



## The Review of Regional Studies

*The Official Journal of the Southern Regional Science Association*



# The Social and Economic Factors Underlying the Incidence of COVID-19 Cases and Deaths in US Counties During the Initial Outbreak Phase\*

Nivedita Mukherji<sup>a</sup>

<sup>a</sup> *Department of Economics, Oakland University, USA*

---

**Abstract:** This paper uncovers the socioeconomic and health/lifestyle factors that can explain the differential impact of the coronavirus pandemic on different parts of the United States during the initial outbreak phase of the pandemic. Using a dynamic panel representation of an epidemiological model of disease spread, the paper develops a Vulnerability Index for US counties from the daily reported number of cases over a 20-day period of rapid disease growth. County-level economic, demographic, and health factors are used to explain the differences in the values of this index and thereby the transmission and concentration of the disease across the country. These factors are also used to examine the number of reported deaths. The paper finds that counties with high median income have a high incidence of cases but reported lower deaths. Income inequality, as measured by the Gini coefficient, is found to be associated with more deaths and more cases. The remarkable similarity in the distribution of cases across the country and the distribution of distance-weighted international passengers served by the top international airports is evidence of the spread of the virus by way of international travel. The distributions of age, race and health risk factors such as obesity and diabetes are found to be particularly significant factors in explaining the differences in mortality across counties. Counties with better access to health care, as measured by the number of primary care physicians per capita, have lower deaths, and so do places with more health awareness as measured by flu vaccination prevalence. Environmental health conditions such as the amount of air pollution are found to be associated with counties with higher deaths from the virus. It is hoped that research such as these will help policymakers to develop risk factors for each region of the country to better contain the spread of infectious diseases in the future.

*Keywords:* COVID-19, US Counties, Vulnerability Index, Socioeconomic Factors

*JEL Codes:* F43, R12, R15

---

\*Acknowledgements. The author wishes to acknowledge the detailed and extensive comments made by the reviewers of the paper. These comments greatly enhanced the quality of the paper. Any remaining errors are solely the responsibility of the author. *Corresponding Author:* Nivedita Mukherji, E-mail: mukherji@oakland.edu

## 1. INTRODUCTION

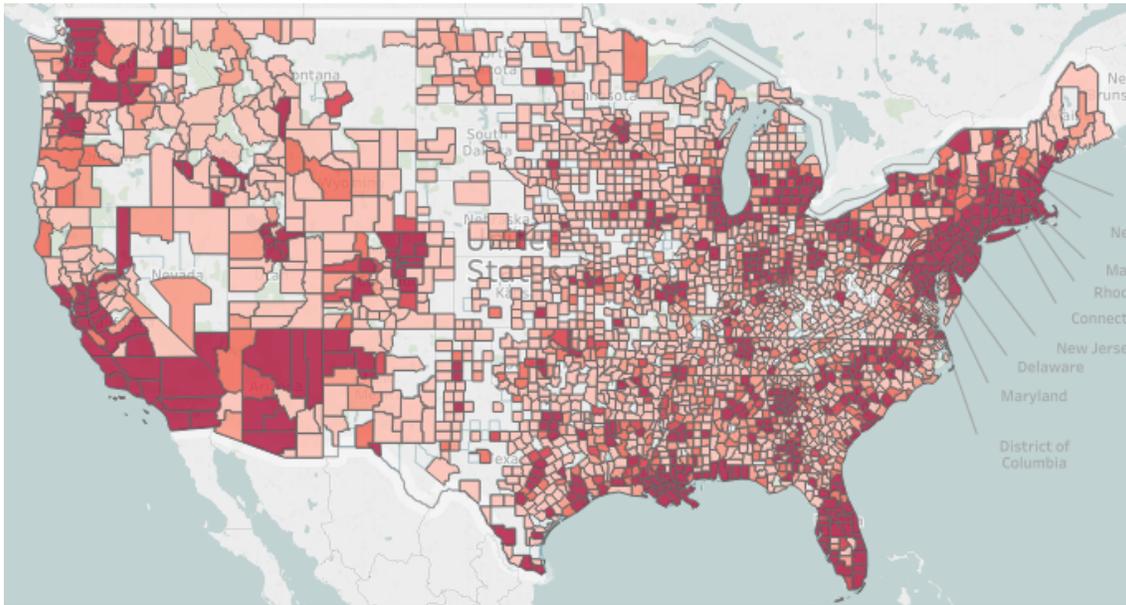
In 2005 the White House Homeland Security Council developed a National Strategy for Pandemic Influenza plan. This plan has been updated several times since then to adapt to the changing environment, and the latest update occurred in 2017. The national strategy paid particular attention to the need for adequate diagnostic testing, advanced vaccine and drug development infrastructures, modernized supply chains for drugs and supplies, and such. In spite of decades of planning, significant deficiencies in all of these areas led to the rapid spread of the coronavirus throughout the US. Moreover, the national strategy paid insufficient attention to the differential impact such a pandemic is likely to have within a country as large and diverse as the US. As the coronavirus spread, new areas of the country emerged as hot spots and public officials were unable to give accurate advance notice to regions across the country to prepare for major outbreaks. The objective of this paper is to explore if it is possible to use local socioeconomic conditions to develop vulnerability indexes for regions across the country to identify places that are most likely to experience adverse effects of an outbreak.

During periods of public health crises, researchers focus on both individual behaviors and societal factors in understanding how a particular region is likely to fare in a crisis. Although no one questions the importance of individual actions and behaviors in protecting against threats to one's health, the combined effects of the socioeconomic factors that shape the environment in which one lives should not be ignored either. The paper makes no attempt to delve deep into the mechanisms by which socioeconomic factors influence individual health outcomes. Rather, it identifies the socioeconomic factors that are found to be most significant in determining the threats to each community. There has been an interest in understanding if and how social and environmental characteristics of the places in which people reside exert influences on the health outcomes of people beyond their individual behaviors. Diez-Roux (1998) offers an argument for paying attention to the macro and group-level factors on individual health outcomes. Recognizing that this is unfortunately unlikely to be the last infectious disease with the potential to spread rapidly in a developed country such as the US, both individual health factors along with macro and socioeconomic factors need to be considered.

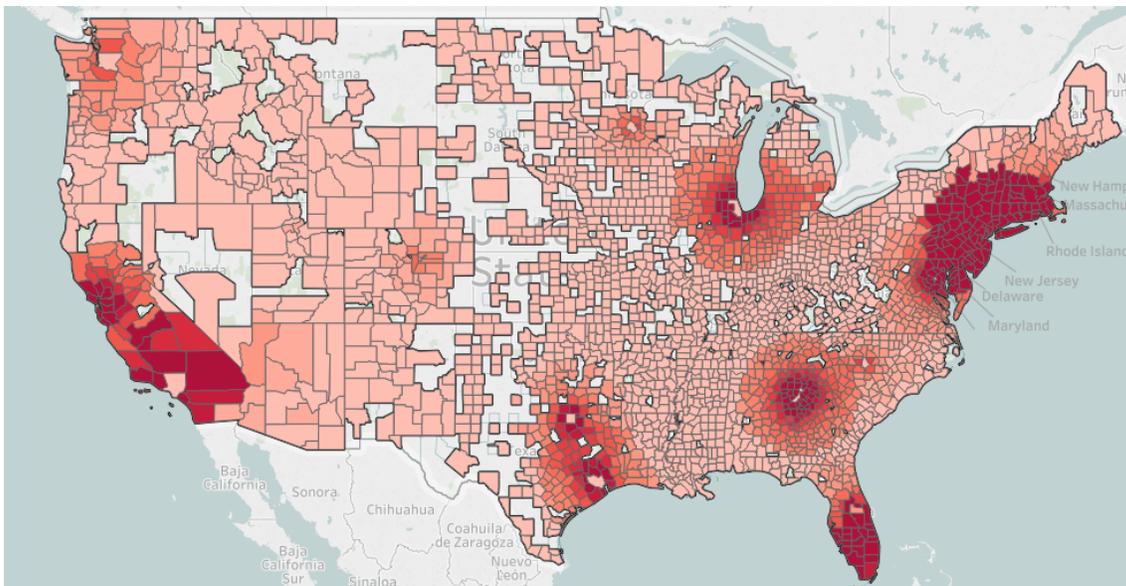
The literature on the transmission of infectious diseases is more heavily focused on the context of developing countries that have traditionally suffered the most from such diseases. These studies find that the highest impact areas have low income, poor sanitary conditions, and poor health care conditions (Campos et al. (2018) for Zika, Redding et al. (2019) for Ebola are recent examples). Moore et al. (2016) used Ebola to develop an Infectious Disease Vulnerability Index for countries in Africa. The literature on the socioeconomic determinants of the spread of infectious diseases in developed countries is not extensive partly because these countries have largely been unaffected by large outbreaks - Adda (2016) is an exception. Using data from France, it offers an extensive analysis of the transmission of three viruses - influenza, gastroenteritis, and chickenpox. The paper asks the important questions of whether the virus spreads more rapidly during periods of economic growth and if its spread follows a "gradient determined by economic factors." Using data from France, Adda (2016) finds that the viruses studied propagated faster during times of economic boom due to

increased economic activity and contact between people. Qiu et al. (2020) has conducted a similar analysis for Wuhan, China. Both papers find a positive relationship between the spread of the virus and economic activity. In the case of the coronavirus, Figure 1 clearly indicates that during the first growth phase, the virus spread to the regions of high economic activity along the two coasts of the US.

