THE UNIVERSITY OF CHICAGO

The Effect of Educational Attainment on COVID-19 Morbidity in the City of Chicago

By

Haoyu Shi

December 2021

A paper submitted in partial fulfillment of the requirements for the

Master of Arts degree in the

Master of Arts Program in the Social Sciences

Faculty Advisor: Donald R. Hedeker

Preceptor: Amit Anshumali

Abstract

The Effect of Educational Attainment on COVID-19 Morbidity in the City of Chicago
By
Haoyu Shi

Coronavirus disease 2019 (COVID-19) has affected everyone but not equally. While many researchers examine the social inequalities of COVID-19 through income, housing, occupation, and health care accessibility, less attention has been paid to education due to lack of information for education associated with test reports. Based on the education-health gradient (Cutler & Lleras-Muney, 2006), education has causal effects on various health outcomes; even with similar income and occupation, people with higher educational attainment generally have better health outcomes. This theory has been widely adopted to research chronic and degenerative diseases, such as cardiovascular disease, diabetes, and high blood pressure. However, there is limited research on the effect of education on communicable diseases. To analyze the potential impacts of education on communicable diseases, this research examines the relationship between educational attainment and COVID-19 morbidity in 56 ZIP Code Tabulation Areas (ZCTAs) of Chicago. COVID-19 statistics were retrieved from the Chicago Department of Public Health (CDPH) and socioeconomic status (SES) statistics were retrieved from the American Community Survey 2019 (ACS-2019). Due to the variation in the test rates across Chicago, the estimates of morbidity based on infection rate tend to underestimate the true morbidity of communities with lower test rates. Therefore, this research measures COVID-19 morbidity by both test positivity rate and infection rate. Educational attainment is measured by the proportion of adults with at least a bachelor's degree. Median income and average household size are used to control accessibility to resources and living conditions. The proportion of workers in service occupations are used to control occupational risk. Several linear regression models were conducted. Because the SES variables are highly correlated, regularization was used to combat multicollinearity. Because infectious diseases spread from one community to the neighboring communities, spatial regression was used to combat the spatial diffusion effect of COVID-19. The results show consistent negative and significant associations between educational attainment and both COVID-19 test positivity rate and infection rate. This indicates that high educational attainment lowers COVID-19 morbidity at community level. Public health experts and policymakers should make education more accessible to mitigate the social inequality in health and achieve better population health. Since this research analyzes cross-sectional data at aggregate level, the relationship between educational attainment and COVID-19 morbidity in this research might not be causal and may not be generalized to individuals. Further research utilizing individual-level and longitudinal data is required to study the causal pathways of education on infectious disease outcomes.

Keywords: COVID-19, Public Health, Social Determinants of Health, Education-Health Gradient, Regularization, Spatial Regression

LIST OF ACRONYMS

ACS	American Community Survey
AIC	Akaike Information Criterion
CDPH	Chicago Department of Public Health
COVID-19	Coronavirus Disease 2019
OLS	Ordinary Least Squares
RMSE	Root Mean Square Error
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SES	Socioeconomic Status
SSR	Sum of Squared Residuals
VIF	Variance Inflation Factor
WHO	World Health Organization
ZCTA	ZIP Code Tabulation Area

Introduction

Coronavirus disease 2019 (COVID-19), an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has spread over the world since first being discovered in December 2019, leading to an ongoing pandemic. As updated in August 2021, more than 200 million cases have been diagnosed with more than 4 million deaths in over 200 countries and territories (WHO, 2021). While COVID-19 has affected everyone, there are systematic disparities in COVID-19 outcomes among different populations. Many researchers examine the social inequalities of COVID-19 outcomes through racial discrimination, income, housing, occupational exposure, and health care accessibility (Chowkwanyun & Reed, 2020; Cordes & Castro, 2020; Credit, 2020; Devakumar et al., 2020; Hawkins et al., 2020; Mutambudzi et al., 2021; Patel et al., 2020). However, less attention has been paid to education due to the lack of education information associated with COVID-19 test reports.

People with higher educational attainment generally have better health outcomes (Cutler & Lleras-Muney, 2006; Kunst et al., 2002). In Healthy People 2020, a set of national health objectives proposed by the Office of Disease Prevention and Health Promotion, education has been included as one of the five social determinants of health. Previous research has shown that education has causal effects on various health behaviors and health outcomes (Cutler & Lleras-Muney, 2010; de Walque, 2007; Sassi et al., 2009). Cutler and Lleras-Muney (2006) summarize the theoretical framework of the education-health gradient that education influences one’s health outcomes mainly through the causal pathways of income and access to health care, labor market, the value of future, information of cognitive skills, preferences on health, relative position or rank in society, and social networks.

While this framework has been widely adopted to research chronic and degenerative diseases, such as cardiovascular disease, diabetes, and high blood pressure, there is limited research on the effect of education on communicable diseases. On the basis of the education-health gradient, education may also have causal impacts on reducing one's risk to get infectious diseases through similar pathways. Bridging the gap between education and infectious diseases not only helps public health experts to detect populations at risk expediently but also provides evidence to invest in education to reduce health inequality. In the US and many European countries, the differences in life expectancy between people with a bachelor's degree and others have been enlarging before the pandemic (Kunst et al., 2002). This pandemic will continue to aggravate the existing gap between less-educated and well-educated (Daly, 2020).

In addition to reducing health inequality, providing better education may have beneficial effects on the population health of both well-educated and less-educated. The effect of education will lead to ripple effects on population health. One with better education can pass their health knowledge and behaviors to their families and communities (Cutler & Lleras-Muney, 2006). For the scenario of COVID-19, with fewer active cases, the spread and mutation of communicable diseases will also be slowed down. As millions of people get COVID-19, the chance to develop variants that breakthrough immunization becomes high. The resurgence of the COVID-19 pandemic caused by the Delta variant warns us that vaccination might not be a once-for-all remedy (Kupferschmidt & Wadman, 2021). To bring society back to normal, we need to find ways in addition to quarantine, social distancing, and vaccination. Education might provide protection against infectious diseases in the long term.

Research Questions:

How does educational attainment as a social determinant of health impact the morbidity of communicable diseases? Does educational attainment impact COVID-19 morbidity? If controlling other SES, will educational attainment still affect COVID-19 morbidity?

Theoretical Background

The Education-Health Gradient

The main theoretical framework of this research is the education-health gradient. Much research, at both individual and population levels, finds that people with higher education generally tend to have better health behaviors and health outcomes, such as smoking, drinking, obesity, cardiovascular diseases, blood pressure, and life expectancy (Kilander et al., 2001; Sassi et al. 2009; Conti et al., 2010; Cutler & Lleras-Muney, 2006; Cutler & Lleras-Muney, 2010). According to the education-health gradient theory, education has causal effects on health outcomes through the pathways of income and access to health care, labor market, the value of future, information of cognitive skills, preferences on health, relative position or rank in society, and social networks (Cutler & Lleras-Muney, 2006). While this theoretical framework has been widely adopted in chronic and degenerative diseases, less attention was paid to infectious diseases. Even though a limited number of researchers (Concepción-Zavaleta et al., 2020; Cordes & Castro, 2020) study the relationship between education and COVID-19 outcomes, they take education as a proxy to income, occupation, and housing, overlooking other pathways, such as behaviors, knowledge, and social networks, that education may alter COVID-19 outcomes.

People with higher educational attainment/years of schooling tend to have better jobs with higher income and health insurance, which provides better living conditions and accessibility to health care facilities (Bathmaker et al., 2016; Conti et al., 2010). A study from the UK shows that financially poor people tend to participate in jobs with limited opportunities to work from home, and they tend to live in poor housing conditions with limited personal space, which increases the risk of getting COVID-19 (Patel et al., 2020). In addition to income, certain occupations are more likely to contract to COVID-19, especially essential workers. A study based on the 2014-2017 Medical Expenditure Panel Survey (MEPS) estimates that among 112.4 million essential workers, only 31.2 million could work at home (Selden & Berdahl, 2021). Many essential workers have lower educational attainment compared to workers in other industries. According to a report based on the American Community Survey 2019 (ACS-2019) five-year estimates, 29.5% of workers in all frontline industries have a college or more advanced degree, which is lower than the proportion (34.2%) of workers with the same educational attainments in all industries (Rho et al., 2020). In all six frontline industries (Grocery, Convenience, and Drug Stores; Public Transit; Trucking, Warehouse, and Postal Service; Building Cleaning Services; Health Care; and Child Care and Social Services), only healthcare workers have higher educational attainment than the overall population (Rho et al., 2020).

In addition to occupation and income, education also influences one's behaviors. Education in early life has an important causal effect in explaining differences in many health behaviors in later life (Conti et al., 2010), such as smoking (De Walque, 2007) drinking, and exercise (Brunello et al., 2016). Education provides necessary knowledge and skills for making health decisions, such as general knowledge in disease prevention, critical thinking skills, and decision-making abilities (Cutler & Lleras-Muney, 2006). At the beginning of the COVID-19 pandemic, the federal government's inadequate, inconsistent, and largely non-evidence-based response, and social media "armchair experts" led to a tremendous amount of misinformation and even disinformation related to COVID-19, which require people to make rational choices based on their knowledge and critical thinking skills. (Jaiswal et al., 2020). In addition to knowledge, people with higher education tend to reduce risky behaviors due to their perceived value of the future and preferences on health and well-being (Cutler & Lleras-Muney, 2006). In a cross-sectional survey from Germany, less educated people were less likely to avoid gatherings, adapt their work situation, reduce personal contacts and meetings, or increase hand hygiene (Lüdecke & von dem Knesebeck, 2020).

