ISSN 1022-8594
JUJS

Vol. 44 No. 1, pp.47-66

Identifying Factors of Spatial Heterogeneity in COVID-19 Incidence for Dhaka Division, Bangladesh

Azizur Rahman[1*], Arifa Tabassum[1], Nujhat Tabassum[1], Mariam Akter[1]

*[1]Department of Statistics and Data Science, Jahangirnagar University,
Dhaka-1342, Bangladesh.*

Abstract

The modern world is dealing with a novel infectious disease known as COVID-19, which is rapidly spreading from one person to another or from the environment, causing illness in our communities. The application of geospatial and statistical methods allows researchers to gain a better grasp of the nature of this disease's spatial and temporal dissemination. However, there is a paucity of research on employing the Bayesian framework of spatial modeling for COVID-19 and identifying sources of regional heterogeneity in reported coronavirus (COVID-19) disease. Sociodemographic factors, environmental factors, socio-economic factors, and contact patterns are among the postulated mechanisms that generate regional variation in reported coronavirus (COVID-19) infections. Through the Bayesian framework of the spatial conditional model, we evaluate the impact of socio-demographic factors on observed variance in COVID-19 standardized morbidity rates (SMRs) across seventeen districts of Dhaka divisions. The conditional approach uses random effects in a generalized linear mixed model (GLMM) to account for geographic similarity between observations. We created two types of intrinsic conditional autoregressive (ICAR) models and discovered that the first type performed better in explaining variation in COVID-19 SMR rates throughout the Dhaka division in terms of lowest deviation information requirement (DIC).The results of this model revealed that COVID-19 SRMs were higher in Dhaka, Faridpur, and Rajbari districts, and that only transmission mechanisms-population density was a relevant factor in explaining the change in COVID-19 standardized morbidity rates across the study period. The outcomes of this study will aid in gaining a better understanding of the COVID-19 situation and will serve as a guideline for developing an effective action plan to control and prevent disease spread.

Keywords: Coronavirus, Heterogeneity, Standardized morbidity rates (SMRs), Spatial analysis.

Introduction

Infectious diseases, caused by bacteria, viruses, or parasites, can spread from one person to another or from the environment, resulting in illness in our communities.

---

[*] Author for Corresponding email: rahman.aziz83@gmail.com

The present world is facing a unique infectious disease, named the novel coronavirus disease, widely known as COVID-19. COVID-19 is a respiratory infectious disease that evoked on December, 2019 in Wuhan, China, and reported first death was in China in January 2020 [1]. Two months later, on March 11, 2020, WHO declared the COVID-19 outbreak as a world pandemic situation when it hits over 114 countries (WHO, 2020d). Statistics have shown that about one million people had contracted the disease in the first three months from its emergence and six months later this number has exceeded 10 million cases and more than 500,000 deaths [2].

Like other countries, Bangladesh is no exception in facing this pandemic. In Bangladesh, on March 8, 2020, the Institute of Epidemiological Disease Control and Research (IEDCR) reported the first three confirmed cases in Dhaka city, and on March 18, 2020, Bangladesh faced her first death due to coronavirus [3]. Until September 23, 2020, 352178 confirmed cases have been reported with 5007 deaths for Bangladesh [4]. Many countries around the world have imposed many public health measures to flatten the curve and to reduce its adverse risk. Parallel to other countries, the Government of Bangladesh declared nationwide ‘lockdown’: a complete shutdown of schools, daycare centers, restaurants, shopping malls, and more. The prevention measure and accompanying restrictions have not been imposed simultaneously; they are taken, loosened, or lifted as the evolving situations demand. As a result, Bangladesh has recently seen an unparalleled quake in terms of social, human, and economic catastrophe.

Poor social, economic, environmental, health, and demographic condition of a society within a country had been identified as key indicators of infectious diseases in general [5]. Review of literature identified that socio-economic and demographic matric acted as potential significant determinants in measuring the efficacy of vaccination for tuberculosis in Germany [6], neighborhood socio-economic indicators such as race, percent below poverty, and percent female-headed household were associated with influenza hospitalization rate for the U.S.A [7], humidity and temperature were associated with seasonal flu onset in Spain [8], for pneumonia hospitalization rate aboriginal population were positively associated in Canada [9]. To be specific, areas with low population density and population over 65+ years showed a high H1N1 mortality rate [10]. In addition, the percentage of school-aged children and ethnic minority group had a positive influence on the incidence of influenza in the U.S.A [11], and the location of the school had an impact on Hand-foot-mouth diseases in China [12]. Hence, these findings suggest that identification of social, economic, demographic, and environmental indicators for emerging coronavirus diseases is crucial to combat the transmission rate from human to human at each phase of the pandemic.

Several studies had been conducted to measure the impact of these indicators on the coronavirus (COVID-19) transmission and their magnitude [5, 13, 14, 15, 16]. A country like Bangladesh, a highly dense developing country, is a perfect study unit for understanding the impact of contributing determinants on the COVID-19 outbreak. All recent studies on COVID 19 disease in Bangladesh have primarily focused on the country context rather than the division level. Our approach, on the other hand, was to think about ways to deal with at a granular level. We've decided on the Dhaka division. The Dhaka division was chosen because of the earliest commencement of COVID-19 cases, a greater rate of COVID-19 cases than other divisions, and characteristics such as high population density, high rate of social-demographic inequality, and so on.

To examine the association between COVID-19 incidence and its contributing factors, the use of geospatial and statistical tools provides a crucial understanding of the nature of the transmission of the disease over space and time, as it is a process that occurs in geographical space [17, 18]. Several spatial modeling approaches so far have been conducted on COVID-19 cases for Bangladesh using the GIS approach and traditional spatial models like spatial lag model (SLM), spatial error model (SEM), geographically weighted regression [19, 20]. One study analyzed infection dynamics via SIRmodels for Bangladesh [21]. Another study used the linear and log-linear diagram of COVID-19 cases in different countries including Bangladesh to present graphically [22]. These models considered a frequentist approach in the estimation of parameter and provided single estimates of model parameters. Moreover, modeling spatial dependence by introducing spatially associated random effects within a traditional setting is challenging. On the contrary, Bayesian hierarchical stetting provides a flexible way of introducing spatially associated random effects in the model development. Also, the Bayesian modeling and inferential framework are flexible and enormously rich in their capabilities to accumulate various scientific hypotheses and assumptions. The advantage of using the Bayesian inference approach is that we can obtain the posterior distribution of model parameter estimates rather than point estimates only. Few research have employed a Bayesian framework in the geographical model for COVID-19 pandemic data till now. In this study, we considered an intrinsic conditional autoregressive (ICAR) model to find out the potential determinants of incidence rates of COVID-19 disease. Moreover, using ICAR, we mapped the COVID-19 disease by smoothening the relative risk of disease.