**Figure 1: Number of Cases**



**Figure 2: Weighted International Passengers**



In uncovering the underlying socioeconomic conditions that made different parts of the country more susceptible to becoming hot spots, we identify the factors that determine the

gradient followed by the virus as it spread through the country in the initial phase of the pandemic. It is worth noting that the virus had an initial phase where it rapidly rose from mid-March to mid-April 2020, stabilized and slightly declined on a per-day case count basis till mid-June and then rapidly increased again. It was in the summer of 2020 that it spread throughout the country and eventually touched every part of the country. The focus of this paper is on the initial outbreak phase since it allows us to determine the factors that make a region vulnerable to an infectious disease that unexpectedly arrives from an international source. It is to be noted that in addition to having a particular socioeconomic composition that makes a region particularly vulnerable to the spread of the disease, it is possible for a major super-spreader event, such as an international biotechnology conference in Boston, to cause a region to suddenly becoming a hot spot. Moreover, the socioeconomic results derived from this analysis are relevant for the regions that experienced the initial brunt of the disease. These regions have the least amount of time to prepare for such a situation. It is to be noted that when the virus spread in the summer of 2020, it impacted more vulnerable populations in the South and in rural areas. However, by that time, these populations were well aware of the rapidly growing public health crisis.

Introducing heterogeneity that captures region-specific uniqueness in an epidemiological model of disease spread, the paper develops a Vulnerability Index for the counties included in the study. The index captures the underlying factors that impact the vulnerability of a region to the virus. Socioeconomic factors are used to explain the observed differences in the vulnerability index. This allows us to answer the following questions in the context of this pandemic:

- Do county-level differences in economic conditions, as measured by household income and its distribution, matter in the transmission of an infectious disease such as COVID-19, and do they impact the severity of the health outcomes of the residents of the counties?
- Do demographic characteristics of a region (county) matter for both disease transmission and mortality?
- Do county-level health conditions influence the transmission of the coronavirus and the mortality of the populations to COVID-19? Do access to health professionals and the prevalence of preventive health measures make a difference? Did environmental factors matter?

The impact of income and its inequality in shaping the health outcomes of the people of a region is a widely debated topic. A 1996 issue of the *British Medical Journal* focused on the observed relationship between income inequality and mortality and considered it to be a “big idea.” In the editor’s words, “The big idea is that what matters in determining mortality and health in a society is less the overall wealth of that society and more how evenly wealth is distributed. The more equally wealth is distributed, the better the health of that society.” This topic has been studied by public health experts and economists alike - Deaton and Lubotsky (2003) and Deaton and Lubotsky (2009) find that impact of income inequality on health is diminished once the racial distribution of the population is accounted for while Ash and Robinson (2009) challenge their finding. Researchers also consider if some of the adverse

socioeconomic factors have more significant long-term impacts than instantaneous impacts. Ronzio (2003) and Zheng (2012) are some studies that demonstrate that adverse effects of inequality may take years to show up in health outcomes. Subramanian and Kawachi (2004) provides a comprehensive study of the issue of income inequality and health and addresses the need to consider both individual and macro effects in understanding whether and how inequality poses a threat to public health.

Since the pandemic began, there has been an explosion of interest amongst researchers of all disciplines and around the world. As anecdotal evidence of the differential impact of the disease on different demographic groups began to emerge, interest in studying the impact of various types of factors on the spread and severity of outcome arose. Baum and Henry (2020) uses a cross-sectional spatial autoregressive model to study the impact of gender, black and Hispanic ethnicities, median income, and a health index score on the number of cases. Using data for May 23, 2020, they find median income has a negative effect, and the percent of black and Hispanic populations has a positive effect on the number of cases. Galea and Abdalla (2020) drives attention to the importance of racial and socioeconomic inequalities in understanding their differential impact on different parts of the society and the unrest that results from it; Hawkins et al. (2020) use a cross-sectional analysis to study the impact of the distressed communities index and its components in the number of cases and fatalities at the county level. Using data from May 2, 2020, they find median income and the percent of the black population both have a positive impact on cases and fatalities, and that socioeconomic factors in general have a stronger impact than age and health-related factors. Kjøllesdal et al. (2022) analyzes the impact of such factors in the context of the immigrant population of Norway and found that crowded housing conditions and low income contribute to more cases and hospitalizations, Aguilar-Palacio et al. (2021) conducted a detailed study of a cohort of individuals in the Aragon region of Spain during three phases of the pandemic and revealed the challenges in reported cases brought about by limited testing data during the early phase of the pandemic. Their analysis also revealed the negative impact of income inequalities. Lassale et al. (2020) find similar challenges faced by ethnic minorities in England. In a cross-country analysis of 50 countries using data as of May 1, 2020, Ang et al. (2021) find a positive association of cases and fatalities with GDP and health conditions such as obesity. They find that lower-income dispersion is associated with lower critical cases and mortality. Using data from 80 countries Chaudhry et al. (2020) find that countries with more personal freedom experienced more cases, and so did countries that have a higher urban population, lower science test scores, and lower access to healthcare.

These studies have primarily used cross-sectional data within a country or across countries to analyze the impact of socioeconomic factors on cases, mortality, and/or hospitalizations. The uniqueness of this paper lies in deriving a vulnerability index for each region once the fundamental mechanisms of disease spread espoused by epidemiological models through the interaction of susceptible and infected individuals are taken into account in a dynamic panel model and then determining the impact of socioeconomic factors on that vulnerability index. This is also a unique way of both understanding the impact of the socioeconomic factors on the spread of the virus and of determining a vulnerability index. Other indexes use socioeconomic factors to build them - Marvel et al. (2021) for COVID-19, Cutter et al. (2003) in the context of environmental hazards, for instance. Spielman et al. (2020) raise

some concerns about these indexes and question if they measure what they are supposed to in a given context. The index derived in this paper stands apart from the traditional approaches by identifying the portion of the cases that cannot be explained by the past number of cases and the basic interactions between infected and susceptible individuals and is thereby context-specific.

Recent experiences with infectious diseases suggest that they are a national security threat (Cecchine and Moore, 2006) and pose a major threat to both the health and economic wellbeing of people around the world. In spite of the experience with H1N1, SARS, and Ebola, countries such as the US did not develop a coherent infrastructure or strategy to determine which parts of the country are at a particularly higher risk of disease transmission. This paper shows that it is possible to utilize the economic, demographic, and lifestyle profiles of regions to develop a risk factor for each geographical area so that when the next epidemic arises, public officials are better prepared to anticipate where the hot spots are likely to arise and take the necessary containment steps. Without advance preparation, the next disease will be just as difficult to contain as this. The large differences within state boundaries show the importance of developing more local strategies that take into consideration a multitude of factors.