Additionally, education plays a vital role in social networks. People tend to tie with those with similar characteristics, such as age, ethnicity, and educational attainment (McPherson et al., 2001). Previous research has shown that approximately 30% of personal networks among Americans are highly homophilous on education (Marsden, 1987). Social networks provide financial, physical, and emotional support. Peers also influence one's health behaviors through recognition or disapproval (Cutler & Lleras-Muney, 2006). More importantly, for infectious diseases, the propagation of an agent that requires serial transfer from host to susceptible through intimate personal contact often follows personal networks (Klovdahl, 1985). Previous research in contact tracing has shown consistencies in social networks and transmission patterns of a variety of infectious diseases that require close contact (Friedman, 2001). Current evidence suggests that the main transmission route of COVID-19 is by respiratory droplets among people who are in close contact with each other (WHO, 2020). Therefore, the spread of COVID-19 may also follow social networks that are influenced by education. As shown in *Figure 1*, based on the education-health gradient, education may influence COVID-19 morbidity through income, occupation, housing, knowledge and behaviors, and social networks.

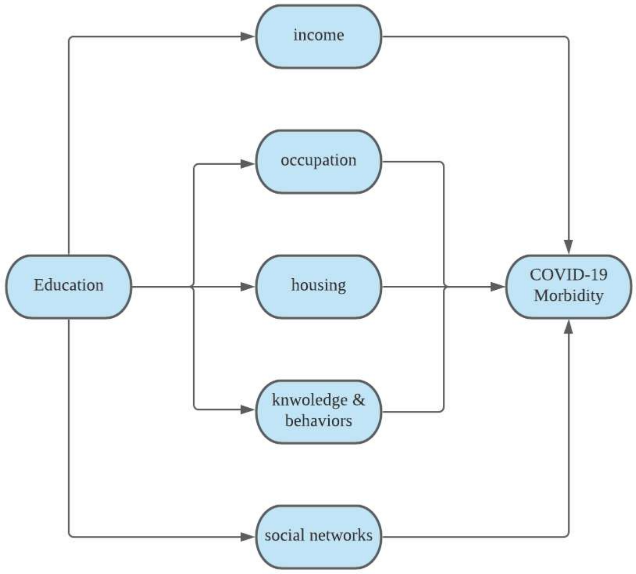


Figure 1. The Causal Pathways between Education and COVID-19 Morbidity

The Spatial Diffusion Effect of Infectious Diseases

While the theoretical framework of the education-health gradient applies to many health outcomes, the spatial diffusion effect of disease transmission is particularly important for aggregate-level data analysis based on geographical regions. COVID-19 is not directly caused by health behaviors, such as smoking and drinking, or persistent psychological burdens, such as lower social ranking and racial discrimination, as many health outcomes analyzed in the education-health gradient (Cutler & Lleras-Muney, 2006). Instead, it is caused by SARS-CoV-2, and certain SES may alter the probability of contacting that agent. Infectious diseases spread from person to person, neighborhood to neighborhood, and city to city (Emch et al., 2012). Places near one another may have similar infection rates due to close proximity and increased social and cultural ties (Arthur et al., 2017).

In addition to the nature of respiratory diseases, people who share similar SES often live closely. The process of gentrification makes the center of a city into affluent neighborhoods for gentry while the peripheral regions of the city into ghettos for lower class (Lees et al. 2008). In the United States, the legacy of retail redlining (a spatially discriminatory practice of not serving certain areas based on mortgage applicants' racial identity rather than on economic criteria) propagates disparities between race and neighborhoods (D'Rozario & Williams, 2005). Therefore, to analyze the impact of education and other SES on COVID-19 morbidity at geographically aggregated level, it is crucial to consider the spatial effect. On the one hand, both SES and COVID-19 morbidity may be highly clustered. The spatial diffusion effect of infectious diseases may underestimate the standard deviations of linear models, leading to more significant estimates of SES on COVID-19 morbidity. On the other hand, because COVID-19 is relatively new, it is likely that isolated neighborhoods tend to have lower COVID-19 morbidity compared to places with more connections at the early stage of the pandemic. Neighborhoods with lower SES may be more isolated from others. If there are fewer active cases visiting those neighborhoods, the morbidities may be lower than expected based on SES.

Data

Scope and Datasets

Education information is usually not associated with COVID-19 test records. It is costly to collect the information on COVID-19 test results and various SES variables for a large sample size. Due to the expense of collecting individual-level data, this research analyzes the relationship between educational attainment and COVID-19 morbidity at the ZIP code level. Previous research has shown that the stability of SARS-CoV-2 varies in different environmental conditions (Chin et al., 2020), and different regulation strategies result in different COVID-19 outcomes (Kaufman et al., 2020). As shown in *Figure 2*, this research focuses on the ZIP code tabulation areas (ZCTAs) in the City of Chicago to reduce the impacts of environmental factors and disease control policy on the spread of COVID-19. The shapefiles of ZCTAs were retrieved from the Chicago Data Portal (City of Chicago, 2020). While ZCTAs 60827, 60707, 60638, 60655, and 60633 are not entirely within the boundary of the City of Chicago, the COVID-19 and sociodemographic statistics include the entire areas of these ZCTAs. ZIP code level sociodemographic statistics, including population by age cohort, people with certain educational attainment, median income, working population in the service category, and average household size, were retrieved from the American Community Survey 2019 (ACS-2019) five-year estimates (Census, 2020). As shown in *Figure 2*, in the northwest, ZCTA 60666, which designates the area of O'Hare International Airport, was excluded from the research due to a lack of SES information and COVID-19 statistics.

ZIP code level COVID-19 statistics, including confirmed cases and testing volume, were retrieved from the Chicago Department of Public Health (CDPH) (City of Chicago, 2021). In Chicago, the COVID-19 vaccine phase 1A began on December 15, 2020. Therefore, this research analyzes the cumulative COVID-19 statistics released on week 51 (until December 19) to eliminate the effect of vaccination on the spread of COVID-19. According to CDPH, as of December 19, 2021, 2,288,229 COVID-19 tests had been done, and 196,192 cases had been confirmed, in which 195,667 (99.73%) cases have ZIP code information. Because one person can be tested multiple times, the overall test rate of Chicago, calculated by the total testing volume divided by the total population, was considerably high (82.78%). The overall test positivity rate and infection rate of the City of Chicago are 8.57 % and 7.10%. ZCTAs 60827 and

60707, were trimmed due to considerably low test rates (2.83% for ZCTA 60827 and 10.18% for ZCTA 60707) and high test positivity rates (28.62% for ZCTA 60827 and 38.21% for ZCTA 60707). Compared to other ZCTAs, such low test rates tend to significantly underestimate the actual morbidity. The enormously high test positivity rates may also overestimate the actual morbidity. After removing ZCTAs 60666, 60827, and 60707, 56 ZCTAs were used to analyze in this research.

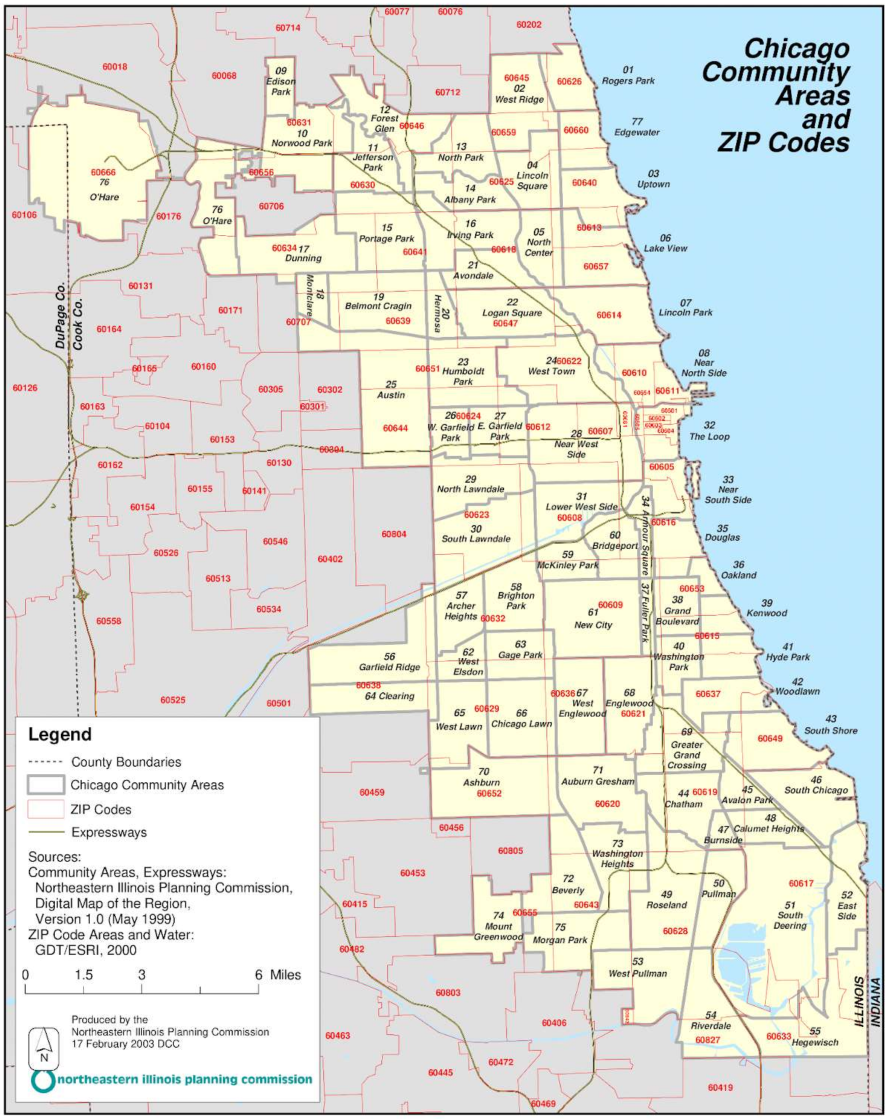


Figure 2. Chicago Community Areas and ZIP Codes (Northeastern Illinois Planning Commission, 2003)

