Administração Pública e Gestão Social
ISSN: 2175-5787
apgs@ufv.br
Universidade Federal de Viçosa
Brasil

The Relationship between Pandemic Circumstances and Socioeconomic Factors and Implications for the Government

Wang, Wenru; A. Vasarhelyi, Miklos; Rosa, Fabricia; Machado de Freitas, Marcelo
The Relationship between Pandemic Circumstances and Socioeconomic Factors and Implications for the Government
Administração Pública e Gestão Social, vol. 13, núm. 4, 2021
Universidade Federal de Viçosa, Brasil

Artigos

The Relationship between Pandemic Circumstances and Socioeconomic Factors and Implications for the Government

La interacción entre circunstancias pandémicas con intervenciones gubernamentales y factores sociales

Implicações governamentais da interação entre o contexto pandêmico e fatores socioeconômicos

Wenru Wang
Rutgers, The State University of New Jersey: Newark, NJ, US, Estados Unidos
ww321@rutgers.edu

Miklos A. Vasarhelyi
Rutgers, The State University of New Jersey, Newark, NJ, Estados Unidos
miklosv@business.rutgers.edu

Fabricia Rosa
Universidade Federal de Santa Catarina: Florianopolis, SC, Brasil
fabriciasrosa@hotmail.com

Marcelo Machado de Freitas
Universidade Federal de Santa Catarina: Florianopolis, SC, Brasil
mmf.marcelofreitas@gmail.com

Redalyc: https://www.redalyc.org/articulo.oa?id=351568433008

Recepción: 04 Octubre 2020
Aprobación: 22 Febrero 2021
Publicación: 01 Octubre 2021

Abstract:

Research Objective: This research explores the role that socioeconomic factors play on the pandemic development and it provides government with public control implications.
Theoretical background: Government of different regions responds to the pandemic with various government policies and interventions. Meanwhile, socioeconomic factors that are long-term results of government policies also impact the development of the pandemic in different regions. We study how these socioeconomic factors are related with the COVID-19 pandemic development.
Research Design: Using data from IBGE - Brazilian Institute of Geography and Statistics, Ministry of Health, and Santa Catarina State Department of Health, we conduct a set of regression analyses at the municipality-level using socioeconomic factors as independent variables, and log-transformed confirmed COVID-19 cases and confirmed COVID-19 deaths as dependent variables.
Results: We identify several significant indicators of pandemic outcomes and we find that municipalities in the Brazilian state with more basic health units, higher bedroom density, higher autonomy level of municipal revenue tend to have more confirmed cases and deaths due to COVID-19. We also find other socioeconomic factors including sanitation condition and age group diversity as important indicators of COVID-19 confirmed cases and deaths.
Originality: This research is one of the first to understand the impact of socioeconomic factors on the development of the COVID-19 pandemic and it utilizes unique datasets that have not been used by other studies to our best knowledge.
Practical and theorical contributions: The deliverables of this research will improve the understanding of the ongoing pandemic, predict the pandemic development trend, and prepare the policymakers with improved information provisioning by pointing out the significant impacts of socioeconomic factors.

Resumen:

Objetivo de la investigación: Esta investigación explora el papel que juegan los factores sociales y las políticas gubernamentales en el desarrollo de la pandemia.
Antecedentes teóricos: el gobierno de diferentes regiones responde a la pandemia con diversas políticas e intervenciones gubernamentales. Mientras tanto, los factores sociales que son resultados a largo plazo de las políticas gubernamentales también impactan el desarrollo de la pandemia en diferentes regiones. Estudiamos como estos factores sociales se relacionan con el desarrollo de la pandemia de COVID-19.
Diseño de la investigación: Utilizando datos del Instituto Brasileño de Geografía y Estadística (IBGE), el Ministerio de Salud y el Departamento de Salud del Estado de Santa Catarina, realizamos un análisis de regresión a nivel municipal utilizando factores sociales como variables independientes, y casos confirmados de COVID-19 y muertes confirmadas por COVID-19 como variables dependientes.
Resultados: Encontramos que los municipios del estado brasileño con más Unidades Básicas de Salud y mayor densidad poblacional tienden a tener más casos confirmados y muertes por COVID-19. También encontramos que el producto interno bruto (PIB) de un municipio es un indicador importante de casos confirmados y muertes por COVID-19.
Originalidad: esta investigación es una de las primeras en comprender el impacto de los factores sociales en el desarrollo de la pandemia COVID-19.
Contribuciones prácticas y teóricas: Los resultados de esta investigación mejorarán la comprensión de la pandemia en curso, predecirán la tendencia de desarrollo de la pandemia y prepararán a los responsables políticos con un mejor suministro de información al señalar los impactos significativos de los factores sociales y las políticas gubernamentales.

Resumo:

Objetivo da pesquisa: esta pesquisa explora o papel que fatores socioeconômicos desempenham no desenvolvimento da pandemia de COVID-19 e oportuniza aos governos essas implicações.
Enquadramento Teórico: os governos de diferentes regiões respondem à pandemia com diversas políticas e intervenções governamentais. Enquanto isso, fatores socioeconômicos que são resultado de longo prazo das políticas governamentais também impactam o desenvolvimento da pandemia em diferentes regiões. A presente pesquisa estuda como esses fatores socioeconômicos estão relacionados ao desenvolvimento da pandemia de COVID-19.
Desenho da Pesquisa: Usando dados do Instituto Brasileiro de Geografia e Estatística (IBGE), do Ministério da Saúde e da Secretaria de Estado da Saúde de Santa Catarina, conduziu-se uma análise de regressão em nível municipal usando fatores socioeconômicos como variáveis independentes e casos confirmados de COVID-19 e mortes confirmadas de COVID-19 como variáveis dependentes.
Resultados: foi identificada uma série de indicadores significantes para o desenvolvimento da pandemia. Foi verificado que os municípios do estado brasileiro com mais Unidades Básicas de Saúde (UBS), mais habitantes por residência e maior autonomia de receita municipal tendem a apresentar mais casos confirmados e óbitos por COVID-19. Também foram evidenciados outros fatores socioeconômicos como importantes indicadores de casos e óbitos confirmados por COVID-19.
Originalidade: esta pesquisa é uma das primeiras a compreender o impacto dos fatores socioeconômicos no desenvolvimento da pandemia COVID-19 e utiliza uma base de dados únicas que ainda não foi utilizada em outros estudos, do que é do conhecimento dos autores.
Contribuições práticas e teóricas: os resultados desta pesquisa irão melhorar a compreensão da pandemia em andamento, prever a tendência de desenvolvimento da pandemia e preparar os formuladores de políticas com melhores informações, apontando os impactos significativos de fatores socioeconômicos.

1. Introduction

On March 11th, 2020, the World Health Organization (WHO) officially declared the Novel Coronavirus Disease, COVID-19, a worldwide pandemic (WHO, 2020). The total number of global confirmed COVID-19 cases has doubled since July 2020, and the number reaches 30 million by the mid of September (Winsor & Shapiro, 2020). Since the first identified outbreak in Wuhan, China, governments of different countries and at different levels respond individually by announcing various policies and interventions to understand the pandemic development and control the transmission. The responding government policies include individual considerations of each country and profoundly impact society socially and economically. It is the responsibility of governments at all levels to protect people's health and life by announcing and implementing government policies in areas such as public health, political and institutional arrangements, and resource allocations effectively and timely (Duan, Hu, Vasarhelyi, Da Rosa, & Lyrio, 2020). Scientists in different fields are devoting to find cures and track sources of the novel disease. Researchers in social science explored the impacts of government policies, such as a full lockdown that are announced and implemented to respond to the unexpected pandemic. For instance, Chaudhry, Dranitsaris, Mubashir, Bartoszko, and Riazi (2020) conducted a country-level study and found that low preparedness of national health policy and actions towards the pandemic, such as a longer time to border closures from the first reported case, leads to higher levels of a national confirmed case. They also found that rapid border closures, full lockdowns, and wide-spread testing were not associated with COVID-19 mortality per million people, but full lockdowns and reduced country vulnerability to biological threats were significantly related to higher patient recovery rates. The conclusion of Chaudhry *et al.* (2020) could help policymakers evaluate the immediate impacts of the COVID-19 related policies and adjust the current policies according to the achieved impacts. Researchers have also investigated the relationship between the current pandemic outcomes and socioeconomic factors, which can be seemed as long-term results of government policies that are not particularly responding to a pandemic. Socioeconomic factors include income, education, employment, and community safety that could influence significantly how well and for how long people live in a community (The County Health Rankings & Roadmaps program, 2020). In the same study, Chaudhry *et al.* (2020) also identified that obesity is associated positively with COVID-19 caseloads. With higher obesity prevalence and higher per capita gross domestic product (GDP), they observed increased mortality per million people. For Horton (2020), addressing COVID-19 means addressing hypertension, obesity, diabetes, chronic cardiovascular and respiratory diseases, and cancer. Paying more attention to Noncommunicable diseases (NTDs) is not an agenda only for richer nations. As NTDs are also a cause of health problems in the poorest countries. Other researchers also tried to identify socioeconomic factors that significantly result in higher COVID-19 morbidity and mortality rates, but the results are mixed. For example, Rocklöv and Sjödin (2020) concluded that high population densities facilitate the spread of COVID-19 through higher contact rates. This study proposed that controlling contact rates should be the primary strategy to control the epidemic outbreak, and such a strategy depends on population densities (Rocklöv & Sjödin, 2020). However, in a study conducted by Hamidi, Ewing, and Sabouri (2020) with data collected from metropolitan counties in the United States, they found that county density was not significantly associated with the infection rate after controlling for metropolitan population. They believed that the metropolitan population, not population density, to be one of the most significant indicators of COVID-19 infection rates (Hamidi, Sabouri, & Ewing, 2020). They interpreted the insignificance of population density as people in the denser area show more adherence to social distancing guidelines.

