

Modeling Severe Acute Respiratory Syndrome Coronavirus 2019 (SARS-CoV-19) Incidence across Conterminous US Counties: A Spatial Perspective

Olawale Oluwafemi^{a*}, Oluseyi Oladepo^b

^a , Spatially Integrated Social Science Program, University of Toledo, OH, United States of America, America, Olawale.Oluwafemi@rockets.utoledo.edu

^b African Regional Institute for Geospatial Information Science and Technology, OAU, Ile-Ife, Nigeria, oluseyioladepo@gmail.com

* Corresponding author

Abstract: This study examines the spatial distribution of COVID-19 incidence and mortality rates across the counties in the conterminous US in the first 604 days of the pandemic. The dataset was acquired from Emory University, Atlanta, United States, which includes socio-economic variables and health outcomes variables (N=3106). OLS estimates accounted for 31% of the regression plain (adjusted R²= 0.31) with AIC value of 9263, and Breusch-Pagan test for heteroskedasticity indicated 472.4, and multicollinearity condition number of 74.25. This result necessitated spatial autoregressive models, which were performed on GeoDa 1.18 software. ArcGIS 10.7 was used to map the residuals and selected significant variables. Generally, the Spatial Lag Model (SLM) and Spatial Error Model (SEM) models accounted for substantial percentages of the regression plain. While the efficiency of the models is the order of SLM (AIC: 8264.4; Breusch-Pagan test: 584.4; Adj. R² = 0.56) > SEM (AIC: 8282.0; Breusch-Pagan test: 697.2; Adj. R² = 0.56). In this case, the least predictive model is SEM. The significant contribution of male, black race, poverty and urban and rural dummies to the regression plain indicated that COVID-19 transmission is more of a function of socio-economic, and rural/urban conditions rather than health outcomes. Although, diabetes and obesity showed a positive relationship with COVID-19 incidence. However, the relationship was relatively low based on the dataset. This study further concludes that the policymakers and health practitioners should consider spatial peculiarities, rural-urban migration and access to resources in reducing the transmission of COVID-19 disease.

Keywords: COVID-19, GIS, Spatial Analysis, United States

1. Introduction

SARS-CoV-19 otherwise known as Severe Acute Respiratory Syndrome Coronavirus 2019 was first reported on December 30, 2019, in Wuhan, China, as a pneumonia-related diagnosis (Xie et al., 2020). Several weeks later, The World Health Organization (WHO) officially tagged coronavirus "2019-ncov" as 2019 novel coronavirus and estimated its incubation period to be about 2 to 14 days. The primary aim of the spatial analytical process is to measure geographic distributions, analyze patterns, map clusters, and model spatial relationships among observed variables. Hence, spatial analysis becomes vital in medical geography because diseases' distribution tends to be intrinsically linked with socio-economic, political, and environmental conditions that affect susceptibility. However, mapping patterns of phenomena have provided tremendous advantages of observing hidden relationships among variables (Oluwafemi et al., 2013). The community transmission was first reported in February, 2020 (CDC, 2020; Desjardin, et al., 2020). It is worthy of note that confirmed cases of COVID-19 had been reported in every state in the U.S by mid- March 2020 (Schuchat, 2020). As at April 7th 2020, the worldwide cases of COVID-19 has risen to over 131 million cases, 2.85 million deaths and over 74.5

million people have recovered from the virus (The New York Times, The COVID-19 Tracking Project, 2020). As of April, 7th 2021, in the US, the virus has infected 30,732,250 million people and resulted in 554,579 deaths (The New York Times, The COVID-19 Tracking Project, 2020). The major transmission route of COVID-19 is through respiratory droplets from close direct contact with symptomatic, pre-symptomatic or asymptomatic people, and indirect contact through objects, or aerosols over longer distances (Zhang, et al., 2020). The Basic Reproduction Number (R₀), is a commonly used epidemiologic measure of transmissibility of an infectious agent. R₀ will be greater than 1 during an outbreak and will drop to less than 1 as the outbreak subsides. This statistic can be used to estimate the proportion to be vaccinated within a population in order to control the spread of the infection (Delamater, et al., 2019). Across different regions of the U.S., the R₀ ranged from 1.3 to 3.8 on March 1, 2020 and from 0.64 to 1.1 on May 1, 2020, according to Rt.live.com website (Rt Live, 2020). Early research suggested that the average number of days from transmission of COVID-19 to case confirmation was 18 (Backer, et al., 2020). According to Desjardins et al. (2020), COVID-19 estimated R₀ equals 2.2 to 6.7 depending on various sources. The challenge of COVID-19 has been of global concern because of it's "unknowns"

