Social Determinants of Health for COVID-19 Diagnosed Patients

Dr. David C. Patton, david.patton@a3.ai

Dr. Xue Yang, xue.yang@a3.ai

Changrong Ji, changrong.ji@a3.ai

08/28/2020

Abstract

Many studies have been conducted on small patient populations regarding the medical conditions and health outcomes for COVID-19 diagnosed individuals. In this paper, we examine patients with health insurance for a large population. We focus on social determinants of health. Our results indicate a correlation between income levels, housing situation, occupation, and ethnicity. We find a strong disparity in the impact on the Hispanic population compared to all other ethnicities.

1.0 Introduction

With access to a large amount of information both social (individual consumer) and medical, we aim to find correlations between COVID-19 diagnosed outcomes and said information. Such correlations can indicate a predictive relationship that may be exploited in practice for possible treatments, preventative practices, and social/economic support.

In the following sections we introduce the data used in this study.

1.1 Data

The source of data for this study was provided by the COVID-19 Research Database[1]. The database is a public-private consortium organized by Datavant, Health Care Cost Institute, Medidata, Mirador Analytics, Veradigm, Change Healthcare, Snowflake, and many others. For this study we utilized a linked database combining consumer's social data from AnalyticsIQ, insurance claims from Office Ally, and mortality data from Datavant. Highlights and limitations of each database include:

AnalyticsIQ

Consumer data for 242,468,278 individuals in the US as of the most recent quarter. The data is further described by the company below. One of the limitations of the data is that some social attributes are derived from multiple variables via the company's proprietary modeling methods.

> *AnalyticsIQ is a leading predictive data and analytics innovator that leverages a blend of publicly available data and custom algorithms informed by cognitive psychology concepts to describe consumers across three areas - People, Behaviors, and Predictors. Headquartered in Atlanta and recently named one of Georgia's Top 10 most innovative companies, AnalyticsIQ's team of data analysts, scientists, and cognitive psychologists have over 100 years of collective analytical experience and expertise. --AnalyticsIQ*

An Anonymous Claims Data Clearing House

Medical claims records for 98,220,768 patients with more than 3 billion service records. These records cover over 7 years of their history. The geographical distribution of this data is heavily concentrated in the western United States.

Mortality

230,183,559 death records of more than 100 years.

> *Mortality data is provided by Datavant and contains obituary data sourced from online newspapers, funeral homes, online memorials, direct submissions and more. --Datavant*

These three datasets are combined via Datavant's privacy-preserving record linkage.[2]
This provided us with 34,834,778 patients in common between Analytics and Office Ally. And the number of patients in all three databases was 1,554,840.

2.0 Method

Identifying COVID-19 Cases

We identified COVID-19 cases via the Office Ally claims data. If a diagnosis contained the American ICD-10-CM code *U07.1*, then they were considered to be positively diagnosed as a COVID-19 patient by a physician.

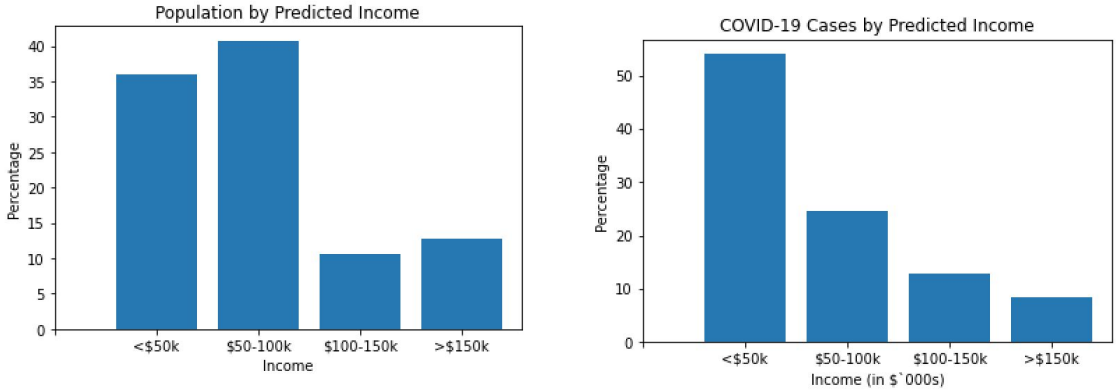
Identifying COVID-19 Deaths

If patients were diagnosed with COVID-19 and they are present in the Mortality Death Index, then we attribute them as a COVID-19 death. One limitation of this method is that the linked dataset does not contain the cause of death. It is possible that a patient may recover from COVID-19 and then soon afterward perish from a completely unrelated cause.