Therefore, this study aimed to identify potential indicators related to COVID-19 incidence focusing on the Dhaka division using an ICAR model. The objectives of this study were i) to find out the spreading nature of COVID-19 in a selected time period for Dhaka Division by disease mapping, ii) to identify the socio-demographic factors which have a significant contribution to the standard morbidity ratio (SMRs) of coronavirus disease and iii) to quantify the spatial relationships between contributing factors and the COVID-19 SMRs.

Methods and Materials

Study Area

In this study, our approach was to analyze the spatial pattern of COVID-19 propagation and its associated risk factors for the broader Dhaka division in Bangladesh using available daily and district-level data. There are 64 districts in Bangladesh under 8 divisions [23]. The divisions are Dhaka, Chottogram, Khulna, Rajshahi, Rangpur, Barishal, Mymensing, and Sylhet which consists of 13, 11, 10, 8, 8, 6, 4, and 4 districts, respectively. We merged Dhaka and Mymensing to create a broader Dhaka division and that consists of 17 districts in total. The rationale for merging was because Mymensing has a close border with Dhaka, and climatically both divisions have the same nature at every season. High population density, better job opportunity, high migrant households, location of economic zone, the concentration of factory/industry workers, etc. gear up the spread of coronavirus and make the Dhaka division more vulnerable to disease compared to other divisions. Seventeen districts: Dhaka (DH), Faridpur (FD), Gazipur (GA), Gopalgonj (GO), Jamalpur (JA), Kishorgonj (KS), Madaripur (MaD), Manikgonj (MaN), Munshigonj (MuN), Nasirabad (NS), Narayangonj (NG), Nershindi (NR), Netrokona (NT), Rajbari (RJ), Shariatpur (SH), Sherpur (SER) and Tangail (TA), under broader Dhaka division considered as spatial units of this study.

Data Collection

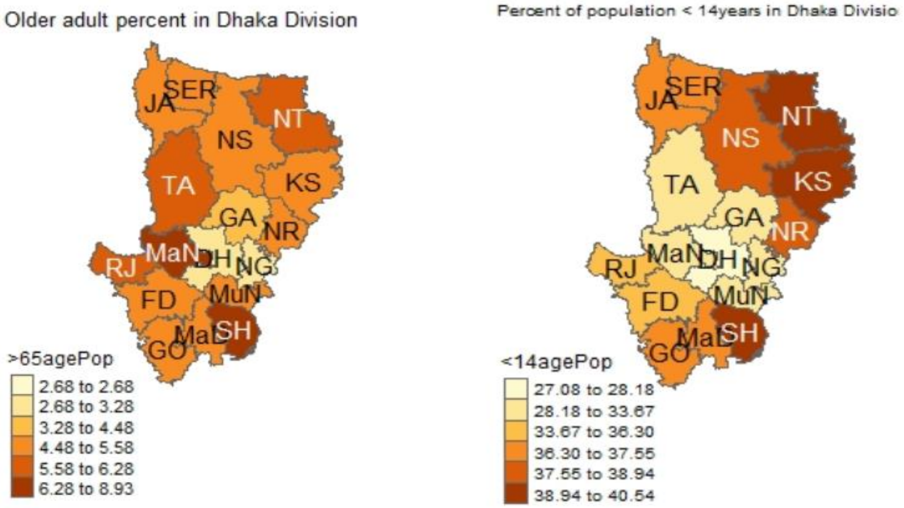
As a government organization, the Institute of Epidemiology, Disease Control and Research (IEDCR) has been responsible for monitoring the spread of the pandemic and updating the database of COVID-19 cases on daily basis at the district and city level across Bangladesh. Moreover, other private research organizations are monitoring the daily COVID-19 cases through their sources. For this study, a number of confirmed cases data at the district level in Dhaka divisions were considered from March 8, 2020 to September 23, 2020, and retrieved from CARE Bangladesh COVID-19 data portal (www.data.carebd.org).

A total of four socio-demographic variables: population size >65 years, population size <14 years, ethnic proportion of the population, population density, were considered as explanatory variables in the model development. The associate hypotheses for all selected predictors are described in table 1. Most of the data were retrieved from the Bangladesh Bureau of Statistics (BBS) database (www.bbs.gov.bd), which is the centralized official authority in Bangladesh for collecting statistics on demographics and socio-economic facts of the country. The geographical distribution of these data is shown in figure 1(a), 1(b), 1(c), and 1(d), respectively. All data are inherently spatial in nature and hence indicate that areas located nearby tend to have more similar values than those separated by greater distance [24].

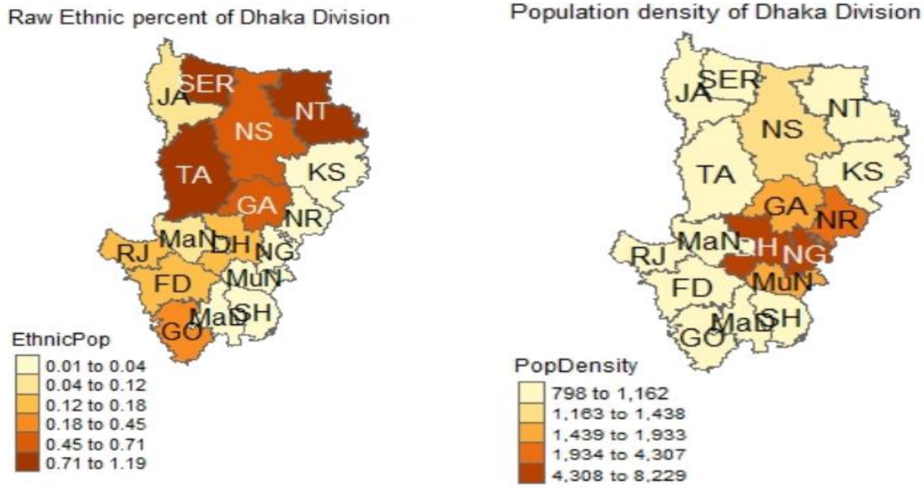
Table 1. Model predictors and associated hypotheses

