Estimates for Model 1.*

Parameter	Estimate	Standard Error	t value	Pr (> |t|)	Adj R-Square
Intercept	0.8498	0.2028	1.189	0.00011	0.3978
PAfA(X)	2.9488	1.0728	2.744	0.00825**	

Significance: 0.001***, 0.05**, 0.01*.

Table 2. *Analysis of Maximum Likelihood Estimates for Model 1 with cubic form.*

Parameter	Estimate	Standard Error	t value	Pr (> |t|)	Adj R-Square
Intercept	0.7328	0.1908	3.841	0.000352	0.5289
X	2.7764	1.4696	1.889	0.0647	
X1	54.3885	13.285	4.093	0.00015	
X2	-91.6092	20.121	-4.553	3.52e-058**	

Significance: 0.001***, 0.05**, 0.01*.

In terms of adjusted R-square value, the cubic model provided better fit to our data. Therefore, we will choose the cubic form of functional relationship between outcome variable and the predictor.

As our next step, we fitted a log-link linear in X model (with and without quasi-likelihood) and log-link cubic in X model.

Table 3. *Analysis of Maximum Likelihood Estimates for Model 2.*

Parameter	Estimate	Standard Error	z value	Pr (> |z|)	AIC
Intercept	-0.757341	0.00150	-482.5	<2e-16***	887549
PAfA(X)	5.013122	0.00744	673.5	<2e-16***	

Significance: 0.001***, 0.05**, 0.01*.

Table 4. *Analysis of Maximum Likelihood Estimates for Model 2 with cubic form.*

Parameter	Estimate	Standard Error	z value	Pr (> |z|)	AIC
Intercept	-3.802e-01	1.32e-03	-287.8	<0.0001	771491
X	3.891e+00	1.267e-02	307.1	<0.0001***	
X1	4.475e+01	1.441e-01	310.8	<0.0001***	
X2	-1.424e+02	4.754e-01	-261.3	<0.0001***	

Significance: 0.001***, 0.05**, 0.01*.

So from table 3, $\widehat{\beta_0} = -0.7573(0.0015)$ and $\widehat{\beta_1} = 5.013(0.0074)$- the relative risk describing the area-based association between incidence of influenza and state wise proportion of African American people (PAfA) was $\exp(5.013) = 150.35$, which is large enough. Moreover, in terms of AIC value, the cubic model provided better fit to our data in Poisson log-link regression. Here, in both models, the covariates are highly statistically significant. But since the Poisson model is restrictive in the sense that the variance is constrained to equal the mean, therefore, we tried to fit quasi-Poisson regression (both for linear in X and linear in cubic of X) to captures the over dispersion of estimates.

Table 5. *Quasi-likelihood Estimates for Model 2.*

Parameter	Estimate	Standard Error	t value	Pr (> |t|)	Over-dispersion parameter	Residual Deviance
Intercept	-0.7573	0.2238	-3.384	0.00138	20330.13	886950
PAfA(X)	5.0131	1.0613	4.723	1.86e-05**		

Significance: 0.001***, 0.05**, 0.01*.

From table 5, we found that proportion of African American people in each state (PAfA), (X_i) had significant ($p<0.0001$) effect on standardized morbidity rate for influenza disease and estimates are identical to table 3 however, standard errors were multiplied by $\sqrt{20330.13} = 142.58$; still a large over-dispersion here (in table 5 and table 6), which indicates that Poisson model in clearly inadequate.

Table 6. *Quasi-likelihood Estimates for Model 2 with cubic form.*

Parameter	Estimate	Standard Error	t value	Pr (> |t|)	Dispersion parameter	Residual Deviance
Intercept	-0.3802	1.3585	-0.280	0.781	18605.62	770546
X	4.6282	2.7452	1.686	0.09781		
X1	44.7461	148.15	0.902	0.0764		
X2	-124.22	488.69	-0.254	0.800		

Significance: 0.001***, 0.05**, 0.01*.

Moreover, according to residual plot of model 2 (see figure 3) we observed that there was some variations or specific pattern in residuals, which indicated that data may have some spatial dependencies.