Looking at the controversy existed in different studies, we believe that one of the interpretations could be that these studies collect data from different countries and some study at the country level and some study at a lower level. According to Horton (2020), unless governments develop policies and programs to reverse deep disparities, our societies will never be truly protected from COVID-19. The controversy also highlights that governments need to consider the deviations of the governing region before adjusting and implementing policies that may receive satisfying results in another region.

What is known at the moment is that the reflexes of COVID-19 are regional, complex and dynamic, because each country or region has its own characteristics, the pandemic involves epidemiological, economic and social issues concomitantly, and the result of the measures adopted to cope with the pandemic are volatile. There is a need for interaction between government and society constantly for the definition and execution of public policies capable of responding in a timely and dynamic way to the problems associated with the pandemic. Therefore, we propose our research question as follows: What socioeconomic factors affect the COVID-19 confirmed cases and mortality and what are the implications for government interventions?

This paper conducts a municipality-level analysis to examine the relationship between socioeconomic factors and pandemic outcomes in one Brazilian state, Santa Catarina. Brazil is one of the global epicenters of the coronavirus outbreak. It has had over four million cases, which is more than the total cases in

Europe, and over 140,000 deaths by September 30, 2020 (Andreoni, 2020). We want to perform the analysis on a municipality-level because a country as big as Brazil has twenty-six states and one federal district. Socioeconomic factors such as health system establishment, demographic information, and income dispersion vary greatly, and state governments have different reactions to the pandemic from the north to the south. We study municipalities in one state to minimize the impacts that other factors may have on the pandemic outcomes. We chose Santa Catarina as the main study object because it was the first state in Brazil to make more flexible and relaxed quarantine and social distancing measures, and the state reopened businesses and most economic activities on April 13, 2020 (Merco Press, 2020). There have been news reports right after the reopening arguing that reopening policy might make Santa Catarina a “Brazil’s test case” despite “the worst phase of the pandemic in in Latin America is likely to be just around the corner” (Soares, 2020). We want to use the analysis to observe the pandemic outcomes in the "Brazil’s test case" after months of business reopening and try to explore the important indicators of COVID morbidity and mortality.

This paper utilizes analytical methodologies to generate a regression model for improved information provisioning in public health decision-making. By targeting one state of the current epicenter of the COVID-19 outbreak, the proposed model aims to identify specific social factors as significant predictors of COVID-19 pandemic development and illustrate why the government should consider and anticipate the possible consequences of public policies. Our regression analyses identify several significant indicators toward the number of confirmed cases and deaths due to COVID-19. We find that more Basic Health Units (in Portuguese, *Unidades Básicas de Saúde*), more households with greater than two people per bedroom, high autonomy level of municipal revenue lead to higher number of confirmed COVID-19 cases and deaths. We find other significant socioeconomic factors that cause higher number of confirmed cases: higher percentage of people aged 15 to 64, lower infant mortality rates, smaller municipality size, more people with inadequate water supply and sanitation, fewer people in self-employed status, and higher Municipal Human Development Index. We also identify that population density has positive relationship with the reported death tolls due to COVID-19. We further apply the conclusion to predict three cities as new hotspots using the state-released COVID-19 data by July 12, 2020, and we validate the prediction by using data released later. We also include Minas Gerais as an additional state to see the generalization of our model and the generated conclusions.

We believe our research is a critical addition to the current social science studies in the COVID-19 pandemic. It also contributes to public administration studies because (1) it incorporates analytical technologies to investigate the relationship between socioeconomic factors and pandemic outcomes; (2) it provides a predictive methodology for the government to identify possible pandemic outbreak hot spots for more suitable decision-makings; (3) it points out the implications that government interventions could make use of.

This article is organized in such a way as to the theoretical and methodological basis of the research, and present results. Thus, after this introduction, we present the literature review, then present the research design, including all the variables that are been used in the research, following for presentation of the analysis results and related discussion, and finally, we present the conclusions and references used.

2. Literature Review

The medical field has been devoting to tracking the origin of the novel coronavirus and inventing vaccines to cure the COVID-19 disease that has made people globally suffer since the start of 2020. Researchers in public health and social science have also committed to studying the relationship between the pandemic outbreak, government measures, and social characteristics of the regions with outbreaks (Duan *et al.*, 2020; Chaudhry, Dranitsaris, Mubashir, Bartoszko & Riazi, 2020; Rocklöv & Sjödin, 2020; Hamidi, Ewing, *et al.*, 2020; Hamidi, Sabouri, *et al.*, 2020). Nevertheless, it seems that most studies about the subject are more concerned to understand the impact of population mobility or population reactions to the pandemic outcomes (Chande *et al.*, 2020). Other studies are also including socioeconomic factors as possible explanations for the pandemic outcomes. Chaudhry *et al.* (2020) conducted a country level exploratory analysis to measure the impacts of national health actions undertaken towards COVID-19 outcomes. They collected data from 50 countries with top accumulative confirmed cases. They also applied multivariable regression to identify factors (such as socioeconomic factors [inequality, income] and healthcare capacity [hospital beds, number of ICUs beds, number of physicians and number of nurses] associated with COVID-19 infection and mortality. Their study found that higher COVID-19 accumulative cases are positively related to factors including higher obesity, median population age, and longer time to border closures from the first reported case. Mortality per million was significantly impacted by higher obesity prevalence and per capita gross domestic product (GDP). The study concluded that low levels of national preparedness, the scale of testing, and population characteristics were associated with increased national confirmed cases and overall mortality. Population density is another socioeconomic factor that many studies are curious about. They examined population density's relationship with the pandemic and came up with inconsistent conclusions. For example, Rocklöv and Sjödin (2020) indicated that high population density is bad for the COVID-19 transmission because high population density means higher contact rates. Therefore, the authors believed that it is possible to control the pandemic development by controlling the contact rates in populated areas (Rocklöv & Sjödin, 2020).

According to Horton (2020), what we have learned so far is that the history of COVID-19 is not so simple. Two categories of diseases are interacting in specific populations - infection with acute severe respiratory coronavirus syndrome 2 (SARS-CoV-2) and a number of noncommunicable diseases (NTDs). These conditions group into social groups according to patterns of inequality deeply rooted in our societies. The aggregation of these diseases in a context of social and economic disparity exacerbates the adverse effects of each separate disease (Horton, 2020).

Meanwhile, Hamidi, Sabouri, & Ewing (2020) conducted a county-level study using data from the United States. They claimed that county density was not a significant predictor for infection rate after controlling for the metropolitan population. Instead, the metropolitan population should be one of the most significant indicators of COVID-19 infection rates (Hamidi, Sabouri, *et al.*, 2020). They interpreted the conclusion as people in the denser area are more compliant with social distancing guidelines. In another study by Hamidi, Sabouri, & Ewing (2020), they further investigated the role of population density and metropolitan size by conducting a longitudinal analysis. This study claimed that county density led to a lower infection rate and death rate of the COVID-19 pandemic, which contrasts with general assumptions, news media, and other studies (Hamidi, Sabouri, *et al.*, 2020). The study admitted that the relationship between density and the pandemic is complicated, and it agreed with the assumption that population density increases the person-to-person contacts. The study also pointed out that the unexpected result of density may be due to better healthcare infrastructures in the denser area. However, there is no evidence to indicate that the denser area in the researched U.S. counties has better healthcare infrastructures.

