

Case study

Determinants of Pneumonia mortality in Bogota, Colombia: A spatial econometrics approach

David Payares-Garcia^{a,*}, Bibiana Quintero-Alonso^b, Carlos Eduardo Melo Martinez^b^a ITC Faculty Geo-Information Science and Earth Observation, University of Twente, Hengelosestraat 99, 7514 AE Enschede, Netherlands^b Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Cra. 7 N° 40b-53, Bogotá, Colombia

ARTICLE INFO

Keywords:

Pneumonia
Respiratory diseases
Spatial autocorrelation
Spatial autoregressive models
Contextual factors

ABSTRACT

Bogota, the capital and largest city of Colombia, constantly fights against easily transmitted and endemic-epidemic diseases that lead to enormous public health problems. Pneumonia is currently the leading cause of mortality attributable respiratory infections in the city. Its recurrence and impact have been partially explained by biological, medical, and behavioural factors. Against this background, this study investigates Pneumonia mortality rates in Bogota from 2004 and 2014. We identified a set of environmental, socioeconomic, behavioural, and medical care factors whose interaction in space could explain the occurrence and impact of the disease in the Iberoamerican city. We adopted a spatial autoregressive models framework to study the spatial dependence and heterogeneity of Pneumonia mortality rates associated with well-known risk factors. The results highlight the different types of spatial processes governing Pneumonia mortality. Furthermore, they demonstrate and quantify the driving factors that stimulate the spatial spread and clustering of mortality rates. Our study stresses the importance of spatial modelling of context-dependent diseases such as Pneumonia. Likewise, we emphasize the need to develop comprehensive public health policies that consider the space and contextual factors.

1. Introduction

Pneumonia is an acute respiratory infection that affects the lung parenchyma and oxygenation. It is characterized by the inflammation of the lung tissue caused by a bacterial infection (Cabre, 2009). According to the World Health Organization (WHO), Pneumonia is one of the leading causes of hospitalization and death among respiratory infections in developing countries (Zar et al., 2013). In Colombia, one out of five patients with Pneumonia dies without early treatment (Forero, 2013), and the figures worsen in the country's main cities where poverty rates, pollution levels, and health care access demand are alarming.

Bogota, the capital of Colombia, is one of the regions most affected by Pneumonia (Isturiz et al., 2010). Due to overcrowding, the growing number of manufacturing and automotive establishments, long working hours, sedentary lifestyle, and other factors that are typical in big cities such as Bogota, Pneumonia receives a particular welcome among its inhabitants (Demougeot et al., 2013; Hemilä et al., 2006; Almirall et al., 2015). As reported by The Bogota Health secretariat (BHS), Pneumonia accounts for 60% of all deaths from respiratory diseases, targeting mainly children under five years of age and elderly adults (Rosselli and Rueda, 2012; Benavides et al., 2012a).

Some studies have focused on associating biological (Taboada B. et al., 2015), medical (Benavides et al., 2012b) and behavioural (Rojas et al., 2016) predisposing factors with the burden of Pneumonia in Bogota. However, the associations found are patient-specific and cannot be extended to a population. Furthermore, while biological and medical factors affect Pneumonia outcomes, literature has established social, economic, and environmental variables as potential driving factors of the disease mortality (Keall et al., 2012; Jahanihashemi et al., 2018; Tasci et al., 2018; Wimalasena et al., 2021).

A class of analytic approaches to integrating population-level health data and risk factors is spatial statistic methods. Since infectious diseases, such as Pneumonia, are highly heterogeneous and volatile across space (Navas et al., 2019), spatial methods can help to explain the spatial variation in the diseases concerning their determinants. For example, a study conducted by Crighton et al. (2007) examined the spatial pattern of Pneumonia and influenza hospitalization in Ontario, Canada, finding spatial associations between the disease rates and local demographic and health care variables. Anon (2014), using Bayesian conditional autoregressive models in an ecologic-level framework, explained variations in Pneumonia hospitalization rates in the North of

* Corresponding author.

E-mail address: d.e.payaresgarcia@utwente.nl (D. Payares-Garcia).<https://doi.org/10.1016/j.sste.2023.100581>

Received 24 November 2021; Received in revised form 23 December 2022; Accepted 30 January 2023

Available online 6 February 2023

1877-5845/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

England through levels of material deprivation, childhood illness, and distance to hospitals.

Generally speaking, statistical methods for modelling spatial health data assume that the observed magnitude of an event (mortality/incidence cases) is the realization of a random variable with a specific probability distribution. As data typically comes as the number of cases, the Poisson or the binomial probability distributions are the most frequent choices, with the former reigning in the literature. Methods such as generalized linear models (GLM) and their extension, generalized linear mixed models (GLMM) (Elliott et al., 2001; Wakefield, 2004) are mainstream in spatial disease modelling. While these methods are flexible to guarantee the data distributional assumptions, they possess multiple drawbacks. GLM are easy to interpret but often fail to account for data intrinsic spatial autocorrelation (Beale et al., 2010). GLMM can successfully incorporate spatial dependence (as well as time and mixed effects) but are highly parametrized and imply complex Bayesian estimation of parameters (Cressie et al., 2009). The parameter estimation procedures, usually via Markov Chain Monte Carlo (MCMC) or integrated nested Laplace approximation (INLA), are computationally demanding and challenging to implement and understand for non-statisticians.

Spatial autoregressive (SAR) models are regression models that consider spatial dependency among spatial units (Anselin et al., 1996). These methods are convenient to model ecological data when a global process is assumed, and the spatial effects can be considered endogenous, exogenous or residuals (Anselin, 1988). Unlike GLM, SAR models account for spatial autocorrelation and use no link function to model the relationship between the dependent and independent variables. Furthermore, since SAR models fall under the umbrella of linear regression models, they are widely known and adopted by practitioners in many disciplines due to the natural interpretation of coefficients (Chi and Zhu, 2008). Other advantages of SAR models include quick estimation of coefficients (via maximum likelihood estimation), a representative parameter for the spillover effects, and minimal computational expenses.

The purpose of our study is to conduct an area-level ecological analysis that links Pneumonia mortality rates and potential risk factors in Bogota's districts for the years 2004, 2007, 2011, and 2014 using spatial autoregressive methods. We adopt a population-level analysis focused on the relevance of contextual physical and environmental conditions that modulate health and disease outcomes. In this work, statistical models are built by applying spatial autoregressive techniques to account for the spatial relationships between the disease mortality rates and socioeconomic, behavioural, environmental, and health care determinants. The results of our approach intend to encourage population and place-centred strategies to prevent Pneumonia incidence in Bogota's population. Furthermore, we seek to highlight the importance of the spatial dimension in epidemiological research.

2. Materials and methods

2.1. Data

This study conducted a retrospective spatial epidemiological analysis to estimate Pneumonia mortality and associations with its underlying risk factors in Bogota, Colombia. As a geographical unit of study, we used the urban administrative divisions of Bogota known as districts. In total, Bogota has 20 districts: 19 are urban and 1 is rural (Fig. 1). The urban districts are considered census divisions that act as main areas for ecological studies. Given the incidence of the demographic, socioeconomic, and environmental factors in disease causation, an area-level ecological analysis allows identifying health determining factors by extending individual risk characteristics to a population context (Levin, 2006; Wakefield and Lyons, 2010). According to the *Departamento Nacional de Estadística* (2015), in 2004, there were approximately 7 million urban residents in Bogota. This number increased by 11.4% in

2014 (7.8 million). The district populations vary from approximately 24,000 inhabitants in La Candelaria (city centre) to 1,174,000 in Suba (North). 48% of Bogota's population locates in districts (Ciudad Bolívar, Usme, Kennedy, and Suba) with limited access to public services, housing, education, employment, and healthcare (*Subsecretaría de información y estudios estratégicos*, 2016). In contrast, only 4% lives in relatively wealthy districts (Teusaquillo and Chapinero) (Guzman and Bocarejo, 2017).

