

SPATIO-TEMPORAL MODEL FOR MAPPING COVID-19 RISK

by

LIA AMELIA

JOHN HIGGINBOTHAM, COMMITTEE CHAIR

ABBAY GREGG

RANDI HENDERSON-MITCHELL

JOHN MCDONALD

JASON PARTON

A THESIS

Submitted in partial fulfillment of the requirements  
for the degree of Master of Science  
in the Department of Community Medicine and Population Health  
in the Graduate School of  
The University of Alabama

TUSCALOOSA, ALABAMA

2021



## ABSTRACT

The COVID-19 was a major threat to public health around the world from the beginning of COVID-19 pandemic. The U.S. was one of the countries with the most COVID-19 cases. Despite the mitigation efforts to control the disease at both local and national levels, the number of COVID-19 cases in the U.S. remained high throughout the pandemic. This study focused on Cook County in Illinois. During the COVID-19 pandemic, Cook County was one of the counties with the highest COVID-19 cases in the U.S. This study described the spatial and temporal dynamics of COVID-19 risk in two-week periods from August 2020 to December 2020 in Cook County. This study also assessed the impact of neighborhood socioeconomic and demographic on COVID-19 incidence.

The Bayesian spatio-temporal model was used to produce COVID-19 risk maps and to evaluate covariates' effects. The results show the spatial heterogeneity in COVID-19 risk from time to time, with the risk peaked in the first weeks of November. Over different time points, some parts of the county exhibited constant COVID-19 high-risk levels. Among these high-risk areas, many of them were majority-Hispanic neighborhoods in Chicago (i.e., Chicago west side) and Cook County suburbs (i.e., Franklin Park and Elgin). The model summary shows that the percentage of Hispanic population, health insurance coverage, and public transit commuters were associated with COVID-19 incidence. The posterior median and the 95% credible interval for the relative risk of a 1% increase in the percentage of Hispanic population was 1.009 (1.007, 1.011), indicating that a 1% increase in the percentage of Hispanic population corresponds to an increase

in COVID-19 risk of 0.9%. The corresponding relative risk for a 1% increase in health insurance was 1.015 (1.006, 1.025), while for a 1% increase in the percentage of public transit commuters, the relative risk was 0.991 (0.987, 0.995).

This study's findings highlight the importance of integrating the geographical information system into disease routine surveillance programs and transforming routinely collected health data into critical information. This information can be used to identify risk factors that could be addressed by allocating resources or implementing health policies.

## DEDICATION

To Rizal and Hana

## ACKNOWLEDGEMENTS

I would like to acknowledge everyone who were instrumental in the completion of my thesis. Firstly, I would like to express my gratitude to my committee chair, Professor John Higginbotham, for his constant guidance and encouragement in completing this thesis. I would also like to acknowledge helpful feedback and advice from my committee members throughout the research process: Professor Abbey Gregg, Professor Randi Henderson-Mitchell, Professor John McDonald, and Professor Jason Parton. Also, I would like to express my appreciation to Barbara Wright for being very helpful in scheduling virtual meetings and to Sally Hulseley for providing administrative support during the completion of the thesis.

I am deeply grateful for Rizal, who supported me with love and understanding, who filled my days with laughter and goofiness. I am very much thankful to Hana, my not-so-little girl anymore, who taught me to appreciate little things in life, live the moment, and slow down. Lastly, my special thank goes to my parents, who inspired me to pursue my academic goals.

## CONTENTS

ABSTRACT.....	ii
DEDICATION.....	iv
ACKNOWLEDGEMENTS.....	v
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
CHAPTER 1 INTRODUCTION.....	1
CHAPTER 2 LITERATURE REVIEW.....	4
CHAPTER 3 RESEARCH PROBLEM.....	14
CHAPTER 4 METHODS.....	16
CHAPTER 5 RESULTS.....	26
CHAPTER 6 DISCUSSION.....	33
REFERENCES.....	39

## LIST OF TABLES

1. The age-specific Illinois COVID-19 incidence rates .....	18
2. Calculating SIR of a zip code in a two-week period .....	19
3. Moran's I statistic .....	23
4. Durbin-Watson statistic .....	24
5. The summary of posterior distribution .....	31

## LIST OF FIGURES

1. John Snow’s map of cholera in Soho, London .....	5
2. The summary of the number of new cases in two-week periods .....	17
3. Average SIR over all ten two-week periods from August to December 2020 .....	20
4. Socioeconomic and demographic characteristics in Cook County .....	22
5. Average COVID-19 fitted risks .....	26
6. The progression of COVID-19 fitted risks from period 1 (August 19th to 31st) to period 10 (December 14th to 26th) .....	27
7. The minimum, the maximum, and the average fitted risk of each zip code .....	29
8. The trend of COVID-19 fitted risk in selected high-risk zip codes .....	30

# CHAPTER 1

## INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by SARS-CoV-2 virus (WHO, 2020a). The first reported cluster of cases of COVID-19 occurred in Wuhan, Hubei Province, China, in December 2019. COVID-19 spread rapidly to almost every country, and on March 11, 2020, the World Health Organization declared COVID-19 as a pandemic. In the U.S., the first case of COVID-19 was reported on January 21, 2020, from a patient in Washington State who just returned from Wuhan on January 15, 2020 (CDC, 2020b). Since late March 2020, the number of COVID-19 confirmed cases in the U.S. had risen sharply and made the U.S. one of the countries with the most COVID-19 cases. Based on the U.S. COVID-19 county-level data (The New York Times, 2020), in March 2020, the virus started to hit hardest in the counties that laid within the nation's metropolitan areas, for example, New York City, Westchester, Nassau, and Suffolk County in New York, Cook County in Illinois, and Los Angeles County in California.

This study focused on Cook County. Cook County, located in the state of Illinois, is the second-most populous county in the U.S. It is the home of one of the largest cities in the nation, Chicago. In September 2020, Cook County remained one of the top counties by the number of confirmed COVID-19 cases, after Los Angeles (California), Miami-Dade (Florida), and Maricopa (Arizona). By mid-September 2020, more than half of positive cases and nearly two-thirds of the deaths due to COVID-19 in Illinois took place in Cook County (Szalinski, 2020).

Several restriction measures to slow the spread of COVID-19 had been considered in Cook County. In March 2020, the governor of Illinois issued executive orders to stay at home, close the schools, and restrict bars and restaurants to serve only for delivery and pick-up orders (State of Illinois, 2020). The mayor of Chicago ordered the closure of public areas such as lakefront trails and bike paths (Cauguiran et al., 2020). At the beginning of May 2020, the governor mandated masks in public places whenever social distancing was not possible. All the measures' success depended on various factors, including when the restrictions began, public compliance, and other socioeconomic related aspects, such as the number of workers who could not perform remote working and the number of people covered by health insurance.

Despite all efforts to prevent the wider spread of COVID-19 in the county, controlling COVID-19 remained a major challenge. Based on COVID-19 data collected by the Illinois Department of Public Health (IDPH), the number of COVID-19 new cases in the county fluctuated from time to time (Illinois Dept. of Public Health, 2020). When the county experienced a decline in the overall number of new cases in a particular period, the decrease may not be uniform across the regions as the number of cases remained high in some parts of the county. Utilizing regularly collected disease data and applying proper analysis and interpretation to the data can improve the current disease control measures. Through this study, using periodically collected COVID-19 data, we aimed to estimate COVID-19 risk and visualize the dynamics of the risk in maps over different periods in Cook County. These estimates can help state and local governments make better decisions in mitigation planning and improve the existing disease surveillance program.

Furthermore, we also evaluated the impact of socioeconomic and demographic factors on COVID-19 incidence. Economic and social aspects are potential determinants of a wide range of

health problems in the U.S., including chronic disease, infectious disease, and injuries (Pamuk E et al., 1998). In the U.S., the benefits of a good health care system are not fairly distributed across different races, income levels, gender, sexual orientations, and other group identities, leading to different health outcomes. Recent studies have found that disparity in social and economic aspects could cause differences in risk for COVID-19 infection within and between societies (Abedi et al., 2020; Khatana & Groeneveld, 2020; Moore et al., 2020). Hence, demographic and socioeconomic factors like race, ethnicity, unemployment rate, and health insurance coverage were included in the analysis to assess if specific subgroups and characteristics were affected disproportionately by COVID-19. The study's findings may help policymakers formulate strategies to lower the transmission rate in high-risk communities and design programs to reduce health inequity.

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1. Spatial epidemiology

Spatial epidemiology focuses on analyzing the geographical variation of health outcomes and its association with potential risk factors. John Snow, a physician in 19th century London, pioneered the implementation of spatial epidemiology by mapping and identifying the death due to cholera during cholera outbreak in London in 1854 (Begum, 2016). Snow demonstrated that the epidemic source was contaminated water rather than bad air, which was previously suspected. He mapped the deaths and found that the people who died mostly lived and clustered around the Broad Street water pump. He showed his investigation results and convinced local authorities about the findings, which led to the removal of the water pump handle. John Snow's map is depicted in Figure 1. In the map, Snow used bars to represent deaths attributed to cholera in a household. Snow's cholera map made a significant contribution in containing cholera. His work inspires how people today find and prevent the spread of the disease.



SIR or SMR is usually used to measure disease risk. The SIR or SMR is obtained by dividing the observed number of cases/deaths by the expected number of cases/deaths in the area. A study involving small-area mapping studies may present an additional source of variability in the map. For example, areas with fewer people where there are almost zero disease cases can have extreme SIR values and cause large variability in disease estimated rates (Anderson & Ryan, 2017). The disease risk estimation in this area can be smoothed by implementing Bayesian methods. Bayesian methods to estimate the disease relative risk have been widely used. Martínez-Bello et al. (2017) used disease mapping techniques for dengue surveillance in Colombia. In this study, the researchers estimated the relative risk of dengue disease per census geographical unit. The relative risk was estimated using Bayesian areal models for disease mapping. Covariates like vegetation index and land surface temperature were included in the models to quantify their impact to high dengue risk. This study's results emphasized the importance of using relative risk maps of dengue for public health planning. This study also found that vegetation index was associated with a high risk of dengue in Colombia.