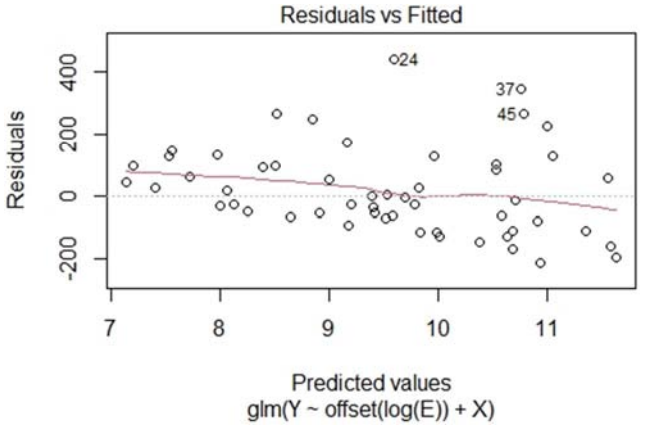

Figure 3. *Residual plot versus predicted values of model 2 (quasi-likelihood) linear in X.*

Empirical Bayes Approach: Two stage Model (Poisson – Gamma model without covariate)

As our next step, we fitted the generalized linear mixed model (GLMM) to capture the over-dispersion problem by introducing the random effects term in the model. In this way we could obtained smooth estimates for disease mapping as it reduced the instability inherent in SMRs based on small-expected numbers.

Table 7. *Empirical Bayes Estimates for Model 3.*

Parameter	Estimate	Standard Error
Intercept	0.204565	
Alpha	1.19672	0.9141

The weights on the SMR for model 3, ranges between 0.892 to 0.956 with median 0.934. Here the estimated standard deviation of the random effects model 3 was 0.9141. We observed that the weight on the observed SMR increases as E_i increases that means the estimate was dominated by the data. In this case, α was not so large.

Empirical Bayes Approach: Two stage Model (Poisson – Gamma model with covariate)

A reduction in standard deviation of random effects was observed when we moved from without covariate (model 3) model to with covariate model (model 4) (see table 7 and 8).

Table 8. *Empirical Bayes Estimates of parameter for Model 4 (with single covariate).*

Parameter	Estimate	Standard Error
Intercept	-0.0976	
PAfA(X)	2.01861	
Alpha	1.3076	0.8745

Table 9. *Empirical Bayes Estimates of parameter for Model 4 with cubic form.*

Parameter	Estimate	Standard Error
Intercept	-0.17314	
X	2.1488	
X1	33.1077	
X2	-46.0456	
Alpha	1.3281	0.8643

In comparison of two models: log-linear model in *PAfA* and log-linear model in cubic of *PAfA*, we might expect the standard deviation of random effects to be reduced in size when we add an important covariate but this did not happen (see table 8 and table 9). However, based on the mean square error (MSE) of the three model, we find cubic model provide lowest MSE (3.90E-08), indicates a better fit compared to others.

An empirical Bayes estimates (RRmedian) shown in figure 4 provide the stable relative risk estimates for different models (emp0: without covariate, emp1: with covariate and emp: with cubic term) of area-level data by assuming that the relative risks arise from a common gamma distribution, which allows smoothing towards a common value. In figure 5, we observed that the log-linear model did not fit well for large values of PAfA, which indicates the usage of a flexible model (cubic model of covariate PAfA to flatten off relative risk for larger values of PAfA).