3.0 Findings

3.1 Income Inequality

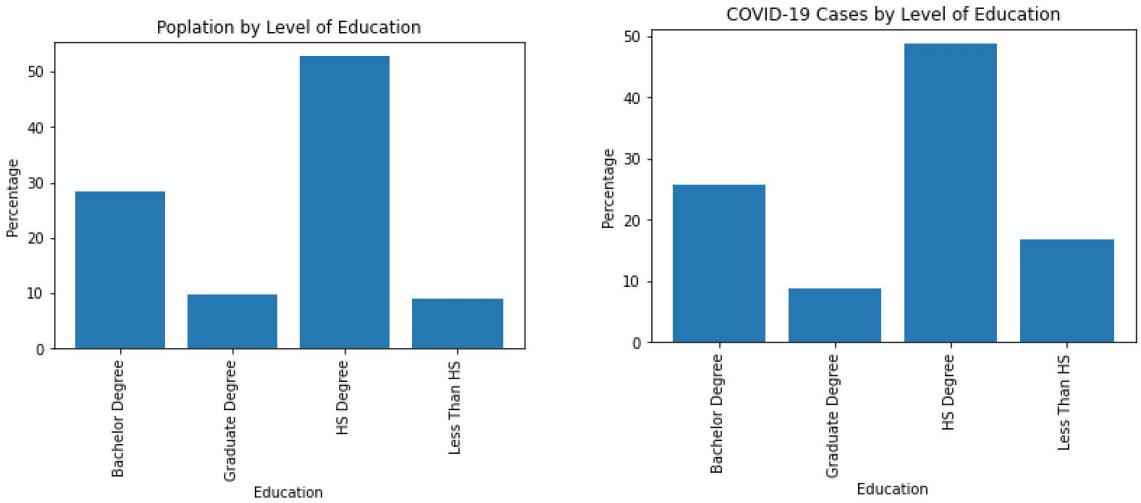
It has been widely reported that COVID-19 is not equally impacting populations in the same manner.[6] One of the inequalities is at the level of income. To analyze this we looked for correlations via AnalyticsIQ's IncomeIQ_Plus_v3 attribute. This attribute predicts an individual's household income level via a proprietary scoring model. Our results are shown below with the general population on the left and COVID-19 cases on the right.



We see a large increase for the lowest income category and a large decrease for the second lowest category.

3.2 Education Attained

Next we wanted to determine if there is a correlation between education level and the transmission of COVID-19. AnalyticsIQ captures an individual's education level in their AIQ_Education_v2 attribute. It has four possible values including Less Than HS, HS Degree, Bachelor Degree, and Graduate Degree. Below are the distributions of education levels for the general AnalyticsIQ population and the COVID-19 diagnosed population.

