Significant Correlates of COVID-19 Risks in the Philippines using Spatial Modeling and Mapping

Pauline Gail Dela Roca[1], Jasper Victor Miranda [2], Shirl Angelee Ocampo[3], Timothy Olvina[4], Regina Tresvalles[5] and Shirlee Ocampo[6*]

Department of Mathematics and Statistics, De La Salle University
[1] email: pauline_delaroca@dlsu.edu.ph
[2] email: jasper_miranda@dlsu.edu.ph
[3] email: shirl_ocampo@dlsu.edu.ph
[4] email: timothy_olvina@dlsu.edu.ph
[5] email: regina.tresvalles@dlsu.edu.ph
*[*6] email: shirlee.ocampo@dlsu.edu.ph*

Abstract: COVID-19 risks are affected by demographic, healthcare, environmental, economic, comorbidities and many other factors. In line with the United Nations (UN) Sustainable Development Goal (SDG) 3 on health and well-being, this research has identified the significant correlates of COVID-19 standard incidence ratio (SIR) in the Philippines using multiple linear regression model (MLRM) and principal component analysis. However, regression models are not appropriate when applied to spatially distributed data. Moran's I showed positive spatial autocorrelation among provinces and National Capital Region (NCR) cities in terms of COVID-19 risks. Thus, this research postulated a spatial weight matrix and applied spatial autoregressive (SAR) model to determine significant correlates of provincial COVID-19 risks. It also integrated the obtained regression-based significant factors to a spatial model that takes into account spatial autocorrelations. SAR model which takes account spatial dependencies resulted in low mean square error (MSE) and mean absolute deviation (MAD). Results show that the significant health care correlates of COVID-19 relative risks are hospital beds allocated for COVID-19 patients and licensed COVID-19 testing centers. The significant demographic correlates are proportions of seniors and unemployed. The significant urbanity correlates are number of municipalities and revenue collection. Air quality index and wind speed are significant environmental correlates. Comorbidities like bronchitis, pneumonia, heart disease and hypertension are significantly correlated to COVID-19 risks. The SAR model also indicated that oral health is significantly related to COVID-19 risks. Spatial maps highlighted clustering of COVID-19 high-risk areas such as "National Capital Region plus bubble", and Benguet with nearby Northern provinces.

Key Words: spatial modeling; spatial mapping; COVID-19 risk; SAR; regression

1. INTRODUCTION

The Coronavirus disease 2019 (COVID-19) has seen an exponential increase of cases and deaths worldwide since it was first reported by the Wuhan Municipal Health Commission. The World Health Organization (WHO) officially declared COVID-19 as a pandemic on March 11, 2020 (WHO, 2020). As of June 12, 2022, the Philippines recorded 3.69 million COVID-19 cases and 60,041 deaths. The COVID-19 pandemic had a devastating impact on the nation's welfare, economy, and social aspects.

Only few factors correlated with COVID-19 cases were identified in the context of the Philippines. Chan et al. (2021) and Medina (2020) applied regression models on Philippines COVID-19 data. The research of Medina (2020) only focused and quantified age as a risk factor that influences COVID-19 mortality in the Philippines. Chan et al. (2021) mentioned that further dependent variables may be considered in their analysis. The incorporation of more significant correlated factors of COVID-19 was also recommended for more accurate and better models. Clouston, et al. (2021) and Kapitsinis (2020) correlated the risk of transmission of COVID-19 with environmental variables. Several papers that produced and evaluated statistical and spatial models recommend better models that could explain the correlation of factors to COVID-19 data and estimate the prevalence of COVID-19 with lesser errors. Thus, this study identified the significant COVID-19 correlates by performing regression and spatial models. It postulated a spatial weight matrix and incorporated it in a spatial model. It created spatial maps of COVID-19 standardized incidence ratios using National Capital Region (NCR) cities and provincial spatial units indicating which are high-risk areas. This research is in accordance with UN Sustainable Development Goal 3 on health and well-being and could aid local government units in better health management and policy implementation.

2. METHODOLOGY

2.1. Data

COVID-19 data were obtained from the Department of Health COVID-19 tracker website. (DOH, 2021). The daily COVID-19 cases from April 1, 2020 to September 15, 2021 were added for the entire period and aggregated per province and NCR city. The aggregation was done from the data downloaded from DOH website where there is an indicated province of residence of each COVID-19 case. COVID-19 cases with no indicated province of residence were not considered in the study.