2.1.1. Outcome variable

We obtained information on Pneumonia mortality from the Main Causes of Death Database provided by Bogota's Health Secretariat. The data consists of the number of deaths from Pneumonia as the principal diagnosis, aggregated by year, patients' district of residence, age, and sex. Pneumonia as a primary cause of death is defined by the Colombian List 105 of Mortality derived from the mortality classification of the Pan American Health Organization (PAHO). The database registers patients' death caused by Pneumonia in all hospitals and clinics of Bogota for ten years, starting from January 1, 2004, to December 31, 2014. In total, 10,656 deaths were analysed for the selected period and locations. The Standardized Mortality Ratio (SMR) was calculated since direct age adjustment is not feasible given that the number of deaths for some age groups is modest. Likewise, the SMR allows us to compare the study population's mortality rates to a standard one (in our case, the standard population of Bogota's districts) (Curtin and Klein, 1995). SMR represents the ratio between the directly observed and expected mortality based on the age and sex rates of both the standard and the study population for a district of Bogota. SMR values greater than 1 express an excess of deaths, i.e., the observed number of deaths is higher than the number of expected deaths. Values below 1 indicate that the observed mortality is lower than the expected one. An SMR of 1 occurs when the expected and observed number of deaths are equal.

2.1.2. Covariates

This paper aims to assess the role of social, economic, behavioural, environmental, and medical care variables as risk factors for Pneumonia mortality. Although biological and medical covariates are essential in determining Pneumonia mortality, they are usually patient-specific and, in most cases, not extendable to population-level analysis (Greenland, 2001). Furthermore, Pneumonia mortality studies involving biological and medical factors are rich in the literature. Our scope is focused on the insufficiently studied or ignored influencing factors in Pneumonia mortality. The independent variables examined in this study are presented in Table 1. Details about the covariates selection and sources can be found in the Supplementary Material document.

In Colombia, national, regional, and local surveys are conducted approximately every four years as a cost-saving measure. Most of the data sources used in the present study come from publicly accessible surveys, and thus, this research is bounded by the surveys' periodicity. Our study selected the years 2004, 2007, 2010, and 2014 considering the data available and that changes in the Pneumonia mortality rates are more apparent in spaced time intervals.

2.2. Analysis

The district-level analysis of Pneumonia risk factors includes 12 candidate variables that threaten Bogota's population's health. An initial analysis was performed to assess the degree of spatial autocorrelation in the dependent variable (SMR) for each year using Moran's Index and Geary's C statistics with a significance level of $\alpha < 0.05$. Both Moran's Index and Geary's C statistics measure the similarity and dependency between an observation at a given location and its neighbouring observations, being both statistics a global measurement. However, Geary's C is typically more sensitive to local spatial autocorrelation. To define the neighbourhood structure, i.e., the spatial relations that encode the variables' mutual influence, the spatial weights matrix **W** is used. **W** is



Fig. 1. Bogotá's districts, Colombia. The numbers represent the names of the districts. The district of Sumapaz is shown in the figure, but it is not considered in the present study.

a symmetric square matrix with element W_{ij} at location i and j . The values of W_{ij} represent the structure among locations, where $W_{ij} = 1$ for neighbours and $W_{ij} = 0$, otherwise. The characterization of W is not a straightforward task; on the contrary, the researcher must identify it based on their experience of the studied phenomenon. For this analysis, neighbourhood structures were defined using the Principal Coordinates of Neighbourhood Matrix (PCNM) approach (Borcard and Legendre, 2002), which detects the W configuration that maximizes Moran's I by performing a multiple regression between the outcome variable and the eigenvectors associated with the locations. The PCNM functions provide a reliable method for determining the most optimal spatial neighbourhood structure when limited knowledge about the variable's spatial relations.

We conducted a statistic and spatial exploratory data analysis to examine the relationships between the dependent and independent variables. In the statistical analysis, we used univariate analysis to determine the variables' probability distribution and descriptive statistics and multivariate analysis to compute correlation coefficients, covariance matrix, and linearity tests. Covariates that did not follow a Gaussian distribution or presented a non-linear relationship regarding

the outcome variable were log-transformed to approximate normality. The spatial analysis included cartographic visualization to identify spatial distributions, spatial outliers, spatial association patterns, and spatial heterogeneity. Quantile maps displayed clear evidence of spatial clustering, hotspots, and spatial patterns in the outcome variable. Results from the exploratory data analysis suggested the presence of strong spatial autocorrelation.

The spatial dependence in the outcome variable was scrutinized using the Moran's Index, the local indicator of spatial autocorrelation (Anselin, 1995), and the Local Getis-Ord G_i^* (Ord and Getis, 1995). These spatial diagnostics help to statistically determine spatial autocorrelation and identify the form of the spatial process in the residuals (Elhorst, 2010). We also computed the bivariate Moran's Index (Czaplewski et al., 1993) to find a spatial association between the SMR and the spatially lagged covariates. Spatially lagged covariates with a statistically significant degree of correlation with the SMR are initial potential candidates for the spatial autoregressive models.

Albeit the exploratory analysis insinuates the existence of spatial autocorrelation in our data, it does not inform about the level of influence of the independent variables or the underlying spatial dependence

Table 1
Selected socio-economic, behavioural, environmental and medical care factors.

Type	Variable	Description	Source
Environmental	Temperature	mean annual temperature (°C)	IDEAM
	Pollution PM10	Mean annual particle air pollution PM10	RMCA
Socio-economic	Unemployment	percentage of unemployed workers in the civilian labour force	ECV, EMB
	Housing quality	Score of housing conditions based on its physical and environmental characteristics	ECV, EMB
	Life quality	Score of living conditions based on socio-economic characteristics	ECV, EMB
	Poverty	Percentage of population living in poverty based on unmet basic needs criteria	ECV, EMB
	School attendance	Percentage of 5 years or older attending education institutions	ECV, EMB
	Water coverage	Percentage of water and sanitation coverage	EAB
Behavioural	Physical activity	percentage of 13 years of older who regularly exercise.	SISCRED
	Incomplete meals	Percentage of homes that eat only one of the three daily meals.	ECV, EMB
Medical care	Density of hospitals	Number of health facilities/100,000 population	REPS
	Influenza Vaccine	Percentage of 1 year or older vaccinated against Haemophilus influenzae type B	SDS

EAB: Empresa de Acueducto y Alcantarillado de Bogotá, EMB: Encuesta Multipropósito de Bogotá, ECV: Encuesta Nacional de Calidad de Vida, IDEAM: Instituto de Hidrología, Meteorología y Estudios Ambientales, RMCA: Red de Monitoreo de Calidad del Aire de Bogotá, SISCRED: Secretaría de Recreación, Cultura y Deporte, SDS: Secretaría Distrital de Salud.

structure. The former can be addressed by performing a classic linear regression, such as an ordinary least-squares (OLS) regression. The latter is by analysing the regression residuals in search of spatial dependence. We conducted a linear regression analysis to determine potential driving factors that might explain Pneumonia's SMR in Bogotá for each year. We assessed our covariates' contribution – in modelling the SMR – through a classic OLS regression, and two regularized variations, the Lasso and Elastic Net regressions (Zou and Hastie, 2005). The Lasso and Elastic Net regressions induce penalization terms to reduce the risk of overfitting, the variance, the correlation effect between variables, and the influence of irrelevant predictors derived from a standard OLS regression. We tested the three regression approaches by adopting a k -fold cross-validation. We selected the regression models whose covariates' linear combination showed the lowest Akaike information criterion (AIC). Subsequently, we applied formal diagnostic tests to inspect each model assumptions such as linearity, multicollinearity, and homoscedasticity. The explanatory variables selected in classic regression with its assumptions were considered candidates for a spatial regression model. Normality and spatial dependence in the residuals were also tested.

Autocorrelation in the model residuals, regardless of its nature, violates the OLS assumptions (Anselin, 1988). If spatial autocorrelation is identified, a model that accounts for spatial dependence has to be used instead. Spatial econometrics literature has developed models to incorporate spatial dependence in three different forms: (i) a spatially lagged dependent variable, (ii) spatially lagged independent variables, and (iii) a spatial structure in the error term (Elhorst, 2010). The simultaneous interaction of the three forms produces a spatial autoregressive model known as the General Nesting Spatial (GNS) Model. The GNS is expressed as,

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}_y + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\theta + \varepsilon \\ \varepsilon &= \lambda \mathbf{W}\varepsilon + \mathbf{u} \\ \mathbf{u} &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \quad (1)$$

Where \mathbf{y} represents a vector consisting of one observation on the dependent variable for every spatial unit, \mathbf{X} the matrix of independent variables, \mathbf{W} is the spatial weights matrix that describes the structure of dependence between units, \mathbf{W}_y denotes spatially lagged dependent variable, $\mathbf{W}\mathbf{X}$ the spatially lagged independent variable, and $\mathbf{W}\varepsilon$ the spatial interaction effects in the error term. The scalar parameters ρ

and λ measure the strength of dependence between units, while θ , like β , is a vector of response parameters. \mathbf{u} is a vector of independently and identically distributed disturbance terms with zero mean and variance σ^2 .

