creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

보건학석사 학위논문

Spatial Analysis of the Association Between the Socio-Economic Variables and the Incidence rate of COVID-19 in Seoul

서울시 COVID-19 발생률과 사회경제적 요인 간 연관성에 대한 공간 분석

2021년 2월

서울대학교 보건대학원

보건학과 시공간역학전공

김 정 헌

Spatial Analysis of the Association Between the Socio-Economic Variables and the Incidence rate of COVID-19 in Seoul

지도교수 황 승 식

이 논문을 보건학석사 학위논문으로 제출함
2020년 11월

서울대학교 보건대학원
보건학과 시공간역학전공
김 정 헌

김정헌의 석사 학위논문을 인준함
2020년 12월

위 원 장 조성일 (인)

부위원장 이우주 (인)

위 원 황승식 (인)

Abstract

Spatial Analysis of the Association Between the Socio-Economic Variables and the Incidence rate of COVID-19 in Seoul

Jeongheon Kim
The Graduate School of Public Health
Seoul National University

Introduction: In general, COVID-19 is become pandemic and threatened entire the world. Although the COVID-19 National Prevention Regulations were implemented, the outbreak kept generating and occurring a variety of places. Many studies did analyze the COVID-19 to identify the relationships among factors, but a few studies used the spatial regression model to investigate. The purpose of this study is to provide and utilize the basis for health policies appropriate to the district units by studying the factors affecting the rate of occurrence of COVID-19.

Methods: Seoul COVID-19 data were used. Global and local Moran's I was used to explore the spatial distribution and to identify the clusters with the spatial autocorrelation by using GeoDa software. OLS, SLM, SEM model were implemented on the global scale to identify the relationship between the incidence rate and socio-economic variables.

Results: Across the 25 districts of Seoul, the incidence rate of COVID-19 depicted a negative spatial pattern (I=-0.049) but captured both hotspots and coldspots across the districts. In terms of model fitting, the adjusted R^2 of OLS, SLM, SEM was 0.365, 0.369, 0.365, and each of AIC was 157.519, 159.398, 157.517.

Conclusion: Through the COVID-19 incidence rate, space patterns in the districts of Seoul were identified and hotspots and coldspots were detected. Although no other variables have been significant, it suggests that there is no choice but to move to engage in economic activity. In this regard, health policies for each autonomous district should be established, and it is necessary to curb the occurrence of COVID-19 by supplementing public medical personnel.

keywords : **COVID-19, Incidence Rate, Spatial Analysis, Socio-Economic variables**

Student Number : 2019-27919

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

보존용 학위논문 정오표

페이지	정정 전	정정 후
p. ii : 12 p. 5 : 17-18	space patterns Geographically Weighted	spatial pattern Geographically Weighted Regression ~ 이어서 붙이기

1. INTRODUCTION

1.1 Background

Coronavirus disease-19 (COVID-19) is the disease infected by the SARS-CoV-2 virus, and it became the important issue globally through the fast spread (WHO, 2020a). COVID-19 is classified to 'Group 1 infectious diseases' and can be infected droplets by coughing or touching contaminated objects. COVID-19 has respiratory symptoms (cough, fever, feebleness, difficulty with breathing, and pneumonia) with a wide range of severities, and other symptoms such as sore throat, headache, or diarrhea can be revealed. On the other hand, it can be also occurred asymptomatic patients, which means there is no symptoms but infected (KCDA, 2020).

In COVID-19 Weekly Epidemiological Update report on August 31 in World Health Organization (WHO), a cumulative total case of COVID-19 is approximately 25 million cases and 800,000 deaths has been estimated globally. Also, it is still increasing the infected and dead every day. On the American regions, the cumulative cases were about 53%, and the cumulative deaths were nearly 55% of the global cases each. About 16% were estimated in the cumulative cases and about 9% were estimated in the cumulative deaths on South-East Asia. For the region of Europe, the cumulative cases and deaths were 17%, 26%. 4% of the global cases were the cumulative cases of African region and 3% of the global deaths were the cumulative deaths. The cumulative cases of Eastern

Mediterranean were estimated about 8% and the cumulative deaths were estimated approximately 6%. In terms of Western Pacific region, the cumulative cases and deaths were estimated 2% and 1% of the global cases (WHO, 2020b).

In case of South Korea, the total cumulative patients of COVID-19 on September 1, 2020 were 20,182 and 324 patients has been estimated for the cumulative deaths since the start of the outbreak. In the capital area, the cumulative patients of Seoul, Incheon, and Gyeonggi were 3,961, 740 and 3,323. On the other hand, Daegu had the most confirmed cases among the non-capital areas, which has 7,049 confirmed cases (KCDA, 2020).

According to WHO, COVID-19 was discovered and occurred in Wuhan, China at the first time in 2019. On March 11, 2020, WHO has declared the pandemic of COVID-19 officially (WHO, 2020c). For South Korea, the first patient was reported on January 20, 2020, and since then, the number of infected people has increased rapidly due to the outbreak of patients among Sincheonji church members in Daegu city. The number of newly confirmed patients was decreased due to the implementation of the COVID-19 National Prevention Regulations, but as the outbreak of Itaewon club, Saranjeil church spread once again, prevention rules such as intensive social distancing kept implementing (Kim et al, 2020).

In the United Nations report, it is explained that COVID-19 is the social and economic crisis for humans. It is clearly revealed a socioeconomic impact and a burden of disease on both developing countries and advanced countries (United Nations, 2020). In terms of economic aspects, OECD predicted the economic growth rate will be decreased to 1.5% and IMF anticipated it will be decreased down to

−3.0% on this year. IMF expected a real GDP to recover to −1.2% year-on-year in 2020 and 3.4% in 2021 (IMF, 2020). In research report of South Korea, COVID-19 intensified the risk such as the decline of house income and lifetime of the vulnerable. For causing house income decline, it included reduction in sales, close temporarily or close business, unemployment, and reduction in working hours (Korea Institute for Health and Social Affairs, 2020).

Many studies are published related to COVID-19 with Geographical Information System (GIS). GIS is an important tool to perform a spatial distribution of infectious diseases and it is helping to analyze and visualize the spread of COVID-19 (Mollalo et al, 2018, 2019). For example, Johns Hopkins University Center for System Science and Engineering (JHU CSSE) utilized a GIS dashboard that provides real-time data on the global spatial distribution of COVID-19, including the total number of confirmed cases, mortality, and recovered case. This real-time database is easily accessible to the public and can track diseases that spread over time (JHU CSSE, 2020). Following the GIS, the spatial model is an important factor for statistical investigation of geographical relationships between several explanatory variables and disease outbreaks like COVID-19 (Mollalo et al, 2020). One of COVID-19 with GIS studies, it reviewed COVID-19 articles based on the geospatial and spatial-statistical analysis. This study categorized spatiotemporal analysis, health and social geography, environmental variables, data mining, and web-based mapping. On the health and social geography category, some studies examined the correlation between the geographic, socio-economic status, but few studies examined using the spatial model (Franch-Pardo et al., 2020).