Factor	Index	Plot label	Spatial Scale	Hypothesized Effect
Transmission mechanisms				
Density-dependent	Population density	PopDensity	district	+
Diffusion mechanisms				
Local spread	% <14 years child population	<14 agepop	district	+
	% ethnic population	ethnicpop	district	+
Important risk	% >65 years population	>65 agepop	district	+

In this study, we considered this conditional nature of data as prior information in developing models. Therefore, we developed intrinsic conditional autoregressive (ICAR) models for spatial dependence in the full Bayesian approach, which smooths the model estimates by borrowing information from neighboring areas.



(a) (b)



(c) (d)

Figure 1 Measurements of four demographic variables: (a) population size >65 years, (b) population size <14 years, (c) ethnic proportion of population and (d) population density across Dhaka division.

Methodology

The best approach for modeling health events may not always be a linear model, particularly when the interested outcome variables are counts or rates, or when we are interested in estimating health risk from binary data. In such cases, it is customary to use generalized linear models (GLMs), with Poisson and logistic regression models in special cases. In our dataset, the outcome variable is daily COVID-19 cases (Y_i) reported by the public health service provider, for each district of the Dhaka division from the selected time period. As of relevant predictors of coronavirus disease burden, we considered socio-demographic metrics: the proportion of ethnic population size, proportion of older (>65+ years), and child (<14 years) population and population density based on a literature review [19]. We performed the following data processing steps for district-level COVID-19 cases: i) we assumed that our data follows Poisson distribution as$Y_i \sim Pois(E_i \lambda_i)$ $i = 1, \cdots, m$, m represents the number of districts or areas under consideration; ii) we assumed that the expected COVID-19 cases (E_i) were calculated through direct standardization process as $E_i = N_i \left(\frac{\sum_{i=1}^{m} Y_i}{\sum_{i=1}^{m} N_i} \right)$, where, N_i is the population at risk for $i\,th$ area; iii) we calculated standardized morbidity ratio (SMR's) which is the ratio of observed to expected cases; iv) we plotted the histogram of SMRs to have an idea of how SMRs are distributed; v) as an exploratory analysis, we plotted SMR values against each covariate and vi) as a starting point of modeling strategy, we then fitted a loess curve to SMR versus different explanatory variables. This fitted curve will assist us to identify the appropriate functional form of relationship between outcome and predictors. Our model-building strategy was as follows:

Step 1: We conducted the four simple linear regression model considering SMRs as continuous dependent variable and proportion of the ethnic population, the proportion of older and child population and population density as single covariate respectively, (see equation 1 in model structure section) assuming that all of the observations are independent, and residuals follow a normal distribution.

Step2: After that, we conducted four simple Poisson log-link linear in the covariate model defined in equation 2 (see equation 2 in model structure section) assuming dispersion parameter equal to 1 (mean=variance). At this stage, we checked the statistical significance of the covariates and looked for an area-based association between incidence and covariates separately. Variables that were found to be insignificant in the univariate analysis of Poisson regression models were not considered further as explanatory variables.

Step 3: We then conducted quasi-Poisson regression with only significant variables as it provided us the general solution of over-dispersion assuming that variance is a function of the mean. If we find large values for over-dispersion parameters and residual analysis shows any patterns, we may conclude that only the Poisson model is not adequate. That means data may have a spatial dependency.

Step 4: At this stage, we considered the spatial generalized linear mixed model (GLMM) to capture over-dispersion and the autocorrelation induced in the data by the latent spatial process and introducing the random effects in the model (see equation 3 in model structure section). To obtain estimates of the parameters of this model, we used the Bayesian inference approach by applying the full Bayesian technique. We modeled district spatial dependence with an intrinsic conditional autoregressive (ICAR) model, which smooths model predictions by borrowing information from neighboring areas.

Model structure

At first, we developed the most common version of our simple linear regression model with the dependent variable SMR_i and the single covariateX_i. Here X_i's are: the rate of ethnic population, population density, rate of older-aged population, and child population for each area, entered into the model separately. The model can be represented as follows:

Model1: $SMR_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \ (i = 1, \cdots, m = 17)$ (1)

Since our underlying dataset is count data, therefore the best approach for modeling this type of event may not be a usual linear regression model. Hence, it is customary to use generalized linear models (GLMs), with Poisson log-link linear in the covariate model.

Now, we assumed that our observeCOVID-19 cases (Y_i) follows Poisson distribution with mean, μ_i. Then the log-linear model with single covariate (X_i) was defined as follows:

$Y_i \sim_{ind} Poisson(\mu_i),\ (i = 1, \cdots, m = 17)$

with $\mu_i = E[Y_i]$, and

Model 2: $\log \mu_i = \log E_i + \beta_0 + \beta_1 X_i$ (2)

We carry out likelihood analyses using the *glm* function in R-software and the log-linear mean function defined in equation (2). Here, $\exp(\beta_1)$ represents the relative risk describing the area-based association between the incidence of COVID-19 and the covariate X_i. Since the Poisson model is restrictive in the sense that the variance is constrained to equal the mean; therefore we conducted the estimation with the quasi-likelihood approach to captures the over dispersion of estimates. At this stage, we selected only significant predictors to conduct further analysis. Moreover, if we find the over dispersion parameter too high, it suggests that we need to go to step 4: build the spatial generalized linear mixed model (GLMM) to capture the over-dispersion problem by introducing the latent spatial process in the random effects term of the model. In this way, we could obtain smooth estimates of the final model with significant predictors for disease mapping as it reduced the instability inherent in COVID-19 SMRs based on small expected numbers.