Other spatial autoregressive models can be obtained by restricting the GNS model spatial interactions, that is, omitting a form of spatial dependence by setting $\rho = 0$, $\lambda = 0$ and/or $\theta = 0$. For example, the spatial lag model is a particular specification in which $\rho \neq 0$; only the endogenous interaction – spatially lagged dependent variable – is considered (Eq. (2)) (Anselin, 1988).

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}_y + \mathbf{X}\beta + \varepsilon \\ \varepsilon &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \quad (2)$$

We built the seven possible spatial autoregressive models for each year derived from the GNS model using the explanatory variables preserved in the linear regression and those that showed significant spatial autocorrelation with the outcome variable (Bivariate Moran's I). We fitted the models using the maximum likelihood estimation except for the Spatial Lag of X (SLX) model. For further details on the models, see LeSage and Pace (2009), Elhorst (2010). We employed the Lagrange Multiplier tests (Anselin et al., 1996) to identify the appropriate spatial autoregressive model and its form or forms of spatial dependence. However, since the Anselin et al. (1996) test ignores models with exogenous spatial interaction (spatially lagged independent variables), we supported its results with the explanatory character (AIC) of each possible spatial model. In this context, we followed the model selection “combined” approach suggested by Elhorst (2010). After selecting the spatial autoregressive model for each year, we performed formal diagnostics to assess their suitability. We also computed the goodness of fit test such as R^2 and Nagelkerke pseudo R^2 to evaluate the fit of the models. We interpreted the model parameters, and marginal effects through spillover impacts and conventional linear model interpretation (LeSage and Pace, 2014).

2.3. Implementation

We implemented the analysis in the open source software R, using the integrated development environment (IDE) RStudio (RStudio Team, 2020). Most of the analysis and modelling were performed using functionalities from noted spatial statistical R packages. The exploratory

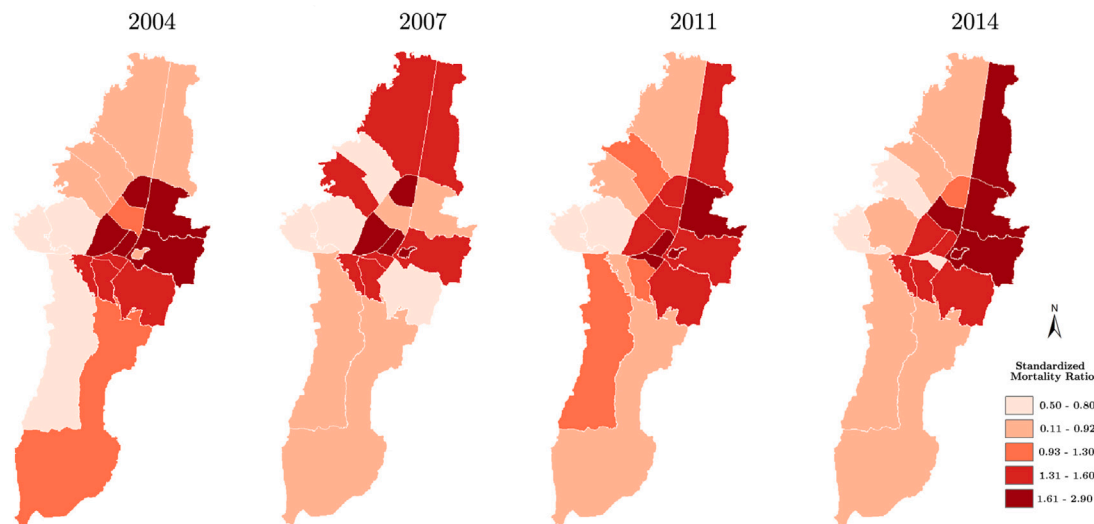


Fig. 2. SMR of Pneumonia by year in Bogota districts.

analysis, matrix selection and OLS regressions were performed using functions of the `spdep`, `adespatial` and `caret` packages as well as some customized functions we developed. The spatial autoregressive models were built and tested with the `spatialreg` and `lmtest` packages. The spatial data management was conducted using the package `sf`.

The spatial models were estimated via maximum likelihood (Anselin, 1988; LeSage and Pace, 2009). Estimating the model parameters is quick because the methods use numerical optimization. All models were estimated, and none suffered from convergence issues.

The analysis implementation and customized functions, including data exploration, model selection and model evaluation, can be found at <https://github.com/DavidPayares/Pneumonia-SAR>.

3. Results

The overall mortality rate is 137/100,000 population for the study period. In the four years analysed, the overall SMR is between 1.1 and 1.3, indicating an excess of deaths every year. At least 14 (70%) of the districts exceed an SMR of 1.0. 2011 and 2014 present the most alarming figures, reaching a maximum SMR of 2.0 and 2.9, respectively. Fig. 2 shows district-level SMR for each study year. District-wise, SMR rates range from 0.5 to 2.9. The most affected districts are located in Bogotá's centre Chapinero, Santa Fé, Los Martires and Puente Aranda. They display high SMRs with values greater than 1.6, above the city's average rate. Initially, Teusaquillo and Usaquen districts displayed low SMRs values in 2004 (0.7 and 0.9), but these increased markedly over the years until they reached high SMRs in 2014 (1.5 and 1.7). The southwestern districts of Bosa and Kennedy have the lowest SMR values in the four years. They never surpassed an SMR of 0.9. The SMR for the districts of Suba and La Candelaria fluctuates from low to high and vice-versa. However, by 2014 both districts exhibit high values. For other districts, the SMR behaviour is constant with rates around 1.1 and 1.3, the global average of the study period. Overall the rates have increased over the years, the lowest SMRs are predominately found in the southwestern districts, and high SMRs locate in the centre and southern districts of Bogotá. Across the years, the spatial pattern of the SMR is consistent except in 2007, when the highest values are located in the city's northern districts (Suba and Usaquén).

3.1. Spatial statistical analysis

We screened the neighbourhood structures of two main criteria: distance (k -nearest neighbours, Delaunay triangulation, Sphere of influence, and Gabriel graph) and contiguity (queen and root) since

Table 2
Spatial weights matrices.

Year	Matrix	Orders
2004	Queen	1, 3
2007	K-2 nearest neighbours	1
2011	K-4 nearest neighbours	1, 3
2014	Queen	1

we made no prior assumptions about the spatial dependence of the SMR in Bogotá's districts. The spatial weights matrix W is selected for each year using the PCNM approach. These matrices represent the neighbourhood structure that optimizes the Moran's Index of our outcome variable, the SMR. In order to find spatial autocorrelation structures of greater orders when the distance or spatial lags increase, we computed the spatial correlograms. The first-order spatial structure W_1 is significant ($p < 0.05$) for all years. Notably, in 2004 and 2011 also the third spatial lag W_3 is significant. These results indicate that the spatial autoregressive model for 2004 and 2011 must contain the outcome variable lagged for both W_1 and W_3 .

Table 2 shows the optimal spatial weights matrices for each year. The spatial weights matrices for 2004 and 2014 – Queen – have the greatest number of neighbours. The Queen criterion considers a neighbour as the region that shares at least one common boundary point with other neighbouring regions. For 2007 and 2011, neighbours are those connected by the shortest Euclidean distance of their centroids. The spatial weights matrix for 2007 only acknowledges the first two nearest neighbours ($k = 2$), while for 2011, the matrix contemplates the four first nearest ones ($k = 4$). Because of the long and narrow geography of the study area, the nearest neighbour spatial matrices K-2 in 2007 and K-4 in 2011 imply that the districts located in the inner area of Bogotá possess neighbours with shorter distances and more directions than the districts in the suburbs, particularly those in the northern and southern regions. For example, consider the location and shape of the Usme and Ciudad Bolívar districts (Fig. 1). Their nearest neighbours are northward and far from the Usme and Ciudad Bolívar centroids.