and the impacts of its emergence on virtually all aspects of life. Early studies on the COVID-19 have suggested that preexisting health conditions, air pollution, and socio-economic variables could be pointers to COVID-19 Incidence and mortality (Petroni et al.;2020; Xie et al.; 2020; Wu et al.; 2020 and Pansini and Fornacca, 2020). However, understanding the spatial distribution of COVID19 cases, and modelling variables that influences its transmission relies mainly on big data. The quality of the big data allows digital manipulations using spatially-inclined software that allows exploratory spatial analysis. A number of studies have explored the relationship between COVID-19 transmission and socio-economic variables in New York City and Chicago, United States. Maroko et al. (2020) showed a relationship between COVID-19 transmission and socio economic and demographic characteristics. The authors affirmed certain socio-economic variables as predictors of COVID-19 disease. Safray et al., (2020) adopted both global and local spatial correlation statistic to examine dependencies in US counties. They affirmed that race, and certain health outcomes are predictors of COVID-19 transmission. Traditional statistical models do not account for spatial dependence (Smirnov & Anselin, 2001; Anselin, 2003; Anselin, et al., 2006). When spatial dependence is present, Ordinary Least Squares (OLS) regression produces biased parameter estimates (Smirnov & Anselin, 2001). In order to mitigate this, spatial autoregressive models are used that uses Maximum Likelihood (ML) estimation and accounts for the presence of spatial dependence in the data. The Incidence of COVID-19 has been found to vary over space at the global, national, and local scales depending on the risk factors (Franch-Pardo et al., 2020; Desjardins et al., 2020; Petroni et al., 2020; Adekunle et al. 2020 and MFF..Sobral et al., 2020). In particular, there is a notable geographical variation in the distribution of COVID-19 cases across the US's counties (Desjardins et al., 2020). In addition, COVID-19 mortality also varies across counties in the US (Zhang and Schwartz, 2020). In the spatial autoregressive models, spatial dependence is incorporated using two different methods; either in a spatially lagged dependent variable or through the error term. The former method is known as a spatial lag model and the later as spatial error model (Anselin, 2003). In the case of unaccounted spatial error, regression will have inefficient results (Anselin, et al., 1996), potentially giving incorrect standard error, wrong significance, or wrong model fit. When the spatial lag term is not treated as an endogenous variable under a proper estimation method, it will produce biased and inconsistent results in the regression model (Anselin, 1988; Baltagi, et al., 2007; Fotheringham & Rogerson, 2008; Lee & Yu, 2010; Badr, et al., 2020). To address the inconsistent, inefficient, and biased results of traditional statistical models, We employed spatial autoregressive models to examine the effects of selected socioe-economic and health outcomes variables on COVID-19 cases in the contiguous U.S. over the months of March 2020 and August, 2020 using Emory University Datasets.

2. Study Area

The United States is one of the North American continent countries; it is believed to be the most powerful nation globally in terms of Gross Domestic Product (GDP). According to the United States Census Bureau 2019 projection, the population is 329, 256,465 million, with the Capital city in Washington DC. The contiguous United States has 3,143 counties and 5 administrative regions (Figure 1). The land area has 3,796,725 square miles (9,833,517 square kilometers) with 50 states within the contiguous United States (ThoughtCo.). The temperature is mostly temperate, tropical in Florida, semi-arid along the Mississippi River, and arid in the southwest's Great Basin (ThoughtCo). This study area includes all counties in the contiguous United States, where COVID -19 incidence and mortality data are available for the study period from January, 21 to September 16, 2020.

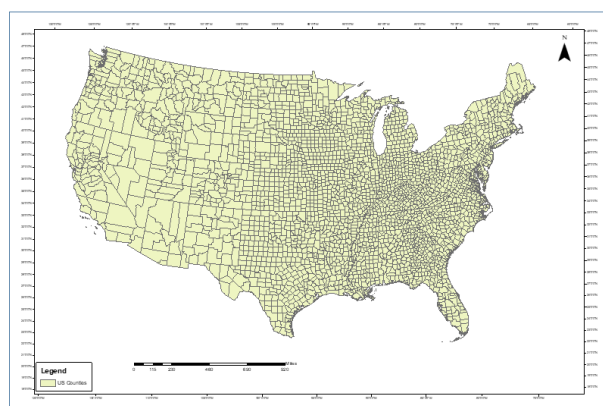


Figure 1: Study Area, Conterminous United States

3. Dataset and Descriptive Statistics

The secondary data was used for this study. The dataset comprises over 40 variables, which includes COVID-19 cases count (7 and 14 days), mortality rate, COVID-19 Incidence, Race (Black, White, Hispanic), % Male, % Female, Household Income, Community Vulnerability Index, Population density, % Insured, % Uninsured, Age over 65, Poverty, Diabetes, Obesity among others. These datasets were acquired from Emory University COVID-19 Health Equity Interactive Dashboard, sourced from government and non-government agencies (Table 1).

Table 1: Data Sources

	Data	Source
1	Total COVID-19 Cases	The New York Times Coronavirus (Covid-19) Data in the United States
2.	Total COVID-19 Cases per 100,000	Derived from the New York Times Coronavirus (Covid-19) Data in the United States and Bridged-race population estimates by The National Center for Health Statistics
3.	Total COVID-19 Deaths per 100,000	Derived from the New York Times Coronavirus (Covid-19) Data in the United States and Bridged-race population estimates by The National Center for Health Statistics
4	Percent Positive	The COVID Tracking Project
5	Cases per 100,000 Persons by Race	Derived from The COVID Racial Data Tracker and American Community Survey by the US Census Bureau
6	% African American	American Community Survey by the US Census Bureau
7	% Hispanic or Latino	American Community Survey by the US Census Bureau
8	% American Natives	American Community Survey by the US Census Bureau
9	% Minority	American Community Survey by the US Census Bureau
10	% in poverty	American Community Survey by the US Census Bureau
11	% Uninsured	American Community Survey by the US Census Bureau
12	% Diabetes	CDC's Division of Diabetes Translation
13	% Obesity	CDC's Division of Diabetes Translation
14	%over 65 y/o	American Community Survey by the US Census Bureau
15	% Male	American Community Survey by the US Census Bureau
16	Socio-economic Vulnerability	CDC's Social Vulnerability Index data 2018 database
17	Population	Bridged-race population estimates by The National Center for Health Statistics
18	Population Density	American Community Survey by the US Census Bureau
19	Household Income	American Community Survey by the US Census Bureau

4. Data Summary

The datasets used for this study were summarized and variable statistics were calculated using STATA 13. The descriptive statistics for the explanatory variables were provided in Table 2.