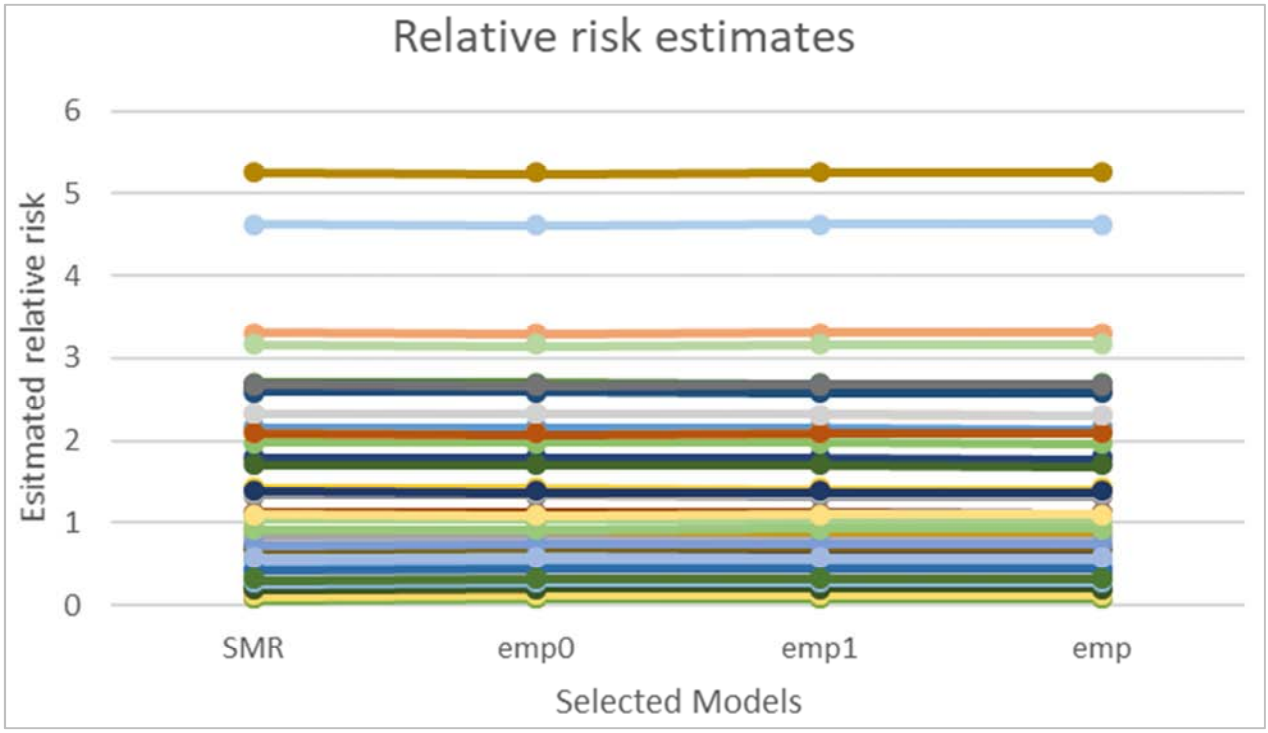

Figure 4. *Model's relative risk estimates along with raw SMRs values.*

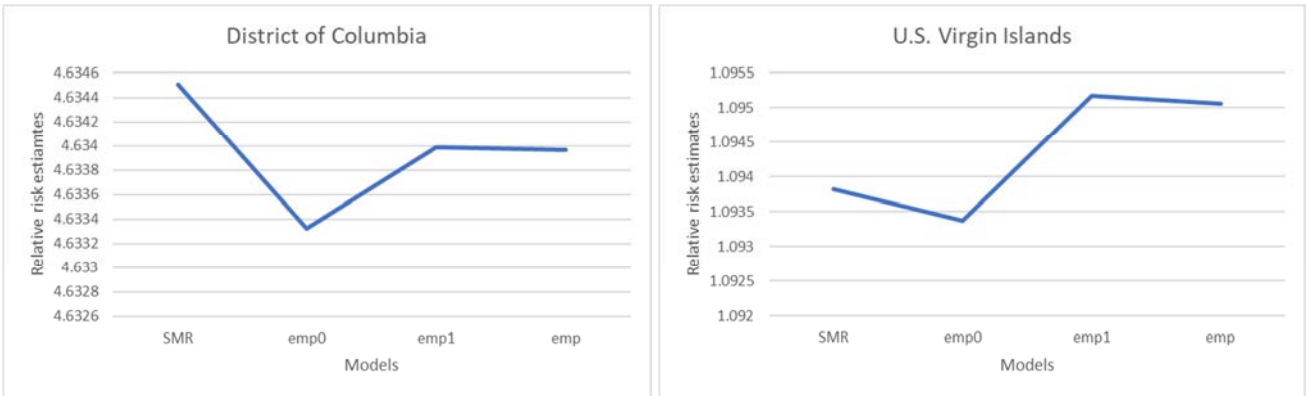

Figure 5. *For specific states (with large PAfA values see figure 1) cubic model flattens off relative risk (showed in original scales).*

Disease Mapping for U.S states ILI dataset (with raw and estimates values)