According to Elliot and Wartenberg, the purpose of geographic correlation studies is to investigate geographic discrepancy across population groups regarding socioeconomic, demographic, lifestyle, and environmental factors to health outcomes, measured on a geographic scale. For example, Maheswaran et al. (1999) investigated magnesium's role in drinking water in the association with mortality from acute myocardial infarction at a small area geographical level in northwest England. This study found no relation between magnesium in drinking water and mortality from myocardial infarction. A study conducted by Neuberger et al. (2009) studied the impact of heavy metal exposures at the Tar Creek Superfund site, Ottawa County, Oklahoma, on the health problem (e.g., excess mortality, low birth weight, and children's blood lead levels) of

the people near the site. The study's result suggested an association between heavy metal exposure and children's health and adults' chronic disease.

## 2.2. Spatial analysis in COVID-19 related studies

Spatial epidemiology has been one of the topics of interest in COVID-19 related studies. Disease maps have been used in studies to analyze the spread of COVID-19 and assess the disease's burden. Spatial models have been widely used to create disease maps. Spatial models can provide more accurate risk estimation as the models consider the structure of spatial autocorrelation. For example, a study conducted by Azevedo et al. (2020) used a direct block sequential simulation method to model and predicted the geographical distribution of COVID-19 infection risk in Portugal. The direct block sequential simulation method was expected to account for the error by spatial uncertainty as a function of population size in each municipality in this study.

A study by Cordes et al. (2020) performed a spatial analysis to identify COVID-19 clusters of high testing rates and the high proportion of positive tests and investigate determinants associated with the clusters across New York City. This study's findings suggested high testing rates and high positive rates occurred more in predominantly black neighborhoods and without health care insurance. At a country level, Mollalo et al. (2020) explored the relationship between 35 socioeconomic, demographic, and environmental variables with COVID-19 incidence in the U.S. using global models (ordinary least squares, spatial lag model, and spatial error model) and local models (geographically weighted regression (GWR) and Multiscale GWR). This study found that multiscale GWR had the highest goodness-of-fit among other models. The result from modeling suggested that median income, income inequality,

percentage of nurse practitioners, and percentage of the black female population were significant variables on spatial variability of COVID-19 incidence rates in the U.S.

### 2.3. Space and time elements in disease data

One of the objectives of this study was to characterize the spatial distribution of COVID-19 risk from different periods. Considering both space and time aspects can be important as it may identify the disease's trend and reveal unusual patterns. The interaction term of time-space in the model can show that different locations may have different temporal trends in the number of cases. Numerous studies have investigated the evolution of health outcomes in both space and time simultaneously. From a non-related COVID-19 study, a study evaluated the change in ischemic heart disease (IHD) risk in New South Wales, Australia, from an eight-year-period between 2006 and 2013 (Anderson & Ryan, 2017). In this study, the authors compared seven spatio-temporal models. The final model was selected based on the Moran's *I* statistic test score and the computing time. A spatio-temporal model using within-cluster smoothing proposed by Lee and Lawson was selected as the best model for the IHD dataset. This model helped to describe the variation in the temporal trend across the area. Using the final model, the results suggested the risk of IHD in the region dropped from 2006 to 2013 but not all the areas in New South Wales experienced the decreasing IHD risk. The risk of IHD remained constant and even higher in more isolated areas.

A recent COVID-19 related study by Gayawan et al. (2020) used a two-component hurdle Poisson model to investigate the spatial and temporal distribution of COVID-19 cases and healthcare capacities in Africa a few weeks after the first COVID-19 case was identified in the continent. The results from a spatio-temporal model indicated that COVID-19 cases varied

geographically across Africa. In particular, a high number of cases were found in neighboring countries in the western and northern parts of the continent. In terms of the burden of healthcare capacities, the study found that countries with higher healthcare capacity had more cases. It suggested that countries with greater healthcare capacity were more likely to identify cases.

Based on the study's purposes and the method used in previous research, this study implemented Bayesian spatio-temporal modeling to estimate COVID-19 risk. The risk was evaluated for each zip code in the county.

#### 2.4. Spatio-temporal modeling

The disease risk in maps is not represented by the observed number of cases alone, considering that cases are mainly distributed based on the underlying population. One way to measure disease risk in an area is by using the SIR. According to Bivand et al. (2013), for area  $i, i = 1, 2, 3, \dots, n$ , SIR can be defined as the ratio between the observed number of cases ( $Y_i$ ) and the expected number of cases ( $E_i$ ) as

$$SIR_i = \frac{Y_i}{E_i} \quad (1)$$

The expected number of cases is obtained based on the data from a larger reference population, like country, state, or county. If SIR equals 1, then there are no differences between the two groups. The SIR that is more than 1 indicates an excess disease incidence compared to what is expected.

SIR calculation in disease maps considers each area independent of others and does not assume any spatial structure in the data. It can also be unstable in areas consist of a smaller number of populations. Consequently, an extremely small SIR may be found in these areas.

Therefore, a generalized linear model is commonly used to estimate the risk. Incorporating spatial elements in the model can be an essential step to improve local risk estimates due to spatial autocorrelation. To assess the presence of spatial autocorrelation, Moran's  $I$  statistic (Moran, 1950) can be applied. Moran's  $I$  is formulated as

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

where  $n$  is the number of spatial units denoted by  $i$  and  $j$ ,  $y$  is the variable of interest with mean  $\bar{y}$ , and  $w_{ij}$  is the spatial weight of the link between  $i$  and  $j$ .

In the case of a spatio-temporal model, space and time elements are considered in the model. Just like a spatial model, the number of observed cases in location  $i$  at time  $k$  in a spatio-temporal model is assumed to be drawn from a Poisson distribution with mean  $E_{ik}\theta_{ik}$ . The formula of a spatio-temporal model is the expansion of a spatial model, which is defined as the following

$$\log(\theta_{ik}) = s_i + x_{ik}\beta + t_k + q_{ik} \quad (3)$$

where  $s_i$  denotes the spatial correlation function,  $\beta$  is the set of coefficients for variable  $X$ ,  $t_k$  indicates the temporal correlation, and  $q_{ik}$  refers to the time-space interaction term (Anderson & Ryan, 2017).

## 2.5. Bayesian disease mapping

The use of Bayesian methods in the areas of disease mapping has been well known. Bayesian approaches can process complex spatial information consisting of data points near each other that share similar characteristics, like socioeconomic and demographic (DiMaggio, 2014).

The Bayesian inference multiplies the prior distribution on model parameters and the data likelihood to obtain the posterior distribution that summarizes the parameters given observed data (Lawson, 2018). The Bayesian hierarchical model can improve the risk estimation as it manages to work with the variables simultaneously as borrowing strength from neighboring areas (Bivand et al., 2013). The model is fitted using the Markov Chain Monte Carlo (MCMC) techniques. The MCMC techniques are commonly used to obtain posterior distributions. These techniques can be very time-consuming and computationally intensive. Principally, MCMC techniques run simulations of the model's parameters to obtain samples that are more likely realizations of the targeted distribution.

## 2.6. Social determinants of health and COVID-19

Social determinants of health, including socioeconomic and demographic status like race, ethnicity, education, employment, and living condition, may significantly affect COVID-19 outcomes. Therefore, this study included some of these variables in the analysis to measure their impact on COVID-19 incidence.

Recent studies have shown that COVID-19 has disproportionately affected minority groups. In the U.S., the COVID-19 infection rate was high among Hispanics, who made up 18.5% of the U.S. population (U.S. Census Bureau) but accounted for 26% of COVID-19 cases (CDC, 2020a). Although African Americans represented 30% of the Chicago population, they made up more than half of Chicago's COVID-19 cases (Yancy, 2020). A similar condition also occurred in New York City, where the number of cases was high among African Americans and Hispanics (The Official Website of the City of New York, 2020). Some studies investigating the association between COVID-19 infection rate and socioeconomic and demographic factors have found that being COVID-19 positive was associated with the African American race (Muñoz-

Price et al., 2020). People of low-socioeconomic status have a higher probability of contracting COVID-19 due to several factors, such as overcrowded housing, limited access to outdoor space, and working conditions that make it harder for them to work from home, such as supermarket and warehouse workers (Patel et al., 2020).

## 2.7. COVID-19 in Cook County

Throughout the pandemic, Cook County remained one of the counties with the most COVID-19 cases in the U.S. Historically, Cook County experienced a surge of COVID-19 cases beginning in March 2020 and peaked on May 1 of that same year. The number of new cases gradually decreased until May 11 and increased again on May 12. After this, the incidence of COVID-19 had slowly declined until mid-June 2020. However, the number of new cases resurged until it reached another peak in early September of the same year (USA Facts, 2020). The demographic characteristics of persons with confirmed cases changed over time. In late May, according to IDPH COVID-19 data, cases in Cook County occurred predominantly among people in their 40s to 50s. The composition shifted in early September as there were more cases among people in their 20s to 30s, indicating that the number of cases rose among younger people over summer 2020. The number of new cases fluctuated across regions of the county over different periods.

State and local authorities have worked together to contain COVID-19 in Cook County. In March 2020, the governor of Illinois issued multiple executive orders to maximize the containment. At the local level, Chicago west suburb areas like Oak Park, River Forest, and Forest Park imposed a shelter-in-place order to their community by March 20, 2020 (Schering, 2020). To respond the surge in new cases among Hispanic residents in Chicago, the mayor of

Chicago along with the governor of Illinois worked to focus on the mitigation measures in predominantly Hispanic areas, such as Little Village and Archer Heights by launching a multilingual digital and video campaign and holding virtual town hall meetings for seniors and immigrant youth in Hispanic communities (Spielman & Sfondeles, 2020).