Measurements and Variables

The response variable in this research is COVID-19 morbidity. The choice of morbidity rather than mortality was made due to the high correlation between the death of COVID-19 and age and comorbidities (Barek et al., 2020; Singh et al., 2020). Chicago is a highly socially segregated city where racial minority communities often have lower life expectancies (Hunt et al., 2015), more people with existing health problems (Hunt et al., 2014; Mayne et al., 2018; Credit, 2020), and less healthcare accessibility (Chowkwanyun & Reed, 2020; Cordes & Castro, 2020; Credit, 2020; Kang et al., 2020). Thus, analyzing mortality will introduce various confounders which makes models less interpretable. While morbidity is less affected by age, comorbidities, and healthcare accessibility, it is impossible to accurately measure the true morbidity. Places with lower test rates often have higher test positivity rates and lower infection rates (Credit, 2020). To avoid that the variation in test rates alters COVID-19 morbidity estimates, this research uses two measurements for morbidity: 1) test positivity rate, which is calculated by confirmed cases divided by testing volume; and 2) infection rate, which is calculated by confirmed cases divided by population. If the results from both test positivity rate and infection rate are consistent, then it is valid to use these two variables. The distributions of COVID-19 test positivity rate and infection rate are shown in *Figures 3a* and *3b*.

The main predictor variable in this research is education. At individual level, many studies (Conti et al., 2010; Cutler & Lleras-Muney, 2006; Cutler & Lleras-Muney, 2010) measure education by educational attainment, which is the highest education level one has reached. At the population level, because young people tend to achieve higher educational levels in later life, this research measures educational attainment by the proportion of people greater than 25 years of age who have received a bachelor's degree (including populations with the highest educational attainment of bachelor's degree, and with graduate or professional degrees; but excluding population with the educational attainment of some college or associate's degree). The distribution of educational attainment is shown in *Figure 3c*.

Education is intertwined with income. People with higher educational attainment tend to have higher incomes (Bathmaker, 2016; Conti et al., 2010). On the one hand, schooling provides necessary skills and knowledge for jobs with high income; on the other hand, people from families with higher income often receive better education (Brown et al., 2010). Cutler and Lleras-Muney (2010), in their research, found that in most cases, education has a stronger correlation to health outcomes among non-poor than among the poor, suggesting that education and income are complementary in the production of health. Therefore, it is necessary to incorporate income as a control variable. The median income in the past 12 months is used to measure income in this research. The distribution of median income is shown in *Figure 3d*.

Given that less educated people tend to have jobs with less opportunity to work at home and lack of individual housing for quarantine (Patel et al., 2020; Rho et al., 2020; Selden & Berdahl, 2021), this research also includes variables for occupational risk and household size. While it is difficult to measure the occupational risk of exposure to COVID-19 at the aggregate level, the proportion of workers in the service sector is used as a proxy for occupational risk because many service occupations work in crowded settings and involve high levels of public interaction. The proportion of workers in service occupations is calculated by the population of workers in service occupations divided by the total workers in all occupations using the data from ACS-2019. The average household size is used for measuring the overall housing density for each ZCTA. The distributions of occupation and average household size are shown in *Figures 3e* and *3f*.

For ease of interpretation, test positivity rate, infection rate, the proportion of adults with at least a bachelor's degree, and the proportion of workers participating in a service occupation, are measured per hundred people. Median income is measured per thousand US dollars. Table 1 shows the variables of interest. The descriptive statistics of variables are shown in Appendix 1. The pair plot of variables is shown in Appendix 2.

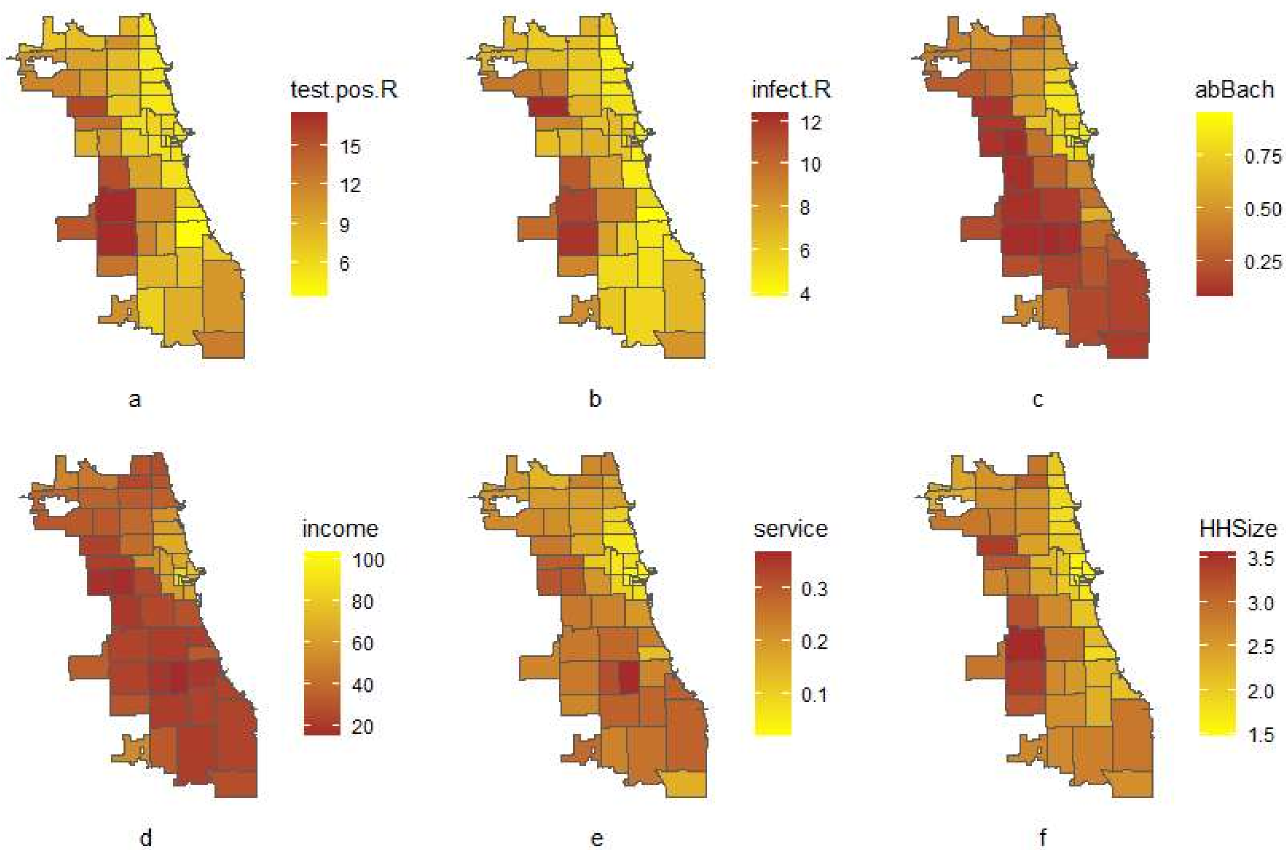


Figure 3. The Distributions of Variables of Interest

Table 1

List of Measurements and Variables

Measurement	Variable	Acronym	Letter
COVID-19 Morbidity	Test positivity rate	test.pos.R	*Y*
	Infection rate	infect.R	*Y*
Education	The proportion of people greater than 25 years old who have attained at least a bachelor's degree	abBach	*E*
Income	Median income in the past 12 months	income	*I*
Occupation	The proportion of employed who are in service occupations	service	*O*
Housing	Average household size	HHSize	*H*

Analytical Methods

Much behavioral and social science research separates the indirect impact of an independent variable through path analysis, which assumes that a variable cannot be both a cause and an effect of another variable (Streiner & Norman, 2007). However, in this research, many SES variables are both the causes and results of each other, as shown in *Figure 4*. Therefore, it is not reasonable to use path analysis for this research. Alternatively, this research uses several regression models to estimate the effect of educational attainment on COVID-19 morbidity. Regularization, including Ridge, Lasso, and Elastic-Net, were used to combat multicollinearity. Spatial regression, including spatial lag and spatial error, were used to mitigate the diffusion effect of infectious diseases. All data cleaning, manipulation, analysis, and visualization were done in R 3.6.2.