The correlates were obtained from Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA, 2020), Department of Environment and Natural Resources (DENR, 2020), Philippine Statistics Authority (PSA, 2020), and Field Health Service Information System (FHSIS, 2020). The study is limited to the available data. The accurate recording of COVID-19 cases and their province of residence is beyond the scope of the study.

A list of the variables used in this study with their corresponding codes and descriptions are enumerated in Table 1.

Table 1. Variables and Descriptions

Variable Code	**Variable Descriptions**
Dependent Variable	
COVID-C	Number of COVID-19 cases
SIR	Ratio of observed number of COVID-19 cases by the expected number of COVID-19 cases
LN_SIR	Value of the natural logarithm of SIR
Healthcare-Related Correlates	
HB	Number of hospital beds
HW	Number of healthcare workers
LCTL	Number of Licensed COVID-19 Testing Laboratories
Healthcenter	Number of health centers and barangay health stations
Demographic and Urbanity Correlates	
Sen	Number of senior citizens
Minors	Number of individuals aged 18 and below
M	Number of male individuals
Ue	Number of unemployed persons
NB	Number of barangays
NM	Number of municipalities
NC	Number of cities
NBB	Number of births of both sexes
RC	Revenue collection
Environment-Related Correlates	
RF	Amount of rainfall
T	Mean temperature

RH	Relative humidity
WindSpeed	Wind speed
AQI	Air quality index
Disease-Related Correlates	
ORAL_H	Number of individuals who received Basic Oral Health Care
TB	Number of Tuberculosis Cases
LRD	Number of lifestyle-related diseases (tobacco and alcohol use, unhealthy diet) cases
CancerBC	Number of women who screened for cervical and breast cancer
Cardio_Diabetes	Number of individuals with with hypertension and type 2 diabetes
ImmuSen	Number of senior citizens with Pneumococcal Polysaccharide Vaccine and Influenza Vaccine
Bronc	Number of bronchitis Cases
COPD	Number of Chronic Obstructive Pulmonary Disease Cases
Hyp	Number of hypertension cases
ILI	Number of cases of Influenza-like Illness
Influ	Number of influenza cases
Pneu	Number of pneumonia cases
Spatial Unit	
Prov	Name of province

2.2. Theoretical Framework and Analysis

2.2.1. Standard Incidence Ratio (SIR)

SIR functions to ascertain if the occurrence of a disease in a population is high or low. It predicts if the number of observed COVID-19 cases in a particular i^{th} geographic area, i = 1, 2...m, is higher or lower than expected. A geographic area or spatial unit refers to a province or NCR city in the Philippines.

The common risk r was calculated by

$$r = \frac{y}{N} \quad , \qquad \text{(Eq. 1)}$$

where:

y = the total count of COVID-19 cases, and
N = the total population exposed to risk.

SIR is the most common statistic to estimate the relative risk θ_i for spatial unit i. It is given by

$$\theta_i = \frac{y_i}{e_i} \quad , \qquad \text{(Eq. 2)}$$

$$e_i = rN_i \qquad \text{(Eq. 3)}$$

where:

e_i = the expected count of spatial unit i,
y_i = the count of cases in spatial unit i,
N_i = the population exposed to risk in spatial unit i, and
m is the total number of spatial units.

If the SIR is one, the actual number of cumulative cases in a spatial unit equals the expected number of COVID-19 cases for that area. If the SIR is less than one, the actual number of cumulative COVID-19 cases in a spatial unit is fewer than the projected number of COVID-19 cases. If the SIR is larger than one, the actual number of cumulative COVID-19 cases in NCR city or province exceeds the expected number of COVID-19 cases in that spatial unit. A risk ratio greater than one indicates an elevated or higher risk of a person from the exposed population to be positive for COVID-19.

2.2.2. Spatial Mapping and Moran's I

Spatial maps are used for data visualization and investigating clustering of diseases. A Philippine map with provincial boundaries was downloaded and integrated in Geoda. Spatial maps of COVID-19 cases and SIR values were created, and clustering of high-risk provinces were observed. The locations with the same shade in color indicate clustering, meaning that there is a high risk for COVID-19 due to the nearness of the values between COVID-19 cases, and SIRs of the provinces and NCR cities.

Moran's I is an index that measures the overall spatial autocorrelation of the data. It includes a test of significance for spatial dependencies between locations. Rejection of the null hypothesis in this test will lead to the conclusion that there exist spatial dependencies among the spatial units.