## CHAPTER 3

### RESEARCH PROBLEM

This study investigated the spatial and temporal pattern of COVID-19 risk in Cook County. This study also assessed the impact of neighborhood socioeconomic and demographic characteristics on COVID-19 incidence as factors like unemployment, race, and ethnicity can significantly affect COVID-19 outcomes.

To the best of our knowledge, there is limited research investigating the spatial distribution of COVID-19 risk from different time points using the fine-scale geographical unit, such as zip codes. This thesis applied spatial analysis to identify the geographical distribution of COVID-19 risk from time to time specified to 168 zip codes in Cook County. Many COVID-19 related studies concentrated their investigation on the metropolitan area, like New York City and Chicago (Gottlieb et al., 2020; Maroko et al., 2020). This study extended to the entire county to account for COVID-19 cases and different restriction measures in Chicago metropolitan area and its surrounding regions.

The following research questions were used to guide the study:

1. How did the spatial distribution of COVID-19 risk change over time in Cook County?
2. Was the risk of COVID-19 the same across the county from time to time?
3. Did COVID-19 risk decrease? If yes, were the decreases in COVID-19 risk uniform across the county?

4. What areas in the county remained at high risk of COVID-19 over different periods?
5. What were the effects of socioeconomic and demographic factors on COVID-19 incidence in Cook County?

## CHAPTER 4

### METHODS

#### 4.1. Data

This study collected Cook County's COVID-19 data from the IDPH website (Illinois Dept. of Public Health, 2020). This public data consists of the cumulative daily counts of COVID-19 cases in the zip code level. Besides reporting the number of daily cumulative cases, this data also reports the demographic characteristics of COVID-19 cases, such as age group, race, and gender. The data were collected daily from August 2020 to December 2020.

A zip-code level of demographic and socioeconomic data was derived from the 2018 five-year-estimates American Community Survey (ACS) (U.S. Census Bureau, 2020). Spatial data were used to locate COVID-19 new cases in Cook County. These data consist of zip code level shapefiles obtained from the U.S. Census Bureau (*TIGER/Line Shapefile*, 2017).

In this study, 168 zip codes in Cook County were included in the analysis. Seven zip codes with no population information in ACS data, such as total population and number of populations at specific age groups, were removed from the analysis.

#### 4.2. Standardized incidence ratio (SIR)

The SIR was used to identify if the number of COVID-19 new cases in a zip code was higher or lower than the predicted number of new cases. In this study, the SIR was expressed as the ratio of the number of new cases (observed) and the predicted number of new cases (expected) per zip code.

The observed case was defined as the number of new positive cases in two weeks (13 days). The expected case was an estimate based on national baseline data. It was calculated by multiplying the age-specific Illinois COVID-19 incidence rate by the zip code age-specific population size. The two-week period was chosen to account for the Coronavirus incubation period (CDC, 2020c). The first two-week period was August 19-31, 2020. This period was named ‘period 1’. The last two-week period (period 10) was December 14-26, 2020. Figure 2 summarizes the number of new cases in two-week periods.

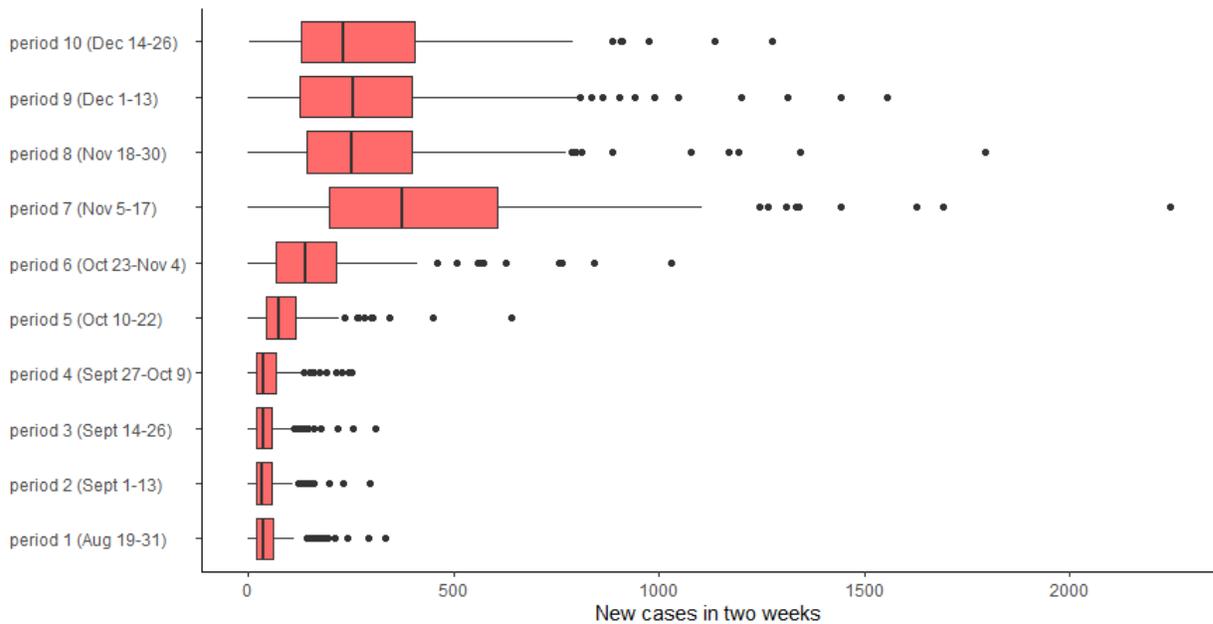


Figure 2. The summary of the number of new cases in two-week periods

Indirect standardization was used in the SIR calculation as the state’s rates were applied to each zip code’s population weights. Before calculating SIR for each zip code, the Illinois age-specific COVID-19 incidence rate was calculated. The calculation started with the count of Illinois COVID-19 new cases in two weeks by the end of November 2020 for each of the eight age groups. The number of new cases for each age group was divided by each age-group estimated population to calculate crude rates. The crude rate was expressed as the number of new cases per 100,000 population at risk. The age group distribution from the U.S. 2000 standard population was used (Klein & Schoenborn, 2001). Each age group crude rate was multiplied by the standard population's proportion to obtain the age-specific Illinois COVID-19 incidence rates. Table 1 shows the age-specific COVID-19 incidence rates in Illinois.

Table 1. The age-specific Illinois COVID-19 incidence rates

Age	Total COVID-19 new cases in two weeks (by November 2020)	Illinois ACS 2018 Estimated population	Crude rate (per 100,000 population)	Proportion of U.S. 2000 standard population	Age-specific state rate (per 100,000 population)
0 to 19	8,128	3,263,837	249.03	0.29	71.44
20 to 29	9,684	1,773,113	546.16	0.13	71.55
30 to 39	8,991	1,723,062	521.80	0.15	79.21
40 to 49	8,480	1,645,914	515.22	0.15	79.33
50 to 59	7,910	1,739,885	454.63	0.11	50.54
60 to 69	5,868	1,400,062	419.12	0.07	30.62
70 to 79	3,615	786,166	459.83	0.06	27.02
80+	2,692	489,458	550.00	0.03	18.34

The details of SIR calculation for a zip code in a particular two-week period are illustrated in Table 2. Figure 3 presents the spatial distribution of average SIR over all ten two-week periods.

Table 2. Calculating SIR of a zip code in a two-week period

Age	Age-specific state rate	Zip code's population size	Expected number of new cases	Observed number of new cases
0 to 19	0.000714	18,367	13.12	117
20 to 29	0.000716	10,683	7.64	109
30 to 39	0.000792	8,451	6.69	94
40 to 49	0.000793	7,324	5.80	110
50 to 59	0.000505	7,061	3.56	85
60 to 69	0.000306	5,176	1.58	53
70 to 79	0.00027	3,378	0.91	25
80+	0.000183	1,055	0.19	18
Total			39.49	611
Zip code's SIR in a period			$611/39.49 = 15.47$	

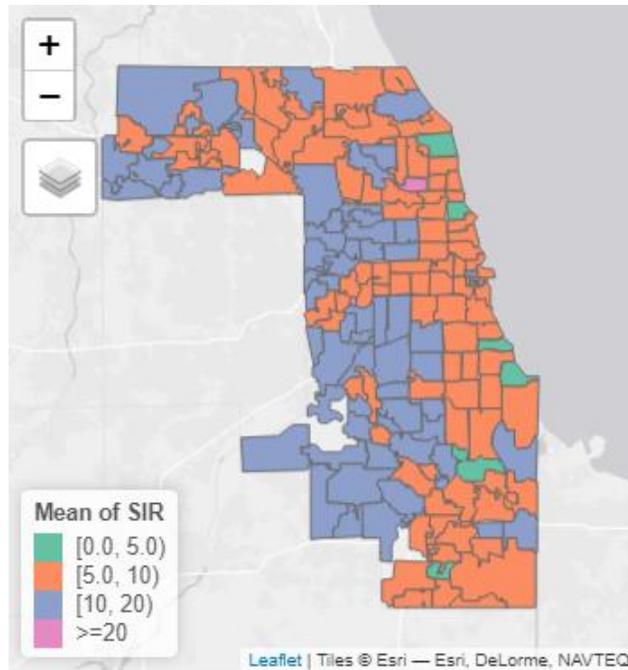


Figure 3. Average SIR over all ten two-week periods from August to December 2020

### 4.3. Statistical analysis

#### 4.3.1. Assessing spatial and temporal autocorrelation

The presence of spatial autocorrelation in the COVID-19 data was assessed using residuals from a simple over-dispersed Poisson log-linear model. This model included some independent variables. These independent variables were selected through a variable selection process.