1.2 Literature review

As many explanatory variables such as socio-economic status, demographic, geographic variables are correlated to COVID-19, the literature review focused on the variables mentioned above and spatial model. For searching some studies, the keywords were set up 'COVID-19', 'SARS-CoV-2', '2019-nCoV' or 'coronavirus disease 2019' together with different conjugations of the words 'spatial analysis', 'spatial regression', 'geographic information system', 'demographic', 'environmental', 'socioeconomic status' in PubMed and EMBASE.

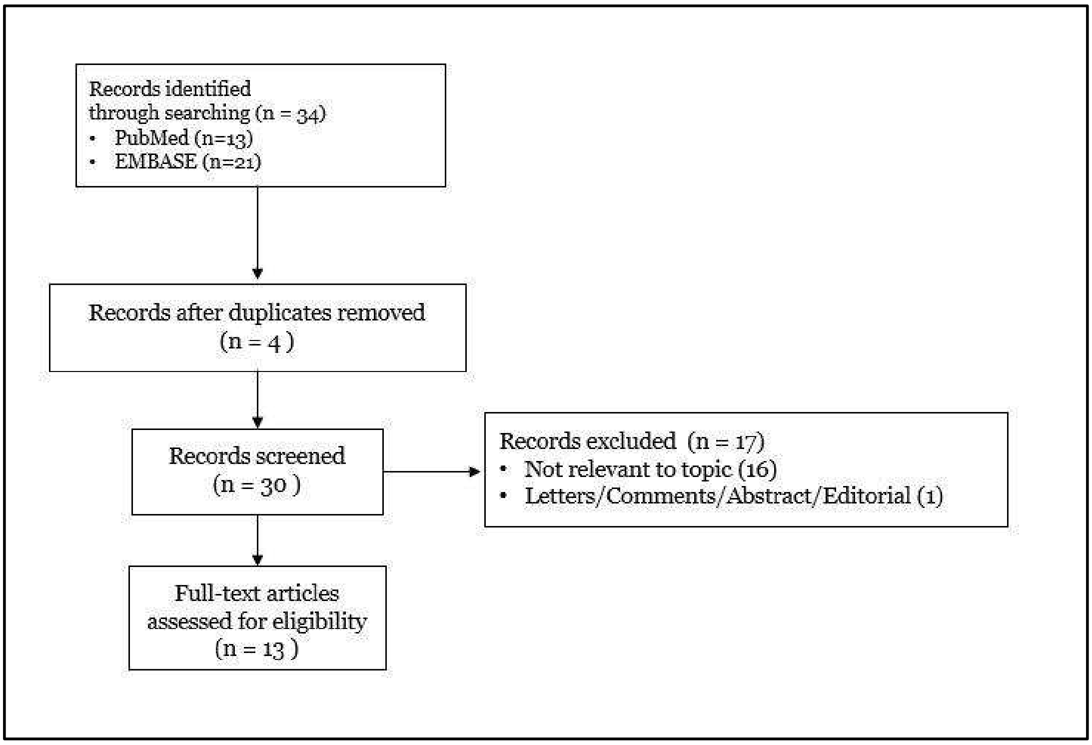

Figure 1. Flow chart for searching literatures

For instance, Ehlert et al., 2020 was examined the COVID-19 incidence rate and socio-economic status using spatial model; Spatial Autoregressive model (SAR), Spatial Error Model (SEM), and Spatial Autocorrelation model (SAC). It resulted the case and death rates were positively associated with early cases, mean age, population density, and the number of people employed in elderly care. On the other hand, they were negatively associated with not only the density of doctors but also the density of schoolchildren and infant care. de Souza et al., 2020 were shown that it analyzed the association between the incidence rate of COVID-19 and Municipal Human Development Index (MHDI), Social Vulnerability Index (SVI). As a result, SVI were positively associated with COVID-19 incidence rate, whereas MHDI were negatively associated with the incidence rate. According to Mollalo et al., 2020, the analysis of the association between socio-economic status and the incidence rate was performed by using spatial model; Spatial Error Model (SEM), Spatial Lag Model (SLM), Geographically Weighted

Regression (GWR), Multiscale GWR (MGWR). The result was shown that the incidence rate was associated with income inequalities, household income, medical expertise such as the proportion of practitioner and nurse, and the spatial variation was also showed with those significant variables. You et al., 2020 was examined the spatial analysis between the COVID-19 morbidity rate and social and economic factors; built environment, economic activities, and public service status. The results showed an increased COVID-19 morbidity associated with the density of increasing population and aged population, construction land area proportion, public green space density with positive estimated coefficients.

Otherwise, increasing average building scale, GDP per unit of land area, and hospital density were the negative estimated coefficients with the decreased morbidity.

Other studies (Lou et al., 2020, Maciel et al., 2020, Zhang et al., 2020) have shown the spatial correlation, distribution, and nonlinear relationship. In case of Lou et al., 2020, the study was employed geographically weighted random forest (GW-RF) to estimate the nonlinear relationship between COVID-19 death rate and risk factors such as going to work by walking, airborne benzene concentration, householder with a mortgage, unemployment, airborne $PM_{2.5}$ concentration, and percent of the black or African American. Those variables were highly associated with the spatial distribution of the death rate. Maciel et al., 2020 was analyzed the univariate and bivariate spatial correlation between the COVID-19 incidence rate and MHDI, and the result showed the moderate positive correlation and high-high clusters were formed and located in the metropolitan area and municipalities in the north region. According to Zhang et al., 2020, multiple regression model was used to investigate the risk factors. This study showed the positive correlation between the incidence and mortality of COVID-19 and socio-economic factors (population density, proportions of elderly residents, poverty, and percent population tested.

According to background and reviewed literatures, the associations between the various variables affecting the COVID-19 incidence rate were largely attributed to socioeconomic status (SES), environmental factors, and medical personnel resources. Ahmed et al., 2020 suggested that there would be socioeconomic disadvantages and inequalities in the occurrence of an epidemic. However, there are only a few examples of factors that occur in each administrative district of geographical variation using spatial models. Therefore, it needs to utilize the global/local spatial model to reflect regional characteristics in the Seoul Metropolitan Government's COVID-19 risk factor identification. The purpose of the study is to provide and utilize the basis for health policies appropriate to the district units by studying the factors affecting the rate of occurrence of COVID-19. On the hypothesis of this study, there is a significant spatial heterogeneity of COVID-19 incidence rate in the level of Seoul's district unit, socio-economic variables, and there is a significantly spatial difference in terms of the association between the incidence rate of COVID-19 and socio-economic variables.