Nevertheless, we think the discussion of density is intriguing, and the evidence the socioeconomic factors' roles are still insufficient to provide support for social control decisions. It is worth exploring the relationship of socioeconomic factors, including population density and the pandemic using data from a South American country, where also suffers from the pandemic. In other words, governments need to develop policies and programs to reverse deep social disparities, using an integrated approach to understand and treat diseases collectively (Horton, 2020). Although we do not know if the assumption of Hamidi, Sabouri, & Ewing (2020) that denser areas have better healthcare infrastructures is valid in Brazil, we include data of Basic Health Units (BHU) in our model, which provides elementary healthcare to its citizens, to investigate its relationship with the pandemic outcomes.

Administrative state capacity (such as BHUs availability) are also important features that should be included in a model that explores contributing factors of the pandemic. The pandemic nature demands governments to act quickly, especially in test capacity, medical resources supplies, and regions most affected. A literature review by Aguiar & Lima (2019) presented different definitions and dimensions about state capacity. The literature suggested that state capacity could be measured in different ways. Reviewing different authors, it is inferable that adequate state capacity could imply better infrastructure (among others), helping governments achieve their goals. For instance, Cárdenas (2010) indicated that state capacity could be defined as governments' capability to generate tax revenues. In the present paper, we are concerned with these state capacities (combined with other variables) to ensure better outcomes related to the COVID-19 pandemic.

During the 2009 H1N1 pandemic, several socioeconomic factors were recognized as important indicators of pandemic development. For instance, in the USA, the absence of workplace policies, such as paid sick leave, and other factors related to the population working in more impoverished conditions, were determinants for the increased risk of getting sick in the 2009 pandemic (Kumar, Quinn, Kim, Daniel, & Freimuth, 2012). We add more evidence of how different socioeconomic factors are getting uneven shares of the COVID-19 burden in this study. While higher-income jobs can often be done remotely and in confinement, lower-income jobs often cannot. For instance, in France, a survey reported that 39% of low-income workers still went to their workplace during lockdowns, whereas only 17% of high-income workers stayed in the old routines (Pullano, Valdano, Scarpa, Rubrichi, & Colizza, 2020). In the United States, African Americans and Hispanics were struck hard by the COVID-19 pandemic. This result could be happening because elevated rates of chronic conditions and the lack of health insurance of this particular population (The Economist, 2020). Although it is expected that lower-income municipalities face a greater number of cases and deaths (Lamb, Kanudla & Shaman, 2020), it is also expected that access to care and adequate state capacity could help decrease the problems faced by this inequality. Lamb, Kanudla & Shaman (2020), in fact, found that COVID-19 cases were related to median household income, the proportion living in households with four or more individuals, and the proportion of adults uninsured.

3. Research Design

This study aims to explore the relationship between socioeconomic factors and COVID-19 pandemic outcomes. We construct two regression models to explore important indicators of pandemic outcomes in Santa Catarina, Brazil. The result of the regression analysis identifies specific socioeconomic factors as indicators of COVID-19 spreading and mortality. Since the socioeconomic factors are either long-term effects of government policies or are immediate results of government responding policies against the pandemic, we believe the study results could provide supports for government decisions in social controls, and the analysis methodology could be adopted to predict potential impacts of specific government policies.

3.1. Data Collection

We first collect socioeconomic factors of 290 municipalities in Santa Catarina, Brazil, with related pandemic outcome data released by the State Department of Health as of July 12, 2020 (Santa Catarina State Department of Health, 2020). The dependent variables in the regression model are the confirmed cases of COVID-19 and confirmed deaths due to COVID-19, both at the municipality level. The independent variables include municipal-level statistics and indices such as area size, total population (2010), population density (inhabitants/km^2), GDP per capita (2017), and Gini index and they are collected from the Brazilian Institute of Geography and Statistics (IGBE) and the Brazil Ministry of Health (IGBE, 2020; Brazil Ministry of Health, 2020). The independent variables are categorized into the following categories: Access to

Care, Housing, Employment, Physical Environment, Income, Length of Life, Family, Education, and State Capacity.

A similar categorization of socioeconomic factors is adopted in the County Health Ranking Model, a program collaborated by the Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute. They use different socioeconomic factors into their model to come up with health ranking scores, and the reasoning is that these factors can influence people's ability to "make healthy choices, afford medical care and housing, manage stress, and more" (The County Health Rankings & Roadmaps program, 2020). We agree with the argument that socioeconomic factors can impact health, so we choose the different socioeconomic factors to explore their influences on the COVID-19 pandemic. Of the 20 independent variables, some variables such as GDP, Gini index, population density, age group diversity, and hospital beds capacity have been included in previous studies and explored their impacts on health outcomes (The County Health Rankings & Roadmaps program, 2020; Chaudhry *et al.*, 2020; Teixeira da Silva & Tsigaris, 2020). Municipal Human Development Index is another variable adopted by various studies to evaluate life quality using three dimensions, including health (Pereira, Mota, & Miranda, 2013).

On the other hand, some of the independent variables collected in our study are not commonly explored in previous literature discussing the relationship between socioeconomic factors and health outcomes. For instance, the Basic Health Unit (BHU) is a unique feature of Brazil's health system, and they are the primary approach to the entire healthcare network. BHU could help people with prevention and treatment actions and caring for minor emergencies, including fever, headache, and sore throat (Secretaria de Saúde do Distrito Federal, 2020). Therefore, our interest in including the number of Basic Health Unit is to see if a higher number of such primary healthcare agencies in a municipality reduce the risk of COVID-19 infections and related deaths.

3.2. Characteristics of the Municipalities

The illustration and basic statistics of the dependent and independent variables are shown in Table 1. Looking at Table 1, it is noticeable that the pandemic outcomes and the socioeconomic factors vary significantly across municipalities in Santa Catarina. For the dependent variables, the mean value for confirmed cases in Santa Catarina is around 145, but the highest value is over 3,000 with a standard deviation of over 400. We also observe a significant gap in confirmed deaths due to COVID-19 across the 290 municipalities. The mean value of deaths from COVID-19 is a little over 1, but the highest value for deaths is 56. The significant variances in both dependent variables indicate that the pandemic transmission in Santa Catarina is limited in several municipalities as of early July when some municipalities have not identified any confirmed case or death from COVID-19. As shown in Figure 1, the histograms of both dependent variables are highly right skewed. We therefore transform the two dependent variables to make them closer to normal distributions by computing the natural logarithm of the dependent variables' values plus one. We apply this computation instead of a simple computation of natural logarithm because we have many zeros in confirmed cases and deaths. We do not want to delete the records with zero value since the deletion will limit the size of the sample data. As the result shown in Figure 1(b), the histogram distribution of the log-transformed confirmed cases becomes more like a normal distribution. However, the histogram of the log-transformed confirmed deaths does not improve significantly.

The summary in Table 1 also shows that socioeconomic factors vary in the individual municipality for the independent variables. For instance, the mean value of density is 82.71 inhabitants per square kilometers, while the highest density for a municipality is 2,337.67 inhabitants/km^2, and the lowest is 31.96 inhabitants/km^2. This fact indicates that Santa Catarina has municipalities that are highly populated and other municipalities that are little resided. Another independent variable of our interests is the number of Basic Health Units (BHU), which provide primary care to municipalities. The summary shows that some cities do not even have BHUs when the summary shows the minimum number is 0, but some municipality has as many as 79 BHUs. As we mentioned previously, BHU is a primary construction of Brazil's health system. The fact that some municipalities do not have BHU is striking. The government of all levels should notice the problem since it presents a nonnegligible loophole that may have long existed. Another independent variable, Hospitals referred to as dealing with any severe cases of the new coronavirus, has even more limited availability in most municipalities. The mean value of the hospital variable is 0. The standard deviation is also smaller than 1, which means that many municipalities do not have public hospitals referred by the State Health Secretariats as COVID-19 treatment hospitals. We also notice that GDP per capita has significant variances with a standard deviation of 14,967.66 across the state of Santa Catarina. Realized revenue and committed expenditures are two more independent variables that have significant variance. Both Revenue and expenditure (in R$ ×1000 as of 2017) have a standard deviation over 200,000. We want to point out one more independent variable that is categorized into state capacity, IFGF_OwnRevenue, which is an index that indicates the autonomy level of municipal revenues. The closer the index is to 1, the less dependent the municipality is on other entities' aid, including the state and Federal Government. From Table 1, the mean value of this autonomy index is 0.299, and the standard deviation is 0.227, which means that most of the municipalities in Santa Catarina have some level of dependency on aid from higher-level authorities.