Table 2: Descriptive statistics of model variables

Variable	Obs	Mean	Std. Dev
Pop. density	3106	269.7114	1808.224
Male	3106	50.02443	2.671016
Poverty	3106	15.61238	6.475496
Diabetes	3106	10.3886	3.81996
Obesity	3106	32.73374	5.74366
Black	3106	50.02443	2.671016
Caserate (Newcasqrt)	3106	1095.914	328.0724

4.1 Research Questions

The study will provide answers to the following research questions;

- Does the COVID-19 occurrence shows spatial variation and spatial dependencies across counties?
- What is the spatial effects that present in the datasets?
- What is the relationship between COVID-19 occurrences and selected socio-economic and health outcomes variables across US counties?

4.2 Aim and Objectives of the study

The aim of this cross-sectional study is to use the spatial analytical models to statistically investigate the spatial relationship between selected explanatory variables and COVID-19 transmissions with the view of solving public health problems and providing a framework for resources allocation to the deprived counties.

The objectives of the study are to:

- to use Ordinary Least Square estimation as diagnostics tool to determine pattern of correlation and spatial dependence of COVID-19 incidence across the counties;
- to use spatial autoregressive models to explain the spatial relationship of between COVID-19 case rate incidence and selected variables in the continental United States;

4.2.1 Working Hypothesis

The study sets to test the following hypotheses

- There is spatial variation and spatial dependencies in the COVID-19 incidence.
- COVID-19 incidence variation can be statistically explained with selected socio-economic and health outcomes variables.

4.2.2 Spatial Analytical Procedure

In order to isolate the real predictor variables that influence COVID-19 incidence was transformed by squaring the incidence variable. I subjected the data to stepwise and exploratory analysis. In this case, COVID-19 incidence is the dependent variable while other variables (population density, male, poverty, diabetes, black, obesity and both urban and rural counties were represented as dummy variables; 1 for urban and 0 for rural counties) were entered into the model as potential predictor variables (regressors).

4.2.3 Spatial Weight Matrix

We first characterized the spatial relationships and identify the neighborhood structure by defining who neighbors are among all observation in the dataset this was done by creating spatial weight matrix (Anselin, 2003; Fotheringham and Rogerson, 2008). The spatial weight matrix expresses the existence of a neighbor relations and quantifies neighborhood structure between observations using $n \times n$ matrix, W (Anselin, 2003). The spatial weight matrix were calculated using Queen contiguity binary weight matrix, which defines the neighbor as any spatial unit that shares a common edge or vertex using GeoDa 1.18 software. The spatial weight is 0 if any two unit i and j are not neighbors and 1 if they are neighbors. The diagonal cells of a spatial weights are also represented by 0 because a geographic units is not considered neighbor of itself (Anselin, 2003). The spatial weights matrix is row standardized (Equation 1), where the given weights, W_{ij} , are divided by the row sum resulting in the sum of all weights to equal n , that is the total number of observations. (Anselin, 2003).

$$W_{ij} = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}} \quad (1)$$

4.2.5 Estimation method & Models adopted

To establish a relationship between the isolated predictor variables and the dependent variable (COVID-19 case rate), I adopted (Ordinary Least Squares estimates (OLS), Spatial Lag Model (SLM), Spatial Error Model (SEM)). For a start, I subjected the potential predictor variables to OLS regression in GeoDa 1.18 platform (geodacenter.github.io). The OLS is a regression method that investigates the relationships between a set of explanatory or independent variables and dependent variable and has the general form (Ward and Gleditsch, 2018).

$$y_i = \beta_0 + x_1\beta + \varepsilon_i \quad (2)$$

Where at county i , y_i is the COVID-19 incidence, β_0 is the intercept, x_1 is the vector of the selected variables, β is the vector of regression coefficient, and ε_i is a random error term.

$$COVID-19 \text{ Case rate}_i = b_0 + b_1 \text{pop density}_i + b_2 \text{male}_i + b_3 \text{poverty}_i + b_4 \text{diabetes}_i + b_5 \text{obesity}_i + b_6 \text{urban dummy}_i + \varepsilon_i$$

The outcome of the OLS regression was shown in Table 2 and OLS residuals was squared and latter mapped (See Figure 2).

4.2.6 Determining the Spatial Dependence

To achieve the first hypothesis of the study, with the null hypothesis saying. There is spatial variation and spatial dependence in the COVID-19 infection cases. Moran's I test which was captured in the OLS diagnostics for spatial dependence shown in Table 3 below indicated Moran's I value (0.4225, $p < 0.00000$) which implies clustering pattern and significant positive spatial dependence in the

number of COVID-19 incidence across the counties in the U.S. Since Moran's I statistic is diagnostic tool, this result actually pointed me to the direction to go for testing for the marginal effect of spatial dependence using Spatial Lag Model (SLM) or Spatial Error Model (SEM).