The spatial GLMM is an example of a hierarchical model or a model defined in stages. At the first stage of the model, we define the distribution of the data given values of the random effects. In the second stage, we define the distribution of the random effects. By combining the first and second stages, we obtained inference about the data and considering the distribution of random effects which leads to the use of the Bayesian inference approach, as an attractive alternative of likelihood approach in the estimation of parameters of the GLMM model. Inference was made through the complete posterior distribution and we summarized it in terms of posterior moments, in most cases we used posterior median. We applied full Bayes (prior distribution on regression and prior parameter of spatial dependence) as of spatial model-fitting approach.

The generic-model for district-wise observed COVID-19 cases y_i is:

$$y_i|\ \mu_i, \tau \sim f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\tau}) \qquad (3)$$

where $\boldsymbol{y} = (y_{\mathbf{1}}, \cdots, y_n)'$denotes the vector of all observations. We modeled the mean(μ_i) of the observed disease cases, where $f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\tau})$ is the distribution of the likelihood of the observed total COVID-19 cases, parameterized with mean, $\boldsymbol{\mu} = (\mu_{\mathbf{1}}, \cdots, \mu_n)'$and precision parameter, τ as appropriate to the likelihood distribution. In this study, we choose log-normal priors for the positive parameters in the form of e^{V_i} since we can specify two quantiles of the distribution and directly solve the two-parameters of the log-normal.

Intrinsic Conditional Autoregressive (ICAR) model (Poisson – lognormal model with covariate)

For our dataset, we assumed that we have area level significant covariate, and therefore assumed that we have the model, given by-

Model 3: $Y_i|\beta, \gamma, U_i, V_i \sim_{ind} Poisson(\mu_i E_i e^{V_i+U_i})$ (4)

with $\log(\mu_i) = f(x_i, \beta) + g(S_i, \gamma)andV_i \sim_{iid} N(0, \sigma_v^2)$

Where $f(x_i, \beta)$ is a regression model; and $g(S_i, \gamma)$ is an expression that may include to capture large scale spatial trend and S_i is the centroid of area *i*. The random effect V_i represents non-spatial over-dispersion and U_i are random effects with spatial structure.

In this model, we specified the distribution of each U_i as if we knew the values of the spatial random effects U_j in “neighboring areas”. We need to specify a rule for determining the “neighbors” of each area. To define neighbors, a number of authors have taken the neighborhood scheme to be such that areas i and j are taken to be neighbors if they share a common boundary. This is reasonable if all regions are of similar size and arranged in a regular pattern. In this study, we took the neighborhood structure to depend on the distance between area centroids and determine the extent of the spatial correlation. Here we assigned the spatial random effects an intrinsic conditional autoregressive (ICAR) prior and the model is defined in equation (4).

Here, we assumed that $U_i|V_j, j \in \delta_i \sim N\left(\overline{U_\iota}, \frac{\omega_u^2}{m_i}\right)$, where δ_i is the set of neighbors of area i, m_i is the number of neighbors and $\overline{U_\iota}$ is the mean of spatial random effects of these neighbors. The parameter ω_u^2 is a conditional variance and its magnitude determines the amount of spatial variation. If majority of the variability is non-spatial, inference for this model might incorrectly suggest that spatial dependence was present. For this model, we applied full Bayesian approach, hence, we need to specify priors both for i) the regression coefficient β and ii) the variance of the random effect σ_v^2. For regression parameters we choose non-informative priors, flat priors and we choose two gamma priors for σ_v^{-2}, assumed that we had enough information for σ_v^2. This information was utilized in precision $\tau_v = \sigma_v^{-2}$ formula. Here to find and summarize the posterior distributions of the regression coefficient $\boldsymbol{\beta}$ from model 6, the following steps were followed: **Step 1**: We choose two informative priors for spatial random component suggested in literature as follows: $tau.T \sim Gamma(1, 0.0260)$, and $tau.T \sim Gamma(1, 0.1399)$. Then used this information in the ICAR model in terms of $Sigma_z = sqrt(p/(tau.T))$. **Step2**: We choose a prior distribution for the proportion(p) of the variance from beta (1,1). **Step 3**: We calculated the ω_u as follows: $Omega_u = {}^{Signma_z}/_{sqrt(1.164)}$. **Step 4**: We then calculated $tauomega.U$ as follows: $tauomega.U = 1/(Omega_u * Omega_u)$.

Results

As the first step of our analysis, we conducted a simple linear regression model with four explanatory variables and plotted their fitted values with actual values. (see supplementary file). Exploratory analysis clearly showed that simple linear models were not a good approach to dealing with count data. As of the second step, we conducted the Poisson-log link linear model for count data as our underlying data is the number of confirmed COVID cases. We conducted the univariate analysis for each of the covariates, to select our preliminary main effects model. We selected all those variables which showed statistically significant at 5% level of significance.Results of the univariate analysis are presented in **Table 2** based on the Poisson-log link linear model (with and without quasi-likelihood estimation approach). From **Table 2**, we observed that factor of transmission mechanism was found highly significant (p<0.0001) and two factors out of three from diffusion mechanism were found highly statistically significant in Poisson-log link linear model with quasi-likelihood approach. Also, from the Table 2 significant factors have lower residual deviance which means that factors could explain a good portion of the variation in the COVID-19 SMRs across Dhaka division. Therefore, variation in COVID-19 SMRs could mostly described by demographic factors.

Table 2. Results of univariate analysis on Poisson-log link linear model (Model 2).

Factor/Variables	**Intercept**	**Estimates**	**Standard Error**	**P-values**	**AIC/Residual Deviance**
		Transmission mechanisms			
PopDensity	-1.205	2.626e-04	8.665e-07 (3.16e-05)	<2e-16*** (5.38e-07)***	16067(15897)
		Diffusion mechanisms			
<14agepop	5.834	-0.186	0.0007 (0.0318)	<2e-16*** (3.3e-05)***	31041(30871)
ethnicpop	0.347	-1.411	0.1102 (0.9137)	<2e-16*** (0.143)	97087(97917)
>65agepop	2.417	-0.612	0.0024 (0.1338)	<2e-16*** (0.0003)***	36649(36471)

Significance: 0.001***, 0.05**, 0.01*, within braces the information are from Poisson-log link with quasi-likelihood approach.