## 2. METHODOLOGY AND DATA

The coronavirus pandemic impacted all 50 states in the US. The experiences of the states, counties, and cities have been anything but homogeneous. To understand this differential effect across counties in the US, we consider two sets of factors - epidemiological factors that explain the spread of infectious diseases and socioeconomic factors that enhance or mitigate the spread of the disease. Epidemiological models explain how an infectious disease evolves in a region based on population and the size of the pool of infected individuals. We use epidemiological models such as the SIR model to determine the fundamental differences in cases based on population size and the number of infections. These factors alone cannot explain the entire heterogeneous outcomes across the country. We expect differences in types and amounts of economic activities, living conditions, demographic makeups, and lifestyle choices to determine the vulnerabilities of communities in the spread of a highly contagious virus such as the coronavirus.

We conduct this analysis in two steps. In the first step, an epidemiological model of disease spread is used to generate estimates of a vulnerability index for each county once population and infections are accounted for. In the second step, county-level economic, demographic, and health data are used to explain differences in the vulnerability indexes across counties.

Epidemiological models of the SIR type describe disease spread dynamics based on three main factors - the size of the population, the number of susceptible individuals, and the number of infected individuals. With a population of size  $N$ , if  $I$  denotes the number of infected individuals,  $R$  denotes the number of recovered individuals, the number of individuals susceptible to the disease is given by  $S = N - I - R$ . At each time  $t$ , the number of new infections depends on the interactions of the susceptible ( $S$ ) and infected ( $I$ ) individuals. The infected individuals are non-infectious during the latent period and asymptomatic

but infectious from the end of the latent period to the end of the incubation period and infectious with symptoms after the end of the incubation period. If  $j$  denotes the number of days it takes to become infectious, at time  $t$ , the interactions of susceptible people with people infected  $t - j$  days earlier lead to new cases. There are many variations of this basic model in the context of the COVID-19 pandemic - Biswas et al. (2020), Chen et al. (2020), Simha et al. (2020), Pandey et al. (2021), and Liu (2021) are a few examples. Other models account for exposed individuals (SEIR - Yang et al. (2020)), exposed and asymptomatic individuals (SEIRA - Buchwald et al. (2020)), and models that account for multiple groups (Contreras et al., 2020). Yang et al. (2021) offers an evaluation of various mathematical and statistical models in the context of China. A fundamental goal of most of these models is to understand the future trajectory of an epidemic, and many of these models are useful for determining the infectiousness of a disease and the phase of the epidemic. They are also useful for determining the impact of intervention strategies - Giordano et al. (2020), Rainisch et al. (2020), Chowdhury et al. (2020), Ferguson et al. (2020), Eubank et al. (2020), Tuite et al. (2020), Gatto et al. (2020) are some examples.

The objective of this paper is to utilize an epidemiological model to determine what made some regions more vulnerable to the virus than others at the initial outbreak stage of the pandemic. Following Contreras et al. (2020) we enhance the basic SIR model and introduce the impact of daily commuters into a region to capture the transmission of the disease from neighboring regions through the daily mobility of labor. Using daily reports of coronavirus cases for counties across the US, we generate a panel dataset of US counties over a 20 day period from March 30 to April 19. This period captures the initial phase of the disease when the daily number of cases became a significant concern for hundreds of counties within the US. The panel data approach in estimating the growth of the virus in different parts of the US allows us to introduce county-specific fixed effects in the estimation. The panel estimates the number of cases as a function of the potential pool of susceptible and infected individuals and time and county-specific fixed effects and is given by the following equation:

$$C_{it} = \beta_0 + \beta_1 C_{it-1} + \beta_2 \frac{S_{it} I_{i,t-j}}{N} + \theta T_{s_i,t} + \gamma_i + \delta_t D_t + u_{it} \quad (1)$$

where,  $C_{it}$  denotes the number of reported cases in county  $i$  at time  $t$ ,  $S$ ,  $I$ , and  $N$  are defined above. To account for the transmission of the disease resulting from the daily inflow of people from neighboring counties, the  $I$  includes the number of infected individuals in county  $i$  and the potential number of infected individuals commuting to region  $i$  from the surrounding counties. The number of commuting individuals is multiplied by the fraction of the population that is infected in each of the regions from which the individuals are commuting. Moreover, the number of reported cases,  $C_i$ , is dependent on access to testing capacity. At the early stage of the pandemic, testing was not uniformly available throughout the country. To control for differences in testing capacity, the variable  $T_{s_i,t}$  is introduced. It is the number of tests administered at time  $t$  in the state  $s$  to which the county  $i$  belongs<sup>1</sup>  $\theta$

<sup>1</sup>Ideally, one should include testing data at the county level to account for variations in testing capacity across counties. However, comprehensive county-level testing data are not available for the early phase of the pandemic. State-level data are used as a proxy. This may cause some bias in the results. It is possible, however, that the vulnerability index values derived from these results will not be greatly impacted since results without any control for testing whatsoever are not significantly different.

gives the coefficient for the number of tests administered in the region  $s$  to which the county  $i$  belongs,  $\gamma_i$  represents the county fixed effects,  $\delta_t$  are the parameters for the time variables  $D_t$ , and  $u_{it}$  is the error for county  $i$  at time  $t$ .

Estimation of the above regression<sup>2</sup> generates the county fixed-effect value,  $\gamma$ , for each county. From these fixed effects, a vulnerability index for each county is derived. This approach is similar to the one used by Mukherji and Silberman (2013) in studying patent citations between metro areas in the US.

In the second step of the analysis, county-level economic and demographic factors are used to explain how they influence the vulnerability index for each county. The economic factors include income, unemployment rate, income inequality, and access to housing. The set of demographic factors includes the size of the population and its density, the racial profile of the counties, the age distribution of the population, and the percentage of the population that was born outside the US. In addition to the county-level economic and demographic data, spatial factors are considered as well. The contagious nature of the disease compels one to consider the spillover effects on neighboring counties. We introduce inverse-distance weighted values of the number of international passengers served by the top international airports in the US. Since the virus originated in China and then spread to other parts of the world, including Europe, before taking hold in the US, international passenger data is introduced to examine if proximity to international airports is related to the concentration of confirmed cases. While international passengers often arrive at a particular airport and then use domestic airlines to travel to other parts of the country, the locations of the international airports are closely tied to areas with concentrations of activities that are globally oriented. Consequently, the international passengers served by these airports are expected to interact in the regions around these airports in large numbers. Using a 300-mile radius around each county where the airports are located, an inverse-distance matrix is used to assign the number of international passengers in the areas surrounding the airports. Figure 2 displays the weighted distribution of international passengers. While this data is unrelated to the number of confirmed cases, the spatial distribution of the passenger data is similar to the spatial distribution of confirmed cases.

The estimation of the impact of these regional factors in explaining differences in vulnerabilities to the disease will be based on Equation (2).

$$V_i = \alpha + vWP + \sum_k \lambda_k e_{ki} + \sum_m \phi_m d_{mi} + \varepsilon_i \quad (2)$$

In the above equation,  $V_i$  represents the vulnerability index of county  $i$ ,  $e_{ki}$  represents the set of  $k$  economic variables that makes a county susceptible to the spread of the disease due to the enhanced interactions between people and working in close proximity. Although the economic activity of a county changes with time, the general distribution of such activities across the country remains relatively stable within short periods of time;  $d_m$  represents the demographic factors. This equation includes the spatially weighted number of international passengers in a region by multiplying an inverse distance-weighted matrix  $W$  with the number of international passengers,  $P$ , served by an international airport in the neighborhood of county  $i$ . The data and estimation results are reported in the next section.

<sup>2</sup>The statistical package Stata 14 was used for all statistical and regressions analysis used in this paper.

### 3. ESTIMATION

#### 3.1. Data

The data on COVID-19 cases and deaths is obtained from the COVID tracking data provided by the New York Times and Johns Hopkins University. Figure 1 displays the distribution of cases in the 2512 counties.

