Determinants of Pneumonia Mortality: A Spatial Autoregressive Model Analysis.

GeoMundus 2019

David Enrique Payares Garcia[1]

1 Universitat Jaume I, Universidad Distrital Francisco José de Caldas
Vicent Sos Baynat, Castellón de la Plana, España
depayaresg@udistrital.edu.co; al394061@uji.es

Abstract

Most epidemiological studies of diseases show that the behavior of socioeconomic and demographic variables is strongly related to mortality rates. However, these studies rarely consider spatial autocorrelation in the data. This work presents a series of Spatial Autoregressive Models (SAR) as a technique that allows modeling a disease such as pneumonia based on socioeconomic, demographic, environmental, and health care factors in Bogotá (Colombia) from 2004 to 2014. The objective is to bring to light the potential of these methods and improve standard estimation methods like OLS, which indirectly, can propagate uncertainty in epidemiological models.

Keywords: Pneumonia, spatial econometrics, autoregressive spatial model, socioeconomic and demographic data, spatial modelling.

INTRODUCTION

Bogotá, the capital and largest city of Colombia by population constantly fights against easily transmitted and endemic-epidemic diseases that lead to enormous public health problems. One of the groups of diseases that presents relevant concern in the district is the respiratory diseases whose cyclical behavior in their dynamics and its close relationship with the socioeconomic and environmental city factor threaten the well-being of the population and result in alarming mortality rates. Pneumonia is currently the leading cause of mortality of respiratory diseases in Bogota and its recurrence has been partially explained by biological or behavioral factors, which has not allowed mitigating its incidence in Bogota's population. The present study identifies a series of environmental, socioeconomic, behavioral and health determinants [1] whose interaction through space has an impact on the mortality rate from pneumonia in the counties of Bogotá in a discrete period from 2004 to 2014.

1 METHODS AND DATA

1.1 DATA

The Bogota Health Secretariat provided the pneumonia database accounting the number of deaths caused by pneumonia; county, gender and age segregated the information. We calculate the standardize mortality ratio to compare the mortality risk of the study population to that of a standard population.

Table 1. Candidate in dependent variables

Type	Variable	Description	Source
Enviromental	TEM	Average Annual Temperature	IDEAM
	PM10	Average Annual Particulate Matter 10	RMCAB
Social and Economic	UER	Unemployment Rate	ECV, EMB, DANE
	HQI	Housing Quality Index	ECV, EMB, DANE
	LQI	Life Quality Index	ECV, EMB, DANE
	NBI	% poor by NBI Index	ECV, EMB, DANE
	SAT	% School Attendance	ECV, EMB, DANE
	CPW	% Coverage of Pipe Water	EAB
Behavioural	SPI	Sporting Index	SISCRED
	NUT	% Families where at least one member does not eat one of the three daily meals.	ECV, EMB, DANE
Health Care	IPS	Hospital Density by County	REPS
	VAC	% Reporting Influenza Vaccination	SDS

IDEAM: Institute of Hydrology, Meteorology and Environmental Studies; **RMCAB:** Network Monitoring Air Quality Bogotá; **ECV:** National Census of the Quality of Life; **EMB:** Living Standards Measurement Study Bogotá; **SISCRED:** Secretariat of Culture, Recreation and Sports; **REPS:** Special Registry of Health Services Providers; **SDS:** Bogota Health Secretariat

We made the selection of explanatory variables base on previous social epidemiology studies. While biological factors play an outstanding role in determining pneumonia mortality, we decided that the focus in this work would be on the less studied socioeconomic, demographic, behavioral and health care factors and their spatial interaction. Explanatory variables examined in this study are described in Table 1.

1.2 ANALYSIS

All data manipulation and statistical analyses were done using R (v.3.2), SPSS (v.11.0.1) and Geoda. We conduct an analysis using Moran's I statistic to assess the degree of spatial autocorrelation in the independent variable. Significant spatial autocorrelation indicates how much a value at a given location depends on, and is similar to, a value at another location (spatial neighbors). The Neighbor relationships are expressed in a row standardized spatial weights matrix whose elements represent the binary spatial relationship between one location and its neighbors. For this analysis, neighbors were defined using k-nearest neighbor, queen and Delaunay triangulation case adjacency, which considers all counties with common borders as neighbors. Initial variable assessing included calculating correlation coefficients, and checking for linearity in the expected relationships, between dependent and independent variables. In cases where there was a non-linear relationship between the response and an explanatory variable, we applied a log transformation. For each outcome, we include the independent variables selected by an ordinary least squares (OLS) regression. We used a backward stepwise procedure with a significance level of $p = 0.1$ to retain explanatory for the final model. In cases where a significant correlation between explanatory variables was detected, we entered each into the model to assess their relative contributions. We compared the fit of alternative models.