Then we run the multivariable analysis with all the significant variables and from Table 3, found that <14 agepop (p-value =0.2919) and >65 agepop(p-value=0.5515) were not statistically significant in Poisson log-link linear model (with quasi-likelihood). Therefore, our final model will consist of only transmission mechanism factor: population density to explain the variation of COVID-19 SMRs across the Dhaka division.

Table 3. Results of multivariate analysis on quasi-Poisson-log link linear model (Model 2).

Factor/Variables	Intercept	Estimates	Standard Error	P-values	Residual Dev(Dispersion parameter)
		Transmission mechanisms			
PopDensity		2.300e-04	8.055e-07	0.0135**	14171(1390.1)
		Diffusion mechanisms			
<14agepop		-6.183e-02	5.62e-02	0.2919	
>65agepop		1.121	1.833e-01	0.5515	

Significance: 0.001***, 0.05**, 0.01*, within braces the information is for dispersion parameter indicates still over-dispersion.

According to residual plot of final model with only single covariate *Popden* which is centralized for interpretation purpose (see Figure 2) and observed that there was some variations or specific pattern in residuals, which indicated that data may have some spatial dependencies and hence, Poisson model was not adequate.

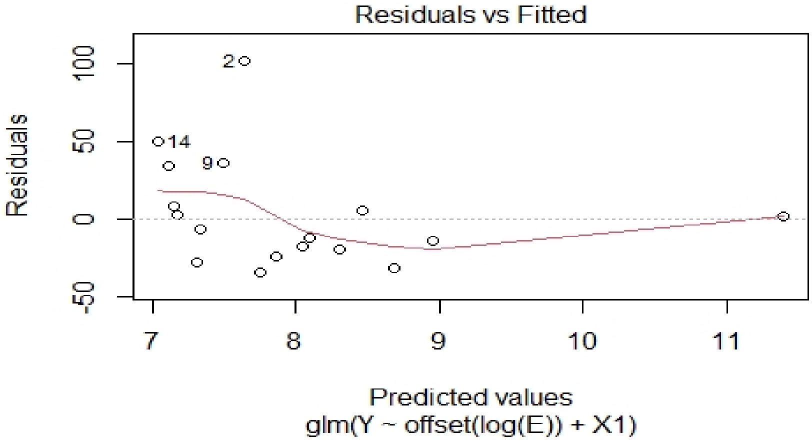


Figure 2. Residual plot versus predicted values of final model 2(quasi-likelihood) linear in centralized value of Population density (*Popden*).

ICAR Model (Poisson – lognormal model with Population density as covariate)

From Table 4, we observedthe posterior medians for two ICAR models with different prior parameters setup for tau.T and proportion (*p*) and run MCMC using two sets of initial values (tau.T=1, p=0.5, beta0=0.01,beta1=0,beta2=0 and beta3=0 and tau. T=2, p=0.5, beta0=0.01, beta1=0, beta2=0 and beta3=0) for 2 chain with 7000 samples in WinBUGS.

Table 4. Sensitivity of ICAR models parameters to prior choice *tau.T* and *p*

Spatial Model	Prior Specification		Posterior medians			DIC
			σ_v	σ_u	p	
ICAR1	$\tau_T \sim Ga(1,0.0260)$	$p \sim Beta(1,1)$	0.7116	1.124	0.867	856.12
ICAR2	$\tau_T \sim Ga(1,0.1339)$	$p \sim Beta(1,1)$	0.0930	1.129	0.781	915.52

From Table 4, DIC value was found lower for ICAR1 model with Gamma (1, 0.02060) compared to other, indicate that it is the most parsimonious model. In addition, the posterior value for proportion of variance *p* explained by spatial correlated error term was 0.87, indicate that this model could explain 86.7% of the variation of COVID-19 SMRs across Dhaka division with population density by spatial component. Therefore, the results were based on ICAR1 model.

The spatial distribution of the coefficients (Relative Risk) of ICAR1 model is presented in Figure 3. For Dhaka, Faridpur and Rajbari districts, population density was found as highly influential factor in explaining the standard morbidity rates of COVID-19 disease. For other districts it has opposite influence.

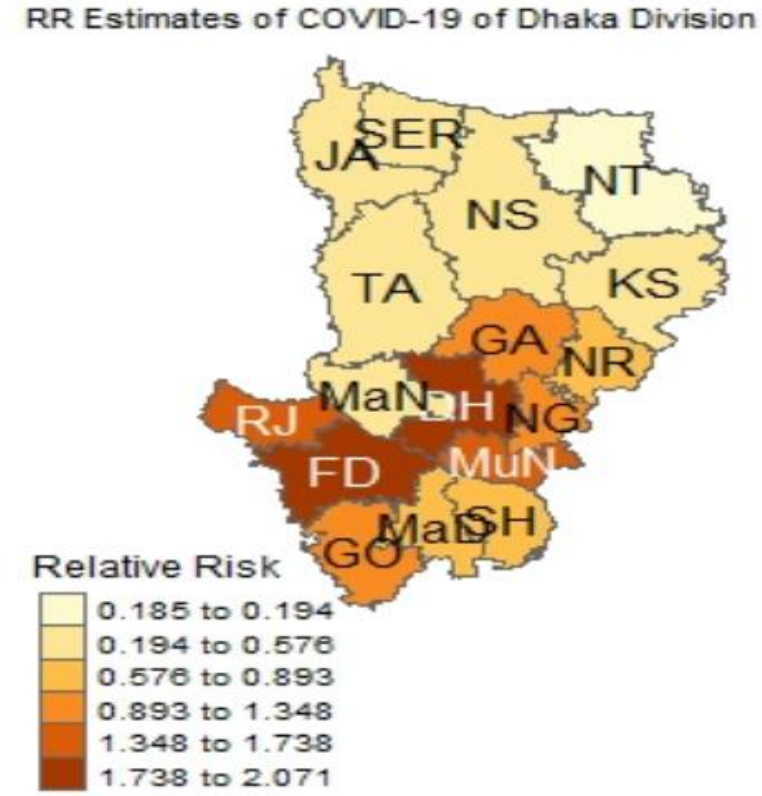


Figure 3. Spatial distribution of the coefficients (Relative Risk) values in describing COVID-19 Standard morbidity rates (SMRs) using ICAR model.