Data for various demographic and economic variables such as population distribution by ethnicity and population density are obtained from data compiled by the USDA's Atlas of Rural and Small Town America and the Federal Communication Commission. The underlying data come from the Census Bureau and the American Medical Association. Some of the demographic data, such as the distribution of the population by race, are from the 2010 census. The total population, per capita personal income, and unemployment data are from 2018. The percentage of the population with various health-related factors such as obesity and diabetes are available from the 2014-18 period. The Wisconsin County Health Rankings data are used for data on flu vaccination and air pollution. County-level daily testing data are not available for all counties during the early period of the pandemic. The US Department of Health & Human Services Health Data website was used to obtain state-level testing data. Data on commuters were obtained from the Census Bureau's Residence County to Commuting County Commuting Flows table for 2015. The inflow of commuters was restricted to the counties within 100 miles. The number of commuters from each of these counties was multiplied by the fraction of the population that was infected and then aggregated to determine the expected number of infected commuters in a county. Data on international air passengers were obtained from the Bureau of Transportation Statistics. This source provides the number of international passengers served by the top 50 international airports in the US. The total number of passengers on international flights was over 109 million in 2018. In order to account for local spillover effects of the virus in the form of increased susceptibility due to a higher prevalence of cases, an inverse distance weighted matrix was created with positive weights assigned up to a 300-mile radius around a county. This radius is just large enough to ensure that each county in the study had at least one other county in the study as a neighbor.

#### 3.2. Estimation of Cases

The previous section explained that the foundation of the analysis of the socioeconomic factors that can contribute to the spread and concentration of the coronavirus in the various parts of the country lies in the epidemiological model of disease transmission. The first step is to generate county-level vulnerability measures from an estimation of equation (1). The daily coronavirus data is available for over 2500 counties. We restrict our analysis to counties that reported an average of 30 cases per day from March 30 through April 19. This generates a panel of 771 counties covering all 50 states. Each of the counties reported at least one confirmed case during the period of analysis resulting in a balanced panel. Equation (1) includes a lagged value of infections in determining the proportion of the population that is susceptible at any time  $t$ . The incubation period for this virus is estimated to be anywhere between 2 to 14 days. People are infected a few days before they develop symptoms and

Table 1: Dynamic Panel Regression of Disease Spread

Variable	Coefficient	St.Err.	t-value	p-value	[95% Conf	Interval]
$\ln Cases_{t-1}$	0.876***	0.055	16.05	0.000	0.769	0.983
$\ln S(I_{it-7} + \sum_j I_{jt-7})/N$	0.054**	0.024	2.26	0.024	0.007	0.101
$\ln TotalTests$	0.02	0.019	1.05	0.294	-0.018	0.059
D9	-0.024***	0.008	-3.06	0.002	-0.039	-0.009
D10	-0.024**	0.010	-2.45	0.015	-0.044	-0.005
D11	-0.03**	0.012	-2.39	0.017	-0.054	-0.005
D12	-0.04***	0.015	-3.01	0.003	-0.072	-0.015
D13	-0.052***	0.017	-3.10	0.002	-0.085	-0.019
D14	-0.047***	0.018	-2.67	0.008	-0.082	-0.013
D15	-0.052**	0.020	-2.59	0.010	-0.092	-0.013
D16	-0.058***	0.022	-2.65	0.008	-0.101	-0.015
D17	-0.049**	0.023	-2.14	0.033	-0.095	-0.004
D18	-0.05**	0.025	-1.96	0.050	-0.098	0.000
D19	-0.055**	0.027	-2.05	0.041	-0.108	-0.002
D20	-0.06**	0.028	-2.17	0.030	-0.117	-0.006
Arellano-Bond test for AR(1) in first differences: z = -7.83						
Arellano-Bond test for AR(2) in first differences: z = 1.49						
Sargan test of overid. restrictions: chi2(25) = 17.04						
Hausen test of overid. restrictions: chi2(25) = 26.47						
Group variable: Fips						
Time variable: Day						
Number of instruments = 41						
F(16, 769) = 6359.17						
Prob > F = 0.000						
Number of Obs = 9228						
Number of Groups = 769						
Obs per Group: Min = 12						
Avg = 12.00						
Max = 12						

Note: t statistics are in parentheses. The 1%, 5%, and 10% levels of significance are given as \*\*\*, \*\*, and \*, respectively.

after they develop symptoms. We assume a seven-day lag for the results reported in the paper. Sensitivity analysis was conducted for different lag lengths. The analysis showed that most of the key relationships are qualitatively unchanged. The variable representing the share of the susceptible population that is exposed to the infected individuals is significant for a lag of seven days and is positive but not significant for the other days. The vulnerability index values derived from the regressions with other lags have correlations above 0.99.

Since cases in period  $t$  depend on the number of cases in period  $t - 1$ , the estimation of equation (1) requires the use of dynamic panel estimation methods. A model with small  $T$  (20) and large  $N$  (771) with a lagged dependent variable is expected to have the Nickell's bias (Nickell, 1981). A difference GMM estimation is found to be the best option for the data. The Arellano-Bond estimation method of Arellano and Bond (1991) that uses lagged values as instruments as implemented by Roodman (2009) was used. The results are reported in Table 1 and show that although autocorrelation of the first order exists, there is no second-order autocorrelation. The Sargan and Hansen tests of no overidentification of instruments are satisfied, and the F statistic shows that the model fits the data well. The table shows that the one period lagged number of cases has a significant impact on the number of cases reported on any day. The interaction between the infected and susceptible population is also significant and positive.

One of the key objectives of this regression is to obtain a set of estimates for the county-level fixed effects. The method of dynamic panel estimation, such as GMM that utilizes first differencing, removes the impact of time-invariant variables such as the time-invariant fixed effects. These are, however, recoverable from the residuals. It is to be noted that for a dynamic panel model of the form,  $y_{it} = \rho y_{it-1} + a_i + e_{it}$ , the residual is given by  $\hat{e}_{it} = a_i + e_{it} + (\hat{\rho} - \rho)y_{it}$ . The average  $\bar{e}_i$  can be used as an estimator of the fixed effects to analyze how the underlying conditions in the various counties impact the fixed effects as long as those factors are uncorrelated with the  $e_{it}$ . That condition is satisfied with the average  $e_{it}$  equalling  $-7.00\text{e-}09$  for the results of the regression of Equation (1).

### 3.3. Estimation of the Vulnerability Index

The estimates of the fixed effects derived from the dynamic panel regression of cases are converted to an index by dividing each fixed value by the mean of these values and multiplying the ratio by 100. This value is interpreted as the Vulnerability Index. A high value of the index indicates that a county is more susceptible to the spread of the disease. The value of the index ranges from a low of 64 for Lincoln, Arkansas to a high of 223 for Manhattan, New York. Table 2 offers a list of the 20 lowest and highest values of the index. The results show that the higher values were in the so-called "hot spots". The table lists the region codes and Urban Influence Codes (UIC) used by the USDA to distinguish between rural and urban areas. Codes 1 and 2 are for metro areas, 11 and 12 are for non-core areas that are not adjacent to any metro area. The table shows the concentration of the high index areas in the Northeast and in large metro areas. The bottom values are found in counties mainly outside the Northeast. There is a large difference in the population densities of the places with high values of the index than the ones with the smallest values. The table shows that there are differences in both location and type of county that distinguish areas with high

**Table 2: Vulnerability Index of Top and Bottom 20 Counties**

County	State	Fips	Vulnerability Index	Population Density	Region Code	UIC
<b>Top 20</b>						
New York City (Manhattan)	New York	36061	223	69468	1	1
Nassau	New York	36059	195	4705	1	1
Suffolk	New York	36103	193	1637	1	1
Westchester	New York	36119	192	2205	1	1
Cook	Illinois	17031	185	5495	2	1
Wayne	Michigan	26163	180	2974	2	1
Los Angeles	California	6037	177	2420	4	1
Bergen	New Jersey	34003	176	3884	1	1
Rockland	New York	36087	175	1796	1	1
Miami-Dade	Florida	12086	172	1315	3	1
Essex	New Jersey	34013	170	6212	1	1
Hudson	New Jersey	34017	170	13731	1	1
Philadelphia	Pennsylvania	42101	169	11379	1	1
Union	New Jersey	34039	167	5216	1	1
Orleans	Louisiana	22071	167	2029	3	1
Fairfield	Connecticut	9001	167	1467	1	2
Orange	New York	36071	166	459	1	1
Middlesex	Massachusetts	25017	166	1838	1	1
Passaic	New Jersey	34031	166	2715	1	1
Suffolk	Massachusetts	25025	166	12416	1	1
<b>Bottom 20</b>						
Richland	Ohio	39139	72	251	2	2
Allen	Ohio	39003	72	264	2	2
St. Francis	Arkansas	5123	72	45	3	3
Delaware	Oklahoma	40041	72	56	3	6
Cass	Missouri	29037	72	143	2	1
Grant	Indiana	18053	72	169	2	3
Washington	Utah	49053	71	57	4	2
Lincoln	South Dakota	46083	71	77	2	2
Madera	California	6039	71	71	4	2
Perry	Missouri	29157	71	40	2	6
Decatur	Georgia	13087	71	47	3	5
Rutland	Vermont	50021	71	66	1	5
Dallas	Iowa	19049	71	112	2	2
Muhlenberg	Kentucky	21177	70	67	3	6
Hawaii	Hawaii	15001	70	46	4	8
Routt	Colorado	8107	70	10	4	8
Martin	Minnesota	27091	69	29	2	11
Dubuque	Iowa	19061	68	154	2	2
Marshall	Iowa	19127	66	71	2	5
Lincoln	Arkansas	5079	64	25	3	2

*Note:* Region Codes R1 - R4 indicate the US regions Northeast, Midwest, South, and West, respectively. The analysis includes 132 counties from R1, 165 from R2, 360 from R3, and 112 from R4.

values of infections from places with smaller outbreaks. We attempt to introduce additional factors that can shed light on why some places experienced significantly higher infection rates than others after controlling for the pool of susceptible individuals. Descriptive statistics of the variables considered are given in Table 3.