The Moran's I and Geary's C statistics revealed the presence of spatial autocorrelation in the years under study except in 2007 (Table 3). The Moran's I's positive values inform a spatial clustering process in the global configuration, i.e., districts with similar SMRs cluster together all over the region. Geary's C statistics also indicate positive spatial autocorrelation, meaning that smaller clusters occur in specific regions of Bogotá; for example, the southern districts display low rates while the central districts are characterized by high SMRs. Both Moran's I and

Table 3
Spatial autocorrelation statistics for SMR.

Year	Moran's I (<i>p</i> -value)	Geary's C (<i>p</i> -value)
2004	0.234 (0.0233)	0.269 (0.0068)
2007	0.221 (0.0617)	0.289 (0.3051)
2011	0.423 (0.0002)	0.248 (0.0039)
2014	0.199 (0.0304)	0.248 (0.0003)

H_0 : No spatial autocorrelation.

Geary's C demonstrate that our variable under study, the SMR, exhibits global and local spatial autocorrelation at the district level at least in three years, reinforcing the idea of using a spatial autoregressive model. Although the statistics did not find spatial autocorrelation in the outcome variable for 2007, we might find it in the linear regression residuals.

We selected initial possible covariates based on the bivariate Moran's I statistic as their spatially lagged version might exert some influence on the SMRs. All the covariates except water coverage present bivariate spatial correlation with the SMR in at least one year. Pollution PM10 is the most frequent variable in the bivariate analysis; it is present in three out of the four years. Poverty, Physical activity, and Influenza vaccine are the less frequent covariates appearing only in one year. In 2007, only two covariates (Incomplete meals and Density of hospitals) were spatially correlated to the SMR, while 2011 and 2014 reach up to six covariates, including Temperature, Housing quality, and Life quality. Further details are available in the Supplementary Material.

3.2. Spatial autoregressive models

The regression analysis determined the candidate covariates to preserve in the spatial autoregressive model. Independent variables for 2004 and 2011 were chosen using a standard OLS regression. For 2007 and 2014, the best criteria for variable selection was the penalized regression, Lasso. There was no evidence of heteroskedasticity, multicollinearity, or nonlinearity in the OLS models (see Supplementary Material). However, the regression model for 2007 suffered from non-normality. We applied a logarithmic transformation to the explanatory variables to ensure multivariate normality.

In the OLS regression residuals, the Moran's I test concluded that only 2011 has spatial error dependence (Moran's I = 0.852, *p*-value = 0.002). Although 2004, 2007, and 2014 showed no residual spatial autocorrelation, we detected spatial dependence in the outcome variable. In other words, introducing a spatially lagged form of the outcome variable might adjust for the spatial autocorrelation in the data. The results from Moran's I in both the dependent variable and the OLS residuals suggest initial spatial autoregressive model configurations that consider the nature of the dependence. For instance, in 2011, we might need a model that accounts for spatial structure in the error term, such as spatial error model (SEM) or spatial lag model with spatial autoregressive error (SARAR).

The linear combination of the independent variables found in the linear regression analysis, the bivariate Moran's I, and the higher-order matrices are used as initial configurations for the spatial autoregressive models. The Lagrange Multiplier tests suggest a spatial error model (SEM) for 2014 and a spatial lag model (SAR) for the remaining years. For 2007 and 2014, the diagnostic test was not statistically significant, which points to non-spatial models such as an OLS regression or spatial model that considers spatially lagged independent variables. Furthermore, since we are also accounting for models with spatially lagged independent variables, Anselin et al. (1996) test limits model selection as it only considers three configurations of spatial autoregressive models. Following Elhorst (2010)'s model selection approach to determine the form or forms of spatial correlation (endogenous, exogenous, or unobserved), we computed the seven model configurations of spatial interactions: spatial lag model (SAR), spatial error

Table 4
AICs for each spatial autoregressive model.

Model	Year			
	2004	2007	2011	2014
OLS	-12.38	-8.43	14.56	3.83
SAR	-14.84	-6.52	12.31	5.44
SEM	-13.86	-6.517	12.84	1.98
SLX	-15.61	-2.904	15.86	-6.01
SDM	-17.90	-1.15	14.63	-4.91
SARAR	-23.16	-4.89	13.08	3.83
SDEM	-27.45	-1.33	13.22	-37.78

model (SEM), spatial lag of X model (SLX), spatial Durbin model (SDM), spatial lag model with spatial autoregressive error (SARAR), spatial Durbin error model (SDEM) and General Nesting Model (GNS) also known as Manski model. Table 4 shows the AICs for the OLS regression and six spatial autoregressive models. We omitted the GNS model as it is known to increase the risk of overfitting (Vega and Elhorst, 2015; Burridge et al., 2016). Likewise, we discarded the GNS model results as they showed non-statistically significant variables and inconsistent parameter estimates.

We selected the models with the most robust explanatory character (lowest AIC) for each year. For 2004 and 2014, we adopted a SDEM model. The SDEM, defined as local spatial spillover specification, considers the spatial correlation in explanatory variables WX and in the error term (disturbances) $W\epsilon$. In this configuration, the outcome variable is affected by the local spillovers of the explanatory variables and the global diffusion of disturbances belonging to neighbouring regions (LeSage, 2014). The classic OLS model was chosen for 2007. The selection of the OLS model was foreseeable as we did not find evidence of spatial autocorrelation in the outcome variable nor the residuals of the OLS regression. The OLS model explains the independent variable as the linear combination of covariates without any spatial interaction. A SAR model better explains 2011. The global spatial specification of the SAR model allows for endogenous spatial autocorrelation W_y only. In the SAR model, a change in the outcome variable of one region influences itself in neighbouring regions, and conversely.

3.3. Results interpretation

We determined that the most relevant independent variables to describe the pneumonia SMRs for duration analysed were the density of hospitals, school attendance, physical activity, housing quality, and poverty. The results showed that the increase in unemployment, poverty, and inadequate housing quality variables, stimulates the increase of the SMRs. It is essential to consider the spatial effects of unemployment and poverty factors, when attempting to understand and address the SMR's outcomes, as these factors may extend beyond the boundaries of a single district and affect the pneumonia mortality rates in neighbouring districts as well. Likewise, a reduction in the density of hospitals and life quality indicators also increases pneumonia rates, suggesting that people living in areas with low living conditions and insufficient healthcare facilities are more affected by pneumonia mortality rates. Furthermore, the incidence of pneumonia mortality was lower in districts with higher influenza vaccine rates. In contrast, school attendance, physical activity, and housing quality are negatively correlated with the SMRs, meaning that higher measures affect the SMR of the districts and their neighbours. Note that not all variables were significant in every year analysed. This might be mainly due to the changes in Bogota's government and, thus, in public policies, resources, and action plans. For example, from 2004 to 2014, Bogota had three different mayors. The discussion elaborates on this point. Another potential explanation on why covariates were not significant for all years might be due to the data sources. The sources do not coincide with every year. Nonetheless, we expect this effect to be

Table 5
Determinants of the SMR of Pneumonia by Bogota's district per year.

Determinant	2004	2007 ^a	2011	2014
	SDEM	OLS	SAR	SDEM
Unemployment	0.010* (0.012)	–	–	–
School attendance	–0.056*** (0.007)	–2.897*** (0.078)	–	–
Physical activity	–0.034*** (0.005)	–0.426** (0.023)	–	–
Housing quality	–0.512*** (0.091)	–	–	–0.054* (0.001)
Poverty	0.048*** (0.009)	–	–	0.013*** (0.007)
Life quality	–	–	–0.015* (0.002)	–
Incomplete meals	–	–	–	0.185*** (0.016)
Density of hospitals	–	–0.030* (0.003)	–0.003** (0.001)	–0.004*** (0.000)
Influenza vaccine	–	–	–	–0.039*** (0.006)
Temperature	–	–	–	–0.348*** (0.045)
Pollution PM10	–	–	–	0.002* (0.001)
W ₃ SMR	2.1425*** (0.147)	–	0.028* (0.029)	–
ρ	–	–	0.658** (0.175)	–
λ	–0.781*** (0.080)	–	–	–0.713*** (0.015)
θ , Unemployment	0.100*** (0.022)	–	–	–
θ , Housing quality	–0.966** (0.043)	–	–	–
θ , Physical activity	–0.077*** (0.017)	–	–	–
θ , Poverty	0.059** (0.024)	–	–	0.351*** (0.025)
θ , Incomplete meals	–	–	–	0.014* (0.005)
θ , Density of hospitals	–	–	–	–0.001** (0.000)
θ , Temperature	–	–	–	–1.593*** (0.182)
θ , Pollution PM10	–	–	–	0.070*** (0.007)
θ , Influenza vaccine	–	–	–	–1.026*** (0.016)
intercept	–8.935* (4.734)	9.60*** (2.792)	1.215* (1.957)	5.072*** (0.823)
σ	0.039	0.167	0.233	0.021
Nagelkerke R ²	0.859	0.528	0.685	0.883
Breusch–Pagan test	0.491	0.352	0.886	0.429
Residuals auto test	0.670	0.980	0.560	0.276

^aThe independent variables were log transformed.
Standard deviations in parenthesis.