TABLE 1
Summary of dependent and independent Socioeconomic factors variables

	Variables		Explanation	Mean; Standard Deviation (Median, Min, Max)	Sources
Dependent variables	Confirmed cases of COVID-19		Santa Catarina released confirmed case numbers for 290 municipalities as of July 12, 2020	145.63; 425.777 (27, 0, 3,431)	(Santa Catarina State Department of Health, 2020)
	Confirmed deaths due to COVID-19		Santa Catarina released confirmed death numbers for 290 municipalities as of July 12, 2020	1.376; 5.549 (0, 0, 56)	
Independent Variables	Access to Care	BHU	Number of Basic Health Units that provide primary cares	6.307; 8.818 (3.5, 0, 79)	(Brazil Ministry of Health, 2020; IGBE, 2020; FIRJAN, 2020)
		Hospital	Number of hospitals for dealing with serious cases of COVID-19	0.17; 0.664 (0, 0, 8)	
		ICU beds	Number of ICU beds available in the municipality	4.745; 19.114 (0; 0; 181)	
	Housing	WaterSewage	Percentage of people in a household with inadequate water supply and sanitation, multiply by 100	1.433; 2.695 (0.555, 0, 28.44)	
		TDens	Percentage of households with greater than 2 people per bedroom, multiply by 100	12.165; 5.276 (11.035; 4.16; 43.65)	
	Employ-ment	CPR	Percentage of people aged 18 or over who are self-employed, multiply by 100	32.13; 1.251 (29.73, 10.64, 62.97)	
		TDes18M	Unemployment rate of the population aged 18 or over, multiply by 100	2.444; 1.800 (2.005; 0; 14.7)	
	Physical Environ-ment	Density	Demographic density, inhabitants/km²	82.71; 191.64 (31.96, 2.06, 2,337.67)	
		Area	The area size of the municipality in km²	325.13; 316.026 (233.02; 35.14; 2637.66)	
		Population	The population amount that live in the municipality as of 2010	24,497; 59,555.03 (7,858; 1,260; 590,466)	
	Income	GDP	GDP per capita in Brazilian Reals as of 2017	32,193; 14,967.66) (29,225; 12,692; 159,237)	
		Gini	Gini Index that measures the degree of inequality, multiply by 100	44.17; 5.344 (44; 28; 62)	
		TRmaxidoso	Percentage of households vulnerable to poverty and dependent on the elderly, multiply by 100	1.085; 0.927 (0.805, 0, 5.28)	
	Length of Life	IdhmL	Municipal Human Development Index - Longevity Dimension, multiply by 100	84.75; 2.92 (85; 77; 89)	
		InfantMortality	Infant deaths per 1,000 live births as of 2017	21.73; 16.745 (16.06, 0, 100)	
	Family	Razdep	Percentage of people under the age of 15 and people aged 65 and over in relation to the population aged 15 to 64, multiply by 100	44.43; 4.899 (44.16, 33.83, 64.01)	
	Education	Schooling	Percentage of children at the age of 6 to 14 years old that go to school, multiply by 100	98.19; 1.292 (98.4, 92.8, 100)	
	State Capacity	IFGF OwnRevenue	The level of autonomy of municipal revenues as of 2016. The closer to 1, the less dependent on aid of other entities (States and Federal Government).	0.299; 0.227 (0.247; 0; 1)	
		Revenue	Realized revenue in R$ (×1000) as of 2017	89,347; 229,408.1 (27,712; 12,912; 2,280,394)	
		Expenditure	Committed expenditure in R$ (×1000) as of 2017	77,066; 201,796.9 (23,542; 10,033; 1,907,864)	

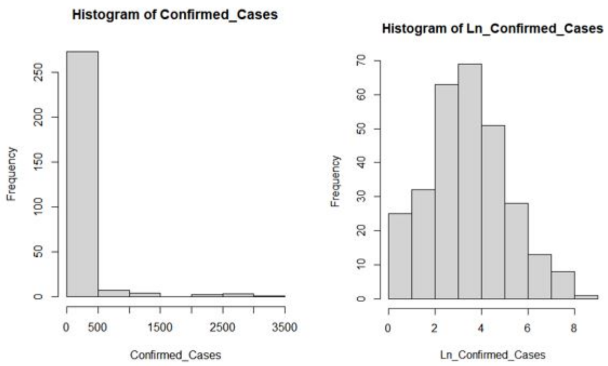

Figure 1(a) - Histogram distribution of confirmed cases (Left) versus the log-transformed confirmed cases (Right)

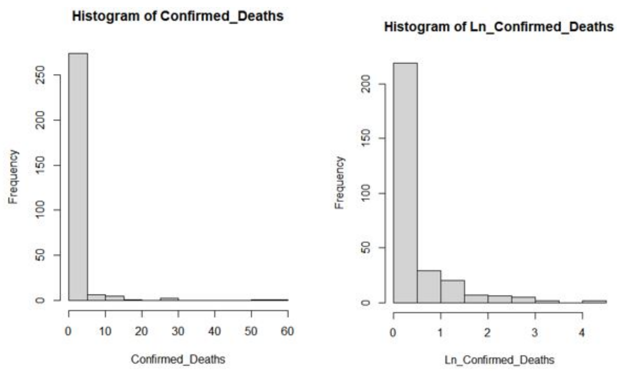

Figure 1(b) - Histogram distribution of confirmed deaths (Left) versus the log-transformed confirmed deaths (Right)

3.3. Regression Analysis

We first run the multiple regression analysis using the above listed 20 independent variables and log-transformed dependent variables. We only show and interpret the results using log-transformed dependent variables in this section because the regression model using the original dependent variables have a larger residual standard error and a non-random residual plot, which means that the model using the original dependent variables is not a good fit for the data and may be problematic. The residual plots showing the non-random pattern using the dependent variables in the original forms are shown in the Appendix. All regression analyses are performed using RStudio (RStudio Team, 2020).

The multicollinearity check of all the 20 independent variables is also performed, and the results of the Variance Inflation Factor (VIF) are shown in Table 2. A commonly used rule of thumb for multicollinearity measurement is that when the value of VIF is ten or more, it is a sign of a serious multicollinearity problem (O'Brien, 2007). Our results show that most of this study's independent variables do not suffer from severe multicollinearity issues. However, five of them (Population, Expenditure, Revenue, Hospital, ICU_Beds) have VIF values higher than 10. Therefore, we delete the five independent variables to eliminate multicollinearity issues and rerun the regression analysis. The new model results using the remaining 15 independent variables are shown in Table 3 and Table 4. The regression model using all 20 independent variables is shown in the Appendix, and we do not interpret the results in this section because of the multicollinearity issues suffered in this model.

Table 2 -Multicollinearity check for all 20 independent variables

Area	1.73
Population	35.20
WaterSewage	1.35
BHU	5.81
CPR	1.94
Density	2.22
Expenditure	271.97
Revenue	273.02
GDP	1.33
Gini	1.45
Hospital	12.71
ICUBeds	25.50
IdhmL	1.73
InfantMortality	1.51
Razdep	3.75
Schooling	1.31
TDens	2.40
TRmaxidoso	1.63
TDes18M	1.60
IFGF_own_revenue	2.13

Table 3 -Regression analysis of significant determinants using 15 independent variables and log-transformed confirmed cases of COVID-19 (dependent variable) in SC, Brazil

Independent Variables	**Estimate Coefficient**	**t-statistic**
Intercept	1.161e+00	0.196
Area	-6.041e-04	-2.552 *
WaterSewage	5.379e-02	2.137 *
BHU	5.989e-02	6.225 ***
CPR	-1.503e-02	-2.328 *
Density	5.661e-04	1.378
GDP	5.471e-06	1.294
Gini	2.129e-02	1.653
IdhmL	5.741e-02	2.193 *
InfantMortality	-1.765e-02	-4.163 ***
Razdep	-1.070e-01	-4.689 ***
Schooling	4.521e-03	0.088
TDens	6.981e-02	4.124***
TRmaxidoso	-4.336e-02	-0.540
TDes18M	2.310e-02	0.565
IFGF_own_revenue	9.136e-01	2.537*
Residual standard error: 0.9925 on 274 degrees of freedom	Multiple R^2: 0.6853 Adjusted R^2: 0.6681	F-statistic: 39.78 on 15 and 274 DF, p-value: < 2.2e-16
*** =significant at the 0.1% level, * =significant at the 5% level		

Table 4 - Regression analysis of significant determinants using 15 independent variables and the log-transformed confirmed deaths due to COVID-19 (dependent variable) in SC, Brazil

Independent Variables	Estimate Coefficient	t-statistic
Intercept	1.161e+00	0.196
Area	-6.041e-04	-2.552 *
WaterSewage	5.379e-02	2.137 *
BHU	5.989e-02	6.225 ***
CPR	-1.503e-02	-2.328 *
Density	5.661e-04	1.378
GDP	5.471e-06	1.294
Gini	2.129e-02	1.653
IdhmL	5.741e-02	2.193 *
InfantMortality	-1.765e-02	-4.163 ***
Razdep	-1.070e-01	-4.689 ***
Schooling	4.521e-03	0.088
TDens	6.981e-02	4.124***
TRmaxidoso	-4.336e-02	-0.540
TDes18M	2.310e-02	0.565
IFGF_own_revenue	9.136e-01	2.537*
Residual standard error: 0.9925 on 274 degrees of freedom	Multiple R^2: 0.6853 Adjusted R^2: 0.6681	F-statistic: 39.78 on 15 and 274 DF, p-value: < 2.2e-16
*** =significant at the 0.1% level, * =significant at the 5% level		

4. Results Interpretation

Table 3 and Table 4 illustrate the regression analyses using 15 independent variables and log-transformed dependent variables. The two tables address the research question that try to identify what socioeconomic factors affect the pandemic outcomes as shown in equations (1) and (2). In equation (1), the independent variable is the log-transformed confirmed COVID-19 case number. In equation (2), the independent variable is the log-transformed confirmed COVID-19 death number.