4.2.7 Spatial Lag Model (SLM)

According to Anselin, (2003); Ward and Gleditsch, (2018), assumes dependency between the dependent variables and incorporates spatial dependence into the regression model with a "spatially-lagged dependent variables". SLM is denoted by:

$$y_i = \beta_0 + x_1\beta + \rho W_i y_i + \varepsilon_i \quad (3)$$

Where,

ρ is the spatial autoregressive variable (i.e. the spatial lag parameter), and W_i is a row of the matrix of spatial weights (that is, vector of the spatial weights). The origin of equation 2 is rooted in the decomposition of the error term in equation 1 (Mollalo et al. 2020; Ward and Gleditsch, 2018). Here, W indicates the neighbors around county i and, thus, accounts for the influence of the predictor variables on the dependent variable at the boundaries around county i (Mollalo et al, 2020; Anselin and Arribas-Bel, 2013).

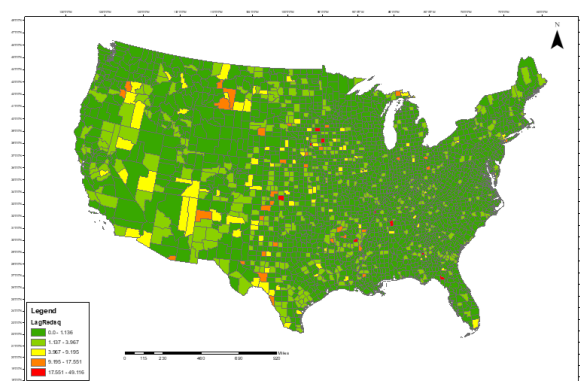


Figure 2: Residual Map of Spatial Lag Model
4.2.8 Error Model (SEM)

The SEM accepts that there is spatial dependence in the error term of OLS and decomposes the error term in Equation 1 above into two terms ($\lambda W_i \xi_i$ and ε_i below) (Anselin, 2003; Chen et al., 2016). The general form for this model is: (Ward and Gleditsch, 2018).

$$y_i = \beta_0 + x_1\beta + \lambda W_i \xi_i + \varepsilon_i \quad (4)$$

where

ξ_i represents the spatial component of the error, λ connotes the existing correlation rate among the components, and ε_i denotes the non-correlated spatial error term. The outcome of the SEM model was shown in Table 3 and SEM residuals was squared and latter mapped (See Figure 3b).

The Spatial Lag residuals in the SEM latter were squared in GeoDa 1.18 platform so as to remove the negative signs

present in the residuals. Figure 3 below shows the SEM squared residuals.

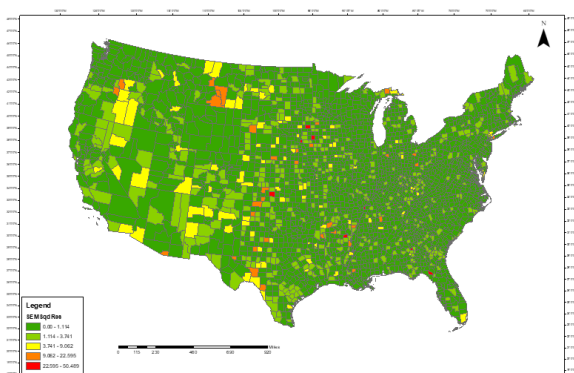


Figure 3: Residual Map of Spatial Error Model

The OLS result indicated a passing model for spatial regression which includes the selected independent variables as population density, male, poverty, black, diabetes, obesity and urban/rural dummies. Meaning that, only the seven variables were taken as the major predictor variables that influenced COVID-19 incidence across US counties between January 21 and September 16, 2020. For uniformity and consistency sake, we subjected the isolated predictor variables to the adopted regression models. For comparison sake, we implemented SLM and SEM on ArcGIS 10.7 GeoDa 1.18 (geodacenter.github.io) software.

SLM and SEM were implemented on GeoDa 1.18 platform. The performances of the models were comparatively evaluated based on the Breusch-Pagan, R^2 , Akaike Information Criterion (AICc) and also the coefficient of the selected independent variables were used to assess the relationships that exist among the variables across counties in the U.S. during the first eight months of the global pandemic.

5.0 Analysis of Results

The initial estimate summary result of OLS with the seven regressors (population density, male, black, poverty, diabetes, obesity and urban/rural dummies) included in the estimation to measure the outcomes of COVID-19 incidence across counties in the US in the first eight month (604 days) of the pandemic is shown in Table 3. While COVID-19 incidence exhibited positive relationships with male, poverty, black, diabetes, obesity and urban/rural dummies poverty, diabetes, and obesity, it was negatively associated with population density (Table 3). The OLS estimates accounted for 31 % of the regression plain (adjusted $R^2 = 0.31$), this implies low R^2 . Although the OLS estimates presented a very low adjusted R^2 , it provides baseline for SLM and SEM. The interpretation to this is that almost 69% of the COVID-19 incidence across the contiguous US are caused by unknown variables to the model and likely due to the local variations which were not captured by the OLS models. The AIC value indicated 9263 while the Breusch-Pang test for heteroskedasticity indicated 472.4 and multicollinearity condition number of 74.25 This implies that selected regressor in the model are

slightly correlated with the error term and demonstrated slight heteroskedastic nature.