2. MATERIALS AND METHOD

2.1. Study area

Seoul is the capital city and most populous city in South Korea. Seoul is one of the most important transportation hubs in South Korea. Seoul has many lines of subway so that many people could easily move from district to district. Since the first patient was occurred on the early of this year, On 16 August 2020, a total of 1,947 confirmed cases of COVID-19 were reported in Seoul. Administrative districts of Seoul have 25 districts: Jongno-gu, Jung-gu, Yongsan-gu, Seongdong-gu, Gwangjin-gu, Dongdaemun-gu, Jungnang-gu, Seongbuk-gu, Gangbuk-gu, Dobong-gu, Nowon-gu, Eunpyeong-gu, Seodaemum-gu, Mapo-gu, Yangcheon-gu, Gangseo-gu, Guro-gu, Geumcheon-gu, Yeongdeungpo-gu, Dongjak-gu, Gwanak-gu, Seocho-gu, Songpa-gu, Gangdong-gu. In this study, the spatial regression analysis was applied in the administrative divisions that refer to 25 districts in Seoul (Figure 2).

2.2. Data selection and preparation

The Korea Disease Control and Prevention Agency (KCDA) is monitoring daily the level of region, and district of COVID-19 across the South Korea. For this study, the COVID-19 data from KCDA daily report were retrieved and reuploaded to the Seoul open data website (data.seoul.co.kr). For the incidence rate of COVID-19, the age-standardized incidence rate was computed for the district and combined to the Seoul district boundary shapefile obtained from the National Spatial Data Infrastructure Portal (data.nsdi.go.kr).

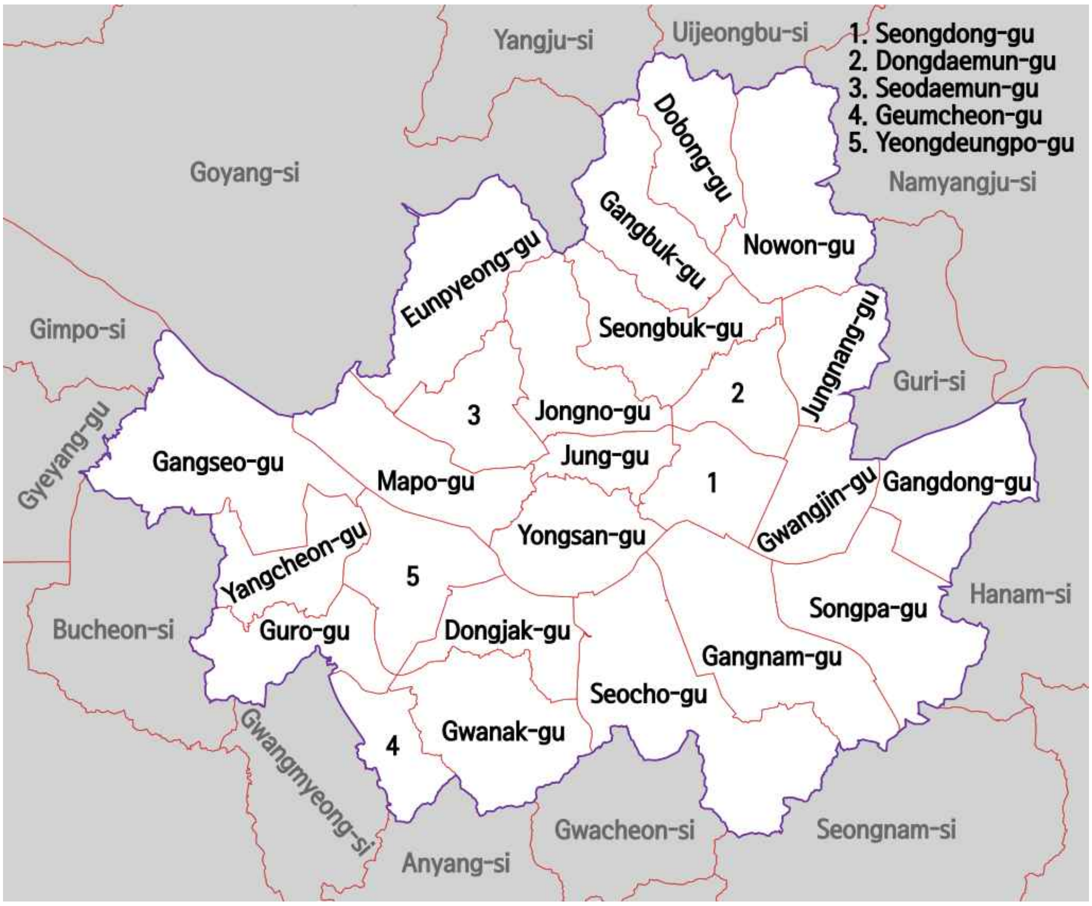

Figure 2. Administrative districts of Seoul

For socio-economic variables, average income, the proportion of doctors and nurses in public health center, and population density are included as explanatory variables, and sex, age of the confirmed patients, education level, economically active population, pure population mobility are included control variables for this study. All explanatory and control variables were retrieved from the Seoul open data website, public data portal (data.go.kr) and merged to the corresponding Seoul district area. For specific description of variables, table 1 provides the detailed explanation of variables.

A total of 1,947 confirmed cases from January 20 to August 16 were included in spatial analysis. The features of COVID-19 data in Seoul were involved infected patients, age-standardized population, and the incidence rate per 100,000 people each 25 districts in table 2. On the infected patients, 'Gwanak-gu' district has the most infected patients, which the number of patients is 156 people, whereas 'Jung-gu' district has the least infected patients, which the number of patients is 28 people. Generally, the number of female patients is more than male patients, and the average age of the COVID-19 patients about 46.1 year on the COVID-19 data features. 'Gwanak-gu' is the highest incidence rate of COVID-19 (31.64 per 100,000), while the lowest incidence rate of COVID-19 is the 'Gwangjin-gu' (12.35 per 100,000).

Table 1. Variables in this study

Variable name	Description	Source of data
Average income	The average income in the level of Seoul district	Public data portal
The proportion of doctors	The proportion of doctors in public health center	Seoul open data website
The proportion of nurses	The proportion of nurses in public health center	Seoul open data website
Population density	The total population of Seoul district/the area of Seoul district	Seoul open data website
Education level population	The proportion of people who graduated above college/university	Seoul open data website
Economic active population	The proportion of people who have occupation in Seoul district	Seoul open data website
Pure population mobility	Number of net migration people/mid-year population in Seoul district	Seoul open data website
Incidence rate of COVID-19	Age-standardized incidence rate in Seoul district (confirmed cases/age-standardized population*100,000)	Seoul open data website

Table 2. Features of COVID-19 data in Seoul district

District	Infected patients (person)	Number of male patients (person)	Number of female patients (person)	Mean age (year)	Age-standardized population (person)	COVID-19 Incidence rate (per 100,000)
Jongno-gu	33	16	17	40.9	146,603.00	22.51
Jung-gu	28	14	14	46.0	121,678.00	23.01
Yongsan-gu	69	39	30	52.0	218,821.00	31.53
Seongdong-gu	67	30	37	51.8	299,044.00	22.40
Gwangjin-gu	43	22	21	43.3	348,116.50	12.35
Dongdaemun-gu	63	32	31	44.9	340,267.50	18.51
Jungnang-gu	59	28	31	51.8	394,293.50	14.96
Seongbuk-gu	133	51	82	51.3	431,427.00	30.83
Gangbuk-gu	41	21	20	41.5	310,925.50	13.19
Dobong-gu	84	28	56	56.7	332,026.00	25.30
Nowon-gu	84	40	44	44.9	532,511.50	15.77
Eunpyeong-gu	85	40	45	44.6	473,512.00	17.95