Variables in the model were selected using a process that considered the combination of background knowledge from relevant studies and statistical assessments. The minority racial and ethnic groups were among variables of interest after several studies showed that these groups were more severely affected by the COVID-19. This study was also interested in examining the effect of low-socioeconomic status variables, such as poverty, education attainment, and Supplemental Nutrition Assistance Program (SNAP) participation to COVID-19 incidence.

Using the correlation matrix, one of the variables from highly correlated variables was removed from the analysis. For example, the unemployment rate was highly and positively correlated with the percentage of African American population, the percentage of households with SNAP participation was also highly and positively correlated with unemployment rate, and the percentage of population with college degree was highly and negatively correlated with SNAP participation. In this case, the unemployment rate was selected to be in the model, while the percentage of African American, SNAP participation, and population with college degree were removed. Although the poverty rate was one of the interest variables, it was excluded due to missing data. In the end, four covariates were selected to be in the model. These covariates were the percentage of Hispanic population, the percentage of population that use public transportation, the percentage of people with insurance coverage, and the unemployment rate.

Figure 4 visualizes the spatial distribution of four independent variables in the model. The model is formulated in equation 4 and 5, with the number of COVID-19 new cases within a zip code  $i$  at period  $k$ ,  $Y_{ik}$ , as the response variable.

$$Y_{ik} = \text{Poisson}(E_{ik}\theta_{ik}) \tag{4}$$

$$\log(\theta_{ik}) = \beta_0 + \beta_1 \text{Hispanic}_{ik} + \beta_2 \text{Insurance}_{ik} + \beta_3 \text{Unemployment}_{ik} + \beta_4 \text{PublicTransit}_{ik} \tag{5}$$

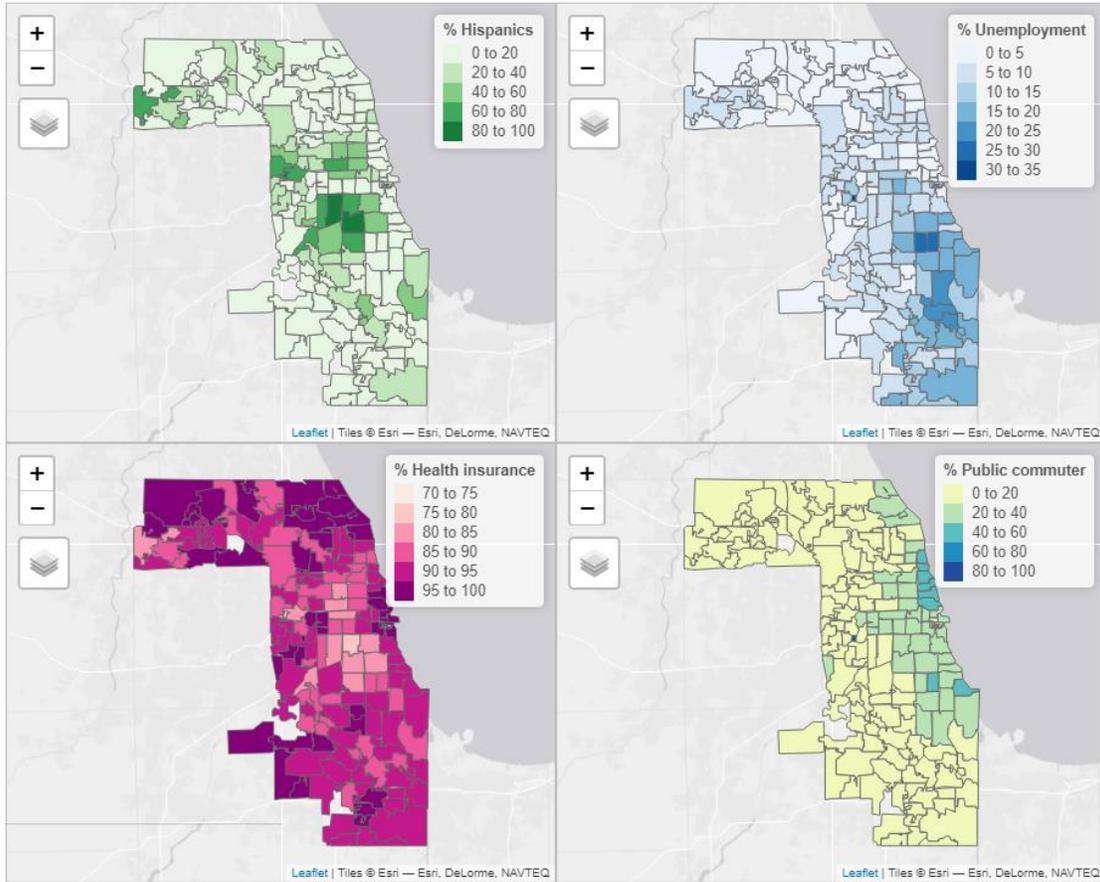


Figure 4. Socioeconomic and demographic characteristics in Cook County

Moran’s  $I$  statistic for each two-week period was computed to identify the presence of spatial autocorrelation from the model’s residuals. The estimated Moran’s  $I$  values from period 1 to 10 ranging from 0.21 to 0.37, with the p-value for each period was less than 0.05, as shown in Table 3. These results indicate strong evidence of unexplained spatial autocorrelation in the residuals.

Table 3. Moran's *I* statistic

Period	Moran's <i>I</i>	P-value
1	0.33693	0.000099*
2	0.37281	0.000099*
3	0.32613	0.000099*
4	0.30535	0.000099*
5	0.25533	0.000099*
6	0.28683	0.000099*
7	0.30905	0.000099*
8	0.21619	0.000099*
9	0.30844	0.000099*
10	0.25695	0.000099*

The temporal autocorrelation for each zip code was assessed using the Durbin-Watson test. There were ten time points per zip code. As shown in randomly selected zip codes in Table 4, the test results indicate temporal autocorrelation in some zip codes

Table 4. Durbin-Watson statistic

Zip code	Durbin-Watson statistic	P- value
60458	1.152676	0.02385190*
60606	1.494530	0.09542648
60655	2.039615	0.36220985
60623	1.131449	0.02140756*
60402	1.421792	0.07436258
60714	1.140752	0.02245555*

#### 4.3.2. Bayesian spatio-temporal modeling

Now that Moran's  $I$  and Durbin-Watson test results suggest the presence of non-independent space and time structure in the data, a model that accounts for spatio-temporal random effects should be allowed. For this purpose, a model proposed by Rushworth et al. (2014) was used in the analysis. This model introduces a single set of random effects ( $\phi_{ik}$ ). These random effects account for both spatial and temporal autocorrelation. The random effects include spatial autocorrelation by imposing a function that depends on the neighborhood matrix ( $W$ ). The neighborhood matrix is defined as 1 when area unit  $i$  shares common borders with area unit  $l$ ,  $W_{il} = 1$ , or otherwise 0, and  $W_{ii} = 0$ . A Gaussian Markov Random Field (GMRF) model is used in the model to represent temporal autocorrelation, where the distribution of random effects at time  $k$  depends on random effects at time  $k - 1$  using Gaussian normal distribution. The model with random effects is formulated as

$$\log(\theta_{ik}) = \beta_0 + \beta_1 \text{Hispanic}_{ik} + \beta_2 \text{Insurance}_{ik} + \beta_3 \text{Unemployment}_{ik} + \beta_4 \text{PublicTransit}_{ik} + \phi_{ik} \quad (6)$$

The random effects at one point depend on the value of random effects at the previous time as described in equation 7.  $\phi_k$  represents the vector of random effect for period  $k$ . It changes from time to time via  $\alpha$ , a multivariate first-order autoregressive process with temporal autoregressive parameter. The temporal and spatial autocorrelation is induced via mean  $\alpha\phi_{k-1}$  and variance  $\tau^2 Q(\rho, W)^{-1}$ , respectively. The precision matrix  $Q(\rho, W)$  was previously proposed by Leroux et al. (2000). The joint prior distribution for  $\phi_1$  is defined in equation 10 where the spatial autocorrelation is induced by matrix  $Q(\rho, W)$ .

$$\phi_k | \phi_{k-1} \sim N(\alpha\phi_{k-1}, \tau^2 Q(\rho, W)^{-1}) \quad k = 2, \dots, N \quad (7)$$

$$\tau^2 \sim \text{Inverse-Gamma}(a, b) \quad (8)$$

$$\alpha, \rho \sim \text{Uniform}(0, 1) \quad (9)$$

$$\phi_1 \sim N(0, \tau^2 Q(\rho, W)^{-1}) \quad (10)$$

The spatio-temporal model was fitted using the `ST.CARar()` function from `CARBayesST` R package (Lee et al., 2020). The model was run for 550,000 MCMC samples with a burn-in period of 50,000. To minimize the autocorrelation of the Markov chain, the samples were thinned by 10. As a result, there were 50,000 samples for inference.

## CHAPTER 5

### RESULTS

#### 5.1. Temporal and spatial trends

The average COVID-19 fitted risk in each two-week period was always above 1, as shown in Figure 5. It indicates that, on average, Cook County had a higher than expected number of new cases in each period.