Table 3. OLS summary statistics for the COVID-19 incidence predictor variables over contiguous US from January 21 to September 16 2020

Variable	Coefficient	t-Statistic	Std. Error	P-Value
Constant	-0.285167	-0.526073	0.542067	0.59893
Pop. Density	-6.1886	-4.73397	1.30729e-005	0.00000
Male	0.051927	5.08593	0.0102099	0.00000
Black	0.0448007	22.7337	0.00197067	0.00000
Poverty	0.0298329	6.58369	0.00453133	0.00000
Diabetes	0.0271132	3.49449	0.00775885	0.00048
Obesity	0.00608548	1.20764	0.00503915	0.22728
Ur-dummy	0.350719	7.04554	0.0497788	0.00000

The SLM and SEM presented slightly improved adjusted R^2 values compared to OLS. This improvement was credited to the incorporation of spatial dependence into the regression analysis of the relationship between dependent variable and the predictor variables. The Rho (ρ) and Lambda (λ) were very significant for SLM and SEM respectively with $\alpha = 0.000$. The adjusted R^2 value computed for SLM (0.29) is slightly higher than that of SEM (0.28). Nevertheless, lower value of standard error was recorded for SLM.

Table 4: SLM and SEM model statistical summary for COVID-19 mortality over contiguous USA

Variable	Coefficient		Z-value		P-value	
	SLM	SEM	SLM	SEM	SLM	SEM
Constant	-2.1274	0.0949	4.88535	-0.212748	0.00000	0.64493
Population density	-3.7217	-6.6536	3.53566	-4.88053	0.00000	0.00000
Male	0.051974	0.051135	6.36422	6.09264	0.00041	0.00000
Black	0.018222	0.028350	10.4907	10.6954	0.00000	0.00000
Poverty	0.026596	0.039904	7.27393	8.99422	0.00000	0.00000
Diabetes	0.002817	0.003109	0.452616	0.492253	0.00000	0.62254
Obesity	0.007956	0.010299	1.97395	2.36401	0.65083	0.01808
Urban dummy	0.203761	0.203913	5.09149	3.28577	0.04839	0.00102
Rho	0.62		0.0177406		0.00000	
Lambda		0.67		0.0179011		0.00000

Generally, the SLM and SEM models accounted for substantial percentages of the regression plain. While the efficiency of the models is the order of SLM (AIC: 8264.4; Breucsh-Pagan: 584.4; Adj. $R^2 = 0.56$) > SEM (AIC: 8282.0; Breucsh-Pagan: 697.2; Adj. $R^2 = 0.56$). In this case, the least predictive model is SEM.

Table 5. Goodness of fit for the OLS, SLM and SEM regression models

Criterion	OLS	SLM	SEM
Adj. R^2	0.31	0.56	0.56
AICc	9263	8264.4	8282.5
Log Likelihood ratio test	-	1000.54	980.4703
Breucsh-Pagan test	472.47	584.43	697.23
Jarque-Bera	863.60	-	-
Moran'I error	0.4225	-	-
Lagrange Multiplier (lag)	1351.12	-	-
Robust LM (Lag)	67.7752	-	-
Lagrange Multiplier (error)	1378.22	-	-
Robust LM (error)	94.87	-	-
Lagrange Multiplier (SARMA)	1445.99	-	-
Number of Observations	3,106	3,106	3,106

5.1 Regression analysis

Based on the model evaluation given above, we adopted SLM. The results of the SLM regression analysis was presented in Table 5. Results showed that Male (0.05), Black (0.01) and Poverty (0.02), Diabetes, (0.002), Obesity (0.007) and urban and rural dummy variable (0.02) exhibited positive and significant relationship with COVID-19 incidence. On the other hand, population density (-3.721) demonstrated negative relationship with COVID-19 incidence across the study area in the first 604 days of the pandemic. The three positively statistically significant variables in the SLM were also mapped to show the spatial variation across US counties between January 21 and September 15 2020 (See Figure, 4,5,6).

