COVID-19: A Midwestern Reporting Project of WNIN and the University of Evansville

Chicago Service Worker Analysis

10/22/2020

Introduction

We answer two questions in this report: who is the Chicago service worker? and what demographics explain the COVID case rate in Chicago ZIP codes. We look at ZIP code level data in order to understand what the service worker looks like in Chicago. We also examine the available COVID data on a ZIP code level and run a statistical linear model to find the demographics that explain the variation we see from the COVID case rate. These demographics do not necessarily drive the COVID case rate, but they do help explain the differences we are seeing in the case rate from ZIP code to ZIP code.

Data Obtained

We obtained Chicago's most recent COVID-19 numbers of individuals tested and those who had confirmed cases on October 1 from the Illinois Department of Public Health. We paired Chicago's COVID-19 numbers with the 5-year American Community Survey (ACS) data ending in 2018. The Census Bureau provides two options: the most recent 1-year data or the more precise 5-year estimates. The 5-year data is from a larger sample, which has less error attached to estimates as well as a higher statistical reliability. However, it is not as current as the 1-year data, because the most recent 5-year data from 2018 had been collected from January 2014 to December 2018. The 2018 estimates were released in December 2019, and the next collection of 5-year data ending in 2019 will be released December 2020. In addition, the 1-year data is not available for smaller geographic areas such as ZIP codes. Thus, we are only left with the 5-year option.

Demographics

The tables we used are as follows:

1. **B05003** - Sex by Age and Citizenship Status, which is broken down into 2 categories for sex, 2 categories for age, 2 categories for nativity, and 2 categories for citizenship.
2. **B01001I** - Sex by Age for the Hispanic and Latino population, which is broken down into 2 categories for sex and 14 categories for age.
3. **B27001** - Health Insurance Coverage Status by Sex by Age, which is broken down into 2 categories for sex, 9 categories for age, and 2 categories for health insurance coverage.
4. **B03002** - Race for the Hispanic and Latino population, which is broken down into 2 categories for origin, 7 categories for race, and 2 categories for multiple races.
5. **C24030** - Sex by Industry, which is broken down into 2 categories for sex, 13 categories for industry, and an additional 13 categories for industry again.
6. **B01001** - Sex by Age, which is broken down into 2 categories for sex and 23 categories for age.
7. **B27010** - Health Insurance by Type, which is broken down into 4 categories for age, 3 categories for types of insurance, and 12 categories for coverages of insurance.

Table 1: Various demographics as they are correlated to the service worker population

Demographic	Correlation to Service Worker
19 to 34 year olds with no health insurance	0.8223515
On Medicaid	0.7604635
19 to 34 year olds	-0.7035059
Not Hispanic or Latino	-0.7241782
Females between 19 and 34	-0.7268381
Females between 18 and 34	-0.7317395
Has only one health insurance policy	-0.7544663
19 to 34 year olds with one health insurance policy	-0.7745397
Has employer-based health insurance	-0.7788886
Females between 19 and 34 that have health insurance	-0.7789289
19 to 34 year olds with any health insurance	-0.7797746
19 to 34 year olds with employer-based health insurance	-0.7846184
Citizens over 18 years old	-0.7943346
Females over 18 years old	-0.8071576
People over 18 years old	-0.8454889
People with health insurance	-0.8965770

Who is a Chicago Service Worker?

As we performed this analysis, much of our focus was on the typical characteristics of the service worker in Chicago. Service worker data was compiled from Census Bureau data on construction, manufacturing, retail trade, transportation, warehousing, utilities, and accommodation and food services. We used correlation analyses, where we compared how strongly correlated one demographic value is with the proportion of service workers for each ZIP code. Correlations are measured from -1 to 1, where a correlation of -1 means that as one variable increases, the other variable decreases. A correlation of 1 means that as one variable increases, the other variable increases.

The table below shows the correlations to the proportion of service workers for each demographic characteristic that we analyzed. Here are a couple examples of how to read this table: the correlation between the proportion of service workers and the proportion of people with health insurance is -.8965770. This means that as the proportion of service workers in a ZIP code increases, the proportion of people with health insurance decreases. As another example, we see a correlation value of 0.76046345 for the correlation between the proportion of service workers and the proportion of individuals with Medicaid. This means that as the proportion of service workers in a ZIP code increases, the proportion of people with Medicaid (as opposed to other kinds of health insurance policies) also increases. The table is ordered by the correlation values.

From these values, we can speculate that service workers appear to be comprised of people on medicaid or between 19 and 34 years old without health insurance. It also appears that service workers are more Hispanic or Latino, male, and are non-citizens than the people who do not work in a service industry.

What Demographics Explain the COVID Case Rate?

For our model we went with a linear regression model and we used a shrinkage technique to be able to identify the variables that did the best job in explaining the variation that we see in the COVID case rate. Based off the many models we ran, it looks as though the proportion of 18 to 34 year olds that are Hispanic/Latino and the proportion of non-Hispanic/Latinos do the best at explaining the variation we see in the COVID case rate. As the proportion of Hispanics/Latinos between 18 and 34 goes up, the rate of COVID goes down. Interestingly enough, we also see that as the proportion of non-Hispanic/Latinos goes up, the case rate goes down at a faster rate.

As a result, it appears that the 18 to 34 demographic really drives COVID case rates down even though

the Hispanic demographic is seeing an increase in their COVID case rate. These other variables also saw a decrease in COVID case rate as their proportion increased: the number of work eligible people, the proportion of citizens over 18, and the proportion of the population that is not Hispanic or Latino and is white. There were two variables that we saw that as the proportion increased, the COVID case rate also increased, and those were; the number of people that are Hispanic, and the proportion of females that were aged 19 to 34.

In all, while building a model that tries to explain the relationship between the COVID case rate between all 34 predictor variables considered, there were seven that did the most to explain that relationship. Those seven were the number of people that are Hispanic; the number of people that are work eligible; the proportion of citizens over 18; the proportion of 18 to 34 year olds that are Hispanic; the proportions of 19 to 34 year olds that are female; the proportion that is not Hispanic; and the proportion that is not Hispanic but white.

Statistical Model

Our model can be represented as:

$$\begin{aligned}\text{COVID Case Rate} = {} & 0.00542(\text{Hispanic Population}) \\ & - 0.0186(\text{Work Eligible Population}) \\ & - 798.1(\text{Proportion that are citizens and over 18}) \\ & - 1330.0(\text{Proportion of Hispanics that are 18 to 34}) \\ & + 180.8(\text{Proportion that are female and 19 to 34}) \\ & + 2553.5(\text{Proportion that is Hispanic}) \\ & - 263.0(\text{Proportion that is not Hispanic and is white})\end{aligned}$$

The coefficients indicate how influential each demographic is on the COVID case rate. For example: A 0.1 increase in the proportion of Hispanic/Latinos that are 18 to 34 years old will correspond to a decrease of 133 per 100,000 in the COVID case rate. Contrast this with the fact that a 0.1 increase in the proportion of Hispanic/Latinos (not necessarily 18 to 34) will correspond to an increase in the case rate of 255 per 100,000. It seems we have an almost 400 cases per 100,000 swing just by focusing on the 18 to 34 age range. If our ZIP code population is more Hispanic or Latino, then our COVID case rate increases, unless that proportion is made up of more and more 18 to 34 year olds. Then our COVID case rate decreases. Looking at a far less significant variable, the population that is Hispanic (note this was a positive correlation), a 1000 increase in this population corresponds to an increase in the case rate by just 5 cases per 100,000.