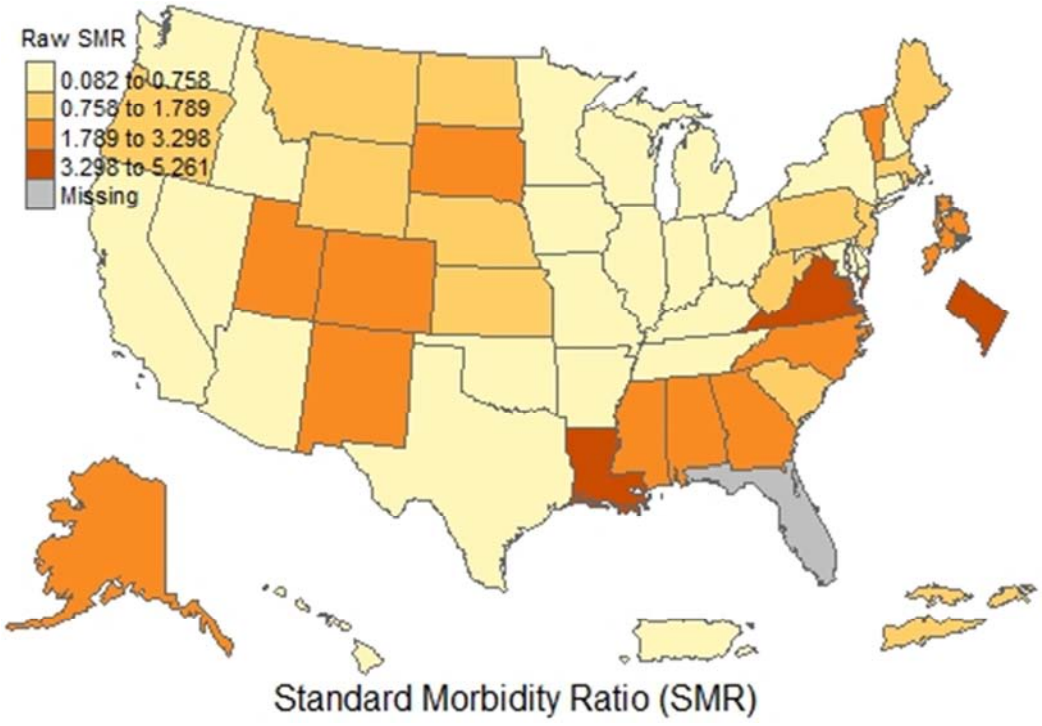

Figure 6. *Mapping of SMRs of influenza disease for U.S states.*

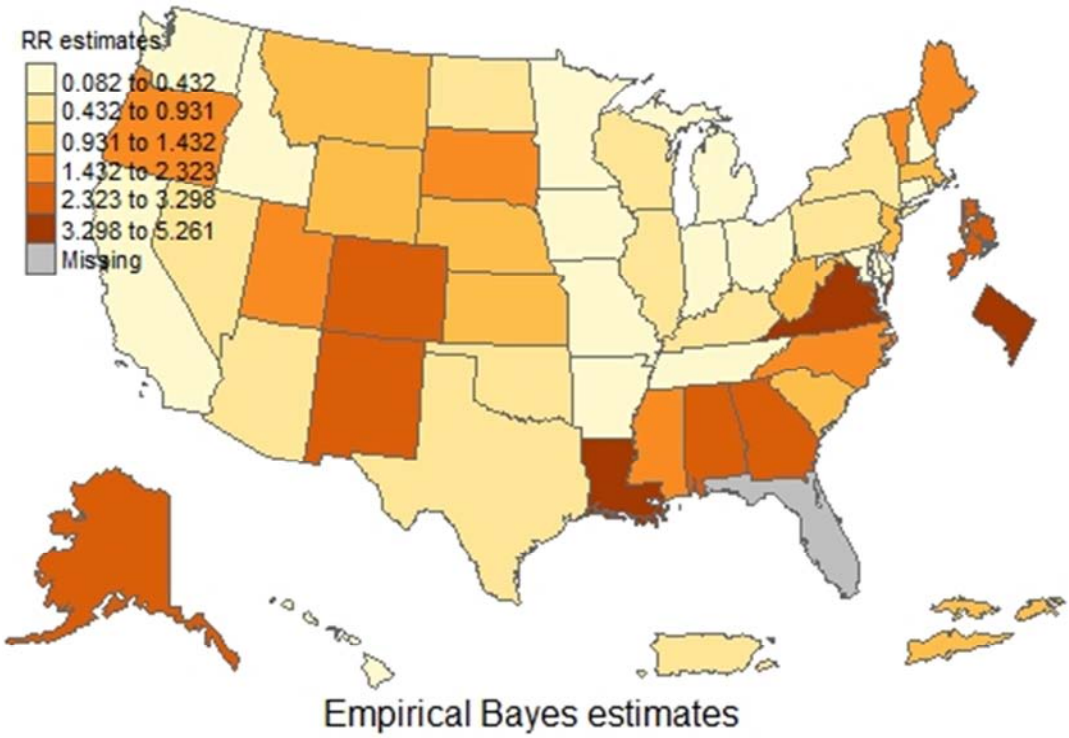

Figure 7. *Empirical Bayes posterior median estimates for cubic model.*

In both figures we observed some of the states had higher relative risk (>2.323).

Full Bayesian Approach: Non-spatial Model (Poisson – lognormal model with covariate)

The Poisson-lognormal non-spatial random effect model (Model 5) with cubic form is given by

$$Y_i|\beta, V_i \sim_{ind} Poisson(\mu_i E_i e^{V_i}),\ V_i \sim_{iid} N(0, \sigma_v^2)$$

with log-link function as follows

$$\log(\mu_i) = \log(E_i) + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + V_i, (i = 1, \cdots, 53)$$

and the log relative risk is defined in question as follows

$$\log(RR_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + V_i, (i = 1, \cdots, 53)$$

where V_i is the area specific random effects that capture the residual or unexplained (log) relative risk of influenza in area i. The covariates are centered here in order to reduce dependence in the parameter estimates. The model was fitted using MCMC via the WinBUGS software.