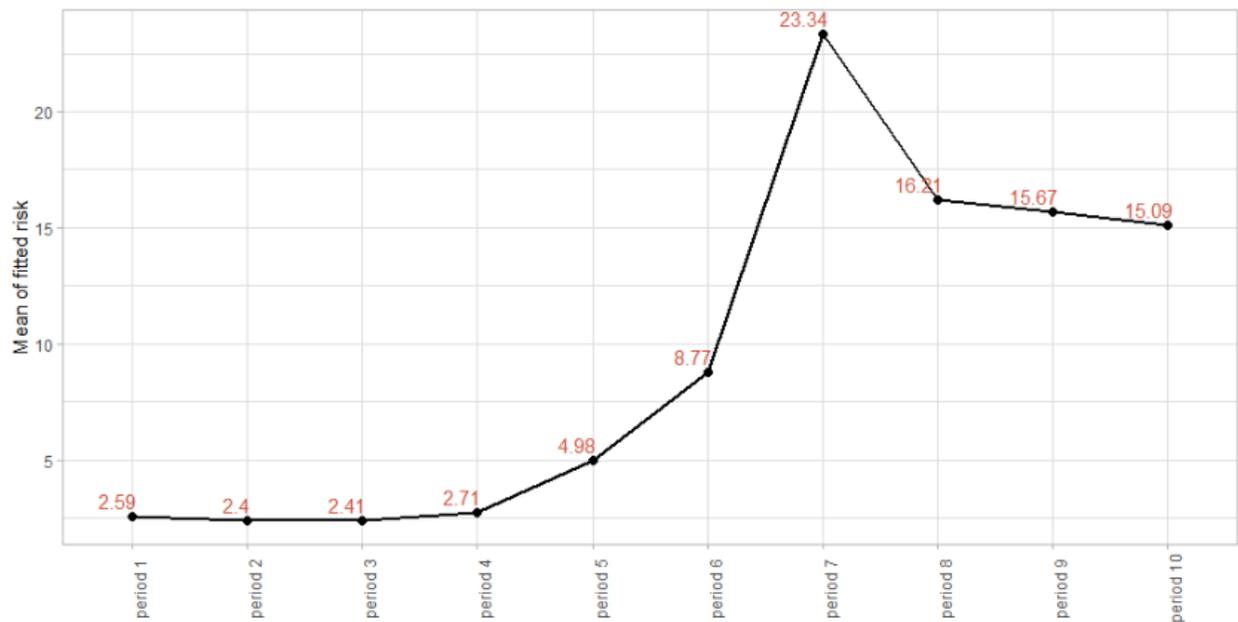


Figure 5. Average COVID-19 fitted risks

Figure 5 shows that the average fitted risk was relatively similar from period 1 to 4, with a slight shift between periods. On average, the risk kept increasing from period 5 to 7. Period 7 had the highest mean of COVID-19 risk. The average risk in this period was 23.34,

suggesting that the average number of new cases between November 5-17, 2020 was 23 times the expected number of new cases. On average, the risk declined in period 8. However, the decreases were not uniform across the county as some areas remained at high risk, as shown in the map of period 8 in Figure 6.

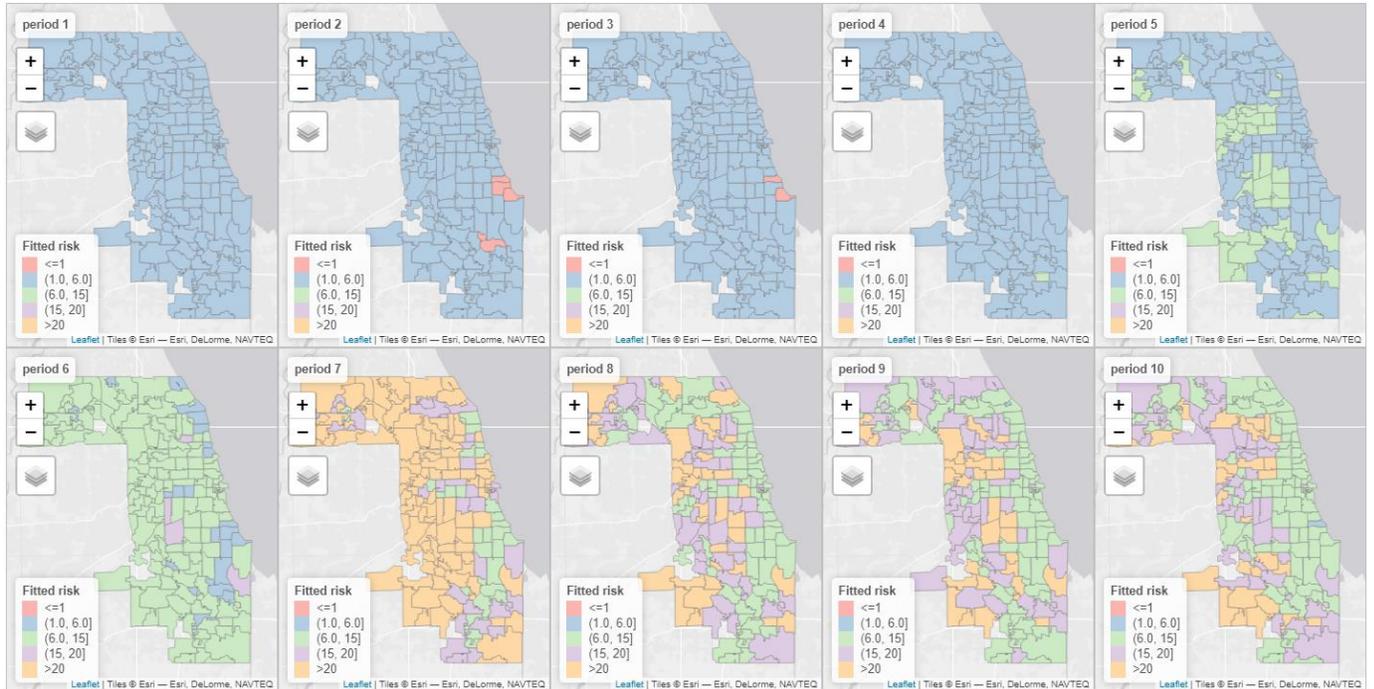


Figure 6. The progression of COVID-19 fitted risks from period 1 (August 19<sup>th</sup> to 31<sup>st</sup>, 2020) to period 10 (December 14<sup>th</sup> to 26<sup>th</sup>, 2020)

Figure 6 shows COVID-19 risks in period 1 were uniform across the county, with risks ranged from above 1.0 to 6.0 in every zip code. Two and four weeks later, although almost all zip code risks were the same as period 1, few zip codes in southside Chicago experienced lower risk (less than or equal to 1), as shown in period 2 and period 3.

The risks started to elevate to above 6.0 in some areas, particularly in Chicago west side, Chicago northwestern side, west Cook County suburbs, northwestern and southwestern parts of the county, as seen in the map of period 5. Two weeks later, the risk of COVID-19

appeared to be higher than 6.0 in almost all areas in the county, except few zip codes in Chicago north suburbs, like Evanston and Wilmette, and some zip codes in Chicago south side area like Hyde Park and South Shore.

By period 7, COVID-19 risks increased to above 20 in most of the zip codes in the county's western part. Some zip codes in the eastern part of the county also experienced elevated risk in the same period, but the risk was lower than 20. The risks in zip codes decreased gradually in period 8. At this period, the risk in some zip codes in the northern part of the county decreased at a faster rate. However, the risk in Chicago west side, west Cook County suburbs, and southwest Cook County suburbs decreased at a slower pace, as shown in period 8.

Despite the growing concern that Thanksgiving gathering could spread the virus, the risk of COVID-19 post-Thanksgiving in period 9 declined, as shown in both Figure 5 and Figure 6. Period 9 map shows that some zip codes in Chicago southwest side and the northwestern and southern parts of the county experienced a decrease in risk from period 8.

The spatial evolution of COVID-19 risk from period 1 to 10 shows that some zip codes in the county had a relatively lower risk. The lower risk zip codes located in Chicago north side area (e.g., Uptown and Ravenswood), Chicago west suburbs (e.g., Oak Park and River Forest), Chicago west side neighborhoods near Oak Park, like Austin, Chicago north suburbs (i.e., Evanston), downtown Chicago, and Chicago south side neighborhoods, like Hyde Park. However, some zip codes constantly experienced high risk and were more negatively impacted by COVID-19.

The zip codes where COVID-19 risk was persistently high located in north Cook County suburbs (e.g., Rolling Meadows and Elgin), inner suburbs of north Chicago, (e.g.,

Lincolnwood), southwest Cook County suburbs, (e.g., Palos Heights and Orland Park), Chicago west side, Chicago southeast side (e.g., South Deering), and west Cook County suburbs (e.g., Franklin Park). Many of these areas, such as Chicago west side, Franklin Park, Elgin, and South Deering are places with a high percentage of Hispanic populations.

## 5.2. COVID-19 in high-risk areas

Figure 7 plots the minimum, the maximum, and the average fitted risk of each Cook County’s zip code. The blue dot represents the average fitted risk. Zip codes with average fitted risk above 13 are highlighted in a dashed red line.