2.2.3. Multiple Linear Regression Model

Multiple linear regression model (MLRM) determines the linear relationship on several predictors with a response variable. The dependent variable Y_i is a function of p predictors $X_{1i}, X_{2i}, \ldots X_{pi}$, and their association can be written as:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \varepsilon_i \,,$$

$$i = 1,2, \ldots m \qquad \text{(Eq. 4)}$$

where:

Y_i = the dependent variable of observation i,
β_i = the regression parameter of X_i,

$X_{1i}, X_{2i}, \ldots X_{pi}$ = the predictors or correlates
m = the total number of spatial units
ε_i = random error, independent from one another and $\varepsilon_i \sim N(0,\sigma^2)$.

2.2.4. Spatial Autoregressive Model

The spatial model is based on the spatial autoregressive model (SAR) of (Pace, et al., 1998) that takes into account errors exhibiting spatial autocorrelation. The SAR model is

$$(\boldsymbol{I} - \boldsymbol{\alpha D})\boldsymbol{Y} = (\boldsymbol{I} - \boldsymbol{\alpha D})\boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad \text{(Eq. 5)}$$

where:
$\boldsymbol{Y}$ = n by 1 vector of observations on the dependent variable,
$\boldsymbol{X}$ = n by k matrix of observations of the independent variables of interest,
$\boldsymbol{\beta}$ = k by 1 vector of parameters,
$\boldsymbol{\varepsilon}$ = n by 1 vector of errors, and
$\boldsymbol{D}$ represents an n x n spatial weighting matrix with zeros on the diagonal and non-negative entries on off-diagonals.

The optimal smoothing constant $0 \leq \alpha < 1$ is 0.5. In this research, a spatial weighting matrix $\boldsymbol{D}$ is formulated to have weights

$$w_{ij} = \begin{cases} 1, if\ spatial\ units\ i\ and\ j\ belong\ to\ the\ same\ region. \\ 0, if\ spatial\ units\ i\ and\ j\ belong\ to\ different\ regions \end{cases}$$

2.3 Statistical Analyses

COVID-19 data and correlates were cleaned, encoded, and merged. Principal component analysis (PCA) was applied to the correlated independent variables to remove multicollinearity. SIR was computed in each spatial unit. Multiple linear regression modeling (MLRM) using forward stepwise variable selection was then applied with SIR as the dependent variable. Diagnostic checking of MLRM assumptions was performed. Natural logarithmic transformation of the response variable SIR was applied so that MLRM assumptions were satisfied. Moran's I was calculated to determine if there is a significant spatial autocorrelation. Spatial maps were created to identify clustering of provinces based on SIR. Spatial autoregressive (SAR) model was fitted to the data. Significant correlates of COVID-19 risks were obtained. The software used are STATISTICA, Microsoft Excel, and SAS.

3. RESULTS AND DISCUSSION

3.1 Spatial Maps and Moran's I

The spatial map of provincial SIRs is shown in Figure 1. A clustering of high-risk spatial units can be seen in 'NCR plus bubble' and spreading to nearby provinces as the shades become lighter from the epicenter which is NCR. Another clustering can be observed in Cebu City and nearby areas. On the other hand, a cluster of high-risk areas can be observed in Benguet and nearby Northern provinces where Baguio City is located. Another clustering of high-risk areas can be noted in Davao and nearby provinces.

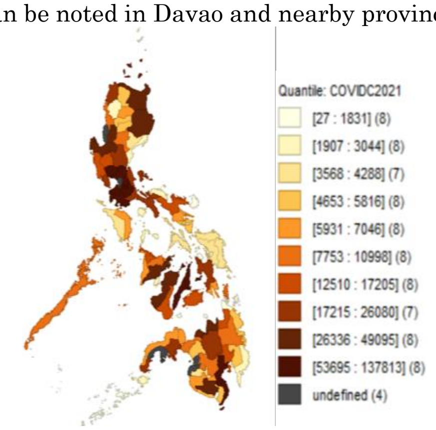

Figure 1. SIR of COVID-19 cases per Province in 2021

Moran's I is equal to 0.391 showing positive spatial autocorrelation. It indicates that the spatial distribution of high COVID-19 risks among provinces and NCR cities are more spatially clustered. Results show that there exist spatial dependencies (p-value < 0.05) on COVID-19 risks across the provinces and NCR cities. Figure 2 shows clustering of COVID-19 high risk areas colored in red shade. Clustering of high-risk areas are shown in 'NCR plus bubble' supporting the reported statement that NCR is the epicenter of COVID-19 epidemic. Another clustering of COVID-19 high risk areas is shown in Northern provinces.