Seodaemun-gu	48	27	21	44.0	303,057.00	15.84
Mapo-gu	69	28	41	44.4	366,806.50	18.81
Yangcheon-gu	81	43	38	42.7	455,764.50	17.77
Gangseo-gu	128	65	63	49.8	587,369.00	21.79
Guro-go	93	47	46	39.5	400,902.00	23.20
Geumcheon-gu	42	23	19	45.3	229,269.50	18.32
Yeongdeungpo-gu	79	41	38	49.7	360,221.00	21.93
Dongjak-gu	81	34	47	46.2	390,350.00	20.75
Gwanak-gu	156	71	85	39.5	493,055.00	31.64
Seocho-gu	86	38	48	50.5	423,950.00	20.29
Gangnam-gu	111	55	56	53.3	531,575.00	20.88
Songpa-gu	123	53	70	43.8	662,019.50	18.58
Gangdong-gu	61	32	29	45.0	425,410.00	14.34

2.3. Spatial pattern and autocorrelation

To identify the spatial pattern and autocorrelation of incidence rate of COVID-19, we draw the choropleth map using natural breaks degree of outcome variable, the incidence rate of COVID-19 and we examine the Global and Local Moran's I to quantify the similarity of the outcome variable among areas that are defined as spatially related and to identify local clusters and local spatial outliers using GeoDa software.

Moran's I generally ranged from +1 to -1. A zero value of Moran's I indicates the null hypothesis of no clustering, a positive Moran's I means positive spatial autocorrelation, which means similar areas of values are clustered. However, if there is a negative value of Moran's I, it indicates negative spatial autocorrelation, which means that neighboring areas do not tend to have similar attribute values.

Global Moran's I

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

Local Moran's I

$$I_i = \frac{n^2}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

Figure 3. Global and Local Moran's I statistics

2.4. Spatial regression analysis

To identify the association between the COVID-19 incidence rate in Seoul and the socio-economic variables, we examined the Ordinary Least Square (OLS) model for the residuals whether there is spatial autocorrelation. If there is a spatial autocorrelation of the residuals in OLS model, we examined the 2 different spatial models on global scale; SEM, SLM.

In the global level of the spatial analysis, the OLS model is the general regression method that examine the relationships between outcome variable and explanatory variables (Ward and Gleditsch et al., 2018). The general form of OLS is showed below.

$$y_i = \beta_0 + x_i\beta + \epsilon_i$$

where y_i is the COVID-19 incidence rate of Seoul as the outcome variable, β_0 is the intercept, x_i is the explanatory variable, β is a coefficient of each term, and ε_i is an error term on the level of Seoul district(i). It is assumed that the variance of the error is the same for all independent variable values. The use of i for slopes and independent variables indicates that in simple regression models, increasing only the slope and independent variables results in a multiple regression model.

The SEM model has the assumption of a spatial dependence in the error term of OLS and is fragmented the error term into two terms. The general form of SEM model is showed below (Anselin et al., 2003).

$$y_i = \beta_0 + x_i\beta + \lambda W_i\xi_i + \epsilon_i$$

where ξ_i means the spatial factor of the error, λ expresses the level of correlation between components at the level of Seoul district i.

The SLM model is assumed dependence of the space between the outcome variable and explanatory variable and incorporated dependence into the regression model with a “spatially lagged dependent variable”. The form of SLM model is showed below (Anselin et al., 2003).

$$y_i = \beta_0 + x_i\beta + \rho W_i y_i + \epsilon_i$$

where ρ indicates the spatial lag parameter, and W_i is the vector of spatial weight. For this weight matrix, it specifies the neighbors at location i, and relates the outcome variable to the explanatory variables at that locations.

To compare the spatial models, all spatial models including OLS were analyzed with the same variables. On the scale of global models, the spatial weight matrix was loaded with Queen’s contiguity and implemented with the matrix using GeoDa software. For local scale models, using AICc method, the optimal number of data is selected and applied to local regression coefficients to apply fixed Gaussian for spatial kernel. The adjusted R^2 and AIC were used to compare the local models when it explains the incidence rate of COVID-19 in Seoul district unit.

2.5. Ethical issues

The Institutional Review Board of Seoul National University approved our study (IRB No. E2010/003-004). Data for this study was downloaded after IRB approved and processed the analysis. Public data portal and Seoul open data are the integrated platforms that provides public data generated or acquired by public institutions in one place. This provides public data in various ways, including file data, open API, and visualization.

3. RESULTS

3.1. Descriptive analysis

The descriptive statistics of variables are shown in table 3. A total of 25 districts of Seoul city were included in analysis. The incidence rate of COVID-19 was implemented as the outcome variable, average income, the proportion of doctors in public health center, the proportion of nurses in public health center, population density, pure population mobility, the number of economic active population, and the number of people who graduated above college/university was examined as the explanatory variables for the spatial regression analysis.

The average of the incidence rate of COVID-19 is 20.50 (SD ± 5.25) per 100,000 population across the Seoul districts. The average income of 25 Seoul districts is about 1.417 (SD ± 0.092) at the level of 1,000,000 won unit. The average of the proportion of doctors and nurses in public health center among 25 districts is 9.28 (SD ± 2.55) and 48.53 (SD ± 4.70). The average population density is approximately 1.741 (SD ± 0.481) among 25 district areas. The average of pure population mobility is -0.57 (SD ± 1.06), and the proportion of economic active population is 46.96 (SD ± 1.806) on average. The mean proportion of people who graduated above college/university is 34.61 (SD ± 6.066) through the 25 districts in Seoul city.

Table 3. Descriptive statistics of variables (N=25)

Variable	Unit	Min	Max	Mean	Std.dev
Incidence rate of COVID-19	Per 100,000	12.35	31.64	20.50	5.25
Average Income	1,000,000₩	1.299	1.670	1.417	0.092
The proportion of doctors in public health center	%	3.00	14.00	9.28	2.55
The proportion of nurses in public health center	%	37.90	55.90	48.53	4.70
Population Density	10,000 Person/km^2	0.677	2.655	1.741	0.481
Pure population mobility	-	-2.70	1.80	-0.57	1.06

The proportion of economic active population	%	43.50	50.50	46.96	1.806
The proportion of people who graduated over college/university	%	25.10	50.40	34.61	6.066

3.2. Spatial pattern and autocorrelation

The spatial pattern of COVID-19 incidence rate in Seoul is shown in Figure 4. In Figure 4, the incidence rate was divided by 5 degrees using natural breaks. There are 6 districts (Seodaemun-gu, Gangbuk-gu, Nowon-gu, Jungnang-gu, Gwangjin-gu, Gangdong-gu) which the incidence rate is below 17.70, 6 districts (Eunpyeong-gu, Mapo-gu, Yangcheon-gu, Geumcheon-gu, Songpa-gu, Dongdaemun-gu) which the incidence rate is more than 17.70 and less than 20.29, 5 districts (Gangseo-gu, Yeongdeungpo-gu, Dongjak-gu, Seocho-gu, Gangnam-gu) which the rate is more than 20.29 and less than 22.40, 5 districts (Dobong-gu, Jongno-gu, Jung-gu, Seongdong-gu, Guro-gu) which the rate is more than 22.40 and less than 30.83, and 3 districts (Seongbuk-gu, Yongsan-gu, Gwanak-gu) which the rate is more than 30.83.