We obtained the following posterior distributions for regression parameters $\beta_0, \beta_1, \beta_2$ *and* β_3 using two sets of initial values (tau.V=1, beta0=0, beta1=0, beta2=0 and beta3=0 and tau.V=2, beta0=0, beta1=0, beta2=0 and beta3=0) for 2 chain and 11000 iterations with 3000 samples.

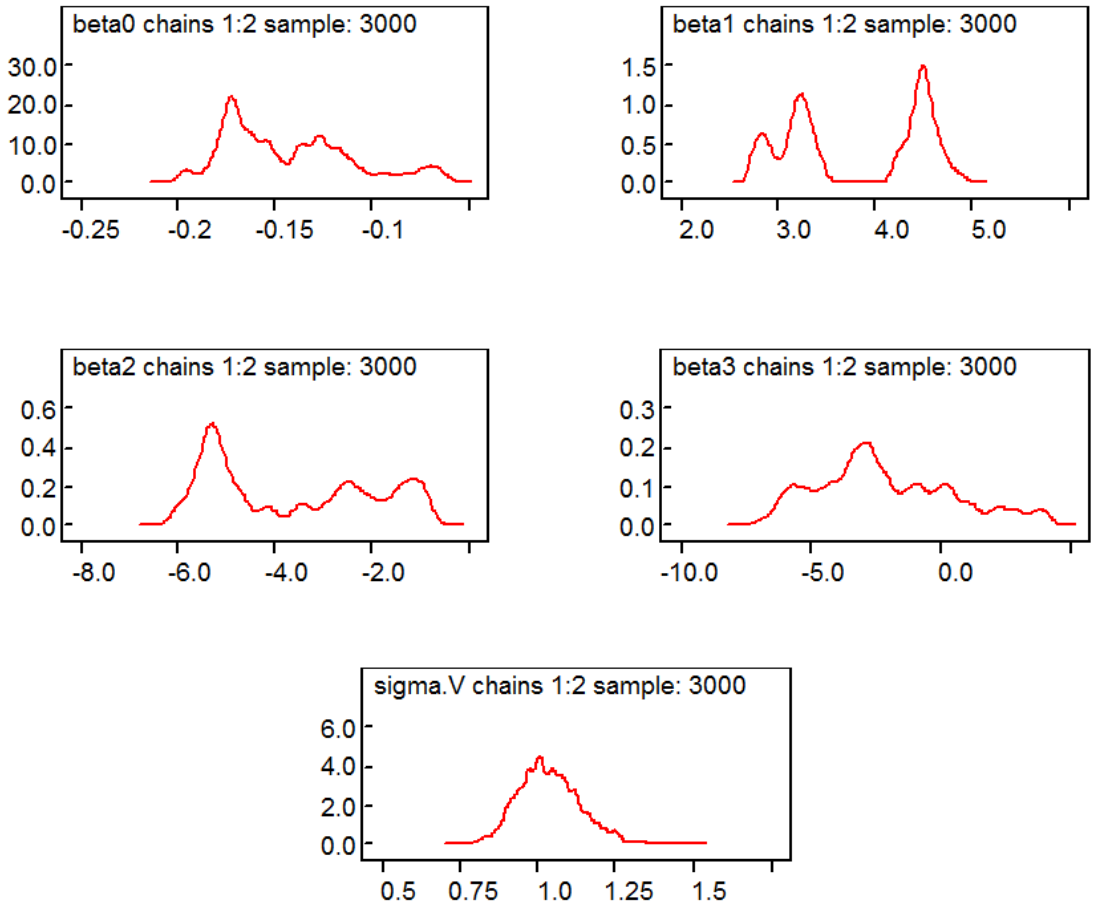

Figure 8. *Posterior distribution of regression parameters (first two rows) and* σ_v^2 *(third row) using two sets of initial values.*

And the associated summary statistics of MCMC output is given in following table 10:

Table 10. *Summary Statistics of MCMC output with selected parameters.*

Node	Mean	SD	MC error	2.5%	Median	97.5%	Start	Sample
beta0	-0.1429	0.033	0.0037	-0.193	-0.151	-0.066	8000	3000
beta1	3.813	0.715	0.0816	2.757	4.166	4.781	8000	3000
beta2	-3.645	1.709	0.1944	-5.943	-4.036	-0.87	8000	3000
beta3	-2.094	2.642	0.2932	-6.291	-2.562	3.698	8000	3000
singma.V	1.038	0.1031	0.0021	0.8612	1.03	1.25	8000	3000

From the above table we observed that regression coefficients has significant effects (as values between lower limit (2.5%) and upper limits (97.5%) did not contain zero value) on ILI incidence, and area specific random effects has significant effect (as values between lower limit (2.5%) and upper limits (97.5%) did not contain zero value). The convergence occurred for parameters of this model in MCMC iterations (see Appendix). Now we need to

incorporate spatial component in our non-spatial model to have a fully spatial model (model 6, the joint model).