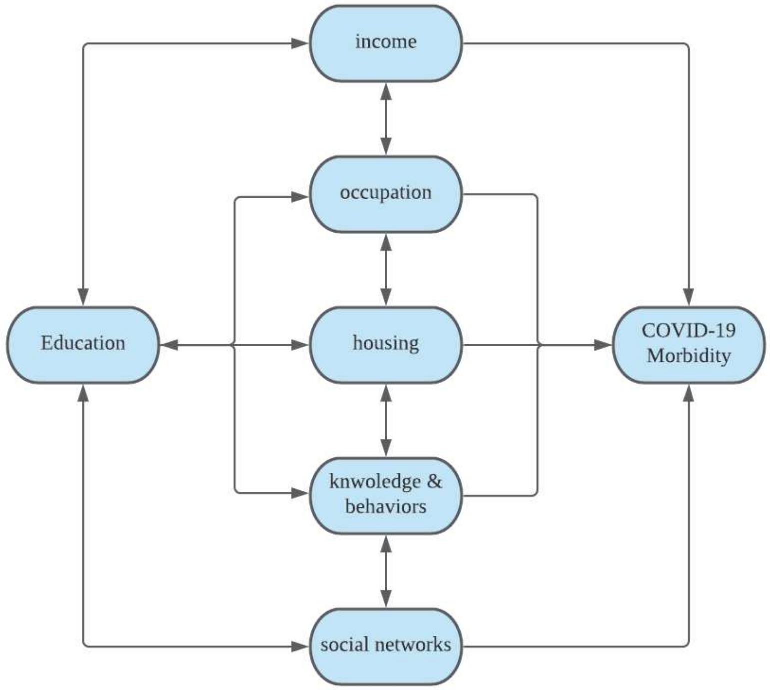


Figure 4. The Causal Relationships between Variables in this Research

Ordinary Least Squares

To assess the total effect of educational attainment on COVID-19, a single OLS regression (1) was conducted. To assess the individual effect of educational attainment on COVID-19 while controlling income, occupation, and housing, a multiple OLS regression (2) was conducted. The results from the OLS are shown in Table 3.

$$Y_i = \beta_0 + \beta_1 E_i + \epsilon_i \tag{1}$$

$$Y_i = \beta_0 + \beta_1 E_i + \beta_2 I_i + \beta_3 O_i + \beta_4 H_i + \epsilon_i \tag{2}$$

Regularization

In this research, many SES variables are often correlated. For example, those with higher educational attainment tend to get higher incomes due to the skills and knowledge gained in school. Those coming from families with higher incomes and living in wealthier neighborhoods tend to have better educational opportunities. As shown in Appendix 2, the correlations between predictor variables are considerably high. The correlation coefficient between educational attainment and median income is .910 ($p < .001$). The correlation coefficient between educational attainment and the proportion of people serving in service occupations is -.940 ($p < .001$). The correlation coefficient between educational attainment and average household size is -.855 ($p < .001$). With multicollinearity, the individual effect of each predictor variable is hard to separate, which may lead to incorrect conclusions (Schreiber-Gregory, 2018).

To assess multicollinearity, the variance inflation factor (VIF) was calculated for each predictor variable. The VIF for each variable is computed by

$$VIF(\beta_j) = \frac{1}{1-R^2_{X_j|X-j}} \tag{3}$$

where $R^2_{X_j|X-j}$ is the R^2 from a regression of X_j onto all of the other predictors. If $R^2_{X_j|X-j}$ is close to 1, then collinearity is present, leading to a large VIF. The VIF for each predictor variable is shown in Table 2. Generally, a VIF > 10 is considered as high multicollinearity. The high VIFs for the predictor variables indicate the existence of strong multicollinearity.

Table 2

The VIFs of Predictor Variables

	abBach	service	HHSize	income

VIF	17.55	4.065	9.900	6.726

To preserve the controlling variables and combat multicollinearity, this research uses regularization, including Ridge, Lasso, and Elastic-Net. The ideas of the three methods are the same. The model for regularization is the same as the OLS (2) except that regularization fits a regression model with penalties for large coefficients of predictor variables. As shown in (4), OLS minimizes the sum of squared residuals (SSR), whereas regularization (5) minimizes the sum of SSR and penalties for coefficients. Therefore, regularization shrinks coefficients toward zero, which can significantly reduce the variance and make the model more stable.

$$\text{OLS: minimize } \sum_{i=1}^{n}(Y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij})^2 \tag{4}$$

$$\text{Regularization: minimize} \sum_{i=1}^{n}(Y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij})^2 + \lambda[(\alpha \sum_{j=1}^{p}\left|\beta_j\right| - (1 - \alpha) \sum_{j=1}^{p}\beta_j^2] \tag{5}$$

where $\lambda \geq 0$ and $\alpha \geq 0$ are tuning parameters, which control the relative impact of the penalty of coefficients. When $\alpha = 0$, the model becomes Ridge (6); when $\alpha = 1$, the model becomes Lasso (7); when λ and α are not constrained, the model becomes Elastic-Net (8).

$$\text{Ridge: minimize } \sum_{i=1}^{n}(Y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij})^2 + (constant) \sum_{j=1}^{p}\beta_j^2 \tag{6}$$

$$\text{Lasso: minimize} \sum_{i=1}^{n}(Y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij})^2 + (constant) \sum_{j=1}^{p}\left|\beta_j\right| \tag{7}$$

Elastic-Net: minimize (8)

$$\sum_{i=1}^{n}(Y_i - \beta_0 - \sum_{j=1}^{p}\beta_j X_{ij})^2 + (constant)\sum_{j=1}^{p}\left|\beta_j\right| + (constant)\sum_{j=1}^{p}\beta_j^2]$$

To pick optimal λ and α, 10-fold cross-validation was conducted. λ was selected from a geometric sequence from 10^3 to 10^3 with a length of 100. α was selected from 0.1 to 0.9 with an increment of 0.1. Root mean squared error (RMSE) is used to select best-tuned models. The tuning parameters and results for the best-tuned models are shown in Table 4.

Spatial Regression

Another important assumption of linear regression is that the error terms, ε_i, are not correlated. However, due to the spatial diffusion effect of infectious disease and that people with similar SES tend to live together, the error terms in this research may be spatially correlated. In such a case, the estimates of variances from OLS tend to be undervalued, leading to more significant estimates for coefficients. As shown in *Figure 5*, the residuals in the west part of Chicago tend to be higher than in other areas, suggesting that the residuals of the non-spatial models may have spatial autocorrelation.

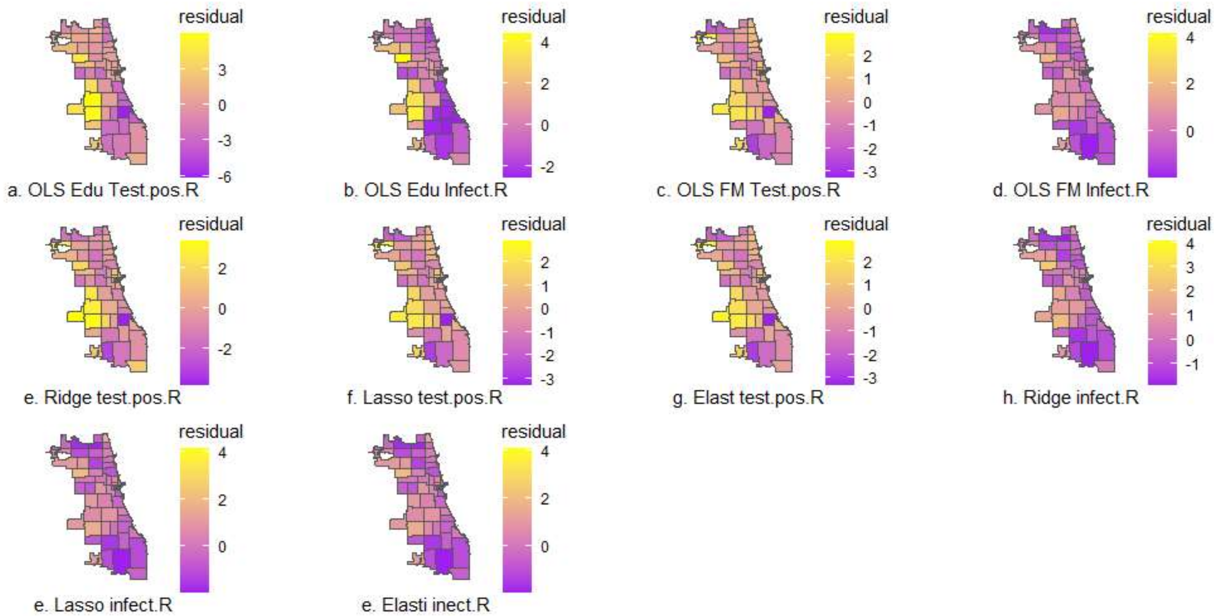


Figure 5. The Spatial Distribution of Residuals for Non-Spatial Models

To access the global spatial autocorrelation, Moran's I (Moran, 1948) for each variable and the residuals for each model was calculated by (9). Moran's I ranges from -1 to 1. -1 indicates complete spatial dispersion; 0 indicates spatial randomness; 1 indicates complete spatial cluster (see Appendix 3).