Discussion

This study aimed to examine the significant factors which have substantial explanatory power to explain the COVID-19 standard morbidity rates (SMRs) at the district-level across the broader Dhaka division with advanced spatial epidemiological models. In this study, we considered only socio-demographic metrics as indicators of disease burden. The results of this study showed that models with only one transmissibility mechanism factor: population density could explain a relatively large amount of variation in the COVID-19 SMRS across the Dhaka division during the selected time period. Among the two possible intrinsic conditional autoregressive (ICAR) models, the ICAR1 model having prior parameter information $\tau_T \sim Ga(1,0.0260)$found to better explain the variation in SMRs of COVID-19 disease.

Only the population density, transmission mechanism, among others found a strongly significant factor in the multivariate model and therefore, included in the final ICAR models. The proximate contact between an infector and infectee increases in high population density areas (You et al., 2020). To be mentioned here, in this study percentage of older age people (>65+years), percentage of children (aged <14 years), and percent of the ethnic population, were found significant in univariate analysis and were not included for developing final main effect models due to insignificant issues in the multivariate model. We found Dhaka district, out of seventeen districts of broader Dhaka division, the epicenter of pandemic and showed the highest estimated SMRs of COVID-19 diseases.

Our study demonstrated that social indicator: percent of the ethnic population in each district, had no significant influence on the occurrence of COVID-19's SMRs. In literature, we found that ethnic groups of people had a positive impact on the onset of infectious diseases [9]. There is a scope of rigorous and detailed study to understand the impact of more related social and economic factors on the occurrence of COVID-19.

One of the limitations of this study was the public availability of data. Due to the unavailability of publicly available data, it was not considered to include economic, weather, built environment, health, and community facilities factors to explain COVID-19 incidence more rigorously. Moreover, it was not logical to draw inference at the individual or community level due to a lack of individual-level data. Furthermore, the influence of lockdown and other associated measures on COVID-19 incidence rate was not considered in this study.

Conclusion

Identification of possible indicators of infectious diseases and their spread is crucial, especially for coronavirus (COVID-19) disease. This study focused to identify potential influential factors affecting the COVID-19 incidence rate at the district level across the Dhaka division. To fulfill the aim, we developed a spatial epidemiological model named intrinsic conditional autoregressive (ICAR) in WinBUGs software, with a significant covariate.

First, univariate simple linear regression (OLS) models were developed with four selected variables. Exploratory analysis of OLS model output indicated that for each variable OLS was not a good choice of modeling approach due to lack of fit. After that, univariate analysis was conducted with the Poisson-log link linear model for all selected variables, assuming outcome variable as count data and we found all four variables were statistically significant. Then, a global multivariate Poisson-log link linear model with a quasi-likelihood approach was considered to determine the significant factors and the associated overdispersion parameter. The multivariate model identified that only transmission mechanism: population density was found to be statistically significant and mode has high over-dispersion value. At this stage, to address the issue of spatial dependency, we employed correlated error structure in the modeling based on neighborhood areas. Hence, intrinsic conditional autoregressive (ICAR) models with different prior parameters of spatial random error components were developed with the transmission mechanism factor. ICAR model with $\tau_T \sim Ga(1,0.0260)$showed improved model performance as an explanatory power with the lowest DIC values compared to another model. Estimates of the ICAR1 model (median relative risk of COVID-19 diseases) showed that the transmission mechanism factor: population density, could be better explain the COVID-19 incidence rate for Dhaka, Rajbari, and Faridpur districts compared to other districts of the Dhaka division. Results from this research offer factors which need to be considered in the development of extensive policies and guidelines aimed at preventing the future pandemic crisis in the Dhaka division.

Credit authorship contribution statement

Azizur Rahman: Conceptualization, methodology, supervision, data curation and software; **Arifa Tabassum:** Writing-original draft preparation, investigation; **Nujhat Tabassum**: Reviewing and editing; **Mariam Akter:** Visualization, software, validation.

Funding

There is no funding issue.

Conflict of Interest

There is no conflict of interest between the authors, and findings of the research.

Acknowledgments

Authors are grateful to the anonymous reviewers for their valuable comments and suggestions. Authors also show their acknowledgment to Bangladesh Agricultural Research Council for providing shapefile of Dhaka divisions.

References

[1] Aylward, B., and Liang, W. (2020). Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). *The WHO-China Joint Mission on Coronavirus Disease 2019, 2019*(February), 16–24. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.

[2] National Institute of Allergy and Infectious disease (NIAID) (2020). COVID-19 vaccine development behind the scene. Retrieved fromhttps://covid19.nih.gov/. Accessed on 20th December, 2020.

[3] Habib, W. B. and Sujan, M. A. (2020, March 19). Coronavirus in Bangladesh: First death amid sloppy response. *The Daily Star.* Accessed on 28th March, 2020.

[4] IEDCR. (2020). Bangladesh Covid-19 Update, Retrieved from https://iedcr.gov.bd/ on September, 30.

[5] Khalatbari-Soltani, S., Cumming, R. C., Delpierre, C., and Kelly-Irving, M. (2020). Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. *Journal of Epidemiology and Community Health*, *74*(8), 620–623. https://doi.org/10.1136/jech-2020-214297.

[6] Bluhm, R., and Pinkovskiy, M. (2020). The Spread of COVID-19 and the BCG Vaccine: A Natural Experiment in Reunified Germany. *Federal Reserve Bank of New York Staff Reports*, 926. https://doi.org/10.2139/ssrn.3604314.

[7] Sloan, C., Chandrasekhar, R., Mitchel, E., Schaffner, W., and Lindegren, M. Lou. (2015). Socioeconomic disparities and influenza hospitalizations, Tennessee, USA. *Emerging Infectious Diseases*, *21*(9), 1602–1610. https://doi.org/10.3201/eid2109.141861.

[8] Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S., and Sugihara, G. (2016). Global environmental drivers of influenza. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(46), 13081–13086. https://doi.org/10.1073/pnas.1607747113.

[9] Crighton, E. J., Elliott, S. J., Moineddin, R., Kanaroglou, P., and Upshur, R. (2007). A spatial analysis of the determinants of pneumonia and influenza hospitalizations in Ontario (1992-2001). *Social Science and Medicine*, *64*(8), 1636–1650. https://doi.org/10.1016/j.socscimed.2006.12.001.