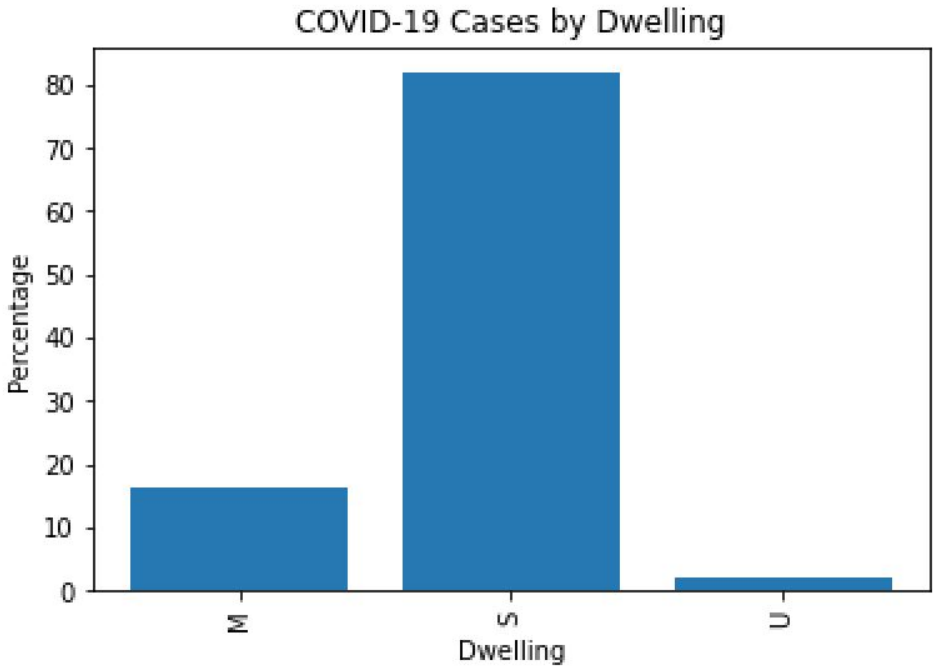


For those individuals with less than a high school degree, the percentage nearly doubled while all others declined (most prominently HS Degree and Bachelor Degree). This indicates a strong correlation between COVID-19 and level of education.

3.3 Housing Situation

Transmission of COVID-19 has been shown to be the most dangerous when inside a structure containing multiple people.[3] So, where people spend their time could be an important factor in COVID-19 outbreaks. Thus we examined the correlation of an individual's housing situation with COVID-19 diagnosis. AnalyticsIQ captures housing via their AIQ_Dwelling attribute. It indicates a single-family unit, a multi-family unit, or unknown. Within the total population of individuals surveyed by AnalyticsIQ, 86% live in single-family units, 12% in multi-family, and 2% were marked as unknown.

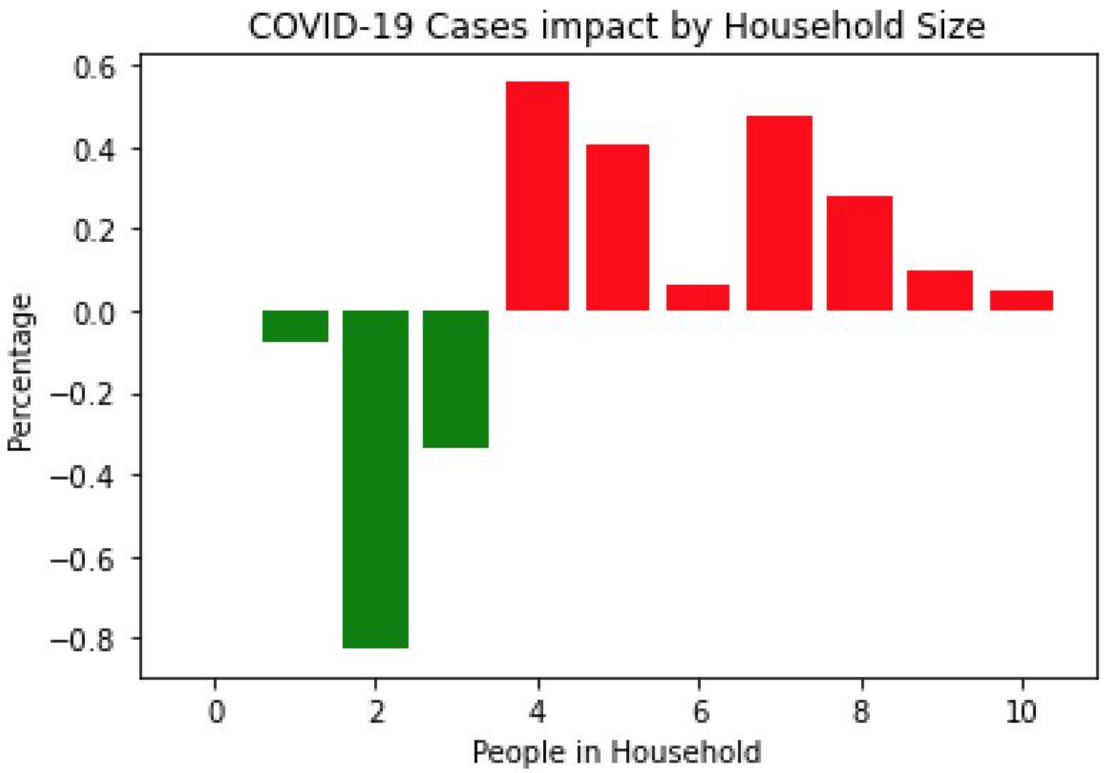
Below are the results for COVID-19 diagnosed patients.