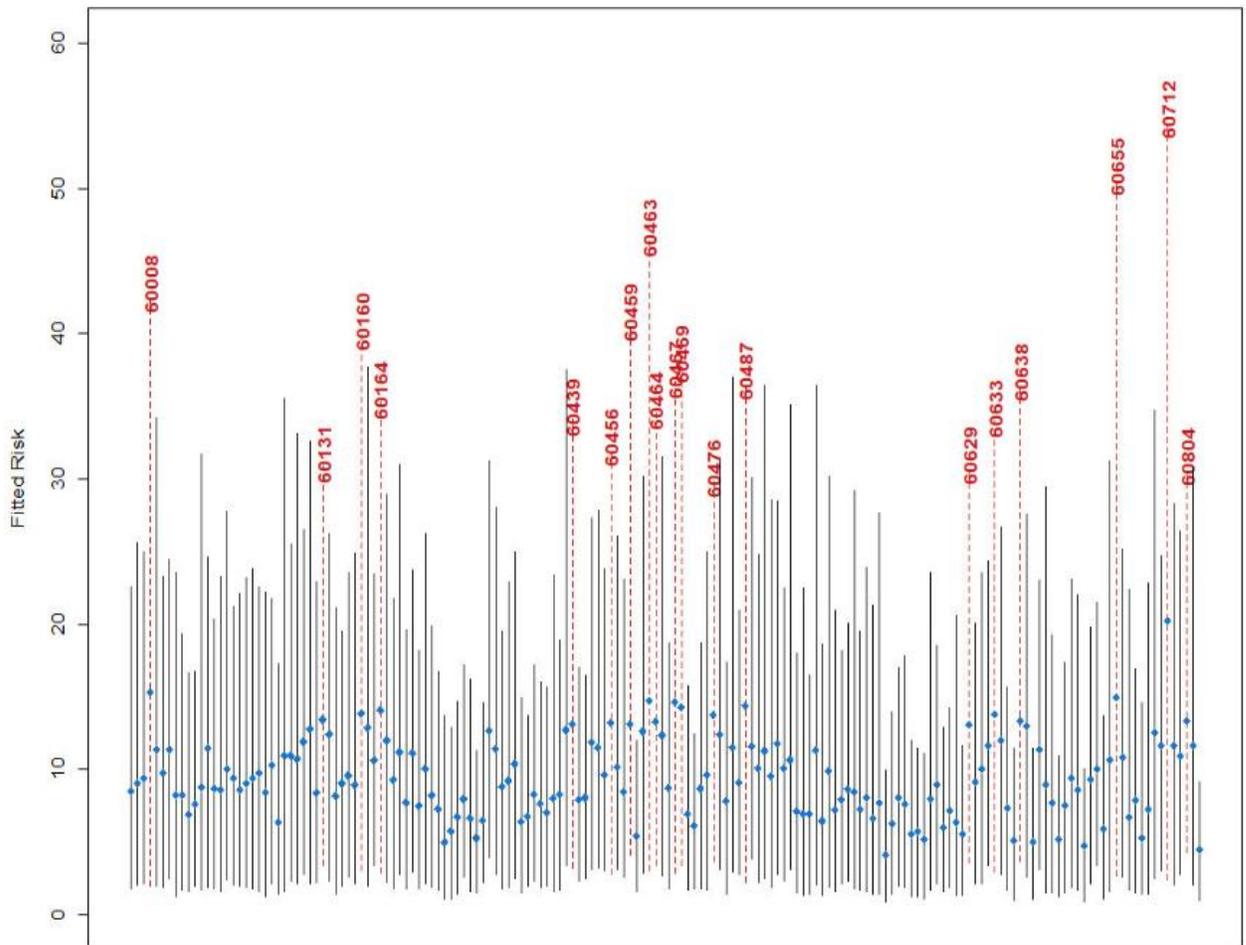


Figure 7. The minimum, the maximum, and the average fitted risk of each zip code

Figure 8 describes further the evolution of COVID-19 fitted risk in four selected high-risk zip codes. The selected zip codes are located in different areas across the county. Zip code 60629 is the only zip code located in Chicago. Zip code 60164, 60464, and 60712 are located in Cook County's suburbs.

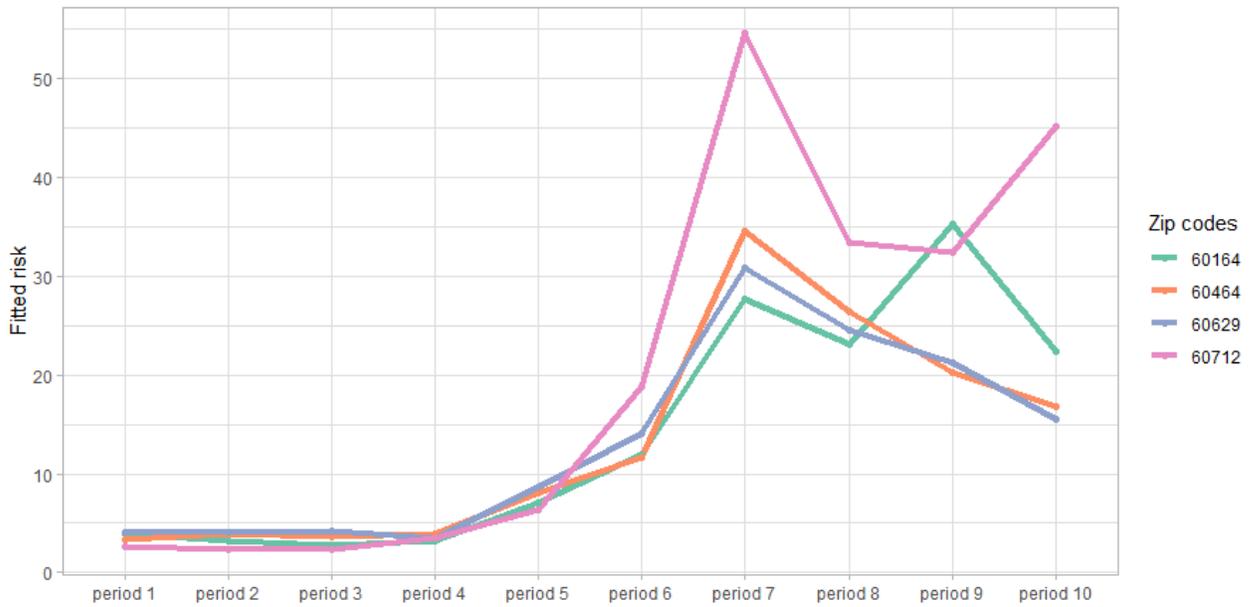


Figure 8. The trend of COVID-19 fitted risk in selected high-risk zip codes

In general, the risk trend in Figure 8 shows an identical pattern. However, individual zip code shows more heterogeneous pattern after period 7. Zip code 60464 and 60629 show a consistent decline in risk after the risk peaked in period 7. But, the risk in zip code 60164 and 60712 remained higher than the other zip codes.

### 5.3. Covariates effects on COVID-19 risk

Model summary from posterior distributions is presented in Table 5. The table consists of the posterior median, the 95% credible intervals (2.5%, 97.5%), and the Geweke's convergence diagnostic score (Geweke, 1992). Geweke's diagnostic with the values between -1.96 and 1.96 indicate convergence. Besides using the Geweke diagnostic score, the convergence was also assessed using convergence plots.

Table 5. The summary of posterior distribution

	Median	2.5%	97.5%	Geweke diagnostic
<i>(Intercept)</i>	0.4547	-0.5268	1.3458	1.0
<i>Hispanics</i>	0.0090	0.0067	0.0112	-1.5
<i>Unemployment</i>	-0.0023	-0.0096	0.0060	-1.3
<i>Health Insurance</i>	0.0147	0.0056	0.0248	-0.9
<i>Public Transit Commuter</i>	-0.0091	-0.0134	-0.0051	-0.7
$\tau^2$	0.1810	0.1631	0.2008	-3.6
$\rho S$	0.9951	0.9898	0.9981	-1.0
$\rho T$	0.6332	0.5758	0.6885	-1.5

The summary of the fitted model above shows that all variables, except *Unemployment*, indicate relationships with COVID-19 incidence as their 95% credible intervals do not contain zero. The spatial ( $\rho S$ ) and temporal ( $\rho T$ ) dependence parameters indicate the strength of spatial and temporal autocorrelations. The spatial and temporal dependence parameters show high values ( $\rho S = 0.9951$ ;  $\rho T = 0.6332$ ), indicating neighboring zip codes are more likely to have a similar risk of COVID-19 than zip codes that are more apart.

Using the exponent of posterior median and the 95% credible interval to obtain the relative risk, the corresponding relative risk for a 1% increase in the percentage of Hispanic

population was 1.009 (1.007, 1.011), indicating that such an increase in Hispanic populations corresponds to a rise in COVID-19 risk by 0.9%. In contrast, a 1% increase in the unemployment rate had no significant effect on COVID-19 risk. The posterior median and the 95% credible interval for the relative risk of a 1% increase in the percentage of people with health insurance coverage was 1.015 (1.006, 1.025), showing that every 1% increase in the health insurance coverage corresponds to 1.5% additional COVID-19 risk. The relative risk for the percentage of people using public transportation was 0.991 (0.987, 0.995), suggesting that such an increase was associated with a decrease in COVID-19 risk by 0.9%.

## CHAPTER 6

### DISCUSSION

This study evaluated the spatio-temporal patterns of COVID-19 risk in Cook County. A Bayesian spatio-temporal model was applied to the zip code level COVID-19 data to estimate the risk of COVID-19 from different time points between August 2020 to December 2020. The model incorporated socioeconomic and demographic covariates (percent Hispanic population, unemployment rate, percent health insurance coverage, and percent public transportation commuters). Furthermore, this study evaluated the impact of the covariates on COVID-19 incidence.

The risk of COVID-19 gradually increased over time. The spatial patterns of COVID-19 risk show a growing number of areas experiencing elevated risk during the study. In period 7 (November 5-17, 2020), the risk of COVID-19 reached its peak in most zip codes of Cook County. COVID-19 risk during this period increased to above 20 in many zip codes located in the western part of the county, indicating the number of new cases was more than 20 times the number of expected new cases. Yet, during this same period, the risk remained lower at 15 at most in some zip codes in Evanston (Chicago north suburb), Chicago north side near the lakeshore, Oak Park and River Forest (Chicago west suburbs), Chicago west side (e.g., West Garfield Park and Austin), and Chicago south side (e.g., Bronzeville and Hyde Park).

The increasing COVID-19 risk in the county in period 7 may be attributed to multiple aspects, including eased COVID-19 restrictions, which allowed gathering with a limited number of people and opening restaurants and other public places. Pandemic fatigue could also cause a spike in new cases as a public response to prolonged public health emergency (WHO, *Pandemic Fatigue*, 2020). In response to the jump of new COVID-19 cases, the state government announced Tier 3 mitigations on November 20, 2020, to restrict restaurants and bars to offer indoor services and close other public places, such as museums and casinos (NBC Chicago, 2020). Furthermore, the mayor of Chicago strongly advised Chicago residents to stay at home as much as possible and avoid Thanksgiving gatherings (City of Chicago, 2020).

After several restriction measures, the risk was lower in period 8, and the risk continued to decline gradually from period 8 to 10. Although the overall risk in period 8 decreased, the decreases occurred partially as some areas in the county remained at high risk of COVID-19.