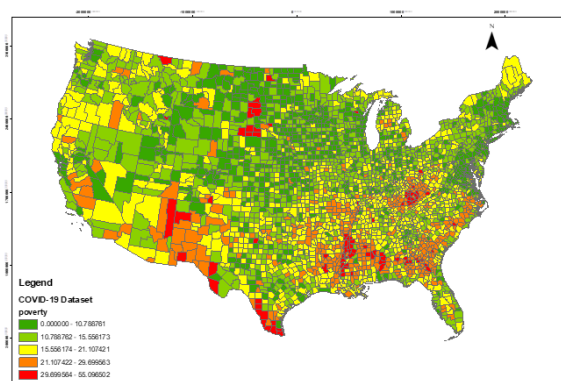


Figure 4: Spatial distribution of poverty at the county level

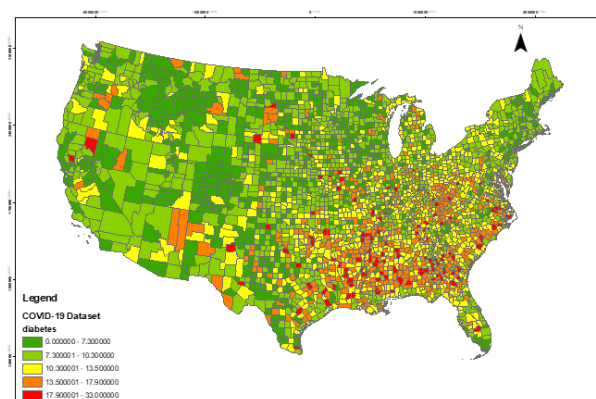


Figure 5: Spatial distribution of Diabetes at the county level

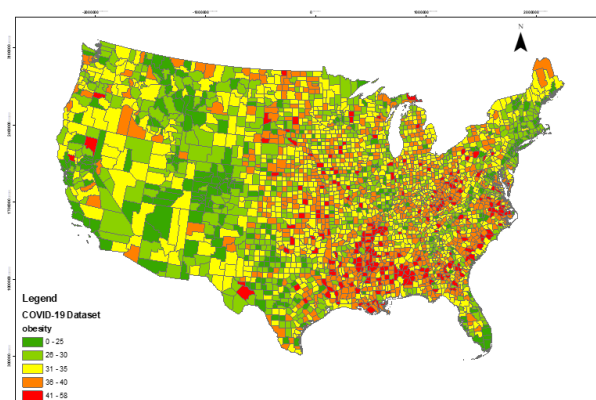


Figure 6: Spatial distribution of Obesity rate at the county level

Discussion and Results

In this study, I attempted to examine the influence of socio-economic and health outcomes variables on COVID-19 incidence across the counties of Contiguous USA. To achieve this, I selected 7 predictor variables (population density, male, black, poverty, diabetes, obesity and urban/rural dummies), I subjected the variables to stepwise and GIS-based exploratory and GeoDa-based regression analysis. Thus, we adopted spatial autoregressive models to establish an explanatory relationship between the dependent and predictor variables. The results provide an

insight to the goodness fit model among the two models that I considered. The study found SLM to be a better and most preferred model above SEM due to low AIC value and Breusch-Pagan test when compared to SEM. The map of the squared Spatial Lag Residual indicated clustering around the centre of the map stretching north and south (Figure 2&3). This pattern reveals an evident cross-border spatial autocorrelation of the examined socio-economic and health variables.

The significant contribution of male, black, race, poverty and urban dummies to the regression plain indicated that COVID-19 transmission is more of a function of socio-economic and rural/ urban conditions rather than health outcomes. Although, diabetes and obesity showed positive relationship with COVID-19 incidence but the relationships was very low. However, available data showed that socio-economic and health conditions are also potent determinant of COVID-19 incidence across the counties of contiguous USA. In particular, the vulnerable and poor populations (i.e., black race, Hispanic, among others) had recorded more COVID-19 incidences in USA (Sun et al., 2020; Saffray et al., 2020).

The results of the regression analysis showed that the prominent influential factors of COVID-19 incidence varied significantly across the counties of USA and this underscores the importance of spatial context in modelling of the outbreak of infectious diseases in space and time. All the outputs of this study revealed an interesting pattern that reflects cross-border spatial autocorrelation of events. Although, researchers had attributed the impacts of COVID-19 to the socio-economic disadvantages and inequalities arising from the pandemic itself (Ahmed et al. 2020; Mollalo et al. 2020). I observed that the severity of the impacts of COVID-19 is a function of cumulative poor socio-economic, rural/ urban interactions and health conditions of the people. For instance, poverty ridden persons may be prone to underlying health conditions such as diabetes, obesity and upper respiratory tract infections these may be as a result of eating habits or inappropriate access to health care. These underlying health conditions in turn expose the sick individuals to complications when infected with COVID-19. Results of correlation analysis show that poor persons and are likely susceptible to the disease conditions. The study also showed that population density had negative effects on the COVID-19 incidence, though our study did not consider the quantity and quality of health-care providers as explanatory variables, the evaluation of the influence of such variables on COVID-19 incidence would be highly revealing. In fact, Buerhaus et al. (2020) and Mollalo et al. (2020) had earlier highlighted the influence of quantity and quality of frontline health workers on COVID-19 incidence.

In the same vein, researchers have earlier speculated that demographic characteristics could be influencing factors on COVID-19 incidence (Bayne et al. 2020; Mollalo et al. 2020). The major challenge of this study has to do with the limitation inherent in the details of the available data. Mollalo et al. (2020) had earlier emphasized the setback

associated with the coarse spatial granularity of the available COVID-19 data. Furthermore, we observed that more influencing factors could be identified, if the data were directly linked to the infected persons. The availability of such data would pave the way for the objective evaluation of the influence of some variables (such as underlying health conditions: obesity, diabetes, upper respiratory tract infections, migration patterns) on COVID-19 incidence. Gupta et al. (2020), Zheng et al. (2020) and Mollalo et al. (2020) had earlier identified the above underlying health conditions as potential factors that have the capacity to aggravate COVID-19 case rate.

Perhaps the most influencing but the most difficult to capture is the influence of behavioral factors on COVID-19 incidence. For instance, COVID-19 occurrence could be influenced by the willingness of the people to comply with rules and regulations regarding COVID-19 pandemic. Also, addiction to certain behaviors or lifestyles could expose some individuals or group of people to infections. Also, Mollalo et al. (2020) highlighted the possible influence of the dichotomy in enforcing COVID-19 guidelines among the states. Therefore, there are still more to learn about the factors influencing COVID-19 pandemic.