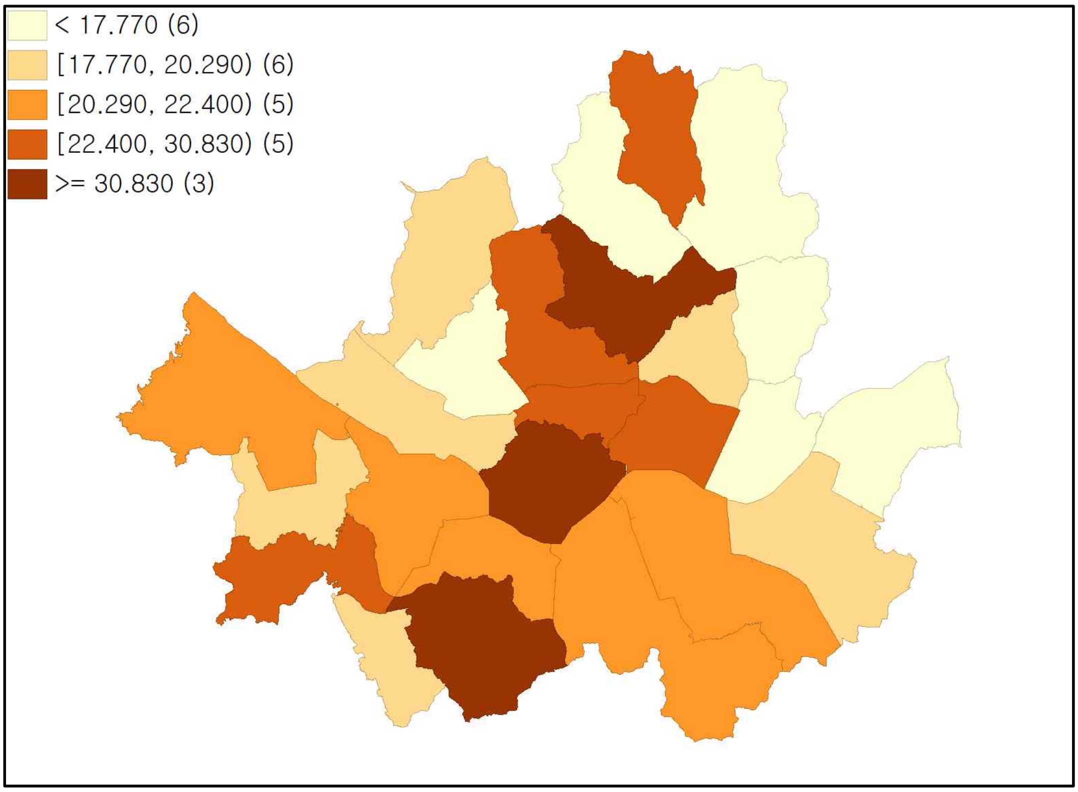

Figure 4. Spatial pattern of the incidence rate of COVID-19 in Seoul

Figure 5 showed the Global Moran's I value of the incidence rate of COVID-19 in Seoul. The result of spatial autocorrelation using Global Moran's I was -0.049 (p=0.498). This indicated it may have similar districts located each other but dispersed spatially due to the negative value of Moran's I.

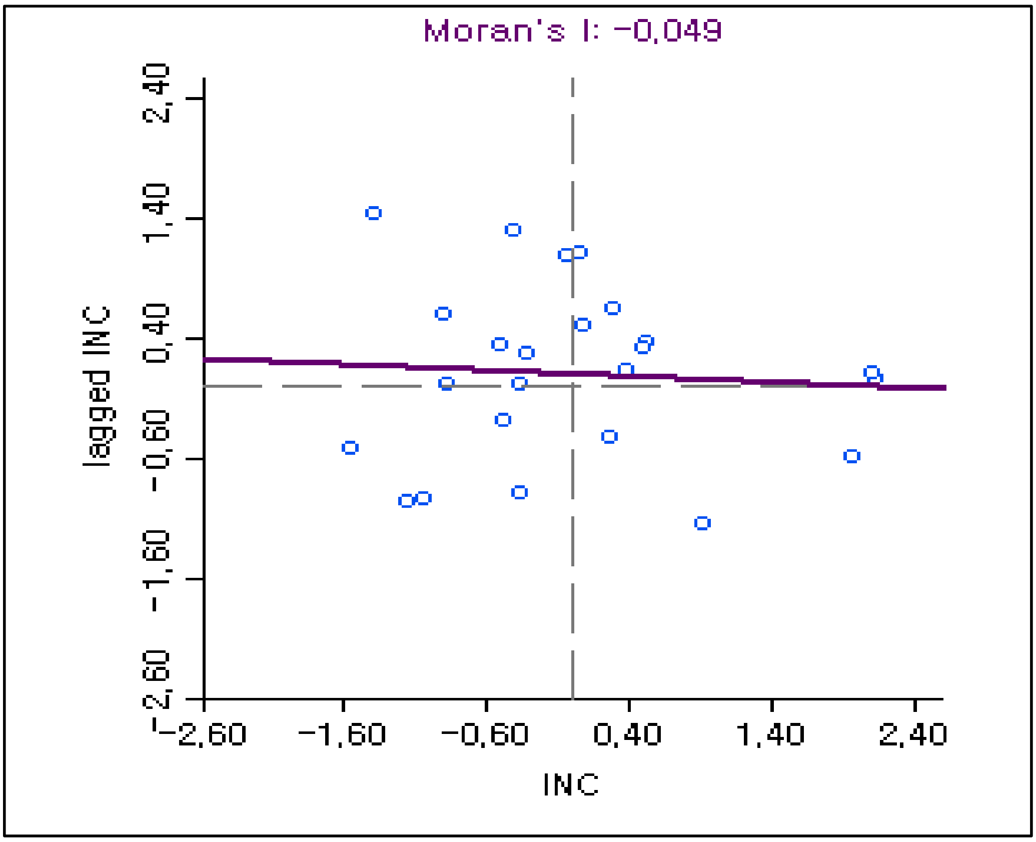

Figure 5. Global Moran's I of the incidence rate of COVID-19

To identify the local spatial correlation across the 25 districts, we also examined the Local Indicator of Spatial Association (LISA). According to LISA analysis within p-value is 0.05, study area which exists the spatial autocorrelation can visualize the hotspot and coldspot. The hotspot (High-High) is the area with a high value similar to the surrounding area, whereas the coldspot (Low-Low) is the area with a low value similar to the surrounding area. The result is shown in Figure 6. For the area of hotspot, Yeongdeungpo-gu and Dongjak-gu are included. In case of coldspot, Jungnang-gu and Songpa-gu are included.