Estimation of Equation (2) sheds light on the factors that contribute most strongly to the differences in the values of the vulnerability index<sup>3</sup> noted in Table 2. The estimation results are reported in Table 4. The results are reported for models that include population and population density separately and together. These two variables play an important role in disease transmission and have a correlation of 0.76. It is important to understand their individual and collective impact on the vulnerability index values of the counties. The other critical variables are classified into three broad groups - economic, demographic, and health/lifestyle. Since many of the conditions in a county are influenced by policies and conditions at the state level, we include a series of state dummy variables to control for state-level influences.

**Table 3: Descriptive Statistics**

Variable	Mean	Std. Dev	Min	Max
Vulnerability Index	100.00	24.55	64.18	222.84
Deaths	22.86	223.35	0.00	9708
Cases Prior 14 Days Per 100,000 People	71.88	173.45	1.00	4147.00
Unemployment Rate	4.02	1.21	1.70	18.10
International Air Passengers	144.36	1035.06	0.00	16459.90
Weighted International Air Passengers	61.10	122.86	0.00	1842.56
ln Median Income	11.00	0.27	10.18	11.85
Gini	0.45	0.04	0.36	0.62
Severe Housing Problems	16.82	4.53	6.80	35.70
ln Population	12.04	1.14	8.76	16.13
ln Population Density	5.57	1.31	1.47	11.15
% Population Over Age 65	14.60	3.97	7.00	51.60
% White Non Hispanic 2010	70.57	18.96	3.33	97.89
% Black	13.52	15.35	0.11	82.95
% Asian	2.65	3.81	0.01	43.01
% Native American	0.90	4.54	0.04	73.30
% Hispanic	10.30	12.73	0.42	95.74
% Multiple Race	2.06	2.06	0.16	35.01
% Foreign Born	8.04	7.18	0.00	52.94
ln Primary Care Physicians Per Capita	0.0007	0.0003	0.00	0.004
Preventable Hospital Stays Per 1000	59.71	18.18	0.00	142.43
% Adult Obesity	29.13	5.04	12.00	45.00
% Diabetes	10.27	2.34	3.90	18.60
Air pollution	9.79	1.93	0.00	19.70
% Flu Vaccination	47.25	6.41	15.00	65.00

<sup>3</sup>Different specifications used in estimating the index led to similar rankings of the index even if the actual values of the coefficients varied. Since it is the relative rankings that matter for the analysis of the underlying socioeconomic factors, the results of the following analysis are largely insensitive to the exact specifications used to derive the index values.

The results show that in the economic group, median household income has a positive and significant effect showing that places of high income are more vulnerable. This is a reflection of the high human interactions that result from high economic activities and is consistent with the findings of Adda (2016). The Gini coefficient measuring the degree of income inequality and severe housing problems is both positive and significant, suggesting that places where large portions of the population experience poor economic, living, and housing conditions transmitted the virus more rapidly. Taken together, these results show that counties with high levels of economic activity but unequal distribution of income and living conditions are more vulnerable to the transmission of the disease.

Moreover, areas of high economic activities are also where the main international airports are located. The number of international passengers served by the airports is positive and significant in most models. This corroborates Figure 2 that shows that the locations of the international airports and the number of international passengers they serve overlap with the regions of high infection rates. The international passengers not only impact the counties in which the airports are located, their effects spill over to the neighboring regions as well. The results show that the distance-weighted number of international passengers served by these airports is positive and significant<sup>4</sup>.

The results related to the demographic variables are consistent with the findings from the health professionals. People living in counties with higher populations and population densities are more vulnerable since they lead to more interactions and hence transmission. Counties with a higher concentration of the elderly are more vulnerable, and so are places with a higher concentration of people of color. Places with a higher concentration of immigrants did not show any significant difference in vulnerability once population was included.

In addition to the economic and demographic factors that can contribute to a region's vulnerability to the disease, we examine if some basic health-related factors play any role. We include three health-related variables - access to health care, individual choices to protect against bad health outcomes, and a measure of the population's success in leading a healthy lifestyle. These are included to examine if the health behavior of the populations in the counties has any direct impact on their vulnerability to an infectious disease such as the coronavirus. Toward this goal, the number of primary care physicians per capita is included as an indicator of access to health care; the percentage of the population that receives the flu vaccine as a measure of the population's voluntary health protection decisions; and the number of preventable hospital stays as an indicator of adverse health outcomes that result from poor health care decisions made by the people of a county. The most significant variable is found to be the preventable hospital stays variable showing that places where people's adverse health care actions are harmful enough to lead to preventable hospital visits, are also places where people are more vulnerable to catching the coronavirus. Vulnerability is not found to be strongly related to having access to health care or taking preventive steps such as getting vaccinated.

The results reported in Table 4 show that collectively the economic, demographic, and to a lesser extent, the health care makeup of the counties can explain a large portion of the

---

<sup>4</sup>The diagonal values of the weight matrix used for the calculation of the weighted international passengers are zeros. Consequently, the weighted values measure the impact in the surrounding areas only.

variance in county-level vulnerability index values. We consider next if these factors can also explain the differences in the number of deaths caused by COVID-19.

**Table 4: Regression Explaining Vulnerability Index**

Variable	Model 1	Model 2	Model 3
<b>Economic Variables</b>			
International Air Passengers	0.00000810*** (3.07)	0.00000593 (0.97)	0.00000833*** (2.97)
Weighted International Air Passengers	0.000195** (2.28)	0.0000362 (0.43)	0.000199** (2.28)
ln Median Income	0.233*** (6.66)	0.252*** (6.73)	0.233*** (6.59)
Gini	0.478** (2.45)	0.295 (1.29)	0.481** (2.45)
Severe Housing Problems	0.00331 (1.59)	0.00396* (1.70)	0.00331 (1.58)
<b>Demographic Variables</b>			
ln Population	0.125*** (20.09)		0.126*** (16.26)
ln Population Density		0.0811*** (12.03)	-0.00218 (-0.31)
% Population Over Age 65	0.00525*** (3.55)	0.00299* (1.95)	0.00525*** (3.54)
% White Non Hispanic 2010	-0.00364*** (-6.48)	-0.00402*** (-6.94)	-0.00366*** (-6.37)
% Foreign Born	-0.000164 (-0.13)	0.00287* (1.95)	-0.000183 (-0.15)
<b>Community Health/Lifestyle Variables</b>			
Preventable Hospital Stays Per 1000	0.000767** (2.08)	0.00119*** (2.88)	0.000757** (2.05)
ln Primary Care Physicians Per Capita	-0.00440 (-0.34)	0.0143 (0.98)	-0.00423 (-0.33)
% Flu Vaccination	-0.000723 (-0.64)	0.000853 (0.66)	-0.000704 (-0.62)
State Fixed Effects	Significant	Significant	Significant
Constant	0.335 (0.76)	1.479*** (3.12)	0.304 (0.70)
Observations	768	768	768
$R^2$	0.779	0.708	0.779
Adjusted $R^2$	0.759	0.682	0.759

*Note:*  $t$  statistics are in parentheses. The 1%, 5%, and 10% levels of significance are given as \*\*\*, \*\*, and \*, respectively.