*Equation (1): log (COVID case number)= a+b1*Area+b2* WaterSewage+b3* BHU+b4*CPR +b5*Density+b6*GDP+b7*Gini+b8* IdhmL+b9* InfantMortality+b10* Razdep+b11* Schooling+b12* TDens+b13* TRmaxidoso+b14* TDes18M+b15* IFGF_own_revenue+e*

*Equation (2): log (COVID death number)= a+b1*Area+b2* WaterSewage+b3* BHU+b4*CPR +b5*Density+b6*GDP+b7*Gini+b8* IdhmL+b9* InfantMortality+b10* Razdep+b11* Schooling+b12* TDens+b13* TRmaxidoso+b14* TDes18M+b15* IFGF_own_revenue+e*

Where a is the intercept of each regression model, b is the coefficient (slope) of each independent variable, e is the residual of each regression model.

We interpret the significant determinants identified in the regression analyses by reading the t-statistics of each independent variable. In Table 3, four independent variables (BHU, Infant_Mortality, Razdep, TDens) are identified as significant indicators at the 0.1% level with three stars addressed at their t-statistics. The result indicates that municipalities with more BHUs, lower infant mortality rate, a higher percentage of people aged 15 to 64, and denser households are more likely to have more COVID-19 reported cases. Five other independent variables (Area, WaterSewage, CPR, IdhmL, IFGF_own_revenue) are identified as significant indicators at the 5% level with one star addressed at their t-statistics. This result provides more characteristics of a municipality that possibly has high reported COVID-19 cases: the municipality is smaller in its area size, more people living in a household with inadequate water supply and sanitation, fewer people in self-employed status, higher Municipal Human Development Index, and higher autonomy level of municipal revenues. The adjusted R squared in Table 2 (a) is 0.6681, which means that the regression model can explain 66.81% of the variation in confirmed cases' values. Lamb, Kanudla & Shaman (2020) also found that socioeconomic factors explained 56% of the variability in case positivity in New York City in the early phase of the pandemic. The identified significant socioeconomic factors provide us with a brief picture of why some municipalities already suffered from the novel coronavirus's high morbidity. We can deduct from the regression analysis that the municipalities provide many employment opportunities since fewer people are working as freelancers. This characteristic and the high percentage of people aged 15 to 64 means that more people have to commute to work, and they are more likely to have less freedom in deciding whether to work at home when the pandemic starts. Other indicators such as smaller area size, more households with greater than two people per bedroom, and more people living in a household with inadequate water supply and sanitation offer a favorable condition for the virus transmitted via respiratory droplets. The rest of the regression analysis conclusion seems interesting yet confusing. The rest of the significant indicators indicate that the municipality is possibly at a relatively highly developed stage because of its independence on external financial aids and its high Municipal Human Development Index. In particular, the municipality is highly developed from the health care perspective because of more access to BHUs and a low infant mortality rate. With better health care conditions, it seems to be counterintuitive that these municipalities have more reported COVID-19 cases than those with a limited number of BHUs and higher infant mortality rates. We will discuss this confusing conclusion later after the conclusion from the regression analysis about the COVID-19 related deaths.

We run a similar analysis using log-transformed deaths related to COVID-19 as the dependent variable and the same set of independent variables. From Table 4, we identify a fewer number of significant indicators than we do in Table 3. BHU is still recognized as a significant indicator of COVID related deaths at the 0.1% level, and density is another indicator that is identified at the same significance level. TDens is significant at the 1% level, and IFGF_own_revenue is a significant factor at the 10% level. To sum up the finding, Table 4 shows that municipalities with a higher number of pandemic-related deaths tend to be equipped with more BHUs, highly dense, having more households with greater than two people per bedroom, and are less dependent on state and federal aids. The regression analysis in Table 4 has an adjusted R squared of 0.5233, which means that this model only explains a little more than half of the death values' variations. We believe that this result is due to the limited data on COVID related deaths. As of July 12, 2020, most of the municipalities in Santa Catarina have not reported any death related to COVID-19. At that time, COVID related deaths were identified in 71 municipalities with a total number of 399. We notice that Table 4 concludes that both municipality density and TDens (household bedroom density) could positively impact the number of deaths due to the virus, so we run a correlation test to make sure the two independent variables are not correlated. We run a Spearman correlation in R to calculate the correlation coefficient between density and TDens. The correlation coefficient is 0.04578, which means that the two variables do not associate with each other. Other studies like Kumar, Quinn, Kim, Daniel, and, Freimuth (2012) also found association between risk exposure to the influenza virus and living in an apartment building with 2 or more people, showing that both pandemics might share similar characteristics about socioeconomic factors affecting the pandemic unfolding. Using data for the COVID-19 pandemic, Lamb, Kanudla & Shaman (2020) also found a positive association between living in a household with 4 or more individuals and case positivity rate.

Table 3 and Table 4 conclude that there are some similarities between the significant indicators identified from the regression analyses about confirmed cases and deaths, respectively. The number of BHU, the household bedroom density, and the autonomy level of municipal revenues could positively impact both COVID-19 cases and deaths. In other words, municipalities with higher confirmed cases and death tolls are more likely to have more BHUs, more percentage of households with more than two people in a bedroom, and these municipalities typically rely less on aids from higher authorities. With more available BHUs in the municipality, both confirmed cases and deaths from COVID-19 are higher than municipalities with a limited number of BHUs. We suspect that the result is associated with the issue of test availability. Since COVID-19 is a novel disease, people need to go to a government-assigned location for testing. People living in a municipality with abundant medical resources such as BHUs are more likely to take tests because they pay more attention to suspected symptoms. They are aware of their easier access to the COVID-19 tests due to the propaganda by the BHUs. Because more people get tested in the municipality with more BHUs, the number of confirmed cases in this municipality increases more than municipalities with limited or zero BHUs. Even if the transmission in municipalities without BHUs are as bad as, or worse than municipalities with many BHUs, the confirmed case numbers and the death numbers could not reflect the real situation without sufficient tests. On the other hand, higher confirmed cases found in some municipalities could also be attributed to the availability of testing because people become gathered at test locations. The gathering provides an opportunity for the respiratory virus to spread further. Our conclusion contradicts the assumption that Hamidi, Sabouri, & Ewing (2020) holds. They assumed that denser areas have better healthcare infrastructures, and the healthcare infrastructures help decrease COVID-19 infections. Our results show that more available basic healthcare units increase the number of COVID-19 reported cases in a municipality. We admit that more BHUs causing more confirmed cases and deaths may be due to the correlation between BHU and other socioeconomic factors, such as the municipality population. We run the correlation test of BHU and the other 19 socioeconomic factors included in this study. We find that BHU is highly correlated (correlation coefficient is greater than 0.5) with the following variables: Population, Expenditure, Hospital, ICU_Beds, and Revenue. However, this correlation result could not overturn the significance of BHU. These four variables are deleted from the original model illustrated in the Appendix and they are not included in Table 3 and Table 4. We still run the regression model with all 20 variables and keep the results in the Appendix. As shown in the Appendix tables, none of the four variables are significant indicators of the confirmed cases and deaths due to COVID-19. Therefore, we retain our conclusion about the critical role that the Basic Health Unit plays in the fight against the pandemic.