Although it was the negative spatial correlation, it should be chosen the spatial regression model because it can occur the bias in estimating OLS. After it identifies the spatial autocorrelation in the residuals of OLS, the spatial regression model was investigated. A proper model selection was used R^2 and AIC.

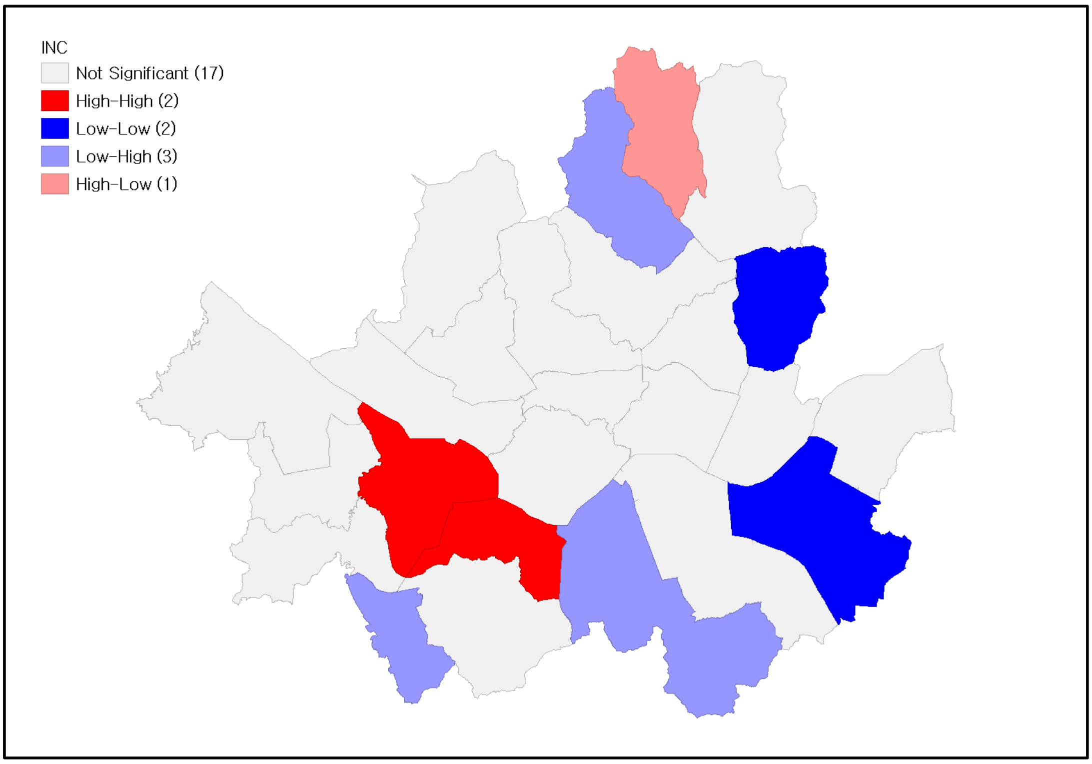

Figure 6. LISA of the incidence rate of COVID-19

3.3. Spatial regression analysis

In the scale of global estimation, the OLS, SLM, and SEM model was examined the association between the incidence rate of COVID-19 and socio-economic variables (table 4). In table 4, there is a positive relationship between the incidence rate of COVID-19 and some of socio-economic variables such as the proportion of doctors, the proportion of nurses, pure population mobility, and the proportion of people who graduated above college/university. It elaborated that high proportion of doctors and nurses in public health center, pure population mobility, and high proportion of people who graduated over college/university increased the incidence rate of COVID-19. On the other hand, the negative relationship between the incidence rate and socio-economic variables like average income, population density, and the proportion of economic active population was identified. It explained that high average income, population density, and economically active population could reduce the incidence rate across the districts. Considering to a selection with the value of R^2 and AIC, the SEM model (R^2: 0.365, AIC: 157.517) was selected in this study.

We investigated the correlations between all pairs of variables (Table 5). The largest Pearson's r was 0.84 between the average income and the proportion of people who graduated above college/university. In case of the incidence rate of COVID-19 and other variables, higher value was 0.36 with the variable, which is the proportion of nurses in public health center.

Table. 4 Regression table of the incidence rate of COVID-19 in association with socio-economic variables

Variable	OLS		SLM		SEM	
	Coeff. (Std.E)	t(p)	Coeff. (Std.E)	z(p)	Coeff. (Std.E)	z(p)
(intercept)	101.643 (64.413)	1.578 (0.133)	104.667 (54.827)	1.909 (0.056)	101.248 (53.158)	1.905 (0.057)
Average Income	-34.343 (28.800)	-1.192 (0.249)	-34.296 (23.856)	-1.438 (0.151)	-34.007 (23.731)	-1.433 (0.152)
The proportion of doctors in public health center	0.719 (0.443)	1.622 (0.123)	0.730 (0.365)	2.001* (0.045)	0.718 (0.366)	1.960* (0.050)
The proportion of nurses in public health center	0.412 (0.226)	1.818 (0.087)	0.404 (0.186)	2.165* (0.030)	0.410 (0.187)	2.192* (0.028)
Population Density	-2.003 (2.486)	-0.806 (0.431)	-2.025 (2.046)	-0.990 (0.322)	-2.004 (2.051)	-0.977 (0.328)
Pure population mobility	0.607 (1.074)	0.565 (0.580)	0.527 (0.885)	0.596 (0.551)	0.580 (0.885)	0.655 (0.512)

The proportion of economic active population	-1.471 (0.934)	-1.575 (0.134)	-1.505 (0.780)	-1.930 (0.054)	-1.470 (0.770)	-1.909 (0.056)
The proportion of people who graduated over college/university	0.399 (0.387)	1.032 (0.317)	0.408 (0.321)	1.270 (0.204)	0.397 (0.319)	1.245 (0.213)
Lamda					-0.016 (0.266)	-0.061 (0.952)
R^2	0.365		0.369		0.365	
Log likelihood	-70.760		-70.699		-70.758	
Akaike info criterion	157.519		159.398		157.517	

$p<0.05$*/OLS (Ordinary Least Square), SLM (Spatial Lag Model), SEM (Spatial Error Model)

Table. 5 Correlation coefficient matrix table of variables

	Incidence rate of COVID-19	Average Income	The proportion of doctors in public health center	The proportion of nurses in public health center	Population Density	Pure population mobility	The proportion of economic active population	The proportion of people who graduated above
Incidence rate of COVID-19	1							
Average Income	0.13	1						
The proportion of doctors in public health center	0.17	−0.14	1					
The proportion of nurses in public health center	0.36	0.15	0.00	1				
Population Density	−0.25	−0.38	0.17	−0.16	1			
Pure population mobility	0.22	0.19	0.00	−0.03	0.07	1		
The proportion of economic active population	−0.25	−0.49	0.33	0.08	0.4	−0.32	1	
The proportion of people who graduated above college/university	0.12	0.84	−0.13	0.2	−0.21	0.16	−0.15	1