[10] Ponnambalam, L., Samavedham, L., Lee, H. R., and Ho, C. S. (2012). Understanding the socioeconomic heterogeneity in healthcare in US counties: The effect of population density, education and poverty on H1N1 pandemic mortality. *Epidemiology and Infection*, *140*(5), 803–813. https://doi.org/10.1017/S0950268811001464.

[11] Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., and Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the

spread of infectious diseases. *PLoS Medicine*, *5*(3), 0381–0391. https://doi.org/10.1371/journal.pmed.0050074.

[12] Huang, J., Wang, J., Bo, Y., Xu, C., Hu, M., and Huang, D. (2014). Identification of health risks of hand, foot and mouth disease in China using the geographical detector technique. *International Journal of Environmental Research and Public Health*, *11*(3), 3407–3423. https://doi.org/10.3390/ijerph110303407.

[13] Atkenson, A. (2020). What Will Be The Economic Impact of Covid-19 in the US?. *NBER Working Paper Series*, *53*(9), 1689–1699.

[14] Corburn, J., Vlahov, D., Mberu, B., Riley, L., Caiaffa, W. T., Rashid, S. F., Ko, A., Patel, S., Jukur, S., Martínez-Herrera, E., Jayasinghe, S., Agarwal, S., Nguendo-Yongsi, B., Weru, J., Ouma, S., Edmundo, K., Oni, T., and Ayad, H. (2020). Slum Health: Arresting COVID-19 and Improving Well-Being in Urban Informal Settlements. *Journal of Urban Health*, *97*(3), 348–357. https://doi.org/10.1007/s11524-020-00438-6.

[15] Liu, J., Zhou, J., Yao, J., Zhang, X., Li, L., Xu, X., He, X., Wang, B., Fu, S., Niu, T., Yan, J., Shi, Y., Ren, X., Niu, J., Zhu, W., Li, S., Luo, B., and Zhang, K. (2020). Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. *Science of the Total Environment*, *726*, 138513. https://doi.org/10.1016/j.scitotenv.2020.138513.

[16] Zhao, Q., Meng, M., Kumar, R., Wu, Y., Huang, J., Lian, N., Deng, Y., and Lin, S. (2020). The impact of COPD and smoking history on the severity of COVID-19: A systemic review and meta-analysis. *Journal of Medical Virology*, *92*(10), 1915–1921. https://doi.org/10.1002/jmv.25889.

[17] Gross, B., Zheng, Z., Liu, S., Chen, X., Sela, A., Li, J., Li, D., and Havlin, S. (2020). Spatio-temporal propagation of COVID-19 pandemics. *MedRxiv*, 2020.03.23.20041517.https://doi.org/10.1101/2020.03.23.20041517.

[18] Mollalo, A., Mao, L., Rashidi, P., and Glass, G. E. (2019). A gis-based artificial neural network model for spatial distribution of tuberculosis across the continental united states. *International Journal of Environmental Research and Public Health*, *16*(1), 1–17. https://doi.org/10.3390/ijerph16010157.

[19] Rahman, M. H., Zafri, N. M., Ashik, F., and Waliullah, M. (2020). Gis-Based Spatial Modeling to Identify Factors Affecting COVID-19 Incidence Rates in Bangladesh. *SSRN Electronic Journal*, 1–36. https://doi.org/10.2139/ssrn.3674984.

[20] Rahman, M. R., Islam, A. H. M. H., and Islam, M. N. (2020). Geospatial modelling on the spread and dynamics of 154 day outbreak of the novel coronavirus (COVID-19) pandemic in Bangladesh towards vulnerability zoning and management approaches. *Modeling Earth Systems and Environment*, *0123456789*. https://doi.org/10.1007/s40808-020-00962-z.

[21] Hasanul, M., Siam, B., Hasan, M. M., Tashrif, S. M., Hasinur, M., Khan, R., Raheem, E., and Hossain, M. S. (2020). *Insights Into the First Seven-months of COVID-19*

Pandemic in Bangladesh: Lessons Learned from a High-risk Country. https://doi.org/10.21203/rs.3.rs-89387/v1.

[22] Sevi, S., Aviña, M. M., Péloquin-Skulski, G., Heisbourg, E., Vegas, P., Coulombe, M., Arel-Bundock, V., Loewen, P. J., and Blais, A. (2020). Logarithmic versus linear visualizations of COVID-19 cases do not affect citizens' support for confinement. *Canadian Journal of Political Science, 53*(2), 385–390. https://doi.org/10.1017/S000842392000030X.

[23] BBS. (2015). Economic census, 2013. *Bangladesh Bureau of Statistic*, Dhaka.

[24] Wheleer, D. C., and Paez, A. (2010). Geographically Weighted Regression. In: Fischer M., Getis A. (eds) Handbook of Applied Spatial Analysis. Springer, Berlin, Heidelberg, 461-486, https://doi.org/10.1007/978-3-642-03647-7_22.

Appendix

Outputs from model 1: **Model1**: $SMR_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \ (i = 1, \cdots, m = 17)$

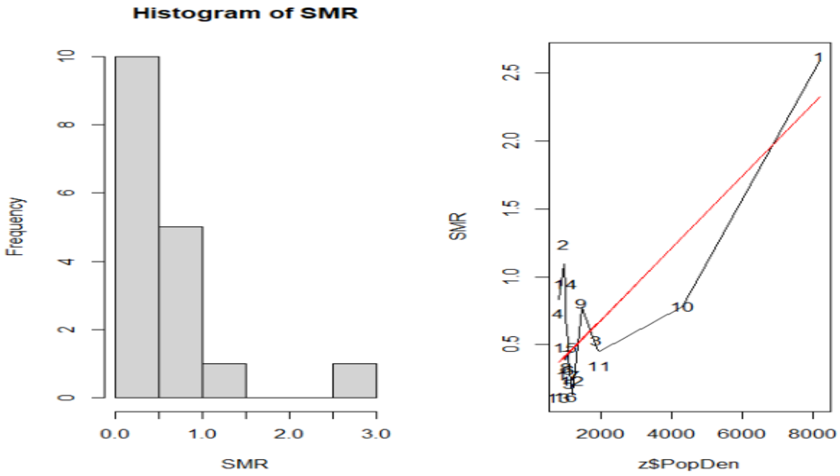


Figure S1. Plot of (Y/E) versus population density (*PopDen*). Solid line represents fitted line for linear in x model and other line represents loess fit.