As we mentioned in the literature review section, previous studies hold different opinions about how population density (of a country) impacts pandemic outcomes (Hamidi, Sabouri, & Ewing, 2020; Rocklöv & Sjödin, 2020). Our analyses show that the population density is a significant indicator for the deaths caused by COVID-19 (at the 0.1% level). However, more importantly, household bedroom density could significantly impact the new disease's transmission and mortality. Rocklöv & Sjödin (2020) concluded that population density is a catalyst of the transmission based on the assumption that virus transmissions are faster in a highly populated area because more contacts are inevitable. Our result further points out that family bedroom density is a vital factor if the government wants to control the transmission and lower the mortality rates of COVID-19. Similar with the conclusions in Lamb, Kanudla & Shaman (2020), household density underscores what is already known from other previous studies, including influenza studies, that majority of infections occur among members of a household (or in workplaces). Family bedroom density and households with inadequate water supply and sanitation are detailed socioeconomic factors that reflect citizens' living conditions but are not commonly discussed in COVID-19 related studies. Our study indicates the vital roles of such factors. It provides support to social control decisions regarding how families, especially those without adequate living space and sanitization, should be more alert to respiratory infectious diseases such as COVID-19. This result is particularly important for countries similar to Brazil. As is commonly known, Brazil has poor communities that usually have a high-density population and small houses with many relatives sharing the same space. The opposite of this situation can be seen in Sweden, a country with a very controversial approach to deal with COVID-19. However, it has seen cases going down during the 2020 summer, despite not having closing schools or doing strict quarantines. On the other hand, Sweden has one of the lowest density population in Europe, and many people live alone in their homes (Erdbrink, 2020).

5. DISCUSSION

In this section, we discuss the possibility of applying and generalizing the conclusion. From the above discussion, it is observable that BHU is a unique variable to Brazil and it distinguishes our study to others. Furthermore, BHU is the only significant indicator at the 0.1% level in both analyses of confirmed cases and deaths due to COVID-19. Therefore, we want to first use the most significant indicator to predict possible COVID-19 hotspots in Santa Catarina to prove the utility of our conclusion. Our conclusion from the regression analyses is that a higher number of BHUs lead to higher numbers of confirmed cases and deaths due to the disease. If the conclusion is valid, then municipalities with more BHUs should either show high cumulative confirmed cases and deaths, or they will become the next outbreak spots.

Figure 3(a) and Figure 3(b) show the top municipalities in Santa Catarina that have high cumulative confirmed cases and deaths, respectively. We include information about the number of BHUs in Figure 4. We pick three cities according to their places on Figure 3 and 4: **Jaraguá do Sul** (confirmed cases ranked 19th, deaths ranked 19th, BHU ranked 6th), **Gaspar** (confirmed cases ranked 23rd, deaths ranked 23rd, BHU ranked 15th), and **Lages** (confirmed cases ranked 25th, deaths ranked 59th, BHU ranked 9th). These three cities did not have many confirmed cases or deaths as of July 12, 2020, but they have many available BHUs, as shown in Figure 4. Most of the high ranked municipalities in Figure 3 (a) and (b) are already identified as highly infected Santa Catarina areas. We think that the three cities are outliers in the regression analyses, and they will be the next outbreak hotspots according to our analyses. We include the municipalities with high cumulative confirmed cases and deaths as of September 30 in Figure 5 (a) and (b).

We have observed that these three cities have more confirmed cases and deaths reported than other cities, reflected by their fast-improved rankings. As for the confirmed cases on September 30, 2020, we observe that Lages is ranked as top as 12th compared to its original 25th position two months ago. Jaragua do Sul is ranked as 13th compared to 19th originally. Gaspar is the 16th highest municipality with confirmed cases when it was only the 23rd in July. As for the COVID-19 related deaths on September 30, 2020, we observe a huge increase for the city of Lages. It is ranked as high as 9th compared to its original 59th position two months ago. The other two cities did not have such big jump in the deaths ranks, but small increases are observed. Jaragua do Sul is ranked as 17th compared to 19th originally. Gaspar is the 19th highest municipality with confirmed cases when it was the 23rd in July. After over two months, our prediction that Jaragua do Sul, Gaspar, and Lages will develop to be the next infection hotspots is validated. We believe the results indicate that our regression analysis method is appropriate for the government to identify possible hotspots and formulate policies that target the predicted areas particularly. If the government could target the predicted areas with adjusted policies, including providing more medical services, the transmission and infection of the COVID-19 may become slower in these areas and the government policies will become more effective.

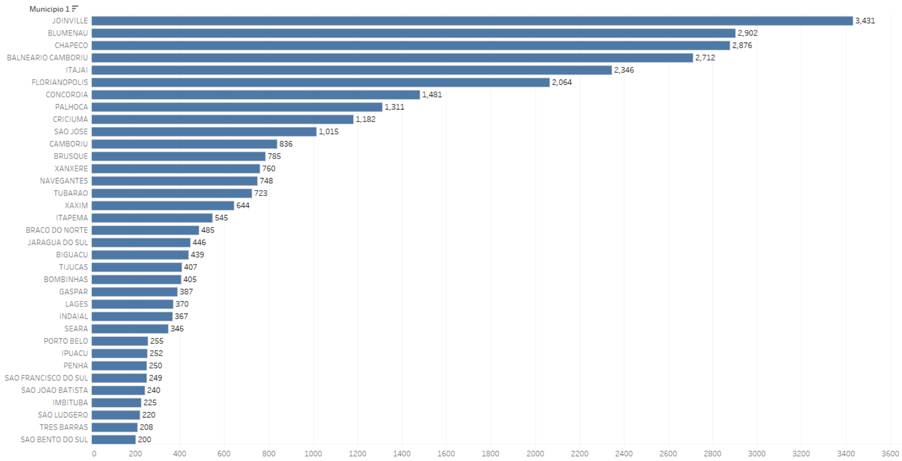

Figure 3 (a) - Municipalities with high cumulative confirmed cases in Santa Catarina as of July 12, 2020.

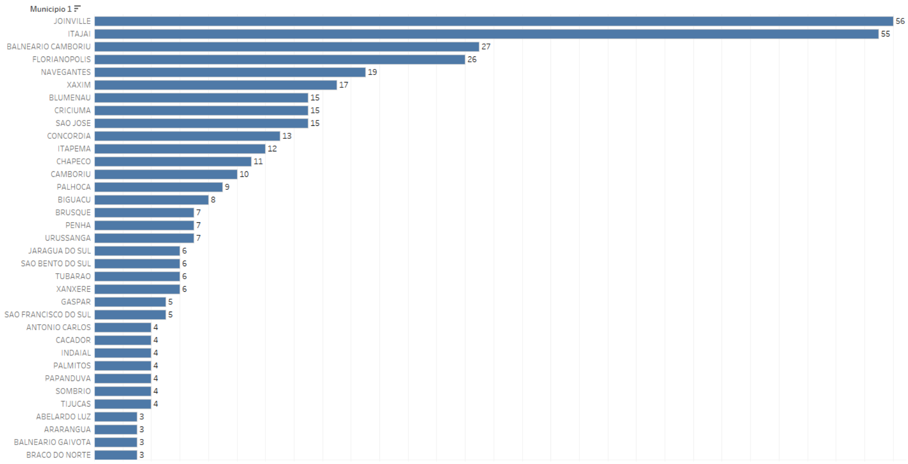

Figure 3 (b) -Municipalities with high cumulative deaths due to COVID-19 in Santa Catarina as of July 12, 2020.

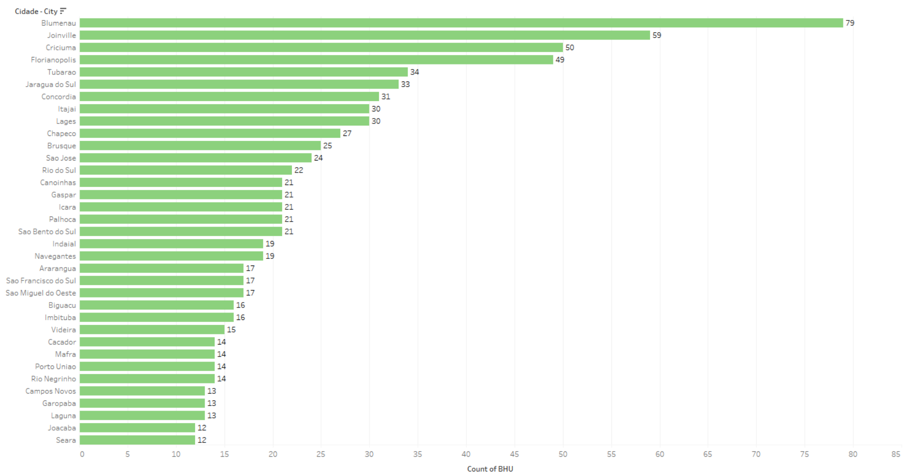

Figure 4 - Rankings of municipalities with Basic Health Units.