Conclusions

The study revealed the footprints of COVID-19 incidences across counties in contiguous United States between January 21 and September 16 2020. This study will help to mitigate the diffusion and the severity of the disease as well as creating early warning surveillances where attention should be focused. This study has shown statistically significant positive dependency in the COVID-19 incidence across counties in the United States. The study further confirms previous findings (Saffray et al., 2020; Maroko et al., 2020) that racial minorities, poverty, and migration pattern between rural and urban locations could explain COVID-19 incidences across US counties. The set of individuals are at high risk of severe COVID-19 infections and deaths which could be explained with poor living standards and poor access to healthcare facilities. This study further concludes that the policy makers should take into cognizance spatial peculiarities, rural-urban migration and access to resources in the transmission of COVID-19 disease in the United States.

Acknowledgements

We are grateful to the One Health Department, Emory University, Atlanta, United States, for providing the COVID-19 data used in this paper.

References

- Abdollahi, E., Champredon, D., Langley, J. M., Galvani, A. P., & Moghadas, S. M. (2020). Temporal estimates of case-fatality rate for COVID-19 outbreaks in Canada and the United States.

- CMAJ*, 192(25), E666-E670.
doi:10.1503/cmaj.200711
- Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., . . . Xia, L. (2020). Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*, 296(2), E32-E40.
doi:10.1148/radiol.2020200642
- Berrigan and Berger, (2019): Geospatial Approaches to Energy Balance and Breast Cancer. Springer nature Switzerland AG. ISSN 2199-2630. <http://doi.org/10.1007/978-3-030-18408-7>
- Bray, I., Gibson, A., & White, J. (2020). Coronavirus disease 2019 mortality: a multivariate ecological analysis in relation to ethnicity, population density, obesity, deprivation and pollution. *Public Health*, 185, 261-263.
doi:10.1016/j.puhe.2020.06.056
- Carico R., Jordan Shepperd and Borden Thomas: Community pharmacists and communication in the time of COVID-19: Applying the health belief model: Research in Social and Administrative Pharmacy. <https://doi.org/10.1016/j.sapharm.2020.03.017>
- Costa F.,(2020):Health belief model for coronavirus infection risk determinantss. *Rev Saude Publica*. 2020;54:47
- Abdollahi, E., Champredon, D., Langley, J. M., Galvani, A. P., & Moghadas, S. M. (2020). Temporal estimates of case-fatality rate for COVID-19 outbreaks in Canada and the United States. *CMAJ*, 192(25), E666-E670.
doi:10.1503/cmaj.200711
- Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., . . . Xia, L. (2020). Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*, 296(2), E32-E40.
doi:10.1148/radiol.2020200642
- Berrigan and Berger, (2019): Geospatial Approaches to Energy Balance and Breast Cancer. Springer nature Switzerland AG. ISSN 2199-2630. <http://doi.org/10.1007/978-3-030-18408-7>
- Bray, I., Gibson, A., & White, J. (2020). Coronavirus disease 2019 mortality: a multivariate ecological analysis in relation to ethnicity, population density, obesity, deprivation and pollution. *Public Health*, 185, 261-263.
doi:10.1016/j.puhe.2020.06.056
- Carico R., Jordan Shepperd and Borden Thomas: Community pharmacists and communication in the time of COVID-19: Applying the health belief model: Research in Social and Administrative Pharmacy. <https://doi.org/10.1016/j.sapharm.2020.03.017>
- Costa F.,(2020):Health belief model for coronavirus infection risk determinantss. *Rev Saude Publica*. 2020;54:47
- Delamater, P. L. et al., 2019. Complexity of the Basic Reproduction Number (R0). *Emerging infectious diseases*, 25(1), pp. 1-4.
- Desjardin et al., (2020): Rapid Surveillances of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters *Journal of Applied Geography*.
<http://doi.org/10.1016/j.apgeog.2020.102202>
- Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F., & Billa, L. (2020). Spatial analysis and GIS in the study of COVID-19. A review. *Sci Total Environ*, 739, 140033.
doi:10.1016/j.scitotenv.2020.140033
- Gayawan, E., Awe, O. O., Oseni, B. M., Uzochukwu, I. C., Adekunle, A., Samuel, G., . . . Adegboye, O. A. (2020). The spatio-temporal epidemic dynamics of COVID-19 outbreak in Africa. *Epidemiol Infect*, 148, e212.
doi:10.1017/S0950268820001983
- Golestaneh, L., Neugarten, J., Fisher, M., Billett, H. H., Gil, M. R., Johns, T., . . . Bellin, E. (2020). The association of race and COVID-19 mortality. *EClinicalMedicine*, 25, 100455.
doi:10.1016/j.eclinm.2020.100455
- Hsing, J. C, Ma, J,Alejandra Barrero, Shilpa J.,Pulendran,P Bea Lin, Monika Thomas-Urbe and Jason Wang(2021): Influence of Health Belief on Adherence to COVID-19 Preventative Practices: International, Social media-based Survey Study.Journal of Medical Internet Research(J Med Internet Res 2021;23(2):e23720) doi: 10.2196/23720
- Janz, Champion, Strecher.,(2002):Health belief model in health care setting: Knowledge, perceived effectiveness, and cues to action on staff behaviors, pp 4 of 38,all academic research document
- Karaye, I. M., & Horney, J. A. (2020). The Impact of Social Vulnerability on COVID-19 in the U.S.: An Analysis of Spatially Varying Relationships. *Am J Prev Med*, 59(3), 317-325.
doi:10.1016/j.amepre.2020.06.006
- Liu, Y., Ning, Z., Chen, Y., Guo, M., Liu, Y., Gali, N. K., . . . Lan, K. (2020). Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature*, 582(7813), 557-560. doi:10.1038/s41586-020-2271-3
- Michelozzi, P., de'Donato, F., Scortichini, M., De Sario, M., Nocchioli, F., Rossi, P., & Davoli, M. (2020). Mortality impacts of the coronavirus disease (COVID-19) outbreak by sex and age: rapid mortality surveillance system, Italy, 1 February to 18 April 2020. *Euro Surveill*, 25(19).
doi:10.2807/1560-7917.ES.2020.25.19.2000620
- May J.M. (1958) The Ecology of Human Disease M.D. Publications Inc. New York (U.S.A.) pp. 1-2
- Meades, M.; Florin, J. and Wesler, W. (1988). Lung Cancer 1978-1981 in the black peoples of South Africa. *Brit. J. Cancer*, 52 pp. 339-346.