***Significant at 1%.

**Significant at 5%.

*Significant at 10%.

minimal as Bogota censuses must follow strict data collection and processing standards.

Table 5 summarizes the results for the study period models. No model suffered from heteroskedasticity, multicollinearity, or residual spatial autocorrelation. The coefficients of the independent variables and the autoregressive terms are statistically significant at least to the

90% confidence level. We chose an $\alpha < 10\%$ as our study focuses on understanding (inferring) the contributions of the contextual factors than predicting Pneumonia outcomes. For 2004, the SDEM configuration shows four socio-economic variables and a behavioural one, their corresponding spatial lags, and the third-order spatially lagged SMR display significant relationships with the SMR. Unemployment,

Poverty, and W_3 SMR is positively related to the SMR within the same district; as these variables increase, so the SMR for Pneumonia does. Similar behaviour occurs with the spatially lagged versions of the determinants; the SMR values in one district are also affected by Unemployment and Poverty figures in neighbouring districts. The spatially autocorrelated error term λ informs that the SMR of a district depends as well on omitted characteristics that affect the mortality rates of all Bogota's districts. This means that unobserved factors may vary systematically over space, resulting in residual spatial error correlation. The negative sign expresses a negative spatial spillover; high disturbances in a district will produce low ones in its neighbour. The SDEM model explained approximately 86% of the variation (Nagelkerke $R^2 = 0.859$).

For 2007, the linear-log OLS model shows that two socio-economic variables and a medical care one influence the SMR for Pneumonia. SMRs were found to decrease with higher percentages of School attendance, Physical activity, and Density of hospitals. Unlike the coefficients of the spatial autoregressive models, the coefficients of the OLS model can be interpreted directly: for every 1% increase in the School attendance and Physical activity variables from their average value, the SMR decreases by 0.02 and 0.004 units, respectively. For the same increase in the Density of hospitals variable, the SMR decreases by 0.003 units. Approximately 52% of the model's variability is explained according to the pseudo- R^2 .

A SAR model is determined better to explain the behaviour of SMR for Pneumonia in 2011. Having a spatial autocorrelation structure in the SMR indicates that the disturbances of first and third-order (W_3 SMR) in one district are transmitted to neighbouring districts, and vice versa; a focal point of SMR for Pneumonia in a district is enough to stimulate the spreading to nearby districts. On average, 66% (ρ) of the SMR disturbances disseminate to neighbouring districts. In the SAR model, we notice a negative relationship between the SMR and two variables: Life quality and Density of hospitals. Inhabitants of areas with low living conditions and insufficient health facilities are more affected by mortality rates for Pneumonia. The Nagelkerke $R^2 = 0.685$ suggests that the model accounted modestly for the variation in SMRs.

The SDEM model was found to be suitable for 2014. Socio-economic, behavioural, environmental, and medical care variables disturb the SMR for Pneumonia of Bogota's districts. The SMR incidence is less severe in districts with more medical care facilities and higher influenza vaccine rates. The SMRs increase in districts with low temperatures and poor air quality levels. We also witness a positive relationship between the mortality rates and the Incomplete meals variable. Furthermore, as Poverty worsens health prospects (Jahanihashemi et al., 2018), people living in districts with Poverty and inadequate housing report higher SMR for Pneumonia. The relationship between the SMR and the explanatory variables also has a spillover effect: the SMR of one district is altered if any of the variables change in neighbouring districts. For instance, the spatially lagged density of hospitals has a negative coefficient estimate, indicating that as the number of hospitals in surrounding counties increases, a decrease in SMR is seen in the home district. This may reflect that population seeks medical care not only in their district of residence but also in neighbouring districts. People are mobile and may travel to another district to attend medical services, notably when these are restricted or limited within their home district. The spatial autoregressive coefficient λ implies that a change in the disturbances of a single district impacts the disturbances of neighbouring regions, and vice versa. In other words, the variations in unobserved factors have a similar spillover effect in the SMRs as the spatially lagged covariates. For 2014, the SDEM model explained 88.3% (Nagelkerke $R^2 = 0.883$) of the SMR variation.

In the presence of a spatially lagged variable (W_y , WX), the coefficients associated with the explanatory variables cannot be interpreted as in a standard linear model. In the spatial autoregressive model framework, the variation of an explanatory variable in a region directly affects its outcome and indirectly affects the outcome of all other

regions. This is true for all the spatial model specifications except the SEM's one, where the marginal effects of the parameters are interpreted as those in linear regression. LeSage and Pace (2009) proposed spillover effects, also called impacts, that quantify the average effect of the estimated parameters when spatial interaction is present.

The variation of an explanatory variable in a given district affects the district itself but indirectly affects all the other neighbouring districts. The former scenario is known as a direct effect and the latter as an indirect effect. The total effect is the sum of the direct and indirect effects. Table 6 presents the average direct, indirect, and total effects of the spatial autoregressive models for 2004, 2011, and 2014. In the SDEM models of 2004 and 2014, the direct impacts mirror the coefficient estimates reported in Table 5. This is a property of spatial autoregressive models that dismiss endogenous spatial interactions (Burridge et al., 2016).

The indirect impacts for every explanatory variable – omitting School attendance in 2004 and, Incomplete meals and Density of hospitals in 2014 – are greater than the direct impacts in all years except 2007 since this has no direct and indirect impacts. This phenomenon implies that variations in the socio-economic, behavioural, environmental, and medical care variables might lead to stronger fluctuations of the SMR in nearby districts rather than in itself (only spatial models). Note that different variables in different years might trigger such fluctuations. For instance, only medical care factors affected the SMR in 2011.

For both 2004 and 2014, the total impact of the SMR for Pneumonia is most affected by the Housing quality variable; the poorer the housing conditions in a district, the higher the mortality rates in the entire region. It also holds for the Life quality variable in 2011, when decreasing the living conditions across Bogota's population will rise the most SMRs across the districts. The ratio between inhabitants and health facilities in a district contributes less to the SMRs for 2011 and 2014. On average, increasing the Density of hospitals and health care centres per 100,000 population in one unit reduces the mortality rates for Pneumonia by 0.008 and 0.005 units. The impacts of School attendance in 2004 are the smallest ones over those of other variables for the same year. Note that the direct impact is identical to the total impact, which means that the variable lacks a spillover effect and that changes in a given district only affect the SMR of such district.

The total impacts of Unemployment, Poverty and Physical activity are similar for 2004. However, the spillover effect of unemployment is greater than its counterparts. The temperature in 2014 is a major factor in Pneumonia mortality rate outcome in Bogotá; the spillover effect to neighbouring districts is approximately 1.6 units expected to increase in SMR if the temperature drops by 1 Celsius degree. The environmental variable Pollution PM10 behaves as the annual mean temperature: it has a more considerable influence on neighbouring districts' mortality rates than in the district it belongs to. Poverty, Incomplete meals, and Influenza vaccine variables display the same order of magnitude impacts for 2014.

4. Discussion

In this research, we focused on understanding the relationships between mortality rates for Pneumonia and potential determining factors in Bogotá, Colombia for 2004, 2007, 2011, and 2014 through spatial statistical models. We found that socioeconomic, behavioural, environmental, and medical care variables contribute to the temporal growth and spatial spreading of Pneumonia mortality in Bogota's territory.

We demonstrated the existence of an underlying spatial process in the distribution of the SMRs across Bogota's districts (Fig. 2). The districts with high mortality rates cluster together as well as the ones with low incidence. High SMRs occur in urbanized commercial and economically dynamic regions, while low rates in mortality localize primarily in rural and suburban areas. Throughout the years, the spatial behaviour of the SMRs remains similar. However, high rates from the

Table 6
Direct, indirect and total impacts on SMR for Pneumonia of Bogota's districts.