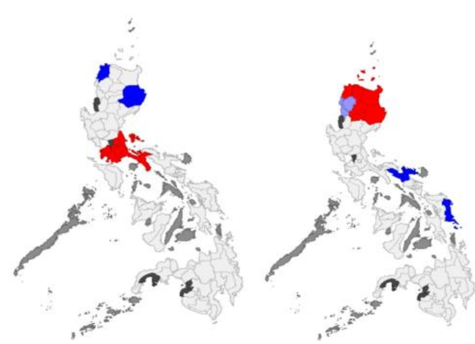

Figure 2. Cluster Maps of COVID-19 Risks

3.2 Multiple Linear Regression Model

MLRM was applied using the computed SIR with natural logarithmic transformation. The results satisfied all the assumptions of MLRM such as normality (p=0.3388), homoscedasticity (p=0.3755), and independence (DW = 1.71) of the residuals. Table 2 shows the MLRM results.

Table 2. MLRM Significant Correlates of COVID-19

Variables	Parameter Estimates	t-value	p-value
Intercept	-2.5	-5.62519	0.000000
P-Ue	2.5	2.37516	0.019819
P-LCTL	55379.7	2.16471	0.033248
P-Bronc	48.3	4.26702	0.000052
P-NM	-14765.1	-4.30476	0.000045
AQI	0.1	4.29474	0.000047
P-Sen	15.2	3.53197	0.000673
P-RC	4.4	3.14278	0.002313
P-CancerBC	16	2.26918	0.02582
WindSpeed	-0.2	-2.22166	0.028996
P-Pneu	-118.3	-2.33678	0.021832
P-HW	74.6	1.98972	0.049875
P-NB	-660.1	-2.09696	0.039004
P-NBB	-36.9	-2.04723	0.043761

3.3 Spatial Autoregressive Model

Since there exist significant spatial dependencies among provinces and NCR cities, a spatial autoregressive (SAR) model was fitted to the data. Table 3 shows the significant correlates of COVID-19 risks under the SAR model.

Table 3: SAR Significant Correlates of COVID-19 Risk

Variable	Parameter Estimate	t-value	p-value
Intercept	-0.6429	-11.4912	<0.001
WindSpeed	-0.18386	-1.74	0.0867*
AQI	0.08309	2.19	0.0321**
P_Bronc	25.18763	1.87	0.0656*
P_HB2	-0.00024	-1.94	0.0566*
P_Hyp	15.08899	1.8	0.077*
P_LCTL	71154	2.46	0.0165*
P_NM	-14458	-3.56	0.0007**
P_ORAL_H	46.93923	2.13	0.0365**
P_Pneu	-137.156	-2.56	0.0127**
P_RC	8.49946	5.55	<.0001**
P_Sen	11.91348	1.91	0.0609*
P_Ue	4.5355	2.18	0.0326**

* significant at 10% level of significance
** significant at 5% level of significance

The study further explored filtering of the significant MLRM correlates by integrating them in the SAR model. Results of MLRM-SAR model are shown in Table 4.

Table 4. MLRM-SAR Model Significant Correlates of COVID-19 Risks

Variable	Parameter Estimate	t-value	p-value
Intercept	-0.20165	-0.72	0.4753
P_Ue	3.68272	2.14	0.0355**
P_LCTL	63128	2.38	0.0195**
P_Bronc	24.95337	2.14	0.0349**
P_NM	-10247	-2.8	0.0063**
AQI	0.07709	2.19	0.0314**
P_Sen	13.03355	2.84	0.0056**
P_RC	9.19731	6.34	<.0001**
P_CancerBC	7.57365	1.09	0.2772
WindSpeed	-0.18524	-1.8	0.0747*
P_Pneu	-114.282	-2.27	0.0255**
P_HW	33.53013	0.83	0.4092
P_NB	-214.454	-0.59	0.5581
P_NBB	-45.2786	-2.32	0.023**

* significant at 10% level of significance
** significant at 5% level of significance

The predictive abilities of the models are compared using mean square error (MSE) and mean absolute deviation (MAD) in Table 5.

Table 5. Predictive Measures of Models

Model	MSE	MAD
MLRM	0.563	0.453
SAR	1.116	3.341
MLRM-SAR	10366.62	22.738

MLRM yielded the lowest MSE and MAD but is not that appropriate due to significant spatial dependencies among the provinces and NCR cities. SAR model is more appropriate with low MSE and low SAR.