Full Bayesian Approach: Spatial Joint Model (Poisson – lognormal model with covariate)

We first considered the model

Model 6:

$$Y_i|\beta,\gamma,U_i,V_i \sim_{ind} Poisson(\mu_i E_i e^{V_i+U_i}) \quad (9)$$

with

$$\log(\mu_i) = f(x_i,\beta) + g(S_i,\gamma)$$

where, $f(x_i,\beta)$ and $g(S_i,\gamma)$ defined in methodology section.

Here to find and summarize the posterior distributions of the β, the partial sill (σ^2) and the range (ϕ), the following steps were followed:

Step 1: Choose an informative prior (assuming we have enough information) for spatial random component as follows:

$$tau.T \sim Gamma(1, 0.0260)$$

Then use this information in joint model in terms of $Sigma_u = sqrt(p/(tau.T))$.

Step 2: Choose a prior distribution for the proportion (p) of the variance that is spatial from beta (1,1):

$$p \sim Beta(1,1)$$

Step 3: Choose a prior distribution for distance half from log-normal distribution as follows:

$$dhalf \sim LN(3.107,0.9106)$$

We obtained the following posterior medians for different joint model with prior parameters setup for tau.T and distances and run MCMC using two sets of initial values (tau.T=1, p=0.5, beta0=0.01, beta1=0, beta2=0 and beta3=0 and tau.T=2, p=0.5, beta0=0.01, beta1=0, beta2=0 and beta3=0) for 2 chain with 3000 samples.

Table 11. *Sensitivity of spatial model (model 6) parameters to prior choice tau.T and p.*

Spatial Model	Prior Specification		Posterior medians			
			σ_v	σ_u	p	distance
Joint	$\tau_T \sim Ga(1,0.0260)$	$d_{1/2} \sim LN(3.107,0.9106)$	0.7116	1.124	0.996	
Joint	$\tau_T \sim Ga(1,0.1339)$	$d_{1/2} \sim LN(3.107,0.9106)$	0.0930	1.129	0.993	
Joint	$\tau_T \sim Ga(1,0.0260)$	$d_{1/2} \sim LN(2.303, 0.4214)$	0.0709	1.077	0.975	

In table 11, we examine the sensitivity of estimates of the non-spatial and spatial contributions of residual relative risk, to the prior choices of random effect parameters and distances. Here, the prior $\tau_T \sim Ga(1,0.1339)$ provides relative risk that follow a log student t distribution with 2 degrees of freedom and fall within the range (0.2, 5) with probability 0.95. And the choice of $d_{1/2} \sim LN(2.303, 0.4214)$ assumes that there is a 5% chance that the correlations lie to 0.5 in less than 5km, and 95% chance that they lie to 0.5 in less than 20km. From the table we see that, most of the residual variability was explained by the spatial component; under the various models, 0.975 to 0.996 of the total variability was spatial in nature. Overall, there was a little sensitivity of these parameters to the choice of priors.

6. Conclusion

The preferred model was a model which includes a cubic term in the proportion of African Americans (PAfA) and a spatial component, since the association with PAfA was strong, there was significant residual spatial dependence. Here, this study considered empirical Bayes as an exploratory analysis of whether to include the residual spatial dependence in our fully Bayesian model. For prior specification of parameters for fully Bayesian spatial model with log-linear cubic term, we conducted a sensitivity analysis in relative estimates. We found a large amount of residual variability of these data, which suggested unobserved risk factors were present and were not surprising since we had no information on other variables that are important (such as temperature, relative humidity, population density, etc.). The limitation of this study was that due to the unavailability of ILI cases from CDC report for each of the weeks in the year 2019, we did not consider ILI cases for Florida states though it is the third largest populated area and not including other important covariates (environmental and socio-demographic, behavioral factors) to determine the relative risk of having Influenza disease. Moreover, the influenza cases are highly related to seasons which was not considered in this study. Hence, conducting a spatial-temporal model could provide an accurate and reliable estimate of relative risk over time.

Conflict of Interest

The authors declare that they have no competing interests.

Acknowledgements

Authors are grateful to the anonymous reviewers for their valuable comments and suggestions. Authors also show their acknowledgment to Centers for Disease Control and Prevention (CDC) for providing weekly report of Influenza.