Year	Determinant	Direct	Indirect	Total
2004	Unemployment	0.010*** (0.012)	0.100** (0.022)	0.110*** (0.019)
	Housing quality	-0.512*** (0.0091)	-0.966*** (0.438)	-1.479** (0.040)
	Poverty	0.048*** (0.009)	0.059** (0.024)	0.107*** (0.022)
	School attendance	-0.056*** (0.007)	0.000 (0.000)	-0.056*** (0.007)
	Physical activity	-0.034*** (0.007)	-0.077** (0.017)	-0.112*** (0.018)
2011	Life quality	-0.015* (0.002)	-0.028* (0.679)	-0.043* (0.719)
	Density of hospitals	-0.003* (0.001)	-0.005* (0.157)	-0.008* (0.166)
2014	Temperature	-0.348** (0.045)	-1.593*** (0.018)	-1.941*** (0.225)
	Pollution PM10	0.002* (0.001)	0.070*** (0.007)	0.068*** (0.006)
	Housing quality	-0.054* (0.154)	-2.684*** (0.589)	-2.739*** (0.456)
	Poverty	0.013* (0.007)	0.352*** (0.025)	0.366*** (0.029)
	Incomplete meals	0.185*** (0.016)	0.014* (0.052)	0.200** (0.066)
	Density of Hospitals	-0.004*** (0.000)	-0.001** (0.000)	-0.005** (0.000)
	Influenza vaccine	-0.039*** (0.006)	-0.102*** (0.016)	-0.142*** (0.013)

Standard deviations in parenthesis.

***Significant at 1%.

**Significant at 5%.

*Significant at 10%.

central districts extended to northern regions. We also noticed that the contiguity relationships for the districts change through the years (Table 2), exposing the changing behaviour of the Pneumonia mortality rates in a complex capital such as Bogotá.

Our results determined that the socioeconomic variable with the greatest association with SMRs vary over the years in Bogotá. In particular, Housing quality was found to be a significant determining factor in 2004 and 2014. This finding aligns with other studies' conclusions in which statistically significant relationships between Poor-quality housing conditions and adverse respiratory health outcomes have been found (Keall et al., 2012; Wimalasena et al., 2021). Although in 2011 we did not detect an association between Housing quality and SMR, we found life quality as a risk factor. Unemployment and Poverty were also identified as contributing aspects to the mortality rates in Bogotá. These associations have been reported (Jahanhashemi et al., 2018) as Pneumonia morbidity and mortality exacerbate in developing countries with mediocre household income and alarming unemployment rates. Although we expected the variable poverty to be a Pneumonia determinant across years, it was only significant in 2004 and 2014. The authors consider three aspects that might have driven this behaviour. Firstly, the values of poverty for 2004 and 2014 are the two highest within the years studied (according to the census data). Secondly, the definition of poverty is based on the unmet basic needs criteria, which relate poverty to low-income levels. This might not be an adequate index to measure poverty. And finally, the reduction of resources and scopes in the social protection and poverty fight policies introduced by the elected mayor in 2014. School attendance achieved significance in the 2004 and 2007 models. School absenteeism has been proven to stimulate unhealthy behaviours in kids, and young adults resulting in poor long-term health outcomes (Allison et al., 2019). We believe this variable was not significant in 2011 and 2014 as the sectoral education plan in Bogota from 2010–2014 re-evaluated and developed

new public policies that sought to increase school attendance levels and mitigate school dropouts. After its introduction, school absenteeism and dropouts were considerably reduced according to the census data of the same period. Districts with historically low school attendance rates, such as Bosa, Kennedy, Usme and Ciudad Bolívar, exhibited higher figures when the sectoral plan was running. Guided by literature, we expected the Water coverage variable to influence the mortality rates. However, it was insignificant for all years. A detailed examination of the variable values showed that most of Bogotá's districts have a 100% access to water and sewer services. Furthermore, the differences in coverage among communities are smaller than a one-percentage point. Behavioural factors such as Physical activity and Incomplete meals were found to raise the SMRs for Pneumonia in 2004, 2007, and 2014. Our findings suggest that limited physical activity and skipping one daily meal increase the death toll of Pneumonia. Williams (2014) discovered that greater exercise energy expenditure reduces mortality associated with Pneumonia as the underlying cause. Furthermore, Pneumonia patients with malnutrition have been found to have higher hospitalization and mortality rates (Ginsburg et al., 2015). Although incomplete meals cannot be directly linked to malnutrition, neglecting meal intake increases the risk for malnutrition and eating disorders.

The Medical care variables examined in this study were found to affect the Pneumonia mortality rates. The density of hospitals was a significant factor in 2007, 2011, and 2014. We expected this association due to the unequal distribution of medical facilities in Bogotá. Most of the hospitals and health care centres are located in the northeastern region of the city. Inhabitants of the Southern and central districts must commute to a district with health care services. In addition to this, since Bogotá's health system is primarily private, access to medical care also relies on socioeconomic factors such as income and employment status. According to the Special Registry of Health Service Providers in Bogota, health care centres increased from 921 in 2007 to

2,109 in 2014. However, this increase was disproportionate concerning Bogota districts and their population. This was evident in 2014 when the density of hospitals acquired a spatially lagged character. As the distribution of hospitals was uneven across districts, the SMR in a specific district strongly depended on the number of hospitals in nearby ones. Vaccination against *Haemophilus Influenzae* type B was negatively related to SMR in 2014. The mandatory vaccination programme established by BHS in 2004 emphasizing the mandatory vaccination against *Haemophilus influenza*, which causes high-risk bacterial diseases such as meningitis and pneumonia, was restructured in 2014 due to budget cuts proposed by the incoming Bogota mayor. Mortality rates for Pneumonia increase dramatically in children unvaccinated against the bacteria *Haemophilus Influenzae* type B (Benavides et al., 2012b).

The environmental factors were inversely associated with SMRs in 2014. The Environmental observatory of Bogota reported the lowest annual mean temperature in 2014 since 2002. This result is supported by epidemiological studies that have proven how cold temperatures weaken the immune system's response to respiratory infections (D'Amato et al., 2018). Results also nominated particulate air pollution (10 μm or less) as a contributing factor in Pneumonia mortality. Increased levels of air pollutants, such as particulate matter up to 10 μm in size (PM10), are well-known risk factors for Pneumonia, asthma, and other pulmonary diseases (Tasci et al., 2018).

This study also explores spatial autoregressive models and confirms their advantages in ecological analysis. We discovered that the mortality rates for Pneumonia in three out of four years adopt a spatial dependence structure. The SAR and SDEM models selected in our analysis understand and represent the spatial interaction between the SMRs and determining factors. We found that the mortality rates of a given district affect and are affected by the mortality rates and socioeconomic, behavioural, environmental, and medical care covariates of nearby districts. Furthermore, in some years, the SMRs across Bogota's districts are also affected by variations in unknown variables (spatial error term).

It is worth noting the impact of the spatial effects on the mortality rates for Pneumonia in Bogota. This study observed a significant spillover effect in most of the determining factors analysed (Table 6). While the covariates impact directly the SMRs of the district they were reported in, the impacts are relatively larger in adjacent districts' rates. This enlightens the need for local (within a district) and global (in neighbouring districts) government policies to mitigate better the burden of Pneumonia through its determining factors.

We outline the contributions of our research in three main different aspects: (i) The understanding and modelling of the spatial variability in mortality rates for Pneumonia in Bogota. We emphasized the importance of the spatial interactions between the mortality rates and district-level determinants to disclose the underlying process that governs Bogota's Pneumonia death outcomes. (ii) The use of spatial econometric methods to explain infectious diseases. We demonstrated the technical and computational convenience of spatial autoregressive models in integrating an infectious disease such as Pneumonia and its driving factors with the spatial domain. The methods herein adopted are advantageous in terms of interpretability and implementation. Statistical and non-statistical epidemiologists can leverage these methods in ecological analysis. (iii) The selection of socioeconomic, behavioural, environmental, and medical factors associated with Pneumonia outcome. We illustrated how a variety of contextual variables excite and modulate the mortality rates for Pneumonia.