### 3.4. Estimation of Deaths

Table 5 displays the results of an estimation of the number of deaths by county. To avoid issues related to the dependent variable being counts of deaths and account for differences in population sizes, the data on deaths and cases are converted to values per 100,000 people. Since the rest of the independent variables are all expressed as percentages or values that do not require population adjustments, the transformation of the deaths and cases to values per 100,000 eliminates the need to include a control for population size. The data for deaths includes values for each of the 20 days of the study. We use a pooled regression methodology with indicator variables for the days for which the data are analyzed and US states to which the counties belong to control for time and same-state impacts.

Unlike the estimation of cases reported in Table 1, a 14-day lag is used from infection to death. The results show that the number of deaths is positively related to the number of cases reported 14 days prior. The indicator variables for the days of the analysis,  $D15$  through  $D19$  show that relative to the 20th day ( $D20$ ), days 15 through 18 had significantly higher deaths indicating that the rise in deaths was slowing down over time. The pandemic peaked around April 23 before spreading through the rest of the country.

The economic, demographic, and health outcomes of people are correlated. Models 1 through 7 differ in terms of the combinations of the variables that are included. Models 1 and 2 focus on the economic variables and include only one race and one health-related variable. Models 3 and 4 focus on the racial composition of the population and report results with and without the income inequality variable. Models 5 through 7 reduce the race-related factors and focus on healthcare factors.

The results of the regression as they relate to the economic variables show that, unlike the spread of the virus, deaths and median household income are negatively related. The Gini coefficient of income inequality is positive and significant. Consequently, places with lower income and higher income inequality had higher fatality rates. This is consistent with our general understanding that people at lower income levels are more vulnerable to serious health shocks. The results are consistent with the literature cited in the Introduction that suggests that inequality is a threat to public health. Once income is accounted for, the unemployment rate and other measures of economic condition are not found to exert any additional effects on deaths.

The demographic variables show that counties with a higher concentration of people above the age of 65 had a significantly higher number of deaths, consistent with what is known about the elevated risk of COVID-19 to the elderly. Models 1 and 2 that only include an indicator for the percentage of Whites in the population show that counties with a higher concentration of Whites had fewer deaths. Once the racial distribution of the non-White population is further disaggregated, the results show that counties with higher concentrations of Asians, Blacks, and Native Americans reported higher deaths. The relationship is negative for counties with more multi-racial populations and immigrants. Once the income inequality measure is included (Model 4), the impact of Blacks and Native Americans is diminished.

Table 5: Regression to Explain Distribution of Deaths

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Cases t-14 Days Per 100000	0.121*** (12.78)	0.119*** (12.69)	0.120*** (12.64)	0.119*** (12.70)	0.118*** (12.59)	0.118*** (12.59)	0.118*** (12.67)
<b>Economic Variables</b>							
ln Med Income	-7.614*** (-5.14)	-4.524*** (-2.63)	-9.285*** (-6.11)	-6.175*** (-3.75)	-1.365 (-0.64)	-0.685 (-0.31)	0.613 (0.28)
Unempl Rate	-0.169 (-1.00)	-0.101 (-0.63)	0.0376 (0.25)	0.0846 (0.56)	-0.399** (-2.51)		
Gini		46.38** (2.44)		43.61** (2.21)	58.29*** (2.85)	58.62*** (2.85)	66.15*** (3.12)
<b>Demographic Variables</b>							
ln Pop Density	1.635*** (3.18)	1.516*** (2.94)	1.505*** (2.77)	1.401*** (2.58)	1.431*** (2.98)	1.487*** (3.11)	1.549*** (3.27)
% Pop Over 65	0.446*** (5.53)	0.417*** (5.79)	0.394*** (5.43)	0.383*** (5.58)	0.370*** (5.87)	0.358*** (5.73)	0.462*** (6.46)
% White	-0.0794** (-2.38)	-0.0654** (-2.12)			-0.0110 (-0.37)	-0.00178 (-0.06)	0.0109 (0.33)
% Asian			0.653*** (5.84)		0.704*** (5.76)		
% Black			0.0796** (2.00)		0.0590 (1.58)		
% Hispanic			0.0167 (0.66)		0.0584* (1.75)		
% Native American			0.0550** (1.97)		0.0455 (1.63)		
% Multiple Race			-0.813*** (-3.72)		-0.692*** (-3.17)		
% Foreign Born	-0.250*** (-6.04)	-0.295*** (-6.13)	-0.317*** (-3.52)	-0.440*** (-4.05)	-0.244*** (-4.90)	-0.242*** (-4.86)	-0.209*** (-4.16)

*continued*

Table 5: Regression to Explain Distribution of Deaths

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
<b>Community Health/Lifestyle Variables</b>							
Air pollution					0.849*** (3.83)	0.795*** (3.60)	0.727*** (3.30)
% Flu Vac					-0.293*** (-4.41)	-0.292*** (-4.40)	-0.302*** (-4.44)
% Diabetes					0.481** (2.49)	0.468** (2.45)	
% Adult Obesity							0.321*** (3.76)
Primary Care Physicians Per Cap	-3.375*** (-5.84)	-4.811*** (-5.03)	-3.880*** (-6.31)	-5.111*** (-5.24)	-3.892*** (-4.65)	-3.775*** (-4.61)	-3.630*** (-4.41)
D15	3.627*** (4.62)	3.549*** (4.55)	3.622*** (4.61)	3.559*** (4.57)	3.491*** (4.50)	3.491*** (4.50)	3.514*** (4.54)
D16	2.409*** (3.49)	2.354*** (3.42)	2.405*** (3.48)	2.362*** (3.44)	2.314*** (3.38)	2.313*** (3.38)	2.330*** (3.41)
D17	1.891*** (2.74)	1.850*** (2.70)	1.888*** (2.75)	1.856*** (2.72)	1.820*** (2.67)	1.820*** (2.67)	1.832*** (2.70)
D18	1.243* (1.84)	1.216* (1.81)	1.241* (1.85)	1.220* (1.83)	1.197* (1.80)	1.197* (1.80)	1.205* (1.81)
D19	0.374 (0.57)	0.363 (0.55)	0.373 (0.57)	0.365 (0.56)	0.355 (0.54)	0.355 (0.54)	0.358 (0.55)
Constant	45.67*** (3.30)	-17.70 (-0.66)	57.93*** (3.71)	-5.973 (-0.20)	-49.25 (-1.53)	-58.21* (-1.73)	-79.90** (-2.33)
State Fixed Effects	Sig.						
Observations	4608	4608	4608	4608	4608	4608	4608
R <sup>2</sup>	0.756	0.757	0.759	0.760	0.761	0.761	0.762
Adjusted R <sup>2</sup>	0.752	0.754	0.755	0.756	0.758	0.758	0.758

Note: *t* statistics are in parentheses. The 1%, 5%, and 10% levels of significance are given as \*\*\*, \*\*, and \*, respectively.

In all of the models reported in Table 5, access to health care in terms of the number of primary care physicians per capita is included. This variable is significant and negative in every model showing that access to physicians to take care of people's everyday health needs is a protection against the most severe outcome of the disease. Consistent with this finding are the results that show that places with higher concentrations of populations with high-risk health outcomes such as obesity and diabetes have higher reported deaths. Using county-level data, papers such as Wu (2020) demonstrate that air pollution is a risk factor for COVID-19 death. The results reported here support their finding. The health results also show that counties with higher rates of flu vaccination have a lower risk of death from the virus. It is to be noted that ours is an observational study, so this result by no means implies that the flu vaccine is a protection against the virus. Rather, the result is to be interpreted in terms of health awareness and action. Counties with higher percentages of the population that get vaccinated against the flu are perhaps places where people, in general, are more responsive to threats to their health and take steps to mitigate them. It may reflect their overall approach to their health and wellbeing.

The results show that the economic factors are important for explaining the differential impacts experienced by counties across the country both in terms of confirmed cases and deaths reported. The demographic and health-related factors are more pronounced in the estimation of deaths than reported cases. This is not surprising since the virus does not discriminate based on any factor other than immunity, but the severity of the disease that can lead to a fatal outcome depends on underlying health and demographic factors.