Appendix

Poisson log-linear non-spatial model parameters convergence (beta0, beta1, beta2 and sigma.V).
BGR plot:

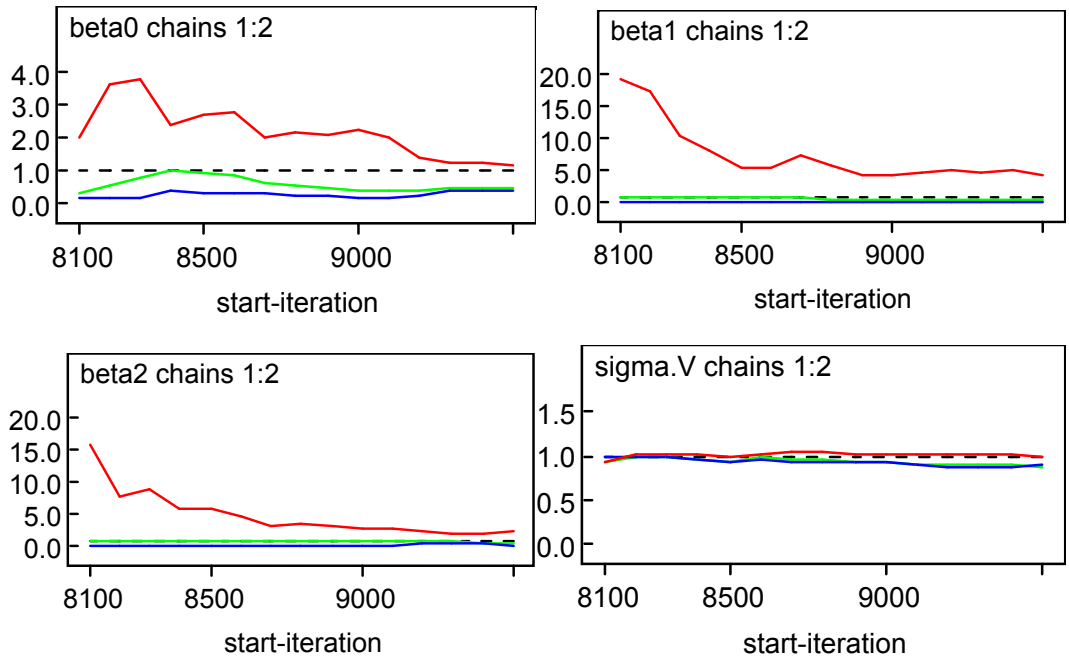

Figure 9. *BGR plot of selected parameters.*

Trace plot:

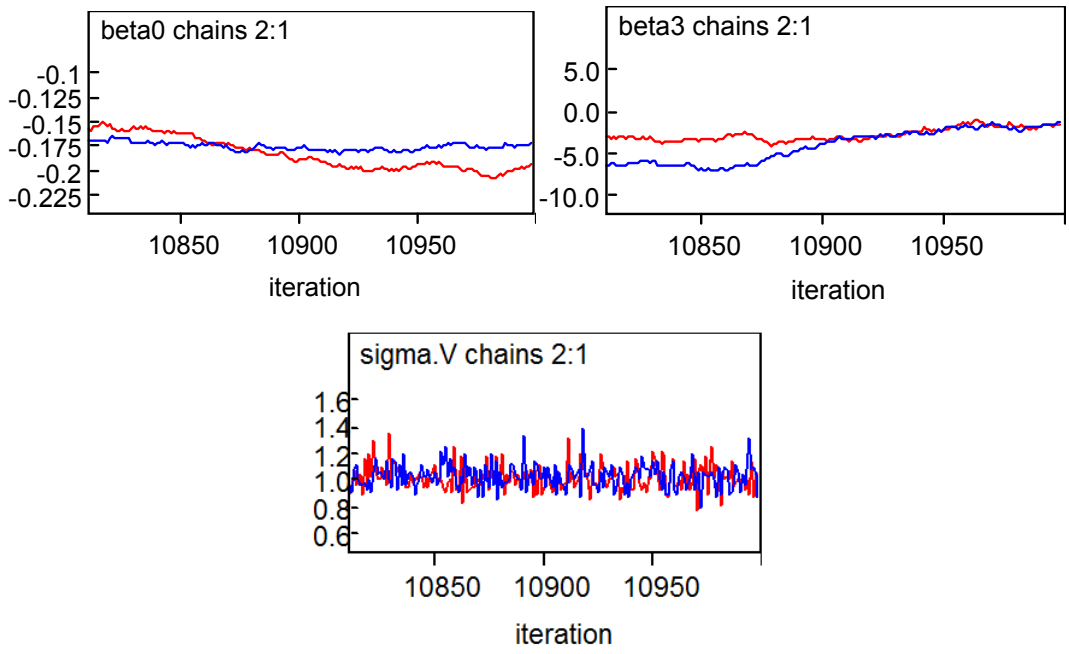

Figure 10. *Trace Plot of corresponding parameters.*

References

[1] "Influenza (flu) - Symptoms and causes - Mayo Clinic." https://www.mayoclinic.org/diseases-conditions/flu/symptoms-causes/syc-20351719 (Access on 24.07.2020).