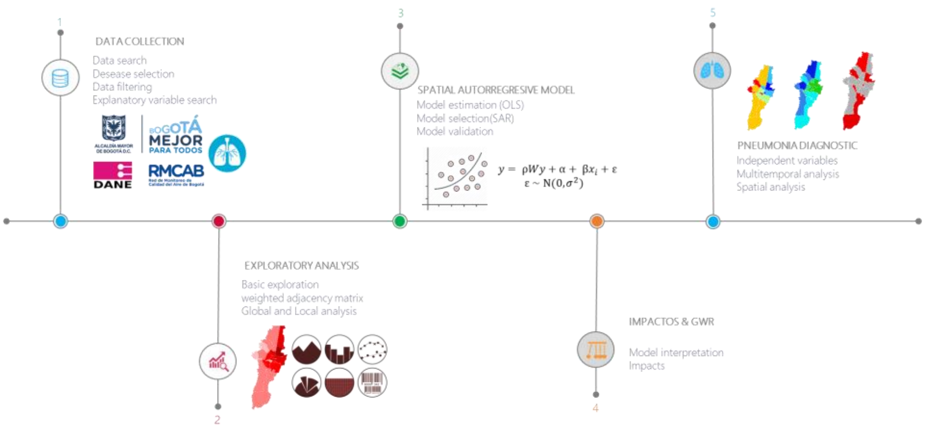

Figure 1. Methods and Data

Formal diagnostic tests for multi-collinearity and heteroscedasticity were conducted. Residuals were also tested for normality as well as spatial dependence using the Moran's I test. We conducted a further analysis using a local indicator of spatial autocorrelation or LISA [3] to assess the degree of localized clustering of model residuals. We considered two alternative models (spatial lag and spatial error models) that incorporate spatial dependence because the spatial dependence in the model residuals represented a violation of OLS assumptions. We found, based on the results of Lagrange Multiplier tests [2] ,that spatial autoregressive, autoregressive error and nested models are the appropriate alternative Models, using the same explanatory variables as in the OLS regression, were fit using maximum likelihood estimation. We performed formal diagnostics tests to assess the suitability of these models. R2 and log-likelihood compare the fit of the models [4]. To interpret the influence of significant explanatory variables we calculated the direct and indirect impacts.

2 RESULTS

There were 30,253 pneumonia deaths in Bogotá over the study period, representing an overall rate of 30/100,000 population. Rates were higher where socioeconomic and health care variables were lowest. The SMR raised in the years, in which, the school attendance decreased and the number of hospital was under the mean value. The Figure 2 indicates that there is a high degree of spatial similarity in the data (clusters are evident), which justify the use of spatial autocorrelation test like Moran's I test [5].

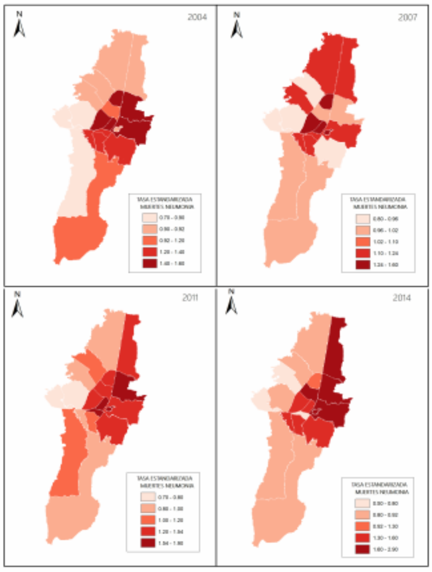

Figure 2. This is a sample figure legend. This is a sample figure legend.

The Figure 3 explain how are the relationships between the counties in Bogotá. In 2004, a county neighbor is considered if shares a boundary with an adjacent county. For 2007 and 2014, the closest county based on the distance between the two counties centroids. In 2011, the neighbors are those, which are connected by triangles [6]. These relationships will allow integrating the spatial disturbance and interactions to the regressive models [7].

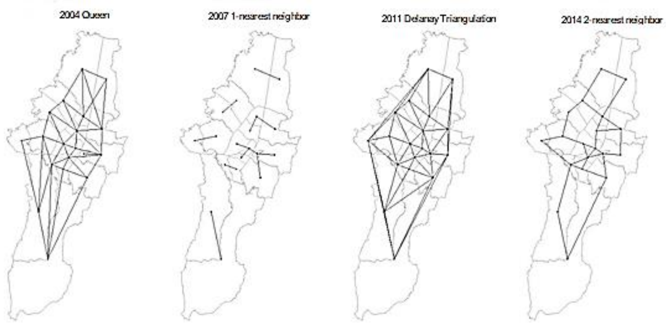

Figure 3. Spatial Adjacency relationships

For 2004, the SARAR (spatial autoregressive model with autoregressive disturbances) [8] model was determined to be appropriate. The model explained a substantial portion of the variation in the SMR for pneumonia (R2= 0.79). SMRs were found to increase with higher Temperatures and low school attendance percentage. The regression coefficient indicates that every 1% decrease in the SAT percentage variable from the mean is associated with an increase of the SRM by 0.07%. For every 1°C increase in TEM, the SMR increases by 0.08 °C. For 2007 and 2014, the GNS (General nesting) model was the best option. The model has high explanatory power, accounting for approximately 92% and 93% of the variation in SMRs in 2007 and 2014 respectively. Low values on variables like school attendance percentage, hospital density, and influenza vaccination increase the SRM. The SAR (Spatial autoregressive error) model is the selection for 2011. The model explained a substantial portion of the variation in the SMR for pneumonia (R2= 0.79) and shows that variables like Life quality index and influenza vaccination have a negative incidence on the SRM for pneumonia.