Under the SAR model, the significant healthcare correlates of COVID-19 risks are hospital beds for COVID-19 patients and licensed COVID-19 testing centers. More hospital beds allocated for COVID-19 patients indicate better health management. A higher number of hospital beds for

COVID-19 patients would help reduce COVID-19 risks. According to Kapitsinis (2020), fewer medical practitioners and hospital beds relate to increased fatality rates, emphasizing the necessity of an effective health care system during a pandemic.

The significant demographic correlates of COVID-19 risks are proportions of seniors and unemployed. Seniors due to aging and comorbidities are of higher risks of COVID-19 infection. Unemployment rate increased due to lockdowns during the COVID-19 pandemic.

The significant urbanity correlates of COVID-19 risks are number of municipalities, and revenue collection. The risk of COVID-19 is significantly related to the number of municipalities. More municipalities indicate more government funding for COVID-19 monitoring and pandemic management. More revenue collections of the local government units imply that there are more people with economic activities, and hence, higher COVID-19 risks. Population density was found to be a major predictor of virus propagation by Han et al. (2021).

Air quality index and wind speed are significant environmental correlates. Slower wind speed may indicate higher risk of COVID-19. This is consistent with Clouston, et al. (2021) that the risk of outdoor transmission of COVID-19 was higher on days with low wind speed. In addition, higher air quality index indicates greater level of air pollution, and thus, greater COVID-19 risk. This supports the findings of the study of Kapitsinis (2020).

Comorbidities like bronchitis, pneumonia, heart disease, and hypertension are significantly correlated to COVID-19 risks. Bronchitis and pneumonia both affect the respiratory system. As such, COVID-19 risk is high if the person is experiencing symptoms like bronchitis and pneumonia (Watt, 2021). Persons with heart disease and hypertension have higher risks of severe COVID-19 infections. Results further show that oral health is significantly related to COVID-19 risks. Better oral health leads to less risk of COVID-19 infection.

4. CONCLUSIONS

The research determined the significant correlates of COVID-19 risks using MLRM, SAR, and MLRM-SAR models. SAR model is more appropriate than MLRM since there exist significant spatial dependencies among NCR cities and provincial COVID-19 risks. It is also better than MLRM-SAR for it has lower MSE and lower MAD. Under the SAR model, the significant correlates are hospital beds for COVID-19 patients, licensed COVID-19 testing laboratories, seniors, unemployed, municipalities, revenue collection, air quality index, wind speed, persons with bronchitis, pneumonia, heart disease, hypertension, and oral health care. Spatial maps showed clustering of COVID-19 high-risk areas in NCR and nearby provinces ('NCR plus bubble'), and in Baguio City, Benguet and nearby Northern provinces.

5. REFERENCES

Chan, S., Chu, J., Zhang, Y., & Nadarajah, S. (2021). Count regression models for COVID-19. *Statistical Mechanics and Its Applications, 563*, Article 125460. https://doi.org/10.1016/j.physa.2020.125460

Clouston, S., Morozova, O., & Meliker, J. (2021). A wind speed threshold and COVID-19. *BMC Infectious Diseases, 21*(1), 1194. https://doi.org/10.1186/s12879-021-06796-z

Han, Y., Yang, L., Jia, K., Li, J., Feng, S., Chen, W., Zhao, W., & Pereira, P. (2021). Spatial distribution of the COVID-19 pandemic and environmental factors. *Science of The Total Environment, 761*, https://doi.org/10.1016/j.scitotenv.2020.144257

Kapitsinis, N. (2020). The factors of the COVID-19 spatially uneven spread. *Regional Science Policy & Practice, 12*(6), 1027–1045. https://doi.org/10.1111/rsp3.12340

Medina, M.A. (2020). Age as a risk factor of COVID-19 mortality in the Philippines. *SSRN Electronic Journal*, 1-6. http://dx.doi.org/10.2139/ssrn.3579145

Pace, R., Barry, R., Clapp, J., & Rodriquez, M. (1998). Spatiotemporal Autoregressive Models of Neighborhood Effects. *The Journal of Real Estate Finance and Economics, 17*(1), 15–33. https://doi.org/10.1023/A:1007799028599

Watt, W. (2021). *COVID-19, Bronchitis.* GoodRx. https://www.goodrx.com/conditions/covid-19/covid-19-vs-bronchitis

World Health Organization. (2021). *WHO Coronavirus (COVID-19) Dashboard.* https://covid19.who.int/