[2] WHO Influenza (Seasonal) [Internet]. Fact Sheet Number 211, 2015 [cited 2016 May 10]. Available from: http://www.who.int/mediacentre/factsheets/fs211/en/index.html.

[3] Molinari M, A et al., (2007). The annual impact of seasonal influenza in the US: measuring disease burden and cost. Vaccine, 25 (27): 5086-96.

[4] World Health Organization (WHO) Global Influenza Strategy, [Internet]. Access on 07.07.202 and available from https://www.who.int/influenza/global_influenza_strategy_2019_2030/en.

[5] Overview of influenza Surveillance in the United States | Seasonal Influenza (Flu) | CDC [Internet]. [Cited 2016 May 31] Available from: http://www.cdc.gov/flu/weekly/overview.html.

[6] Lefgren E., Fefferman N H, Naumor Y N., Groski J., Noumova E N. (2007) Influenza seasonality: underlying causes and modeling theories. *J Virology*, 81 (11): 5429-36, doi: 10.1128/Jul.01680-06.

[7] Teserius D. J., Shaman J., Alonso J. W., Feshbach K. B., Uejio C. K., Conries A., and Viboud C., (2013) Environmental Prediction of Seasonal Influenza Epidemics across Temperature and Tropical climates. *PLOS Pathogens*, https://doi.org/10.1371/journal.ppat.1003194

[8] Wenger B. J., and Naumova N. E. (2010) Seasonal synchronization of influenza in the United States older adult people. *PLOS One*, 5 (4): e10187, doi: 10.1371/journal.pom.0010187.

[9] Schanzer D, Vachon J, Pelletier L. (2011) Age-specific differences in influenza A epidemic curves: do children 690 drive the spread of influenza epidemics? *Am J Epidemiol*. 174 (1): 109–17. doi: 10.1093/aje/kwr037.

[10] Gracies R. G., Ellis J. H., Kress A., and Glass G. E. (2004) Modeling the spread of annual influenza epidemics in the U.S.: The Potential Role of Air Travel. *Health Care Management Science*. 7, 127-134. doi: 10.1023/b:hcms.0000020652.38181.da.

[11] Brownstein S. J., Wolfe C. J. and Mandl D. K., (2006) Empirical evidence for the effect of Airline travel on Inter-Regional influenza spread in the United States. *PLoS Medicine*, 3 (10): 125-135, doi: 10.1371/journal.pmed.0030401.

[12] Shaman J, Pitzer VE, Viboud C, Grenfell BT, Lipsitch M (2010) Absolute Humidity and the Seasonal Onset of Influenza in the Continental United States. *PLoS Biol* 8 (2): e1000316, doi: 10.1371/journal.pbio.1000316.

[13] Deyle R. E., Maher C. M., Hernandez D. R., Sanjay B., Sugihara G., (2016) Global environmental drivers of influenza. *Proc Natl Acad Sci*, 13 (46): 13081-13086, doi: 10.1073/pnas.1607747113.

[14] Mossong J., et al., (2008) Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Disease, *PLoS Medicine*, 5 (3): 381-390, doi: 10.137/journal.pmed.0050079.

[15] Marion Z. H., Ferdyce A. J., and Fitzpatrick M. B., (2018) A hierarchical Bayesian model to incorporate uncertainty into methods for diversity partitioning, *Ecology,* 99 (4): 947-950.

[16] Sloan C., Chandrashekhar R., Mitchel E., Schaffer W., and Lindergren M. L., (2015) Socioeconomic Disparities and influenza Hospitalizations, Tennessee, USA. *Emer Infect Dis*, 21 (9): 1602-1610, doi: 10.3201/eid2109.141881.

[17] Crighton, E. J., Elliott, S. J., Moineddin, R., Kanaroglou, P., and Upshur, R. (2007) A spatial analysis of the determinants of pneumonia and influenza hospitalizations in Ontario (1992-2001). *Social science & medicine* (1982), 64 (8): 1636–1650, doi: 10.1016/j.socscimed.2006.12.001.