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(Y_i - \bar{Y})(Y_j - \bar{Y})}{S^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \tag{9}$$

where S^2 is the sample variance

$$S^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})}{n} \tag{10}$$

and w_{ij} is the weight between ZCTA i and its neighboring ZCTA j. In this research, the weight was calculated by the first-order Queen's contiguity (as shown in Appendix 4). The Moran's I for the response variables, predictor variables, and residuals are shown in Tables 5a and 5b. Analytical and Monte-Carlo (999 simulations) p values were calculated to assess the significance of Moran's I.

Spatial lag model and spatial error model were used to combat the spatial diffusion effect of infectious diseases. As shown in (11) and (12), in the right hand side, there is a spatial lag in addition to the predictor variables. The spatial lag takes the weighted average of the neighboring values of the response variable. In this research, the spatial lag model means that in addition to education and other SES variables, the COVID-19 morbidity of surrounding ZCTAs also contributes to the morbidity of targeted ZCTA due to the spread of COVID-19. As shown in (13) and (14), the error term, ε_i, is dependent on the neighboring error terms, which means that the effects of SES variables on COVID-19 morbidity are different across space. The results from spatial lag model and spatial error model are shown in Tables 6a and 6b.

Spatial Lag:

$$Y_i = \beta_0 + \beta_1 E_i + \rho \sum_{j \in N(i)} w(i,j) Y_j + \epsilon_i \tag{11}$$

$$Y_i = \beta_0 + \beta_1 E_i + \beta_2 I_i + \beta_3 O_i + \beta_4 H_i + \rho \sum_{j \in N(i)} w(i,j) Y_j + \epsilon_i \tag{12}$$

Spatial Error:

$$Y_i = \beta_0 + \beta_1 E_i + \lambda \sum_{j \in N(i)} w(i,j)\epsilon_i \tag{13}$$

$$Y_i = \beta_0 + \beta_1 E_i + \beta_2 I_i + \beta_3 O_i + \beta_4 H_i + \lambda \sum_{j \in N(i)} w(i,j)\epsilon_i \tag{14}$$

where ρ is the coefficient for spatial lag, and λ is the coefficient for spatial error

Results

Table 3

Results from OLS

Response Variable	Predictor Variable	Coefficient	*p* value	R^2	AIC
test.pos.R	(Intercept)	8.531	< .001	.5491	262.8
	abBach	-0.09	< .001		
infect.R	(Intercept)	8.531	< .001	.2848	225.98
	abBach	-0.039	< .001		
test.pos.R	(Intercept)	3.273	.3258	.8605	203.09
	abBach	-0.137	< .001		
	income	0.0851	< .001		
	service	-0.1913	.004		
	HHSize	4.711	< .001		
infect.R	(Intercept)	2.011	.449	.7275	177.95
	abBach	-0.0657	.005		
	income	0.0555	.0022		
	service	-0.1319	.0129		
	HHSize	3.263	< .001		

Table 4

Results from Regularization

			Test Positivity rate			
Model	λ	α	Predictor	Coefficient	RMSE	R^2
Ridge	0	0.2656	(Intercept)	-0.9153	1.569	.8752
			abBach	-0.0601		
			income	0.0429		
			service	-0.0587		
			HHSize	4.712		
Lasso	1	0.02848	(Intercept)	1.277	1.425	.8594
			abBach	-0.1114		
			income	0.0744		
			service	-0.1458		
			HHSize	4.890		
Elastic-Net	0.1	0.0411	(Intercept)	1.289	1.399	.8708
			abBach	-0.1113		
			income	0.0741		
			service	-0.1484		
			HHSize	4.910		

Infection rate						
Model	λ	α	Predictor	Coefficient	RMSE	R^2
Ridge	0	0.1322	(Intercept)	0.3155	1.128	.7059
			abBach	-0.0293		
			income	0.0347		
			service	-0.0618		
			HHSize	3.104		
Lasso	1	0.00705	(Intercept)	1.493	1.106	.6981
			abBach	-0.0592		
			income	0.0529		
			service	-0.1201		
			HHSize	3.308		
Elastic-Net	0.1	0.00869	(Intercept)	1.571	1.084	.7341
			abBach	-0.060		
			income	0.0532		
			service	-0.1218		
			HHSize	3.300		

Table 5a

The Moran's I for Response and Predictor Variables

Variable	Moran's I	Analytical *p* value	Monte-Carlo *p* value
test.pos.R	.5395	< .001	.001
infect.R	.4145	< .001	.001
abBach	.8134	< .001	.001
income	.8142	< .001	.001
service	.7643	< .001	.001
HHSize	.6461	< .001	.001

Table 5b

The Moran's I the Residuals for Each Model

Model	Moran's I	Analytical *p* value	Monte-Carlo *p* value
	Test Positivity Rate		
Single OLS	.3487	< .001	.001
Multiple OLS	-.0058	.442	.434
Ridge	.1670	.0146	.025
Lasso	.0118	.3621	.347
Elastic-Net	.0256	.3033	.291
	Infection Rate		
Single OLS	.3532	< .001	.001
Multiple OLS	.1083	.0638	.074
Ridge	.3532	< .001	.001
Lasso	.1161	.0531	.068
Elastic-Net	.1151	.0544	.079

Table 6a

Results from Spatial Lag Models

	Test positivity rate		
Predictor Variable	Coefficient	*p* value	AIC
(spatial weight)	0.4113	.009	
(Intercept)	7.879	< .001	257.98
abBach	-0.0651	< .001	
(spatial weight)	0.1122	.3239	
(Intercept)	1.799	.5978	
abBach	-0.1246	< .001	
income	0.0834	< .001	204.11
service	-0.1737	.005	
HHSize	4.586	< .001	

	Infection rate		
Predictor Variable	Coefficient	*p* value	AIC
(spatial weight)	0.5201	.0013	
(Intercept)	4.379	< .001	217.62
abBach	-0.0254	.0044	
(spatial weight)	0.2384	.0718	
(Intercept)	0.1460	.9556	
abBach	-0.0541	.0122	
income	0.0523	.0011	176.7
service	-0.1093	.0226	
HHSize	3.028	< .001	

Table 6b
Results from Spatial Error Models

Test positivity rate

Predictor Variable	Coefficient	*p* value	AIC
(spatial weight)	0.5949	< .001	
(Intercept)	13.378	< .001	251.19
abBach	-0.1091	< .001	
(spatial weight)	-0.0157	.9432	
(Intercept)	3.270	.2979	
abBach	-0.1374	< .001	
income	0.0852	< .001	205.08
service	-0.1925	.0016	
HHSize	4.724	< .001	

	Infection rate		
Predictor Variable	Coefficient	*p* value	AIC
(spatial weight)	0.5891	< .001	
(Intercept)	9.004	< .001	214.21
abBach	-0.0494	< .001	
(spatial weight)	0.2811	.1823	
(Intercept)	2.682	.2932	
abBach	-0.0710	.0020	
income	0.0582	< .001	178.17
service	-0.1222	0.0124	
HHSize	2.969	< .001	

Discussion

Based on the results from OLS (Table 3), regularization (Table 4), and spatial regression (Tables 6a and 6b), the proportion of people over 25 years old with at least a bachelor always has negative and significant ($\alpha = .05$) coefficients to both COVID-19 test positivity rate and infection rate. This suggests that at ZIP code level, a higher overall educational attainment leads to lower COVID-19 morbidity, which corresponds to previous research in the education-health gradient that people with higher educational attainment tend to have better health outcomes. Even after controlling for median income, the proportion of working population in service occupations, and average household size, the associations between educational attainment and COVID-19 morbidity remain negative and significant. This suggests that in addition to the material benefits, such as higher income, occupation with less contact to people, and better living environment, education may lower the risk of getting COVID-19 by providing the necessary knowledge for disease prevention, cultivating the critical thinking skills, alternating the value for health, and building social networks with lower concentrations of COVID-19 active cases.

While the relationship between educational and COVID-19 morbidity remains negative for all models, the directions of the effects of income and occupation in full models are different from those in the pair plot (see Appendix 2). As shown in Appendix 2, median income is negatively associated with both COVID-19 test positivity rate ($\rho = -.524$, $p < .001$) and infection rate ($\rho = -.310$, $p < .05$); and the proportion of working population working as service occupation is negatively associated with both COVID-19 test positivity rate ($\rho = .584$, $p < .001$) and infection rate ($\rho = .367$, $p < .001$). These findings correspond to previous individual-level research that financially poor people and essential workers have higher risk of exposure to COVID-19 (Patel et al., 2020). However, as shown in Tables 3, 4, 6a, and 6b, after controlling for other SES, median income is always positively associated with COVID-19 morbidity; and the proportion of people working as service occupations is negatively associated with COVID-19 morbidity. The change in the directions of coefficients is potentially caused by multicollinearity. While regularization methods tend to mitigate this effect, the directions of coefficients from the three regularization methods (See Table 4) are the same as the OLS (see Table 4). Further research with larger sample sizes may reduce the issue of multicollinearity.