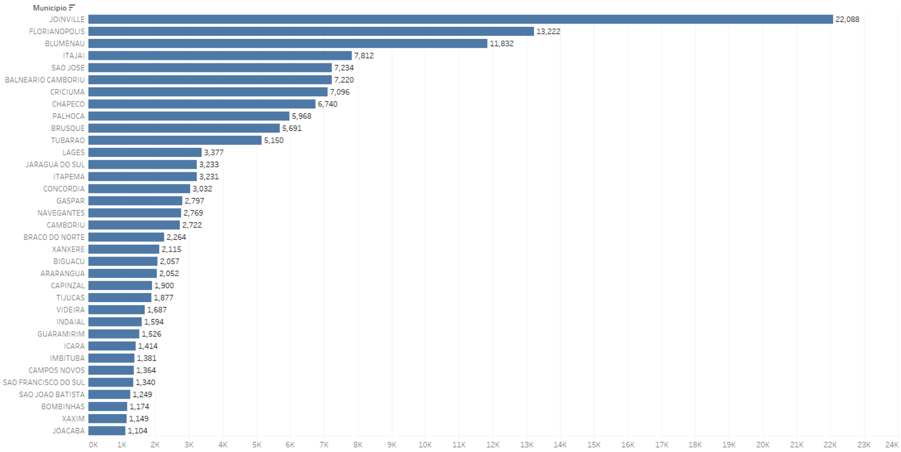

Figure 5(a) -Municipalities with high cumulative confirmed cases in Santa Catarina as of September 30, 2020.

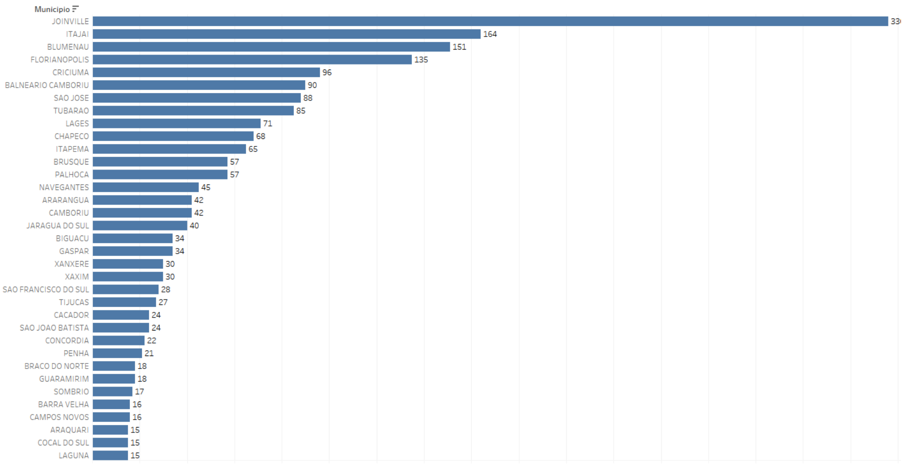

Figure 5(b) - Municipalities with high cumulative deaths due to COVID-19 in Santa Catarina as of September 30, 2020.

The discussion above covers how to utilize the identified socioeconomic factor to predict the pandemic development, and the pandemic development numbers in real life validate the prediction results. We next discuss if the proposed models could be applied to other states in Brazil and generate similar conclusions for significant socioeconomic factors. We choose Minas Gerais as the state to validate our proposed models. According to IGBE (2020), Minas Gerais is ranked as the second-highest state in population, the fourth in area size, and the third in GDP. We collect the related socioeconomic factors used in the proposed models and the COVID-19 case and mortality numbers in Minas Gerais by July 12, 2020. We only collect socioeconomic factors used in the final regression model, making the count of the independent variables 15. After deleting the municipalities with null values in certain socioeconomic factors, our sample contains 608 municipalities in Minas Gerais. The descriptive statistics of the socioeconomic factors for Minas Gerais are shown in Appendix A.3, Table A-3. Because the Minas Gerais sample contains more municipalities than the Santa Catarina sample, we observe that both dependent and independent variables in Minas Gerais are more volatile. For example, the standard deviations for the confirmed cases and death numbers are higher. WaterSewage (Housing variable), Density (Physical Environment variable), Gini (Income variable), and Razdep (Family variable) also have higher variations in Minas Gerais. The descriptive statistics show that comparing to Santa Catarina, both the pandemic outcomes and the social indicators in Minas Gerais have greater variations for its municipalities. Using the Minas Gerais sample, the analysis result could provide us more support to generalize our proposed model and conclusions.We rerun the regression models using Minas Gerais information to observe the pandemic outcomes' significant indicators. The regression analysis results are shown in Appendix A.3, Table A-4, and Table A-5. Table A-4 identifies five independent variables (WaterSewage, BHU, InfantMortality, TDes18M, IFGF_own_revenue) that are significant at the 0.1% level, three independent variables (GDP, IdhmL, TDens) that are significant at the 1% level, and two independent variables (Area, Razdep) that are significant at the 5% level. The adjusted R squared indicates that the regression model could explain 57.91% of the variation in confirmed cases' values. Comparing the results to Santa Catarina results, we observe many similarities in the significant indicators. The Santa Catarina results indicate that municipalities with more BHUs, lower infant mortality rate, a higher percentage of people aged 15 to 64, denser households, higher Municipal Human Development Index, and higher autonomy level of municipal revenues are more likely to have more COVID-19 reported cases. A highly similar conclusion can be conducted from the Minas Gerais results, indicating that municipalities in Minas Gerais with these identified socioeconomic factors (BHU, InfantMortality, IFGF_own_revenue, TDens, Razdep) should be more aware of fast-growing COVID-19 confirmed cases. However, two socioeconomic factors (Area, WaterSewage) show opposite polarity in Minas Gerais results. The Santa Catarina results indicate that smaller area size and more people living in a household with inadequate water supply and sanitation are characteristics of a municipality with possible COVID-19 case outbreak, but the conclusion is opposite in Minas Gerais. Furthermore, we find that TDes18M and GDP are two significant indicators unique to Minas Gerais results, indicating that the high unemployment rate of the population aged 18 or over and high GDP are also characteristics of a municipality with high reported cases. While wealthier population could weather short-term financial losses better (Pullano *et al.*, 2020), we think that the controversy of significant variables in Minas Gerais results and the two newly identified socioeconomic factors could be explained by an assumption that metropolitan area in this state performs more tests to its inhabitants and reports higher confirmed cases. It is reasonable that a municipality with a higher level of development (higher Municipal Human Development Index, fewer people living in a household with inadequate water supply and sanitation) and high GDP has more available financial resources to perform more COVID-19 tests, resulting in more reported confirmed cases. Chaudhry *et al.* (2020) also found positive association between COVID-19 cases and GDP per capita. Their explanation is that countries with better GDP per capita also have more widespread testing, greater transparency with reporting and better national surveillance systems, wich are all plausible explanations for the association we found in Minas Gerais.Comparing Minas Gerais and Santa Catarina results when using the confirmed death number as the dependent variable is also interesting. The conclusion that high numbers of BHU and denser households are important indicators for high death numbers still holds in Minas Gerais, but higher area density results in a lower death number in this case. Other newly identified characteristics that lead to higher death numbers include fewer people living in households with an inadequate water supply and sanitation, higher Municipal Human Development Index, lower Infant Mortality rate, higher unemployment rate of the population aged 18 or over, and higher autonomy level of municipal revenues. Compared to the Santa Catarina sample model, the Minas Gerais sample model identifies more significant indictors when using the COVID-19 death number as the dependent variable. The results seem a lot similar to the results using the confirmed case as the dependent variable. We believe the Minas Gerais sample provides more support to our previous conclusion that high density in households and more available BHUs are the two most important indicators that contribute to a higher number of confirmed cases and deaths due to COVID-19. Minas Gerais's results also suggest that governments of all levels should closely monitor metropolitan areas because of their high possibility of fast-growing COVID-19 cases.

However, this conclusion does not mean that governments should ignore necessary resource allocation to less developed municipalities. The fact that identified important indicators for possible COVID-19 hotspots can be mostly categorized as big cities' characteristics may expose a resource allocation problem. As we stated earlier, the metropolitan area has more available resources such as the BHUs to conduct tests and convince its inhabitants to take the tests. In contrast, the less developed area does not have enough resources to react as its peers.We want to indicate that one limitation of our study is the authenticity of the data source of our dependent variable, Confirmed cases of COVID-19 and Confirmed deaths due to COVID-19. We believe that this is also a limitation that most of the studies that explore the development and the control of the COVID-19 share. We, as similar as other researchers, news media, politicians, and the general public, do not have other resources but the government released numbers to understand and estimate the pandemic circumstances. Our explanation power of the significant indicators may be limited because of the limited nature of the released pandemic data. It is possible that the current pandemic measurements do not completely capture the real pandemic development due to factors such as tests availability and accuracy, medical resources, and the complicated nature of the disease. Although we believe this limitation does not weaken the contribution of our study to government policymaking, it may hinder the estimation on the effects of government policies. We hope that future research could manage to propose a new set of measurements to reflect the pandemic more accurately and timely, so that our established regression model could run using data that actually captures the epidemic development.