4. DISCUSSION

Using 1,947 confirmed cases of Seoul COVID-19 data, this study searched for the exploration of the spatial distribution, autocorrelation, and the relationship between the socio-economic variables and the incidence rate of COVID-19 in the level of districts using spatial regression models. Based on the finding, global Moran's I of the incidence rate of COVID-19 was a negative value (I = -0.049) but could find the hotspots and coldspots under the p-value ($p<0.05$). Combined with the incidence rate and socio-economic variables for using spatial regression model, the proportion of doctors and nurses variables were significantly associated with the COVID-19 incidence rate, indicating that the area which is the high value of proportion of doctors and nurses in public health centers could have a protection from the incidence rate of COVID-19. Examined global-scale spatial regression models on this study, SEM model was shown better to elaborate the association between the incidence rate of COVID-19 in Seoul and socio-economic variables.

Some studies (You et al., 2020, Zhang et al., 2020, Pathak et al., 2020) were shown that the increasing population density can increase the COVID-19 incidence rate. Compared to those studies, the population density variable was not significantly associated with the incidence rate. This is because population density did not impact the movement. In other words, the people of movement could be considered about causing the incidence rate of COVID-19. Seoul had the widest transportation including buses and subways so that people can easily move from district to district. In case of pure population mobility, it was significantly associated with the COVID-19 incidence rate. With the study of Jiang et al., 2020, population mobility was discovered by the rapid transmission, and could likely to be infected by COVID-19 within 11-12 days after people moved from Wuhan to other cities. Also, some studies mentioned the asymptomatic carriers could contribute the rapid transmission of the COVID-19 occurrence (Yu et al.,2020, Bai et al., 2020).

Although the average income variable of each district of Seoul was not significant, it may tend to consider that people in a low average income area could be more occurred the COVID-19 infection. Regarding on the study (Whittle et al., 2020), it showed the COVID-19 positivity rate and socioeconomic predictors (low-income, population density, youth population, proportion of black population) had associated each other, and it addressed the public health management is needed during the pandemic of COVID-19. This is because people living in a low-level income community may have less access to high quality of healthcare, and workers who are also in the level of lower income may go to work outside to get paid and more likely to have a change to get infected the COVID.19 (U.S. Bureau of Labor Statistics, 2018)

In relation of medical experts such as primary care physicians and nurse, the proportion of nurse practitioner was the key to manage the COVID-19 incidence rate (Mollalo et al., 2020). Similarly, the proportion of doctors and nurses in public health center may tend to reduce the COVID-19 incidence rate in this study which the variables were significant. Previous study emphasized the importance to train young professionals on the frontline, but healthcare experts are still lacked compared to OECD average in case of South Korea, especially in public health center. To cope with the COVID-19, the proportion of healthcare professionals are needed and properly deployed against the outbreak.

One of the limitations of this study was data availability. Due to Personal Information Protection Act of COVID-19 patient pathway, a basic demographic information of the confirmed cases was difficult to know and calculate the incidence rate from the Seoul open data portal. Another limitation was the scale of the district of Seoul. Another limitation is that this study did not analyze at the level of local unit so that Modifiable Area Unit Problem (MAUP) could be possible occurred (Jung et al., 2015). In response, the study argued that it is desirable to use as a small-scale unit as possible in small area or micro-data studies because the larger the scale, the greater the degree to which the characteristics of personal information collected from the sensors are lost. For each examined variable, it did not consider the multicollinearity so that the power of explanation of the model and the fitness of model that used in this study could be decreased.

While it showed the relationship between variables had a low explanation, there are some reasons to argue the limitation. According to the report (Kang et al., 2007), point pattern analysis basically is implemented with x, y coordinates but it cannot show that observed events were always associated if the event is occurred at the same place. This is because even if correlations are estimated through spatial concentration, there may be no real significant associations between events that are assumed to be spatially related if there are serious gaps in the time of occurrence for each event. Further direction of this study could be suggested that spatiotemporal point pattern analysis should recognize the limitations of point pattern analysis that emphasize only spatial aspects and expand the causal relationship between events appearing in space to time to find more reliable interrelationships. About the coefficient of spatial regression results, the value of standard error of intercept and average income showed widely. It indicated there is a difference between estimated value and parameter. However, in case of South Korea, even though the structure of the population is very diverse, there was no dependence on socio-economic factors. This suggests the possibility that our social economy is not developing in a way that meets the diverse needs of our communities.

Despite the limitations, this study showed the spatial pattern and captured the hotspot and coldspot using the COVID-19 incidence rate across the Seoul city. It is the first study to examine the relationship between the incidence rate of COVID-19 and socio-economic variables using the spatial regression model even though few variables were significant. According to this study findings, it suggests socio-economic factors related to COVID-19 should consider for health policy and need a variety of experts to cease the pandemic. As social distancing has been reorganized and implemented recently, providing economic supports and health supplements should consider and proceed to people in Seoul where the district is lower income, high density of elderly population.

5. CONCLUSION

This study was focused on the COVID-19 incidence rate across the 25 districts of Seoul city. In relation of spatial distribution, it had negative spatial inequalities but the hotspots and coldspots were captured across the district in Seoul. We employed the spatial regression model in a way of global (OLS, SLM, SEM) scale. The SEM model was a good fitting model to explain the relationship between socio-economic variables and the COVID-19 incidence rate despite many explanatory variables were not significant. To the best of our knowledge, there is a lack of research on using geographic model of COVID-19 in South Korea. So, this study can be regarded as a basis for future geographic modeling of the diseases.

REFERENCES

Ahmed, F., Ahmed, N., Pissarides, C., Stiglitz, J., (2020). Why inequality could spread COVID-19. Lancet Public Health https://doi.org/10.1016/S2468-2667(20)30085-2 (Published online ahead of print).

Anselin, L., (2003). Spatial externalities, spatial multipliers and spatial econometrics. Int. Reg. Sci. Rev. 26 (2), 153 - 166.

Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D. Y., Chen, L., & Wang, M. (2020). Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA*, *323*(14), 1406 - 1407. https://doi.org/10.1001/jama.2020.2565

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr. Anal. 28 (4), 281 - 298.

Das, A., Ghosh, S., Das, K., Basu, T., Das, M., & Dutta, I. (2020). Modeling the effect of area deprivation on COVID-19 incidences: a study of Chennai megacity, India. Public health, 185, 266 - 269. Advance online publication. https://doi.org/10.1016/j.puhe.2020.06.011(Alkhamis et al., 2020)

Eugene, C. (2018). Determinants of Size of Social Economy in Korean Municipalities: Spatial Regression Approach. The Korean Journal of Local Government Studies, 22(2), 455-476.