Table 1. Regression coefficients and standard errors for the total and gender specific pneumonia and influenza models (2004–2014)

Year	2004		2007		2011		2014	
Model	SARAR		GNS		SAR		GNS	
Adjusted-R^2	0.7915		0.9247		0.7919		0.9368	
Variables	Estimated	p-value	Estimated	p-value	Estimated	p-value	Estimated	p-value
ρ	0.82	2.2e-11	-0.90	2.2e-16	0.84	0.07	0.61	1.73e-05
λ	0.93	7.88e-05	0.95	2.2e-16	...	...	-1.10	1.64e-13
Intercept	2.48	4.27e-05	28.74	2.2e-16	-4.22	0.01	8.63	8.26e-03
SAT	-0.07	3.66e-07	-4.37	2.2e-16	...	...	1.19e-01	9.61e-06
TEM	0.08	8.59e-05	...	...	...	...	...	...
SPI	...	...	0.71	2.2e-16	...	...	...	...
IPS	...	...	-0.09	2.2e-16	...	...	6.33e-04	2.2e-16
VAC	...	...	-0.11	0.03	-0.02	0.01	-1.68e-02	0.03
IQI	...	...	...	...	-0.02	0.03	...	...
HQI	...	...	...	...	...	...	-6.5e-01	1.11e-15

The results show that the spatial interaction of environmental factors as temperature, socioeconomic factors such as life quality, school attendance and housing quality, behavioral factors like sporting index and health care factors such as the number of IPS and Influenza vaccination are significant in the regressive models and explain in a certain proportion the morality rate for pneumonia.

3 FUTURE WORK

This research demonstrates how spatial analytic techniques can be applied in studying pneumonia, and more generally, infectious disease. It is evident how spatial autoregressive models, through independent variables, can explain the large proportion of geographic variation of a particular disease. Broad use of spatial modeling techniques should be encouraged. Besides, a better understanding of the geographic variability in pneumonia mortality in Bogotá, and the importance of socioeconomic, behavioral and healthcare factors associated with this outcome; will support the development of more regional and population-specific social, economic, public health and health care programs.

The significance of these findings can be extended to another infectious diseases like bronchitis, HIV/AIDS, cancer, and more specifically, chronic respiratory diseases, where similar factors play a role in determining disease exposure and susceptibility.

4 CONCLUSION

The joining of epidemiological data and the spatial statistical process can help us to understand better disease behavior. Every relationship and geographic variations in disease morbidity and mortality can be explained by the social, environmental and health care factors, allowing us to predict its spread through space and time. The spatial- epidemiological analysis will be performed more and more often due to its appealing visualization, the increasing research in geospatial modeling and technologies and the man’s desire to live (or be located) in healthier environments.

References

[1] Ancot, L., Paelinck, J., Klaassen, L., Molle, W., Albegov, M., Andersson, A. & Snickars, F. (1982), 'Topicsin regional development modelling'(eds, pp. 341-359),Regional Development Modelling in Theory and Practice. Amsterdam: North Holland.

[2] Anselin, L. (1988a), 'Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity',Geographical Analysis20(1), 1–17.

[3] Anselin, L. (1988b), Spatial econometrics: Methods and models, 1st edn, Springer Netherlands., BogotáD.C., Colombia.

[4] Bohorquez, M. (2009), 'Estadística espacial',Universidad Nacional de Colombia, Departamento de Estadística.

[5] Chasco Yrigoyen, C. (2003), Econometría espacial aplicada a la predicción-extrapolación de datos micro-territoriales, 1ra edn, Consejería de Economía e Innovación Tecnológica, Madrid, España.XXVIII Simposio de Estadística (2018)

[6] Dray, S., Legendre, P. & Peres-Neto, P. (2006), 'Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (pcnm)',Ecological Modelling196, 483–493.

[7] Florax, R. & Folmer, H. (1992), 'Specification and estimation of spatial linear regression models: Montecarlo evaluation of pre-test estimators',Regional Science and Urban Economics22, 404–432.

[8] Hyeyoung, L. & Sujit K., G. (2009), 'Performance of information criteria for spatial models',J StatComput Simul79(1), 93–106.Moreno, R. & Vayá, E. (2000), Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial, ub 44 manuelas edn, Edicions Universitat Barcelona, Barcelona, España.