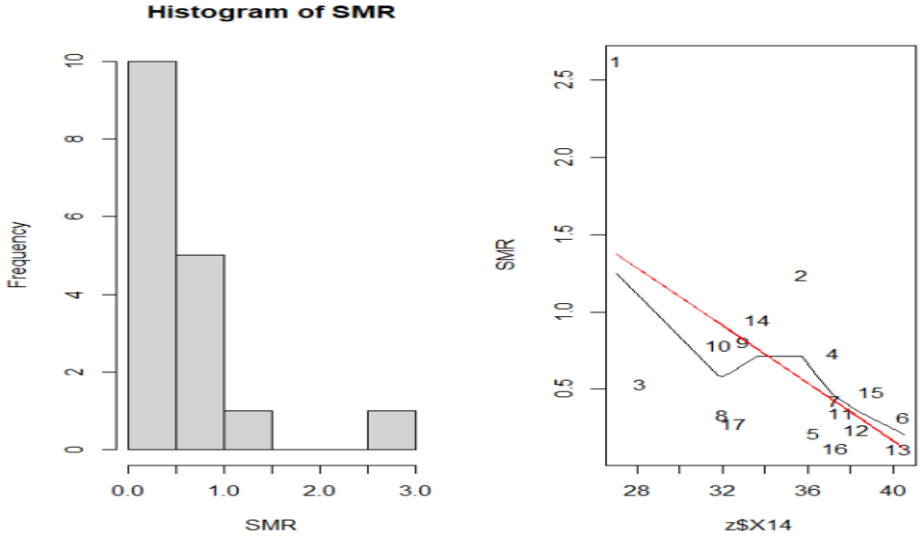


Figure S2. Plot of (Y/E) versus proportion of <14 years population (<*14agepop*). Solid line represents fitted line for linear in x model and other line represents loess fit.

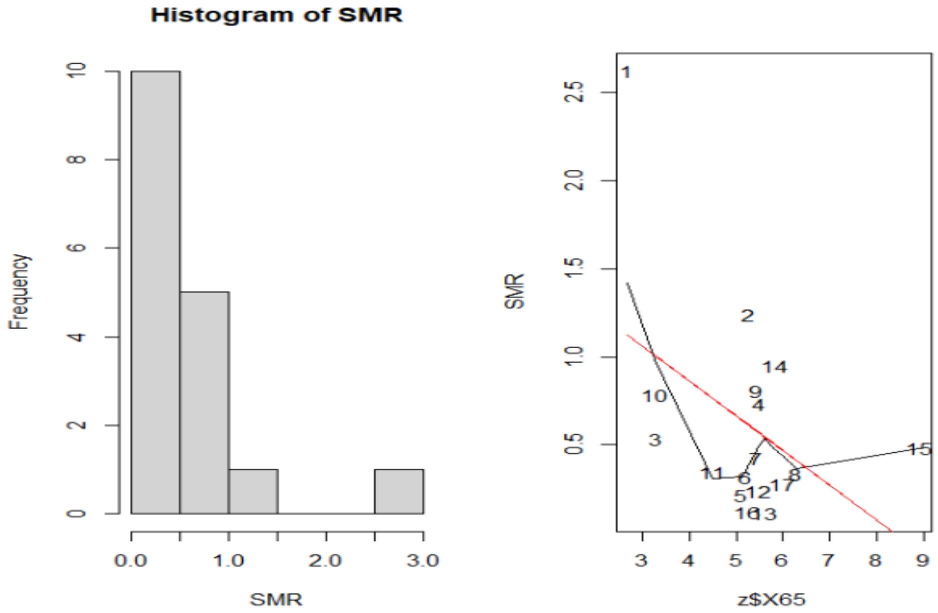


Figure S3. Plot of (Y/E) versus proportion of >65 years people (*>65agepop*). Solid line represents fitted line for linear in x model and other line represents loess fit.

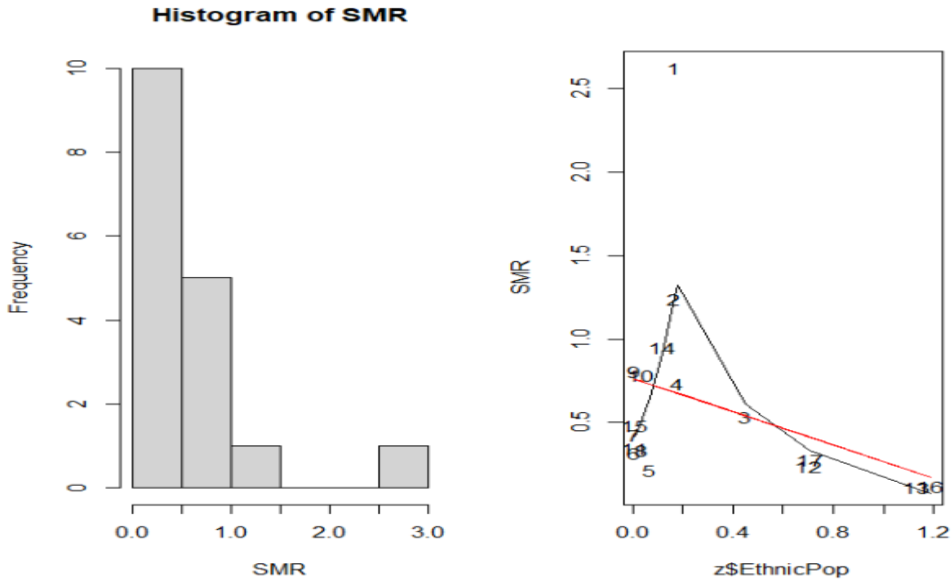


Figure S4. Plot of (Y/E) versus proportion of ethnic people (*ethnicpop*). Solid line represents fitted line for linear in x model and other line represents loess fit.