Based on the risk maps, some parts of the county exhibited constant COVID-19 high-risk from time to time. These high-risk areas, where many of them are areas with a high percentage of the Hispanic population in Chicago and Cook County's suburbs, were not among regions with the highest testing rates, according to IDPH COVID-19 data. During the pandemic, testing is the most important tool to slow the spread of COVID-19. Tests allow us to identify cases, recommend medical treatment for the infected individuals, and instruct close contacts with no symptoms to self-quarantine.

The posterior medians and 95% credible intervals for the relative risks show that a higher percentage of Hispanic residents was associated with higher COVID-19 risk. These

findings align with previous studies focusing on the impact of racial segregation on the COVID-19 infection rate (Hu et al., 2020; Podewils et al., 2020) that minority groups, like Hispanics, were disproportionately affected by COVID-19. Findings from these studies identified that Coronavirus was infecting Hispanic populations at a higher rate due to cultural, socioeconomic, and lifestyle factors. Language barriers and cultural challenges among Hispanic communities may cause vulnerable populations to be uninformed of COVID-19 important information, such as how the virus spreads and where to get free testing. Culturally, many Hispanic households are intergenerational, which creates a larger household size and could increase the risk of catching the virus. Moreover, many Hispanic adults worked in essential industries that require physical presence and worked in jobs with no paid sick leave, making them keep coming to work while sick.

This study found that a higher percentage of population with health insurance coverage was linked to higher COVID-19 risk, suggesting that people with health insurance were more likely to participate in testing, hence more likely to identify cases. In contrast, uninsured individuals may not have a usual place to seek medical treatment and get a referral for a COVID-19 test. People without health insurance may have limited information on where to get a free COVID-19 test. Also, they may be reluctant to take the test due to the fear of paying full out-of-pocket for the test (Tolbert, 2020).

Despite the findings from previous studies that recognized public transportations as high-risk environments for COVID-19 transmission (Buja et al., 2020), this study found that a higher percentage of public transportation commuters was associated with lower COVID-19 risk. In Cook County, according to ACS data, the high percentage of public transportation commuters was concentrated in Chicago, particularly in Chicago north side and downtown

Chicago neighborhoods. The unemployment and poverty rate in these neighborhoods were low. Based on ACS data, many residents in these neighborhoods worked in management, business, science, and art occupations, suggesting that their work may not require a physical presence and can be done remotely. As a result, they have the option to work from home more, avoid using public transport, and minimize the risk of transmission in public places, unlike workers who engaged in manual labor.

### 6.1. Limitations

The findings in this study are subject to some limitations. First, this study used aggregated COVID-19 zip code level with no data on the individual level. Individual-level data would provide better estimates in risk and more accurately quantify the covariate effects on COVID-19 incidence. Second, the number of COVID-19 positive cases in this study used only the cases from known age group cases of IDPH COVID-19 data. The cases from unknown age groups were excluded from the analysis. Third, IDPH provides the number of total positive cases and the number of positive cases per age group. However, there are some discrepancies between the sum of positive cases from all age groups and total positive cases. For the analysis, this study used only the number of positive cases per age group. Fourth, this study's socioeconomic and demographic data do not reflect the change in socioeconomic status that may occur during pandemic; for example, losing job during pandemic may cause people to no longer have health insurance coverage and reducing bus and train passengers capacity may lower the percentage of public transportation commuters. It may cause under or over-estimation of COVID-19 risk. Finally, the period of this study started five months after

the first COVID-19 case was identified in the U.S. Consequently, this study does not provide the whole picture of how COVID-19 risk evolved from the beginning of the pandemic.

## 6.2. Conclusions

Although the level of COVID-19 risk varied across the county, the risk maps show that COVID-19 risk remained high in some areas of the county over different periods. Among these constant high-risk areas, many of them are neighborhoods with a high percentage of Hispanic residents. Lack of access to health care facilities, like free testing sites, limited access to culturally responsive health care, and poor living and working conditions may be attributed to the constant high risk of COVID-19 among Hispanic communities. These findings can present an opportunity to provide better health care for vulnerable populations by providing culturally appropriate programs to address health disparities, opening free testing sites in Hispanic communities, strengthening paid leave policies among low-income families whose work requires physical presence with lack of paid sick leave, prioritizing vaccine to Hispanic community, improving health data collection, and observing disparities in the number of COVID-19 cases among racial and ethnic groups.

The findings in this study provide some insights for future research. There are several additional areas for further development and application. While this study focused on the number of new COVID-19 cases, future studies could further investigate the risk and identify disease patterns using the number of active cases over different time points. This study has demonstrated the importance of considering socioeconomic and demographic covariates that may impact the disease's outcome. As convenient access to COVID-19 testing sites is critical to fighting the pandemic, the future study may look for the effect of access to testing sites

related variables that are still limited in COVID-19 related study. These variables may include the average or median travel time to the nearest testing sites and spatial accessibility to the testing sites that consider the site-to-population ratio.

This study highlights the importance of integrating the geographical information system into disease routine surveillance programs, creating and updating COVID-19 risk maps, and transforming routinely collected health data into critical information. This information can be used to allocate resources in the neighborhoods most impacted by COVID-19 (i.e., providing more free testing sites and implementing effective vaccine distribution toward high-risk groups, including people of color) and assess the effectiveness of health policies.

## REFERENCES

- Abedi, V., Olulana, O., Avula, V., Chaudhary, D., Khan, A., Shahjouei, S., Li, J., & Zand, R. (2020). Racial, Economic and Health Inequality and COVID-19 Infection in the United States. *MedRxiv : The Preprint Server for Health Sciences*.  
<https://doi.org/10.1101/2020.04.26.20079756>
- Adyro Martínez-Bello, D., López-Quílez, A., & Prieto, A. T. (2017). Relative risk estimation of dengue disease at small spatial scale. *International Journal of Health Geographics*, 16, 31.  
<https://doi.org/10.1186/s12942-017-0104-x>
- Alexander, F. E., Williams, J., McKinney, P. A., Ricketts, T. J., & Cartwright, R. A. (1989). A specialist leukaemia/lymphoma registry in the UK. Part 2: Clustering of Hodgkin's disease. *British Journal of Cancer*, 60(6), 948–952. <https://doi.org/10.1038/bjc.1989.396>
- Anderson, C., & Ryan, L. M. (2017). A comparison of spatio-temporal disease mapping approaches including an application to ischaemic heart disease in New South Wales, Australia. *International Journal of Environmental Research and Public Health*, 14(2).  
<https://doi.org/10.3390/ijerph14020146>
- Azevedo, L., Pereira, M. J., Ribeiro, M. C., & Soares, A. (2020). Geostatistical COVID-19 infection risk maps for Portugal. *International Journal of Health Geographics*, 19(1), 25.  
<https://doi.org/10.1186/s12942-020-00221-5>
- Begum, F. (2016). *Mapping disease: John Snow and Cholera — Royal College of Surgeons*.  
<https://www.rcseng.ac.uk/library-and-publications/library/blog/mapping-disease-john-snow-and-cholera/>
- Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). Applied Spatial Data Analysis with R: Second Edition. In *Applied Spatial Data Analysis with R: Second Edition*. Springer New York. <https://doi.org/10.1007/978-1-4614-7618-4>
- Buja, A., Paganini, M., Cocchio, S., Scioni, M., Rebba, V., & Baldo, V. (2020). Demographic and socio-economic factors, and healthcare resource indicators associated with the rapid spread of COVID-19 in Northern Italy: An ecological study. *PLOS ONE*, 15(12), e0244535.  
<https://doi.org/10.1371/journal.pone.0244535>

- Cauguiran, C., Horng, E., & Kirsch, J. (2020). *Coronavirus Chicago: Mayor closes lakefront, 606 Trail, Riverwalk to stop spread of COVID-19*. March 26, 2020. <https://abc7chicago.com/coronavirus-chicago-lakefront-closed-riverwalk-lori-lightfoot/6051898/>
- CDC. (2020a). *COVID Data Tracker*. [https://covid.cdc.gov/covid-data-tracker/?CDC\\_AA\\_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fcases-updates%2Fcases-in-us.html#demographics](https://covid.cdc.gov/covid-data-tracker/?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fcases-updates%2Fcases-in-us.html#demographics)
- CDC. (2020b). *First Travel-related Case of 2019 Novel Coronavirus Detected in United States / CDC Online Newsroom / CDC*. CDC Press Release. <https://www.cdc.gov/media/releases/2020/p0121-novel-coronavirus-travel-case.html>
- CDC. (2020c). *Management of Patients with Confirmed 2019-nCoV / CDC*. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html>
- City of Chicago. (2020). *Chicago - Stay at home advisory*. [https://www.chicago.gov/content/dam/city/sites/covid/health-orders/201112\\_Stay at home advisory and guidance\\_vF.pdf](https://www.chicago.gov/content/dam/city/sites/covid/health-orders/201112_Stay%20at%20home%20advisory%20and%20guidance_vF.pdf)
- Cordes, J., & Castro, M. C. (2020). Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spatial and Spatio-Temporal Epidemiology*, 34, 100355. <https://doi.org/10.1016/j.sste.2020.100355>
- DiMaggio, C. (2014). *Spatial Epidemiology Notes: Applications and Vignettes in R*. [http://www.columbia.edu/~cjd11/charles\\_dimaggio/DIRE/resources/spatialEpiBook.pdf](http://www.columbia.edu/~cjd11/charles_dimaggio/DIRE/resources/spatialEpiBook.pdf)
- Elliott, P., & Wartenberg, D. (2004). Spatial epidemiology: Current approaches and future challenges. In *Environmental Health Perspectives* (Vol. 112, Issue 9, pp. 998–1006). Public Health Services, US Dept of Health and Human Services. <https://doi.org/10.1289/ehp.6735>
- Gayawan, E., Awe, O. O., Oseni, B. M., Uzochukwu, I. C., Adekunle, A., Samuel, G., Eisen, D. P., & Adegboye, O. A. (2020). The spatio-temporal epidemic dynamics of COVID-19 outbreak in Africa. *Epidemiology and Infection*, 148. <https://doi.org/10.1017/S0950268820001983>
- Geweke, J., & Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. *IN BAYESIAN STATISTICS*, 4, 169--193. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.27.2952>
- Gottlieb, M., Sansom, S., Frankenberger, C., Ward, E., & Hota, B. (2020). Clinical Course and Factors Associated With Hospitalization and Critical Illness Among COVID-19 Patients in Chicago, Illinois. *Academic Emergency Medicine*, 27(10), 963–973. <https://doi.org/10.1111/acem.14104>