To analyze the change of the direction of the coefficients, 18 reduced OLS models containing either income or occupation were performed (see Appendix 5 for results from reduced models). Among the 18 reduced models, educational attainment is always negatively associated with COVID-19 morbidity. However, income is always positively associated with both COVID-19 test positivity rate and infection rate. It is likely that when controlling for other SES variables, places with higher median income tend to have higher living expenses. Plus, as shown in *Figure 2*, the ZCTAs with high median incomes are located in downtown Chicago which have higher attractiveness, more public transportation, and higher population density compared to other regions. Therefore, after controlling for other SES variables, communities with higher median income tend to have higher COVID-19 morbidity. Moreover, while much research has shown that income has beneficial effects on various health outcomes (Cutler & Lleras-Muney, 2006; Cutler & Lleras-Muney, 2010; Patel, 2020), the accessibility to material goods might not be essential to COVID-19 morbidity. The inaccessibility of healthcare and pre-existing health conditions related to lower income cause the disparity in COVID-19 mortality rather than morbidity.

Among the 12 reduced models that contain occupation, 10 models have negative associations between the proportion of the working population in service occupation and

COVID-19 morbidity. Only 2 models have positive coefficients between occupation and COVID-19 morbidity. However, the coefficients are not significant and the R^2s for the models are low. It is likely that not all service occupations work in crowded settings and have high levels of public interaction. In comparison, several occupations that are not listed as service in ACS-2019 may have relatively high risks to COVID-19 exposure, such as bus drivers, cashiers, hotel, motel, and resort desk clerks, and butchers and other meat, poultry, and fish processing (IPUMS, 2018). It is likely that the occupations with lower requirements of educational attainment are prone to higher occupational risks of COVID-19 even if these occupations are not categorized as service. Additionally, because many non-essential businesses stopped in-person service during the pandemic, many service workers may be unemployed. According to the U.S. Bureau of Labor Statistics (2021), in Chicago-Naperville-Arlington Heights Metropolitan Division, unemployment surged from 144554 (3.9%) in March 2020 to 569547 (16.4%) in April 2020, and the unemployment rate remains relatively high (8.7% in December 2020 and 8.1% in July 2021). Therefore, the measurement of occupation in this research is not as reliable as educational attainment.

As shown in Tables 5a and 5b, while COVID-19 morbidity tends to be clustered globally, the residuals of the full models do not show significant global clustered patterns. Only single OLS and Ridge models have significant spatial autocorrelation. Consistently, as shown in Tables 6a and 6b, the coefficients of spatial lag and spatial errors for full models are not significant. This is probably caused by the high spatial autocorrelations of and correlations between predictor variables. As shown in Table 5a, while all predictor and response variables are significantly spatially correlated, the Moran's Is for educational attainment (I = .8134), income (I = .8142), occupation (I = .7643), and household size (I = .6461) are greater than test positivity rate (I =

.5395) and infection rate (I = .4145). It is likely that the significant spatial lag and spatial error coefficients for single variable regression are mainly caused by omission of response variables that are correlated together. For the Ridge models and other regularizations, the penalty for coefficients mitigates the effect of multicollinearity. With shrunk coefficients for predictor variables, the residuals tend to be larger, leading to more significant spatial autocorrelation of residuals. Based on the insignificant spatial coefficients, the spatial heterogeneity of COVID-19 morbidity is majorly caused by inequality in SES.

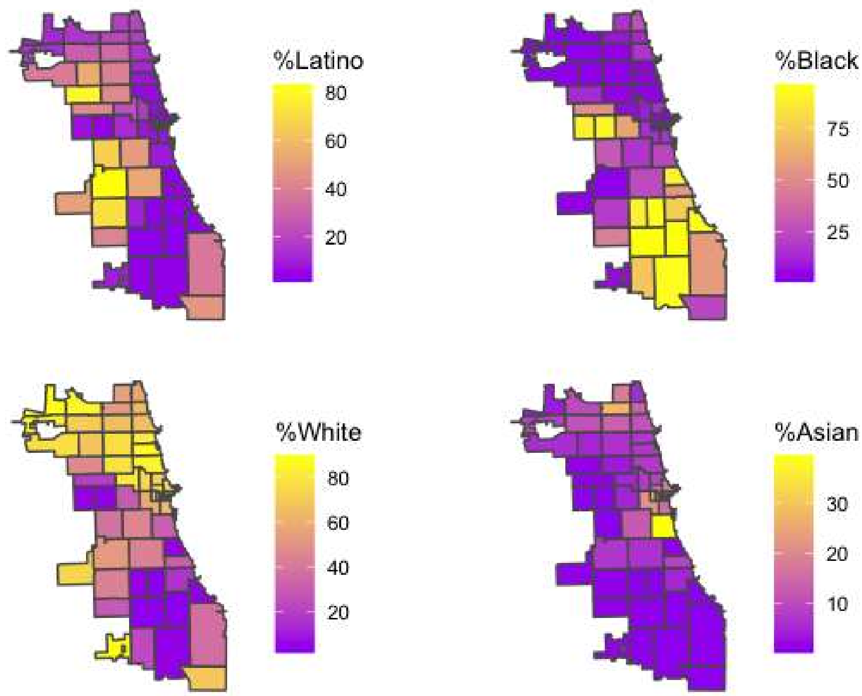


Figure 6. The Distribution of Race/Ethnicity in Chicago

Even though there are no significant global spatial patterns of the residuals from the full models, as shown in *Figure 5*, the west part of Chicago still tends to have higher observed COVID-19 morbidity than expected based on education, income, occupation, and household size. One possible cause of the higher residuals in those ZCTAs is the edge effect that the COVID-19 morbidity is high in the neighboring regions in the west that are not included in this research. Another explanation for the anomaly is that race/ethnicity plays a role in COVID-19 disparity. As shown in *Figure 6*, many Hispanic/Latino live in the west part of Chicago, where both COVID-19 morbidity (see *Figure 2*) and residuals (see *Figure 6*) are high. This corresponds with individual-level COVID-19 statistics by race and ethnicity. CDPH (2021) and CDC (2020) also show that Latino/Hispanic have significantly higher COVID-19 morbidity and mortality than non-Latino/Hispanic. This suggests that 1) in addition to education, income, occupation, and household size, other variables contribute to the high morbidity rate of COVID-19 in Hispanic/Latino communities; 2) the effect of education and other SES factors are different across racial/ethnic groups, suggesting interactions between SES and race/ethnicity. It is likely that many Latino/Hispanic workers, even with similar educational attainment and income as non-Hispanic/Latino counterparts, have a higher opportunity to interact with the public, which increases their occupational risk of getting COVID-19.

Conclusion

People with higher educational attainment tend to have better health outcomes. This also applies to the ongoing COVID-19 pandemic. Neighborhoods with higher overall educational attainment tend to have higher median income, smaller proportion of people working in service occupations, and less crowded housing, which lower their risk to get COVID-19. In addition to the material aspects, education may lower COVID-19 morbidity through the pathways of knowledge, behaviors, and social networks. While COVID-19 spreads spatially, the spatial heterogeneity of COVID-19 morbidity can be mostly explained by inequality in SES. Additionally, race/ethnicity may play an important role in the social inequality of COVID-19. Latino/Hispanic communities tend to have higher COVID-19 morbidity even after controlling for education, income, occupation, and household size. As SARS-CoV-2 keeps mutating, the efficacy of vaccines decreases, which may lead to wider gaps in COVID-19 morbidity and mortality between well-educated and less-educated. For policymakers and public health experts, in addition to quarantine, social distancing, and vaccination, providing more accessible education, especially health education, may reduce the morbidity and mortality of COVID-19 and other infectious diseases in the long term. Since this is an ecological study using cross-sectional data, the results from this research may suffer from ecological fallacy (incorrect assumption about an individual based on aggregate data for a group). Further research analyzing individual-level and longitudinal data is required to study the causal pathways of education on COVID-19 and other infectious disease outcomes.

References

Arthur, R. F., Gurley, E. S., Salje, H., Bloomfield, L. S. P., & Jones, J. H. (2017). Contact structure, mobility, environmental impact and behaviour: The importance of social forces to infectious disease dynamics and disease ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1719), 20160454. https://doi.org/10.1098/rstb.2016.0454

Bathmaker, A.-M., Ingram, N., Abrahams, J., Hoare, A., Waller, R., & Bradley, H. (2016). *Higher Education, Social Class and Social Mobility: The Degree Generation*. Springer.

Brown, P., Lauder, H., & Ashton, D. (2010). *The Global Auction: The Broken Promises of Education, Jobs, and Incomes*. Oxford University Press.