- McGlashan, N.D. (1982). Environmental detectives proberiddle of cancer. *Geographical Magazine* 5 (5). pp. 263-267.
- McGlashan, N.D. (1985). Risk factors in Gastro intestinal cancer. *Geographical Epidemiology*. Vol. 29 No. 1-2 pp. 7-11.
- McGlashan, N.D. and Harington, J.S. (1985). Lung Cancer 1978-1981 in the black peoples of South Africa. *Brit. J. Cancer*, 52 pp. 339-346.
- Meades, M.; Florin, J. and Wesler, W. (1988). Lung Cancer 1978-1981 in the black peoples of South Africa. *Brit. J. Cancer*, 52 pp. 339-346
- Omran, A.R. (1971). The Epidemiological Transition: a theory of the epidemiology of population change. *Milbank Memorial Fund quarterly* 49(4), pp. 509-538.
- Phillips, D.R. (1993) Health and Development: Key issues *Geographia Medica* Vol. 23, No. 2 pp. 132-133
- Oluwafemi, O. A., Babatimehin, O. I., Oluwadare, T. S., & Mahmud, U. M. (2013). Mapping Malaria Case Event and Factors of Vulnerability to Malaria in Ile-Ife, Southwestern Nigeria: Using GIS. *Ethiopian Journal of Environmental Studies and Management*, 6(4). doi:10.4314/ejesm.v6i4.4
- Petroni, M., Hill, D., Younes, L., Barkman, L., Howard, S., Howell, I. B., . . . Collins, M. B. (2020). Hazardous air pollutant exposure as a contributing factor to COVID-19 mortality in the United States. *Environmental Research Letters*, 15(9). doi:10.1088/1748-9326/abaf86
- Shahnazi Hosseini, Mary Ahmad-Livani, Bagher Pahlavanzadeh, Abdolhalim Rajabi, ddurrahman Mohammad Shoaibi and Abdurraham Charkazi (2021): Assesing preventive health behaviours from COVID-19: A cross sectional study with health belief model in Golestan Province, Iran. *Infect Dis Poverty* 9, 157 (2020). <https://doi.org/10.1186/s40249-020-00776-2>
- Shannon, G.W. and Spurlock, C.W. (1975). Urban Ecological Containers, Environmental Risk cell and the use of Medical Services. *Econ. Geography* Vol. 10 pp. 347 - 365.
- Sobral, M. F. F., Duarte, G. B., da Penha Sobral, A. I. G., Marinho, M. L. M., & de Souza Melo, A. (2020). Association between climate variables and global transmission of SARS-CoV-2. *Sci Total Environ*, 729, 138997. doi:10.1016/j.scitotenv.2020.138997
- Sun, F., Matthews, S. A., Yang, T. C., & Hu, M. H. (2020). A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: where geography matters? *Ann Epidemiol*. doi:10.1016/j.annepidem.2020.07.014
- Sun, Q., Qiu, H., Huang, M., & Yang, Y. (2020). Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province. *Ann Intensive Care*, 10(1), 33. doi:10.1186/s13613-020-00650-2
- Tosepu, R., Gunawan, J., Effendy, D. S., Ahmad, O. A. I., Lestari, H., Bahar, H., & Asfian, P. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Sci Total Environ*, 725, 138436. doi:10.1016/j.scitotenv.2020.138436
- Verhasselt, Y. (1977). Notes on Geography and Cancer *Soc. Sei. med.* Vol. 11 pp. 745-749.
- Verhasselt, Y. (1992) Problem on cancer epidemiology *Geographia medica*, Suppl. 8
- Xie, J., Zu, Y., Alkhatib, A., Pham, T. T., Gill, F., Jang, A., . . . Denson, J. L. (2020). Metabolic Syndrome and COVID-19 Mortality Among Adult Black Patients in New Orleans. *Diabetes Care*. doi:10.2337/dc20-1714
- Xu, Y., & Wang, L. (2014). GIS-based analysis of obesity and the built environment in the US. *Cartography and Geographic Information Science*, 42(1), 9-21. doi:10.1080/15230406.2014.965748
- Zhang, R. et al., 2020. Identifying airborne transmission as the dominant route for the spread of COVID-19. 117(26), pp. 14857-14863.
- Zhao, J., Yuan, Q., Wang, H., Liu, W., Liao, X., Su, Y., . . . Zhang, Z. (2020). doi:10.1101/2020.03.02.20030189