- Hu, T., Yue, H., Wang, C., She, B., Ye, X., Liu, R., Zhu, X., Guan, W. W., & Bao, S. (2020). Racial Segregation, Testing Site Access, and COVID-19 Incidence Rate in Massachusetts, USA. *International Journal of Environmental Research and Public Health*, 17(24), 9528. <https://doi.org/10.3390/ijerph17249528>
- Illinois Dept. of Public Health. (2020). *COVID-19 Statistics*. <https://www.dph.illinois.gov/covid19/covid19-statistics>
- Khatana, S. A. M., & Groeneveld, P. W. (2020). Health Disparities and the Coronavirus Disease 2019 (COVID-19) Pandemic in the USA. In *Journal of General Internal Medicine* (Vol. 35, Issue 8, pp. 2431–2432). Springer. <https://doi.org/10.1007/s11606-020-05916-w>
- Klein, R. J., & Schoenborn, C. A. (2001). *Age Adjustment Using the 2000 Projected U.S. Population*. <http://www.census.gov/prod/1/pop/p25-1130/p251130.pdf>.
- Lawson, A. B. (2018). *Bayesian Disease Mapping*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781351271769>
- Lee, D., Rushworth, A., & Maintainer, G. N. (2020). *Package “CARBayesST” Title Spatio-Temporal Generalised Linear Mixed Models for Areal Unit Data*. <https://doi.org/10.1002/sim.4780142112>
- Leroux, B. G., Lei, X., & Breslow, N. (2000). *Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence* (pp. 179–191). Springer, New York, NY. [https://doi.org/10.1007/978-1-4612-1284-3\\_4](https://doi.org/10.1007/978-1-4612-1284-3_4)
- Maheswaran, R., Morris, S., Falconer, S., Grossinho, A., Perry, I., Wakefield, J., & Elliott, P. (1999). Magnesium in drinking water supplies and mortality from acute myocardial infarction in north west England. *Heart*, 82(4), 455–460. <https://doi.org/10.1136/hrt.82.4.455>
- Maroko, A. R., Nash, D., & Pavidonis, B. T. (2020). COVID-19 and Inequity: a Comparative Spatial Analysis of New York City and Chicago Hot Spots. *Journal of Urban Health*, 97(4), 461–470. <https://doi.org/10.1007/s11524-020-00468-0>
- Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of the Total Environment*, 728, 138884. <https://doi.org/10.1016/j.scitotenv.2020.138884>
- Moore, J. T., Ricaldi, J. N., Rose, C. E., Fuld, J., Parise, M., Kang, G. J., Driscoll, A. K., Norris, T., Wilson, N., Rainisch, G., Valverde, E., Beresovsky, V., Agnew Brune, C., Oussayef, N. L., Rose, D. A., Adams, L. E., Awel, S., Villanueva, J., Meaney-Delman, D., ... Westergaard, R. (2020). Disparities in Incidence of COVID-19 Among Underrepresented Racial/Ethnic Groups in Counties Identified as Hotspots During June 5–18, 2020 — 22 States, February–June 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69(33), 1122–1126. <https://doi.org/10.15585/mmwr.mm6933e1>

- MORAN, P. A. P. (1950). NOTES ON CONTINUOUS STOCHASTIC PHENOMENA. *Biometrika*, 37(1–2), 17–23. <https://doi.org/10.1093/biomet/37.1-2.17>
- Muñoz-Price, L. S., Nattinger, A. B., Rivera, F., Hanson, R., Gmehlin, C. G., Perez, A., Singh, S., Buchan, B. W., Ledebor, N. A., & Pezzin, L. E. (2020). Racial Disparities in Incidence and Outcomes Among Patients With COVID-19. *JAMA Network Open*, 3(9), e2021892. <https://doi.org/10.1001/jamanetworkopen.2020.21892>
- NBC Chicago. (2020). *All of Illinois to Enter Tier 3 Mitigations This Week, Gov. Pritzker Announces – NBC Chicago*. <https://www.nbcchicago.com/news/local/all-of-illinois-to-enter-tier-3-mitigations-this-week-pritzker-announces/2373619/>
- Neuberger, J. S., Hu, S. C., Drake, K. D., & Jim, R. (2009). Potential health impacts of heavy-metal exposure at the Tar Creek Superfund site, Ottawa County, Oklahoma. *Environmental Geochemistry and Health*, 31(1), 47–59. <https://doi.org/10.1007/s10653-008-9154-0>
- Pamuk E, Makuc D, Heck K, Reuben C, & Lochner K. (1998). *Socioeconomic Status and Health Chartbook. Health, United States*.
- Patel, J. A., Nielsen, F. B. H., Badiani, A. A., Assi, S., Unadkat, V. A., Patel, B., Ravindrane, R., & Wardle, H. (2020). Poverty, inequality and COVID-19: the forgotten vulnerable. In *Public Health* (Vol. 183, pp. 110–111). Elsevier B.V. <https://doi.org/10.1016/j.puhe.2020.05.006>
- Podewils, L. J., Burket, T. L., Mettenbrink, C., Steiner, A., Seidel, A., Scott, K., Cervantes, L., & Hasnain-Wynia, R. (2020). Disproportionate Incidence of COVID-19 Infection, Hospitalizations, and Deaths Among Persons Identifying as Hispanic or Latino — Denver, Colorado March–October 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69(48), 1812–1816. <https://doi.org/10.15585/mmwr.mm6948a3>
- Rushworth, A., Lee, D., & Mitchell, R. (2014). A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in Greater London. *Spatial and Spatio-Temporal Epidemiology*, 10, 29–38. <https://doi.org/10.1016/j.sste.2014.05.001>
- Schering, S. (2020). *Oak Park issues shelter-in-place order to slow COVID-19 spread as River Forest, Forest Park leaders urge residents to comply as well - Chicago Tribune*. March 19, 2020. <https://www.chicagotribune.com/suburbs/oak-park/ct-oak-mayors-response-covid-19-tl-0326-20200318-tpwctduzjbbbra65stygyk3gmy-story.html>
- Snow, J. (1854). *Snow's cholera map*. Wikimedia Commons. <https://commons.wikimedia.org/wiki/File:Snow-cholera-map-1.jpg>

- Spielman, F., & Sfondeles, T. (2020). *Pritzker, Lightfoot respond to surge of coronavirus cases among Hispanics in Chicago, statewide - Chicago Sun-Times*. May 6, 2020. <https://chicago.suntimes.com/coronavirus/2020/5/6/21249879/pritzker-lightfoot-coronavirus-covid-19-cases-hispanics-african-americans-chicago-illinois>
- State of Illinois. (2020). *Executive Orders Related to COVID-19*. <https://coronavirus.illinois.gov/s/resources-for-executive-orders>
- Szalinski, B. (2020). *COVID-19 in Illinois*. October 26, 2020. <https://www.illinoispolicy.org/what-you-need-to-know-about-coronavirus-in-illinois/>
- The New York Times. (2020). *Coronavirus in the U.S.: Latest Map and Case Count*. <https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html>
- The Official Website of the City of New York. (2020). *Age-adjusted rates of lab confirmed COVID-19*. <https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-deaths-race-ethnicity-04162020-1.pdf>
- TIGER/Line Shapefile. (2017). [https://catalog.data.gov/dataset/tiger-line-shapefile-2017-county-cook-county-il-topological-faces-polygons-with-all-geocodes-co/resource/b9666042-5c3c-4ae8-be27-959e38459889?inner\\_span=True](https://catalog.data.gov/dataset/tiger-line-shapefile-2017-county-cook-county-il-topological-faces-polygons-with-all-geocodes-co/resource/b9666042-5c3c-4ae8-be27-959e38459889?inner_span=True)
- Tolbert, J. (2020). *What Issues Will Uninsured People Face with Testing and Treatment for COVID-19? | KFF*. Kaiser Family Foundation. <https://www.kff.org/coronavirus-covid-19/fact-sheet/what-issues-will-uninsured-people-face-with-testing-and-treatment-for-covid-19/>
- U.S. Census Bureau. (n.d.). *QuickFacts - United States*. Retrieved November 17, 2020, from <https://www.census.gov/quickfacts/fact/table/US/PST045219#PST045219>
- U.S. Census Bureau. (2020). *American Community Survey Data*. <https://www.census.gov/programs-surveys/acs/data.html>
- USA Facts. (2020). *Cook County, Illinois Coronavirus Cases and Deaths | USAFacts*. <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/state/illinois/county/cook-county>
- WHO. (2020a). *Coronavirus*. [https://www.who.int/health-topics/coronavirus#tab=tab\\_1](https://www.who.int/health-topics/coronavirus#tab=tab_1)
- WHO. (2020b). *Pandemic fatigue*. <https://apps.who.int/iris/bitstream/handle/10665/335820/WHO-EURO-2020-1160-40906-55390-eng.pdf>
- Yancy, C. W. (2020). COVID-19 and African Americans. In *JAMA - Journal of the American Medical Association* (Vol. 323, Issue 19, pp. 1891–1892). American Medical Association. <https://doi.org/10.1001/jama.2020.6548>