Brunello, G., Fort, M., Schneeweis, N., & Winter-Ebmer, R. (2016). The Causal Effect of Education on Health: What is the Role of Health Behaviors? *Health Economics*, *25*(3), 314–336. https://doi.org/10.1002/hec.3141

Bureau of Labor Statistics Data. (n.d.). Retrieved October 13, 2021, from https://data.bls.gov/timeseries/LAUDV171697400000004?amp%253bdata_tool=XGtable&output_view=data&include_graphs=true

CDC (2020). *COVID Data Tracker*. Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker

Chin, A. W. H., Chu, J. T. S., Perera, M. R. A., Hui, K. P. Y., Yen, H.-L., Chan, M. C. W., Peiris, M., & Poon, L. L. M. (2020). Stability of SARS-CoV-2 in different environmental conditions. *MedRxiv*, 2020.03.15.20036673. https://doi.org/10.1101/2020.03.15.20036673

Chowkwanyun, M., & Reed, A. L. (2020). Racial Health Disparities and Covid-19—Caution and Context. *New England Journal of Medicine*, *383*(3), 201–203. https://doi.org/10.1056/NEJMp2012910

Conti, G., Heckman, J., & Urzua, S. (2010). THE EDUCATION-HEALTH GRADIENT. *The American Economic Review*, *100*(2), 234–238. https://doi.org/10.1257/aer.100.2.234

Cordes, J., & Castro, M. C. (2020). Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spatial and Spatio-Temporal Epidemiology*, *34*, 100355. https://doi.org/10.1016/j.sste.2020.100355

City of Chicago (2020). *Boundaries—ZIP Codes*. Retrieved August 21, 2021, from https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-ZIP-Codes/gdcf-axmw

City of Chicago (2021). *COVID-19 Cases, Tests, and Deaths by ZIP Code*. Retrieved August 25, 2021, from https://data.cityofchicago.org/Health-Human-Services/COVID-19-Cases-Tests-and-Deaths-by-ZIP-Code/yhhz-zm2v

Credit, K. (2020). Neighbourhood inequity: Exploring the factors underlying racial and ethnic disparities in COVID-19 testing and infection rates using ZIP code data in Chicago and New York. *Regional Science Policy & Practice*, *12*(6), 1249–1271. https://doi.org/10.1111/rsp3.12321

Cutler, D. M., & Lleras-Muney, A. (2006). *Education and Health: Evaluating Theories and Evidence* (No. w12352). National Bureau of Economic Research. https://doi.org/10.3386/w12352

Cutler, D. M., & Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics*, *29*(1), 1–28. https://doi.org/10.1016/j.jhealeco.2009.10.003

Daly, M. C. (2020). *The Unequal Impact of COVID-19: Why Education Matters*. 5.

de Walque, D. (2007). Does education affect smoking behaviors?: Evidence using the Vietnam draft as an instrument for college education. *Journal of Health Economics*, *26*(5), 877–895. https://doi.org/10.1016/j.jhealeco.2006.12.005

Devakumar, D., Shannon, G., Bhopal, S. S., & Abubakar, I. (2020). Racism and discrimination in COVID-19 responses. *The Lancet*, *395*(10231), 1194. https://doi.org/10.1016/S0140-6736(20)30792-3

D'Rozario, D., & Williams, J. D. (2005). Retail Redlining: Definition, Theory, Typology, and Measurement. *Journal of Macromarketing*, *25*(2), 175–186. https://doi.org/10.1177/0276146705280632

Emch, M., Root, E. D., Giebultowicz, S., Ali, M., Perez-Heydrich, C., & Yunus, M. (2012). Integration of Spatial and Social Network Analysis in Disease Transmission Studies. *Annals of the Association of American Geographers*, *102*(5), 1004–1015. https://doi.org/10.1080/00045608.2012.671129

Hawkins, R. B., Charles, E. J., & Mehaffey, J. H. (2020). Socio-economic status and COVID-19-related cases and fatalities. *Public Health*, *189*, 129–134. https://doi.org/10.1016/j.puhe.2020.09.016

Hunt, B. R., Deot, D., & Whitman, S. (2014). Stroke Mortality Rates Vary in Local Communities in a Metropolitan Area. *Stroke*, *45*(7), 2059–2065. https://doi.org/10.1161/STROKEAHA.114.005431

Hunt, B. R., Tran, G., & Whitman, S. (2015). Life Expectancy Varies in Local Communities in Chicago: Racial and Spatial Disparities and Correlates. *Journal of Racial and Ethnic Health Disparities*, *2*(4), 425–433. https://doi.org/10.1007/s40615-015-0089-8

IPUMS (2018). *Occupation Codes for 2018-onward*. Retrieved October 14, 2021, from https://usa.ipums.org/usa/volii/occ2018.shtml

Jaiswal, J., LoSchiavo, C., & Perlman, D. C. (2020). Disinformation, Misinformation and Inequality-Driven Mistrust in the Time of COVID-19: Lessons Unlearned from AIDS Denialism. *AIDS and Behavior*, *24*(10), 2776–2780. https://doi.org/10.1007/s10461-020-02925-y

Kang, J.-Y., Michels, A., Lyu, F., Wang, S., Agbodo, N., Freeman, V. L., & Wang, S. (2020). Rapidly measuring spatial accessibility of COVID-19 healthcare resources: A case study of Illinois, USA. *International Journal of Health Geographics*, *19*(1), 36. https://doi.org/10.1186/s12942-020-00229-x

Kaufman, B. G., Whitaker, R., Mahendraratnam, N., Smith, V. A., & McClellan, M. B. (2020). Comparing Associations of State Reopening Strategies with COVID-19 Burden. *Journal of General Internal Medicine*, *35*(12), 3627–3634. https://doi.org/10.1007/s11606-020-06277-0

Kilander, L., Berglund, L., Boberg, M., Vessby, B., & Lithell, H. (2001). Education, lifestyle factors and mortality from cardiovascular disease and cancer. A 25-year follow-up of Swedish 50-year-old men. *International Journal of Epidemiology*, *30*(5), 1119–1126. https://doi.org/10.1093/ije/30.5.1119

Kupferschmidt, K., & Wadman, M. (2021). Delta variant triggers new phase in the pandemic. *Science*, *372*(6549), 1375–1376. https://doi.org/10.1126/science.372.6549.1375

Loretta Lees, Tom Slater, & Elvin Wyly. (2008). *Gentrification*. Routledge. http://proxy.uchicago.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=e000xna&AN=653933&site=ehost-live&scope=site

Lüdecke, D., & von dem Knesebeck, O. (2020). Protective Behavior in Course of the COVID-19 Outbreak—Survey Results From Germany. *Frontiers in Public Health*, *8*. https://doi.org/10.3389/fpubh.2020.572561

Mayne, S. L., Yellayi, D., Pool, L. R., Grobman, W. A., & Kershaw, K. N. (2018). Racial Residential Segregation and Hypertensive Disorder of Pregnancy Among Women in Chicago: Analysis of Electronic Health Record Data. *American Journal of Hypertension*, *31*(11), 1221–1227. https://doi.org/10.1093/ajh/hpy112

Moran, P. A. P. (1948). The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, *10*(2), 243–251.

Mutambudzi, M., Niedwiedz, C., Macdonald, E. B., Leyland, A., Mair, F., Anderson, J., Celis-Morales, C., Cleland, J., Forbes, J., Gill, J., Hastie, C., Ho, F., Jani, B., Mackay, D. F., Nicholl, B., O'Donnell, C., Sattar, N., Welsh, P., Pell, J. P., … Demou, E. (2020). Occupation and risk of severe COVID-19: Prospective cohort study of 120 075 UK Biobank participants. *Occupational and Environmental Medicine*. https://doi.org/10.1136/oemed-2020-106731

Northeastern Illinois Planning Commission (2003). *Chicago Community Area and Zipcode Map*. City of Chicago, April 4, 2020.

https://www.chicago.gov/content/dam/city/sites/covid/reports/2020-04-24/ChicagoCommunityAreaandZipcodeMap.pdf

Patel, J. A., Nielsen, F. B. H., Badiani, A. A., Assi, S., Unadkat, V. A., Patel, B., Ravindrane, R., & Wardle, H. (2020). Poverty, inequality and COVID-19: The forgotten vulnerable. *Public Health*, *183*, 110–111. https://doi.org/10.1016/j.puhe.2020.05.006

Rho, H. J., Fremstad, S., & Brown, H. (n.d.). *A Basic Demographic Profile of Workers in Frontline Industries*. 10.

Sassi, F., Devaux, M., Church, J., Cecchini, M., & Borgonovi, F. (2009). *Education and Obesity in Four OECD Countries* [Working Paper]. OECD. https://repositorio.minedu.gob.pe/handle/20.500.12799/2530

Schreiber-Gregory, D. (2018, January 1). *Regulation Techniques for Multicollinearity: Lasso, Ridge, and Elastic Nets*.