6. CONCLUSION

The novel coronavirus's high transmission speed and complex symptoms leave governments of all regions and of all levels limited time to understand the global epidemic's development; government needs a tool to better understand the epidemic development, predict the pandemic development and identify possible hotspots. This paper constructs a multiple regression analysis using publicly accessible COVID-19 data and socioeconomic data collected from Santa Catarina, a state in Brazil. We identify significant indicators of pandemic outcomes. We conclude that high density in household and more available BHUs are the two most important indicates that contribute to a higher number of confirmed cases and deaths due to COVID-19. Other significant indicators towards pandemic outcomes are infant mortality rate, age group diversity, municipality area size, households water supply and sanitation, Municipal Human Development Index, and autonomy level of municipal revenues. We also include another state, Minas Gerais, to compare the results with Santa Catarina to see if our conclusion could be generalized to other states. We think the comparing test provides more support to our model and the generated conclusion, and it further reminds the government to consider for a more reasonable resource allocation, especially under the pandemic circumstances. The identified significant socioeconomic indicators include indices and statistics that are already related to public health, and that are not directly associated with the evaluation of population health. One of the contributions of our study is that we point out the importance of associating different socioeconomic factors with health outcomes. We further suggest the government to consider the same methodology to include factors that are unique to the area when establishing and implementing health policies. Our analysis also adds to the ongoing academic discussion that explores the relationship between socioeconomic factors and pandemic outcomes. With the incorporation of the proposed model, policymakers could take appropriate and prompt measures to fight against the current pandemic and effectively prepare for a possible future crisis.

Previous studies include lockdown and border closure as factors to investigate the impact of government policies that are made to fight the pandemic (Chaudhry *et al.*, 2020), but they did not provide validations using their identified significant social factors and government policies. We apply our identified most significant indicator to predict three cities as new hotspots using the state-released COVID-19 data by July 12, 2020. We validate the prediction by showing that the three cities become quickly increased cities with the highest number of confirmed COVID-19 cases and deaths in only two months. We believe this analysis methodology could serve as an illustration and could be adopted by different regions. The methodology could predict the pandemic transmission trend by identifying significant socioeconomic factors using specific data about the region of interest.

Future research could expand our model by including different social factors to examine the impacts of government policies and interventions responding to an unexpected pandemic. Because the demographic situations vary in different countries, we suggest adjusting the variables and checking the independent variables' multicollinearity before applying our regression analysis model. COVID-19 can also be understood as a Complex Intergovernmental Problem (Paquet & Schertzer, 2020) and variables related to that could be included in future research.

References

Aguiar, R. B. D., & Lima, L. L. (2019). Capacidade estatal: definições, dimensões e mensuração. *Bib: revista brasileira de informação bibliográfica em ciências sociais*. São Paulo, SP. N. 89 (ago. 2019), p. 1-28.

Andreoni, M. (2020). *Coronavirus in Brazil: What You Need to Know*. Retrieved from The New York Times: https://www.nytimes.com/article/brazil-coronavirus-cases.html

Brazilian Ministry of Health. (2020). *Coronavirus: See list of hospitals that will be a reference in Brazil*. Retrieved from Coronavirus (COVID-19): https://www.saude.gov.br/noticias/agencia-saude/46257-mapa-hospitais-referencia-novo-coronavirus

Brasil, G. F. (2020). *O que tem na UBS*. http://maismedicos.gov.br/o-que-tem-na-ubs

Brazilian Institute of Geography and Statistics (IBGE). (2020, June 30). *Brazil Unemployment Rate*. Retrieved from Trading Economics: https://tradingeconomics.com/brazil/unemployment-rate

Cárdenas, M. (2010). State capacity in latin america. *Economía, 10*(2), 1-45.

Chaudhry, R., Dranitsaris, G., Mubashir, T., Bartoszko, J., & Riazi, S. (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*, 000, 100464. https://doi.org/10.1016/j.eclinm.2020.100464

Duan, H. K., Hu, H., Vasarhelyi, M. A., Da Rosa, F. S., & Lyrio, M. V. L. (2020). Open Government Data (OGD) driven decision aid: a predictive model to monitor COVID-19 and support decisions in a Brazilian State. *Revista Do Serviço Público, 71*(Special Edition), 140–164. https://doi.org/https://doi.org/10.21874/rsp.v71i0.5009

Erdbrink, T. (2020). *Vilified Early Over Lax Virus Strategy, Sweden Seems to Have Scourge Controlled*. The New York Times. https://www.nytimes.com/2020/09/29/world/europe/sweden-coronavirus-strategy.html

FIRJAN (2020). IFGF 2019. Retrieved from https://www.firjan.com.br/ifgf/Lamb, M. R., Kandula, S., & Shaman, J. (2020). Differential COVID - 19 case positivity in New York City neighborhoods: Socioeconomic factors and mobility. *Influenza and Other Respiratory Viruses*.

Merco Press. (2020). *Virus moving south: Brazilian state of Santa Catarina reimposes strict confinement measures*. Retrieved from Merco Press: https://en.mercopress.com/2020/07/20/virus-moving-south-brazilian-state-of-santa-catarina-reimposes-strict-confinement-measures

Minas Gerais State Department of Health. (2021, February 10). COVID-19 - Confirmed Cases. Retrieved from Minas Gerais State Open Data Portal: http://dados.mg.gov.br/dataset?tags=coronavirus

Hamidi, S., Ewing, R., & Sabouri, S. (2020). Longitudinal analyses of the relationship between development density and the COVID-19 morbidity and mortality rates: Early evidence from 1,165 metropolitan counties in the United States. *Health and Place*, 64(102378), 1–9.

Horton, R. (2020). COVID-19 is not a pandemic. Comment, 396 (10255) 874-875.https://doi.org/10.1016/S0140-6736(20)32000-6

Kumar, S., Quinn, S. C., Kim, K. H., Daniel, L. H., & Freimuth, V. S. (2012). The Impact of Workplace Policies and Other Social Factors on Self-Reported Influenza-Like Illness Incidence During the 2009 H1N1 Pandemic. *American Journal of Public Health, 102*(1), 134–140. https://doi.org/10.2105/AJPH.2011.300307

IGBE. (2020). *Cities and States*. Retrieved from IGBE: https://www.ibge.gov.br/en/cities-and-states/sc/blumenau.html

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, 41(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Pereira, D. V. S., Mota, C. M. M., & Miranda, J. A. G. (2013). Municipal human development index for Brazilian cities based on ELECTRE TRI-C. *Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013*, 736–740. https://doi.org/10.1109/SMC.2013.130

Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S., & Colizza, V. (2020). Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. *The Lancet Digital Health,* 2(12), e638-e649.

Rocklöv, J., & Sjödin, H. (2020). High population densities catalyse the spread of COVID-19. *Journal of Travel Medicine*, 27(3), 1–2. https://doi.org/10.1093/jtm/taaa038

RStudio Team. (2020). *RStudio: Integrated Development for R*. Boston, MA, USA: PBC.

Santa Catarina State Department of Health. (2020, July 14). *COVID-19 - Confirmed Cases*. Retrieved from Coping with COVID-19: http://dados.sc.gov.br/dataset/covid-19-dados-anonimizados-de-casos-confirmados

Paquet, M., & Schertzer, R. (2020). COVID-19 as a Complex Intergovernmental Problem. *Canadian Journal of Political Science/Revue canadienne de science politique*, 1-5.

Secretaria de Saúde do Distrito Federal. (2020, August 20). *Unidades Básicas de Saúde (UBS)*. Retrieved from Secretaria de Saúde do Distrito Federal: http://www.saude.df.gov.br/unidades-basicas/

Soares, M. (2020). *Re-opening for business, Santa Catarina will be Brazil's test case*. Retrieved from The Brazillian Report: https://brazilian.report/business/2020/04/23/re-opening-business-santa-catarina-brazil-test-case/

Teixeira da Silva, J. A., & Tsigaris, P. (2020). Policy determinants of COVID-19 pandemic–induced fatality rates across nations. *Public Health, 187*, 140–142. https://doi.org/10.1016/j.puhe.2020.08.008

The County Health Rankings & Roadmaps program. (2020). *Social and Economic Factors*. Retrieved from The County Health Rankings & Roadmaps program: https://www.countyhealthrankings.org/explore-health-rankings/

The Economist. American inequality meets covid-19. (2020). https://www.economist.com/united-states/2020/04/18/american-inequality-meets-covid-19 (accessed November 30, 2020).

WHO. (2020). *Coronavirus disease 2019 (COVID-19) Situation Report – 73*. World Health Organization.