#### 4. CONCLUSION

This paper has examined the differential experience of infections and deaths across the US due to the COVID-19 pandemic during its initial outbreak phase. The analysis of the number of cases is based on an epidemiological model in which we included a county fixed effect. This is a novel way to introduce heterogeneity in such a model. A dynamic panel regression of the number of cases included the potential number of interactions between susceptible and infected individuals as a proportion of the population along with county fixed effects. The results of the model were used to construct a Vulnerability Index for each county. Economic, demographic, and health/lifestyle factors were used to explain the differences in the Vulnerability Index across the counties. The results showed that counties with higher economic activity have higher vulnerability. The results show that regions around international airports experienced higher numbers of cases than ones that are over 300 miles away. This is consistent with the fact that the virus has arrived on the US shores through travelers coming to the US from abroad. Counties with more elderly and non-White populations are more vulnerable, and so are counties with higher income inequality and housing problems.

The results related to deaths show that counties with lower income and higher cases experienced higher deaths. It is to be noted that counties with higher income reported more cases, but when it came to fatality, lack of income is a risk factor. Counties with higher population density and higher income inequality also experienced more deaths. Counties with higher percentages of non-Hispanic Blacks, Native Americans, and Asians are more likely to die relative to counties with non-Hispanic Whites. Counties with more personal care

physicians per capita experienced lower deaths, and so did counties with a lower percentage of the population with health-risk factors such as obesity and diabetes. Air pollution is also found to be associated with higher deaths. While studies show that long-term exposure to air pollution can cause long-term vulnerability to lung-related diseases, it is to be noted in an observational study such as this one should not draw conclusions about health-risk factors from observed results. This is true about the result related to flu vaccination as well. The results suggest that counties with more health-conscious populations that take preventive actions have a better outcome in terms of surviving the COVID-19 disease.

The economics literature is not extensive in the area of pandemics and epidemics in developed countries. This study contributes to the growing literature on how the various socioeconomic conditions made different regions more susceptible to the COVID-19 pandemic and how those lessons can help regions better prepare for such events in the future. It is to be noted that this paper's focus has been on the initial outbreak phase of the disease. The disease spread throughout the country in the summer of 2020 and reached other periods of rapid growth. It remains for future studies to compare the differences in the factors that impacted the vulnerabilities of the US regions in the different phases of the pandemic. A national strategy to prepare the infrastructure for controlling the spread of infectious diseases should consider these factors and develop Vulnerability Indexes for regions across the country.

## REFERENCES

- Adda, Jérôme. (2016) “Economic Activity And The Spread Of Viral Disease: Evidence from High Frequency Data,” *The Quarterly Journal of Economics*, 131(2), 891–941. <http://doi.org/10.1093/qje/qjw005>.
- Aguilar-Palacio, Isabel, Lina Maldonado, Sara Malo, Raquel Sánchez-Recio, Iván Marcos-Campos, Rosa Magallón-Botaya, and M Rabanaque. (2021) “COVID-19 Inequalities: Individual and Area Socioeconomic Factors (Aragón, Spain),” *International Journal of Environmental Research and Public Health*, 18(12). <http://doi.org/10.3390/ijerph18126607>.
- Ang, Joshua Ping, Fang Dong, and Jason Patalinghug. (2021) “COVID-19: Effectiveness of Socioeconomic Factors in Containing the Spread and Mortality,” *International Review of Applied Economics*, 35(2), 164–187. <http://doi.org/10.1080/02692171.2020.1853078>.
- Arellano, Manuel and Stephen Bond. (1991) “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *The Review of Economic Studies*, 58(2), 277–297. <http://doi.org/10.2307/2297968>.
- Ash, Michael and Dean E Robinson. (2009) “Inequality, Race, and Mortality in US Cities: A Political and Econometric Review of Deaton and Lubotsky (56: 6, 1139–1153, 2003),” *Social Science & Medicine*, 68(11), 1909–1913. <http://doi.org/10.1016/j.socscimed.2009.02.038>.
- Baum, Christopher and Miguel Henry. (2020) “Socioeconomic Factors influencing the Spatial Spread of COVID-19 in the United States,” Boston College Department of Economics: Boston, US.
- Biswas, Kathakali, Abdul Khaleque, and Parongama Sen. (2020) “Covid-19 Spread: Reproduction of Data and Prediction using a SIR model on Euclidean Network,” *arXiv preprint arXiv:2003.07063*. <http://doi.org/10.48550/arXiv.2003.07063>.
- Buchwald, Andrea G, Jimi Adams, David M Bortz, and Elizabeth J Carlton. (2020) “Infectious Disease Transmission Models to Predict, Evaluate, and Improve Understanding of COVID-19 Trajectory and Interventions,” *Annals of the American Thoracic Society*, 17(10), 1204–1206. <http://doi.org/10.1513/AnnalsATS.202005-501PS>.
- Campos, Monica C, Jamille G Dombrowski, Jody Phelan, Claudio RF Marinho, Martin Hibberd, Taane G Clark, and Susana Campino. (2018) “Zika Might Not Be Acting Alone: Using an Ecological Study Approach to Investigate Potential Co-Acting Risk Factors for an Unusual Pattern of Microcephaly in Brazil,” *PLoS One*, 13(8), e0201452. <http://doi.org/10.1371/journal.pone.0201452>.
- Cecchine, Gary and Melinda Moore. (2006) *Infectious Disease and National Security: Strategic Information Needs*. Rand Corporation, Santa Monica, CA.
- Chaudhry, Rabail, George Dranitsaris, Talha Mubashir, Justyna Bartoszko, and Sheila Riazi. (2020) “A Country Level Analysis Measuring the Impact of Government Actions, Country Preparedness and Socioeconomic Factors on COVID-19 Mortality and Related Health Outcomes,” *EClinicalMedicine*, 25, 100464. <http://doi.org/10.1016/j.eclinm.2020.100464>.
- Chen, Yi-Cheng, Ping-En Lu, Cheng-Shang Chang, and Tzu-Hsuan Liu. (2020) “A Time-Dependent SIR Model for COVID-19 With Undetectable Infected Persons,” *IEEE Transactions on Network Science and Engineering*, 7(4), 3279–3294. <http://doi.org/10.1109/TNSE.2020.3024723>.
- Chowdhury, Rajiv, Kevin Heng, Md Shajedur Rahman Shawon, Gabriel Goh, Daisy Okono-

- fua, Carolina Ochoa-Rosales, Valentina Gonzalez-Jaramillo, Abbas Bhuiya, Daniel Reidpath, Shamini Prathapan, et al.. (2020) “Dynamic Interventions to Control COVID-19 Pandemic: A Multivariate Prediction Modelling Study Comparing 16 Worldwide Countries,” *European Journal of Epidemiology*, 35(5), 389–399. <http://doi.org/10.1007/s10654-020-00649-w>.
- Contreras, Sebastián, H Andrés Villavicencio, David Medina-Ortiz, Juan Pablo Biron-Lattes, and Álvaro Olivera-Nappa. (2020) “A Multi-Group SEIRA Model for the Spread of COVID-19 Among Heterogeneous Populations,” *Chaos, Solitons & Fractals*, 136, 109925. <http://doi.org/10.1016/j.chaos.2020.109925>.
- Cutter, Susan L., Bryan J. Boruff, and W. Lynn Shirley. (2003) “Social Vulnerability to Environmental Hazards,” *Social Science Quarterly*, 84(2), 242–261. <http://doi.org/10.1111/1540-6237.8402002>.
- Deaton, Angus and Darren Lubotsky. (2003) “Mortality, Inequality and Race in American Cities and States,” *Social Science & Medicine*, 56(6), 1139–1153. [http://doi.org/10.1016/S0277-9536\(02\)00115-6](http://doi.org/10.1016/S0277-9536(02)00115-6).
- Deaton, Angus and Darren Lubotsky. (2009) “Income Inequality and Mortality in US cities: Weighing the Evidence. A Response to Ash,” *Social Science & Medicine*, 68(11), 1914–1917. <http://doi.org/10.1016/j.socscimed.2009.02.039>.
- Diez-Roux, Ana V. (1998) “Bringing Context Back into Epidemiology: Variables and Fallacies in Multilevel Analysis,” *American Journal of Public Health*, 88(2), 216–222. <http://doi.org/10.2105/AJPH.88.2.216>.
- Eubank, Stephan, I Eckstrand, B Lewis, Srinivasan Venkatramanan, Madhav Marathe, and CL Barrett. (2020) “Commentary on Ferguson, et al., “Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand,” *Bulletin of Mathematical Biology*, 82(4), 1–7. <http://doi.org/10.1007/s11538-020-00726-x>.
- Ferguson, Neil, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, et al.. (2020) “Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand,” Imperial College of London: London, UK.
- Galea, Sandro and Salma M Abdalla. (2020) “COVID-19 Pandemic, Unemployment, and Civil Unrest: Underlying Deep Racial and Socioeconomic Divides,” *JAMA*, 324(3), 227–228. <http://doi.org/10.1001/jama.2020.11132>.
- Gatto, Marino, Enrico Bertuzzo, Lorenzo Mari, Stefano Miccoli, Luca Carraro, Renato Casagrandi, and Andrea Rinaldo. (2020) “Spread and Dynamics of the COVID-19 Epidemic in Italy: Effects of Emergency Containment Measures,” *Proceedings of the National Academy of Sciences*, 117(19), 10484–10491. <http://doi.org/10.1073/pnas.2004978117>.
- Giordano, Giulia, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, and Marta Colaneri. (2020) “Modelling the COVID-19 Epidemic and Implementation of Population-Wide Interventions in Italy,” *Nature Medicine*, 26(6), 855–860. <http://doi.org/10.1038/s41591-020-0883-7>.
- Hawkins, Robert B, Eric J Charles, and J Hunter Mehaffey. (2020) “Socio-Economic Status and COVID-19-Related Cases and Fatalities,” *Public Health*, 189, 129–134. <http://doi.org/10.1016/j.puhe.2020.09.016>.
- Kjøllestad, Marte, Katrine Skyrud, Abdi Gele, Trude Arnesen, Hilde Kløvstad, Esperanza