Selden, T. M., & Berdahl, T. A. (2021). Risk of Severe COVID-19 Among Workers and Their Household Members. *JAMA Internal Medicine*, *181*(1), 120. https://doi.org/10.1001/jamainternmed.2020.6249

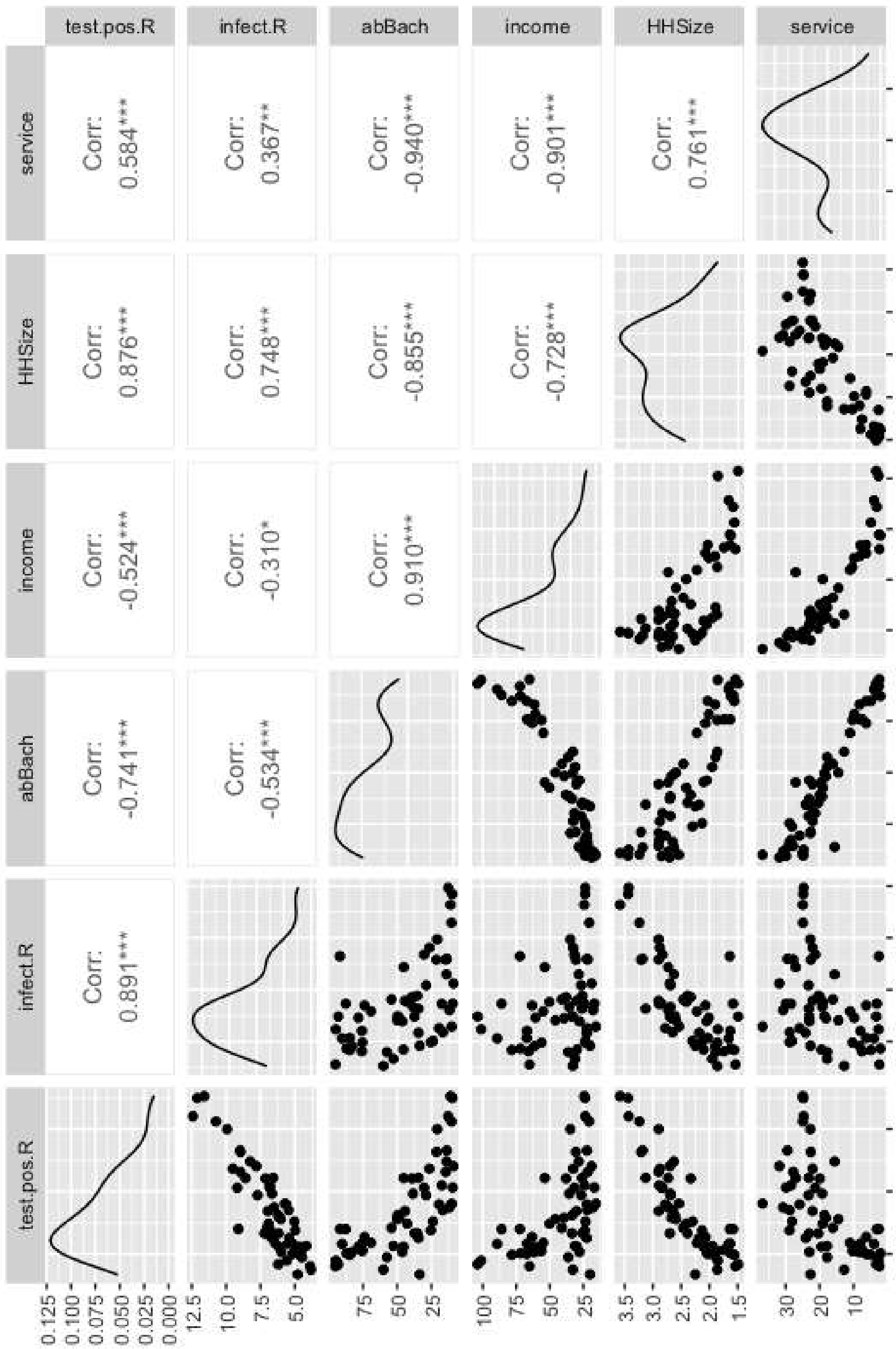
WHO (2021). *Coronavirus (COVID-19) Dashboard*. Retrieved August 6, 2021, from https://covid19.who.int

APPENDICES

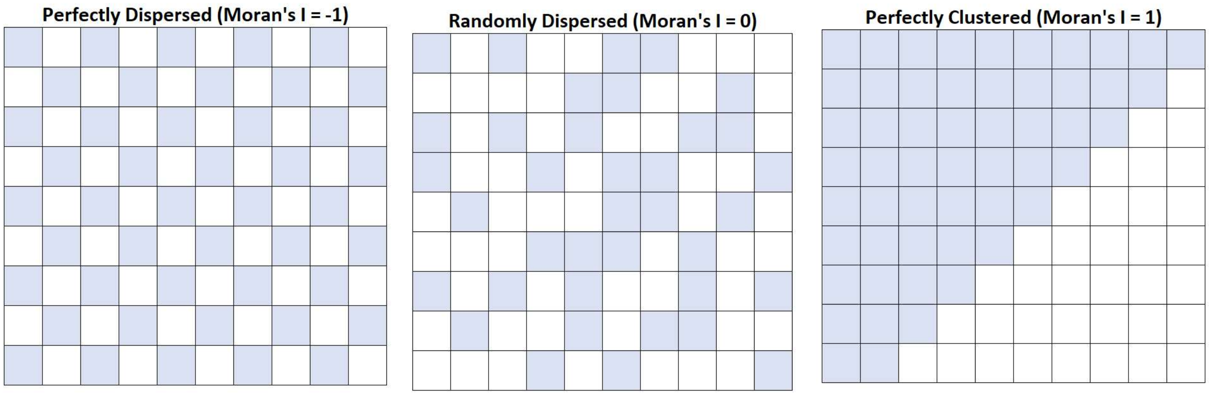
Appendix 1. The Descriptive Statistics of Variables

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
test.pos.R (%)	3.359	5.561	7.414	8.368	10.479	17.622
infect.R (%)	3.829	5.170	6.227	6.749	7.714	12.417
abBach (%)	8.87	20.40	40.90	45.47	74.28	95.19
Income (thousand dollars)	15.67	23.79	32.63	41.44	57.34	103.77
HHSize	1.490	1.938	2.435	2.407	2.792	3.580
service (%)	2.021	10.171	20.121	18.288	24.846	36.821

Appendix 2. The Pair Plot of Variables



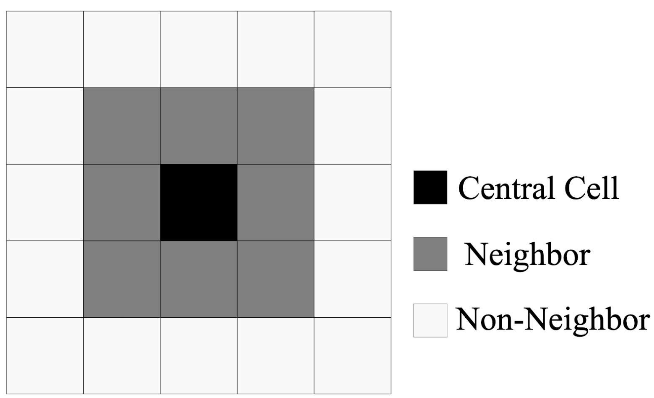
Appendix 3. Examples of Spatial Pattern



Source: What is Moran's I? (Definition & Example). (2021, January 28). *Statology*.

https://www.statology.org/morans-i/

Appendix 4. First-Order Queen's Contiguity



Appendix 5. Results from Reduced Models

Test Positivity Rate

Predictor Variable	Coefficient	p value	R^2
(Intercept)	11.577	< .001	
abBach	-0.1972	< .001	.682
income	0.1397	< .001	
(Intercept)	4.086	.243	
income	0.2315	.025	.3407
service	0.0019	.963	
(Intercept)	-9.710	< .001	
income	0.0384	.0101	.7948
HHSize	6.864	< .001	
(Intercept)	24.871	< .001	
abBach	-0.2108	< .001	.6582
service	-0.3765	< .001	
(Intercept)	-6.303	< .001	
service	-0.076	.0514	.7835
HHSize	6.690	< .001	

(Intercept)	-4.226	.0716	
abBach	-0.0854	< .001	
income	0.1033	< .001	.8362
HHSize	5.081	< .001	
(Intercept)	20.546	< .001	
abBach	-0.2589	< .001	
income	0.1102	< .001	.7323
service	-0.2702	< .001	
(Intercept)	-9.785	< .001	
income	0.0394	.0966	
service	0.0034	.9552	.7948
HHSize	6.852	< .001	
(Intercept)	4.812	.2003	
abBach	-0.0892	.0029	
service	-0.2632	< .001	.8177
HHSize	5.176	< .001	

	Infection Rate		
Predictor Variable	Coefficient	*p* value	R^2
(Intercept)	7.778	< .001	
abBach	-0.1074	< .001	.4659
income	0.0932	< .001	
(Intercept)	4.435	.0554	
income	0.0101	.7057	.137
service	0.1041	.1185	
(Intercept)	5.086	< .001	
income	0.0452	< .001	.6764
HHSize	4.143	< .001	
(Intercept)	16.829	< .001	
abBach	-0.1182	< .001	.4402
service	-0.2568	< .001	
(Intercept)	-1.275	.1242	
service	-0.1067	< .001	.6562
HHSize	4.149	< .001	
(Intercept)	-3.160	.0852	

abBach	-0.0300	.1104	
income	0.0680	< .001	.692
HHSize	3.517	< .001	
(Intercept)	13.970	< .001	
abBach	-0.1500	< .001	
income	0.0729	.0015	.5392
service	-0.1865	.0058	
(Intercept)	-4.244	.0139	
income	0.0336	.0479	
service	-0.0386	.3732	.6813
HHSize	4.287	< .001	
(Intercept)	3.014	.2944	
abBach	-0.0344	.1221	
service	-0.1788	.0015	.6718
HHSize	3.565	< .001	