Our results are of direct interest to governmental health institutions and epidemiological researchers. This study might support programmes to improve population housing and living conditions and access to health care to reduce Pneumonia mortality rates in Bogota. Public policies such as the Distrital Policy of Housing and Habitat and the National Plan for Social Development and Life Quality must adequately evaluate social protection actions' effect on health outcomes. So far, these policies have no emphasis on population health. Likewise, public

health measures should encourage physical activity and healthy dietary habits. We believe that exposure and susceptibility to Pneumonia can be further restrained by establishing geographically directed policies that prioritize districts whose characteristics stimulate the mortality rates in Bogota's territory. For example, we suggest expanding the population at risk delineated in the Action Plan against Respiratory Infectious proposed by the Bogota Health Secretariat in 2016. Currently, only age-related risk groups (infants and elderly) are targeted. We recommend including additional risk groups based on the spatial susceptibility structures of the SMR and the determinants presented in this study. Furthermore, the methods used in this work are not only limited to Pneumonia; researchers can naturally extend these methods to other infectious diseases such as HIV, malaria, and Zika, diseases closely related to poor socioeconomic, environmental, and health care factors. We hope our analysis can guide future spatial econometrics research dealing with disease outcomes.

Future studies may improve the results by reshaping the spatial units into a finer partition or using a multilevel modelling approach. Ecological-level analyses are susceptible to the ecological fallacy (Jargowsky, 2004) derived from aggregating individual-level information to represent a population. Reducing the spatial scale of an ecological-level analysis might better decodes the variability of a health condition and its determinants. An extension of our research might be a comprehensive Spatio-temporal analysis if data is available. For example, a Spatio-temporal autoregressive model that considers both the interaction in space and time for Pneumonia mortality rates given explanatory variables. This alternative could enrich the selection of global determinants and allow predictions for a given Spatio-temporal window. Another approach might consider the hierarchical Bayesian framework due to its flexibility in modelling space interaction in epidemiological studies (Navas et al., 2019). In this context, the Bayesian models assume that the number of cases (deaths) is Poisson distributed and express the relative risk as a linear combination of variables (environmental, socioeconomic, demographic, and biological) and spatial random effects.

5. Conclusions

In conclusion, this paper introduces spatial autoregressive models to explain the mortality rates for Pneumonia based on socioeconomic, behavioural, environmental, and medical care indicators in Bogota. One of the main benefits of our approach is the inclusion of contextual determinants and the ability to quantify their spatial effects in Pneumonia death outcomes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported by: Core Spatial Data Research (Faculty of Engineering, COL0013969, Universidad Distrital Francisco José de Caldas), and Applied Statistics in Experimental Research, Industry and Biotechnology (COL0004469, Universidad Nacional de Colombia).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.sste.2023.100581>.