- Diaz, and Thor Indseth. (2022) “The Correlation Between Socioeconomic Factors and COVID-19 among Immigrants in Norway: A Register-Based Study,” *Scandinavian Journal of Public Health*, 50, 52–60. <http://doi.org/10.1177/14034948211015860>.
- Lassale, Camille, Bamba Gaye, Mark Hamer, Catharine R Gale, and G David Batty. (2020) “Ethnic Disparities in Hospitalisation for COVID-19 in England: The role of Socioeconomic factors, Mental Health, and Inflammatory and Pro-Inflammatory Factors in a Community-Based Cohort Study,” *Brain, Behavior, and Immunity*, 88, 44–49. <http://doi.org/10.1016/j.bbi.2020.05.074>.
- Liu, Xiaoping. (2021) “A simple, SIR-like but individual-based epidemic model: Application in comparison of COVID-19 in New York City and Wuhan,” *Results in Physics*, 20, 103712. <http://doi.org/10.1016/j.rinp.2020.103712>.
- Marvel, Skylar W, John S House, Matthew Wheeler, Kuncheng Song, Yi-Hui Zhou, Fred A Wright, Weihsueh A Chiu, Ivan Rusyn, Alison Motsinger-Reif, and David M Reif. (2021) “The COVID-19 Pandemic Vulnerability Index (PVI) Dashboard: Monitoring County-Level Vulnerability Using Visualization, Statistical Modeling, and Machine Learning,” *Environmental Health Perspectives*, 129(1), 017701. <http://doi.org/10.1289/EHP8690>.
- Moore, Melinda, Bill Gelfeld, and Christopher Paul Adeyemi Okunogbe. (2016) *Identifying Future Disease Hot Spots: Infectious Disease Vulnerability Index*. RAND Corporation, Santa Monica, CA. <http://doi.org/10.7249/RR1605>.
- Mukherji, Nivedita and Jonathan Silberman. (2013) “Absorptive Capacity, Knowledge Flows, and Innovation in US Metropolitan Areas,” *Journal of Regional Science*, 53(3), 392–417. <http://doi.org/10.1111/jors.12022>.
- Nickell, Stephen. (1981) “Biases in Dynamic Models with Fixed Effects,” *Econometrica*, 49(6), 1417–1426. <http://doi.org/10.2307/1911408>.
- Pandey, Prashant, Yu-Ming Chu, JF Gómez-Aguilar, Hadi Jahanshahi, and Ayman A Aly. (2021) “A Novel Fractional Mathematical Model of COVID-19 Epidemic Considering Quarantine and Latent Time,” *Results in Physics*, 26. <http://doi.org/10.1016/j.rinp.2021.104286>.
- Qiu, Yun, Xi Chen, and Wei Shi. (2020) “Impacts of Social and Economic Factors on the Transmission of Coronavirus Disease 2019 (COVID-19) in China,” *Journal of Population Economics*, 33(4), 1127–1172. <http://doi.org/10.1007/s00148-020-00778-2>.
- Rainisch, Gabriel, Eduardo A Undurraga, and Gerardo Chowell. (2020) “A Dynamic Modeling Tool for Estimating Healthcare Demand from the COVID19 Epidemic and Evaluating Population-Wide Interventions,” *International Journal of Infectious Diseases*, 96, 376–383. <http://doi.org/10.1016/j.ijid.2020.05.043>.
- Redding, David W, Peter M Atkinson, Andrew A Cunningham, Gianni Lo Iacono, Lina M Moses, James LN Wood, and Kate E Jones. (2019) “Impacts of Environmental and Socio-Economic Factors on Emergence and Epidemic Potential of Ebola in Africa,” *Nature Communications*, 10(1), 1–11. <http://doi.org/10.1038/s41467-019-12499-6>.
- Ronzio, Cynthia R. (2003) “Urban Premature Mortality in the US between 1980 and 1990: Changing Roles of Income Inequality and Social Spending,” *Journal of Public Health Policy*, 24(3), 386–400. <http://doi.org/10.2307/3343384>.
- Roodman, David. (2009) “How to do xtabond2: An Introduction to Difference and System GMM in Stata,” *The Stata Journal*, 9(1), 86–136. <http://doi.org/10.1177/1536867X0900900106>.

- Simha, Ashutosh, R Venkatesha Prasad, and Sujay Narayana. (2020) “A Simple Stochastic Sir Model for Covid 19 Infection Dynamics for Karnataka: Learning from Europe,” *arXiv preprint arXiv:2003.11920*. <http://doi.org/10.48550/arXiv.2003.11920>.
- Spielman, Seth E, Joseph Tuccillo, David C Folch, Amy Schweikert, Rebecca Davies, Nathan Wood, and Eric Tate. (2020) “Evaluating Social Vulnerability Indicators: Criteria and their Application to the Social Vulnerability Index,” *Natural Hazards*, 100(1), 417–436. <http://doi.org/10.1007/s11069-019-03820-z>.
- Subramanian, Subu V and Ichiro Kawachi. (2004) “Income Inequality and Health: What Have We Learned So Far?,” *Epidemiologic Reviews*, 26(1), 78–91. <http://doi.org/10.1093/epirev/mxh003>.
- Tuite, Ashleigh R, David N Fisman, and Amy L Greer. (2020) “Mathematical Modelling of COVID-19 Transmission and Mitigation Strategies in the Population of Ontario, Canada,” *Canadian Medical Association Journal*, 192(19), E497–E505. <http://doi.org/10.1503/cmaj.200476>.
- Yang, Wuyue, Dongyan Zhang, Liangrong Peng, Changjing Zhuge, and Liu Hong. (2021) “Rational Evaluation of Various Epidemic Models Based on the COVID-19 Data of China,” *Epidemics*, 37. <http://doi.org/10.1016/j.epidem.2021.100501>.
- Yang, Zifeng, Zhiqi Zeng, Ke Wang, Sook-San Wong, Wenhua Liang, Mark Zanin, Peng Liu, Xudong Cao, Zhongqiang Gao, Zhitong Mai, et al.. (2020) “Modified SEIR and AI Prediction of the Epidemics Trend of COVID-19 in China Under Public Health Interventions,” *Journal of Thoracic Disease*, 12(3), 165. <http://doi.org/10.21037/jtd.2020.02.64>.
- Zheng, Hui. (2012) “Do People Die from Income Inequality of a Decade Ago?,” *Social Science & Medicine*, 75(1), 36–45. <http://doi.org/10.1016/j.socscimed.2012.02.042>.