Fotheringham, A.S., Yang,W., Kang,W., (2017). Multiscale geographically weighted regression

(MGWR). Annals of the American Association of Geographers 107 (6), 1247 - 1265.

IMF. (2020). World Economic Outlook, 2020 Apr Chapter 1

Jiang, J., Luo, L. Influence of population mobility on the novel coronavirus disease (COVID-19) epidemic: based on panel data from Hubei, China. *glob health res policy* 5, 30 (2020). https://doi.org/10.1186/s41256-020-00151-6

Jung, J., Park, J., (2015). Test of the Scale Effect of MAUP in Crime Study: Analyses of Sex Crime Using Nation-Wide Data of Eup-Myon-Dong and Si-Gun-Gu, 15(10), 150-159.

Maciel, J. A. C., Castro, S., II, & Farias, M. R. (2020). Initial analysis of the spatial correlation between the incidence of COVID-19 and human development in the municipalities of the state of Ceará in Brazil. Rev Bras Epidemiol, 23, e200057. doi:10.1590/1980-549720200057

Mollalo, A., Mao, L., Rashidi, P., Glass, G.E., 2019. A GIS-based artificial neural network model for spatial distribution of tuberculosis across the continental United States. Int. J. Environ. Res. Public Health 16 (1), 157.

Mollalo, A., Sadeghian, A., Israel, G.D., Rashidi, P., Sofizadeh, A., Glass, G.E., 2018. Machine learning approaches in GIS-based ecological modeling of the sand fly Phlebotomus papatasi, a vector of zoonotic cutaneous leishmaniasis in Golestan province, Iran. Acta Trop. 188, 187 - 194.

Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. The Science of the total environment, 728, 138884.

https://doi.org/10.1016/j.scitotenv.2020.138884

Qi, H., Xiao, S., Shi, R., Ward, M. P., Chen, Y., Tu, W., Su, Q., Wang, W., Wang, X., & Zhang, Z. (2020). COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis. The Science of the total environment, 728, 138778. https://doi.org/10.1016/j.scitotenv.2020.138778

United Nations, (2020). The Social Impact of COVID-19.

U.S. Bureau of Labor Statistics (2018). Available online at: https://www.bls.gov/charts/american-time-use/work-by-ftpt-job-edu-p.htm (accessed June 6, 2020)

Ward, M.D., Gleditsch, K.S., (2018). Spatial regression models. 155. Sage Publications.

Weinstein, Bayarmagnai & Silva, Alan & Kouzoukas, Dimitrios & Bose, Tanima & Kim, Gwang-Jin & Correa, Paola & Pondugula, Santhi & Kim, Jihoo & Carpenter, David. (2020). A methodological blueprint to identify COVID-19 vulnerable locales by socioeconomic factors, developed using South Korean data. 10.1101/2020.10.26.20218842v1.

World Health Organization (WHO), 2020a. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19).

World Health Organization (WHO), 2020b. Coronavirus Disease 2019 (COVID-19) Situation Report – 192

World Health Organization (WHO), 2020c. Rolling Updates on Coronavirus Disease (COVID-19).

Wu, X., Nethery, R.C., Sabath, B.M., Braun, D., Dominici, F., (2020).

Exposure to Air Pollution and COVID-19 Mortality in the United States. https://doi.org/10.1101/2020.04.05.20054502.

You, H., Wu, X., & Guo, X. (2020). Distribution of COVID-19 Morbidity Rate in Association with Social and Economic Factors in Wuhan, China: Implications for Urban Development. International journal of environmental research and public health, 17(10), 3417. https://doi.org/10.3390/ijerph17103417

Yu, P., Zhu, J., Zhang, Z., & Han, Y. (2020). A Familial Cluster of Infection Associated With the 2019 Novel Coronavirus Indicating Possible Person-to-Person Transmission During the Incubation Period. *The Journal of infectious diseases*, *221*(11), 1757 - 1761. https://doi.org/10.1093/infdis/jiaa077

Zheng, YY., Ma, YT., Zhang, JY. (2020). COVID-19 and the cardiovascular system. *Nat Rev Cardiol* 17, 259 - 260. https://doi.org/10.1038/s41569-020-0360-5

강호제. (2007). 시공간분석기법 : 시공간분석의 변화와 응용. 국토, 104-111.

김태완, (2020). 코로나19로 인한 영향 및 사회정책 대응 방안. 보건복지 ISSUE & FOCUS 제385호 1-8

질병관리본부(KCDC), (2020). 코로나바이러스감염증-19 국내 발생 현황(8월 11일 정례브리핑)

국문초록

서울시 COVID-19 발생률과 사회경제적 요인 간 연관성에 대한 공간 분석

김 정 헌
보건학과 시공간역학전공
서울대학교 보건대학원

연구배경: 코로나19는 전 세계적으로 유행병이 되어 많은 사람들이 위협받고 있다. 코로나19 국민예방수칙이 시행되었지만, 계속되는 집단발병은 다양한 장소에서 발생되고 있다. 코로나19를 분석한 많은 연구가 사회경제적 요인 간의 관계를 확인했지만, 공간회귀모형을 사용하여 분석한 연구는 거의 없다. 이에 본 연구는 코로나19 발생률에 미치는 요인을 연구함으로써 시군구 단위에 적합한 보건 정책의 근거를 제공하고 활용하고자 한다.

연구방법: 서울시 코로나19 확진자 현황 데이터를 사용하여 공간패턴을 수행하였다. GeoDa 소프트웨어를 사용하여 전국적/국지적 Morans' I를 통한 공간자기상관성을 탐색하였다. 그리고 전국적 공간 회귀 모형의 경우, GeoDa 소프트웨어를 사용하여 OLS, SLM, SEM을 수행하여 분석하였다.

연구결과: 서울시 25개의 자치구를 통하여 코로나19 발생률이 음의 공간자기상관성(I=-0.049)으로 나타났지만, 25개 자치구 내 핫스팟과 콜드스팟을 포착할 수 있었다. 코로나19 발생률과 사회경제적 요인 간 공간회귀 분석을 수행하였고, 자치구별 보건소 내 의사 및 간호사 비율에 대한 변수가 유의하게 나타났다. 그리고 모형 적합 측면에서 OLS, SLM, SEM의 결정계수는 0.365, 0.369, 0.365이며, AIC값은 157.519, 159.398, 157.517로 나타났다.

결론: 코로나19 발생률을 통해서 서울시 25개의 자치구 내 공간패턴을 핫스팟과 콜드스팟으로 확인할 수 있었다. 비록 다른 변수들이 유의하게 나타나지는 않았으나 경제활동을 위하여 사람들은 이동을 할 수 밖에 없다는 점을 시사하고 있다. 이와 관련하여 코로나19의 발생을 억제하기 위해서 서울시는 25개의 자치구별 건강 정책이 구축되어야 하고, 공공의료 인력도 확대되어야 한다.

주요어 : 코로나19, 발생률, 공간 분석, 사회경제적 요인
학 번 : 2019-27919