References

- Allison, M.A., Attisha, E., Lerner, M., De Pinto, C.C.D., Attisha, C.E.E., Beers, N.S., Gibson, E.J., Gorski, P., Kjolhede, C., O'Leary, S.C., Schumacher, H., Weiss-Harrison, A., 2019. The link between school attendance and good health. *Pediatrics* 143 (2), 20183648. <http://dx.doi.org/10.1542/peds.2018-3648>, <https://pediatrics.aappublications.org/content/143/2/e20183648> <https://pediatrics.aappublications.org/content/143/2/e20183648.abstract>.
- Almirall, J., Serra-Prat, M., Bolibar, I., Palomera, E., Roig, J., Boixeda, R., Bar-tolomé, M., de la Torre, M., Parra, O., Torres, A., 2015. Professions and working conditions associated with community-acquired pneumonia. *Arch. de Bronconeumologia* <http://dx.doi.org/10.1016/j.arbr.2014.10.018>.
- Anon, 2014. Spatial variation in the risk of hospitalization with childhood pneumonia and empyema in the north of England. *Epidemiol. Infect.* 142 (2), 388–398. <http://dx.doi.org/10.1017/S0950268813001015>, URL <https://www.cambridge.org/core/journals/epidemiology-and-infection/article/spatial-variation-in-the-risk-of-hospitalization-with-childhood-pneumonia-and-empyema-in-the-north-of-england/EABD1505054FA2EEF1371BAACED054F1>.
- Anselin, L., 1988. A typology of spatial econometric models. In: *Spatial Econometrics: Methods and Models*. Springer Netherlands, Dordrecht, pp. 32–40. http://dx.doi.org/10.1007/978-94-015-7799-1_4.
- Anselin, L., 1995. Local indicators of spatial association-LISA. *Geogr. Anal.* 27 (2), 93–115. <http://dx.doi.org/10.1111/j.1538-4632.1995.tb00338.x>.
- Anselin, L., Bera, A.K., Florax, R., Yoon, M.J., 1996. Simple diagnostic tests for spatial dependence. *Reg. Sci. Urban Econ.* 26 (1), 77–104. [http://dx.doi.org/10.1016/0166-0462\(95\)02111-6](http://dx.doi.org/10.1016/0166-0462(95)02111-6).
- Beale, C.M., Lennon, J.J., Yearsley, J.M., Brewer, M.J., Elston, D.A., 2010. Regression analysis of spatial data. *Ecol. Lett.* 13 (2), 246–264. <http://dx.doi.org/10.1111/j.1461-0248.2009.01422.x>, [arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1461-0248.2009.01422.x](https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1461-0248.2009.01422.x).
- Benavides, J.A., Ovalle, O.O., Salvador, G.R., Gray, S., Isaacman, D., Rodgers, G.L., 2012a. Population-based surveillance for invasive pneumococcal disease and pneumonia in infants and young children in Bogotá, Colombia. *Vaccine* <http://dx.doi.org/10.1016/j.vaccine.2012.03.054>.
- Benavides, J.A., Ovalle, O.O., Salvador, G.R., Gray, S., Isaacman, D., Rodgers, G.L., 2012b. Population-based surveillance for invasive pneumococcal disease and pneumonia in infants and young children in Bogotá, Colombia. *Vaccine* 30 (40), 5886–5892. <http://dx.doi.org/10.1016/j.vaccine.2012.03.054>, URL <https://pubmed.ncbi.nlm.nih.gov/22484295/>.
- Borcard, D., Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol. Model.* [http://dx.doi.org/10.1016/S0304-3800\(01\)00501-4](http://dx.doi.org/10.1016/S0304-3800(01)00501-4).
- Burridge, P., Elhorst, J.P., Zigova, K., 2016. Group interaction in research and the use of general nesting spatial models. *Adv. Econom.* 37, 223–258. <http://dx.doi.org/10.1108/S0731-90532016000037016>.
- Cabre, M., 2009. Pneumonia in the elderly. In: *Current Opinion in Pulmonary Medicine*. <http://dx.doi.org/10.1097/MCP.0b013e328326f571>.
- Chi, G., Zhu, J., 2008. Spatial regression models for demographic analysis. *Popul. Res. Policy Rev.* 27 (1), 17–42. URL <http://www.jstor.org/stable/41217935>.
- Cressie, N., Calder, C.A., Clark, J.S., Hoef, J.M.V., Wile, C.K., 2009. Accounting for uncertainty in ecological analysis: The strengths and limitations of hierarchical statistical modeling. *Ecol. Appl.* 19, <http://dx.doi.org/10.1890/07-0744.1>.
- Crighton, E.J., Elliott, S.J., Moineddin, R., Kanaroglou, P., Upshur, R.E., 2007. An exploratory spatial analysis of pneumonia and influenza hospitalizations in ontario by age and gender. *Epidemiol. Infect.* 135 (2), 253–261. <http://dx.doi.org/10.1017/S095026880600690X>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2870578/>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2870578/?report=abstract> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2870578/>.
- Curtin, L.R., Klein, R.J., 1995. Direct standardization (age-adjusted death rates), Vol. 6. *Healthy People 2000 Statistical Notes / National Center for Health Statistics*.
- Czaplewski, R.L., Reich, R.M., Forest, R.M., Range Experiment Station (Fort Collins, C., 1993. Expected value and variance of Moran's bivariate spatial autocorrelation statistic for a permutation test [microform] / Raymond L. Czaplewski, Robin M. Reich. U.S. Dept. of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station Fort Collins, Colo., p. 13, [2].
- D'Amato, M., Molino, A., Calabrese, G., Cecchi, L., Annesi-Maesano, I., D'Amato, G., 2018. The impact of cold on the respiratory tract and its consequences to respiratory health. *Clin. Transl. Allergy* 8 (1), 1–8. <http://dx.doi.org/10.1186/s13601-018-0208-9>, URL <https://ctajournal.biomedcentral.com/articles/10.1186/s13601-018-0208-9>.
- Demougeot, L., Rolland, Y., Gérard, S., Pennetier, D., Duboué, M., Vellas, B., Cesari, M., 2013. Incidence and economical effects of pneumonia in the older population living in French nursing homes: Design and methods of the INCUR study. *BMC Public Health* <http://dx.doi.org/10.1186/1471-2458-13-861>.
- Departamento Nacional de Estadística, 2015. *Informe de Coyuntura Económica Regional*. Technical Report, DANE- Banco de la República.
- Elhorst, J.P., 2010. Applied Spatial Econometrics: Raising the Bar, Vol. 5, no. 1. Taylor & Francis Group, pp. 9–28. <http://dx.doi.org/10.1080/17421770903541772>, URL <https://www.tandfonline.com/doi/abs/10.1080/17421770903541772>.
- Elliott, P., Wakefield, J., Best, N., Briggs, D., 2001. *Spatial Epidemiology: Methods and Applications*. In: Oxford medical publications, Oxford University Press, <http://dx.doi.org/10.1093/acprof:oso/9780198515326.001.0001>.
- Forero, J.C., 2013. Neumonía adquirida en comunidad en el adulto Es hora de implementar las guías de manejo clínico institucionales. *Acta Medica Colombiana* 38, 206–207, URL http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-24482013000400004&nrm=iso.
- Ginsburg, A.S., Izadnegahdar, R., Berkley, J.A., Walson, J.L., Rollins, N., Klugman, K.P., 2015. Undernutrition and pneumonia mortality. *Lancet Glob. Health* 3 (12), e735–e736. [http://dx.doi.org/10.1016/S2214-109X\(15\)00222-3](http://dx.doi.org/10.1016/S2214-109X(15)00222-3), <http://www.thelancet.com/article/S2214109X15002223/fulltext> [https://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(15\)00222-3/abstract](https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(15)00222-3/abstract).
- Greenland, S., 2001. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *Int. J. Epidemiol.* <http://dx.doi.org/10.1093/ije/30.6.1343>.
- Guzman, L.A., Bocarejo, J.P., 2017. Urban form and spatial urban equity in Bogotá, Colombia. In: *Transportation Research Procedia*. <http://dx.doi.org/10.1016/j.trpro.2017.05.345>.
- Hemilä, H., Kaprio, J., Albanes, D., Virtamo, J., 2006. Physical activity and the risk of pneumonia in male smokers administered vitamin e and β -carotene. *Int. J. Sports Med.* <http://dx.doi.org/10.1055/s-2005-865670>.
- Isturiz, R.E., Luna, C.M., Ramirez, J., 2010. Clinical and economic burden of pneumonia among adults in Latin America. *Int. J. Infect. Dis.* <http://dx.doi.org/10.1016/j.ijid.2010.02.2262>.
- Jahanihashemi, H., Babaie, M., Bijani, S., Bazzazan, M.M., Bijani, B.D., 2018. Poverty as an independent risk factor for in-hospital mortality in community-acquired pneumonia: A study in a developing country population. *Int. J. Clin. Pract.* 72 (5), <http://dx.doi.org/10.1111/ijcp.13085>, URL <https://pubmed.ncbi.nlm.nih.gov/29665161/>.
- Jargowsky, P.A., 2004. Ecological fallacy. In: *Encyclopedia of Social Measurement*. Elsevier Inc., pp. 715–722. <http://dx.doi.org/10.1016/B0-12-369398-5/00112-2>.
- Keall, M.D., Crane, J., Baker, M.G., Wickens, K., Howden-Chapman, P., Cunningham, M., 2012. A measure for quantifying the impact of housing quality on respiratory health: A cross-sectional study. *Environ. Health: Glob. Access Sci. Source* 11 (1), <http://dx.doi.org/10.1186/1476-069X-11-33>, URL <https://pubmed.ncbi.nlm.nih.gov/22583775/>.
- LeSage, J.P., 2014. What regional scientists need to know about spatial econometrics. *Rev. Reg. Stud.* 44 (1), 13–32. <http://dx.doi.org/10.52324/001c.8081>, URL www.srsa.org/rrs.
- LeSage, J., Pace, R.K., 2009. Introduction to spatial econometrics. pp. 1–341. <http://dx.doi.org/10.1111/j.1467-985x.2010.00681.13.x>.
- LeSage, J.P., Pace, R.K., 2014. Interpreting spatial econometric models. In: Fischer, M.M., Nijkamp, P. (Eds.), *Handbook of Regional Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1535–1552. http://dx.doi.org/10.1007/978-3-642-23430-9_91.
- Levin, K.A., 2006. Study design VI – ecological studies. *Evid.-Based Dent.* <http://dx.doi.org/10.1038/sj.ebd.6400454>.
- Navas, A.L., Osei, F., Leonardo, L.R., Magalhães, R.J., Stein, A., 2019. Modeling Schistosoma Japonicum infection under pure specification bias: Impact of environmental drivers of infection. *Int. J. Environ. Res. Public Health* 16 (2), <http://dx.doi.org/10.3390/ijerph16020176>, URL <https://pubmed.ncbi.nlm.nih.gov/30634518/>.
- Ord, J.K., Getis, A., 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geogr. Anal.* 27 (4), 286–306. <http://dx.doi.org/10.1111/j.1538-4632.1995.tb00912.x>.
- Rojas, J.P., Leal, A.L., Patinó, J., Montañez, A., Camacho, G., Beltrán, S., Bonilla, C., Barrero, R., Marín, C., Ramos, N., 2016. Characterization of patients who died of invasive pneumococcal disease in the child population of Bogotá, Colombia. *Revista Chilena de Pediatría* 87 (1), 48–52. <http://dx.doi.org/10.1016/J.RCHPE.2015.10.005>.
- Rosselli, D., Rueda, J.D., 2012. Burden of pneumococcal infection in adults in Colombia. *J. Infection Public Health* <http://dx.doi.org/10.1016/j.jiph.2012.04.003>.
- RStudio Team, 2020. *RStudio: Integrated Development Environment for R*. RStudio, PBC, Boston, MA, URL <http://www.rstudio.com/>.
- Subsecretaría de información y estudios estratégicos, 2016. *Encuesta Multipropósito Bogotá EMB 2011 - 2014. Technical Report*, Secretaría de Planeación.
- Taboada B., L.B., Leal Castro, A.L., Caicedo V., M.P., Camargo B., C.B., Roa B., J.H., 2015. Etiología de la neumonía adquirida en la comunidad en un hospital de cuarto nivel en Bogotá: estudio descriptivo de un registro institucional durante los años 2007 a 2012. *Infectio* 19 (1), 10–17. <http://dx.doi.org/10.1016/J.INFECT.2014.11.005>.
- Tasci, S.S., Kavalci, C., Kayipmaz, A.E., 2018. Relationship of meteorological and air pollution parameters with pneumonia in elderly patients. *Emerg. Med. Int.* 2018, 1–9. <http://dx.doi.org/10.1155/2018/4183203>.
- Vega, S.H., Elhorst, J.P., 2015. THE SLX MODEL. *J. Reg. Sci.* 55 (3), 339–363. <http://dx.doi.org/10.1111/JORS.12188>.
- Wakefield, J., 2004. A critique of statistical aspects of ecological studies in spatial epidemiology. *Environ. Ecol. Stat.* 11, <http://dx.doi.org/10.1023/B:EEST.0000011363.12720.38>.

- Wakefield, J., Lyons, H., 2010. Spatial aggregation and the ecological fallacy. <http://dx.doi.org/10.1201/9781420072884-c30>.
- Williams, P.T., 2014. Dose-response relationship between exercise and respiratory disease mortality. *Med. Sci. Sports Exerc.* 46 (4), 711–717. <http://dx.doi.org/10.1249/MSS.0000000000000142>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4073613/pmc/articles/PMC4073613/?report=abstract> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4073613/>.
- Wimalasena, N.N., Chang-Richards, A., Wang, K.I., Dirks, K.N., 2021. Housing risk factors associated with respiratory disease: A systematic review. *Int. J. Environ. Res. Public Health* 18 (6), <http://dx.doi.org/10.3390/ijerph18062815>.
- Zar, H.J., Madhi, S.A., Aston, S.J., Gordon, S.B., 2013. Pneumonia in low and middle income countries: Progress and challenges. In: *Thorax*. <http://dx.doi.org/10.1136/thoraxjnl-2013-204247>.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67 (2), 301–320.