In the COVID-19 patient population, those residing in a multi-family unit are 16% and those residing in a single-family unit are 82%. This is a 4% increase for multi-family and 4% decrease for single-family.

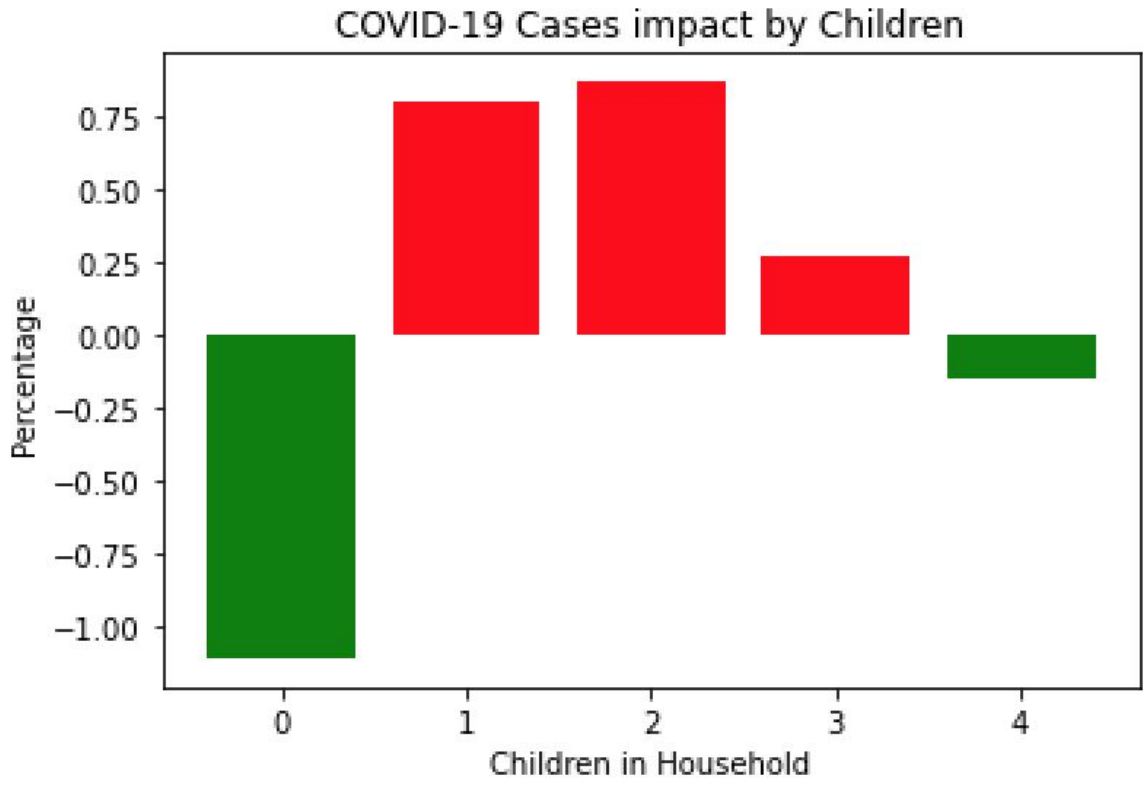
Household Size

AnalyticsIQ captures further information about the living conditions of individuals. With the AIQ_People_In_HH attribute, the number of people residing in the individual's household is tracked. When we analyzed the difference of the total AnalyticsIQ population vs those diagnosed with COVID-19 as shown below, it was very evident that living in a household with more people correlates with COVID-19.



Children in Household

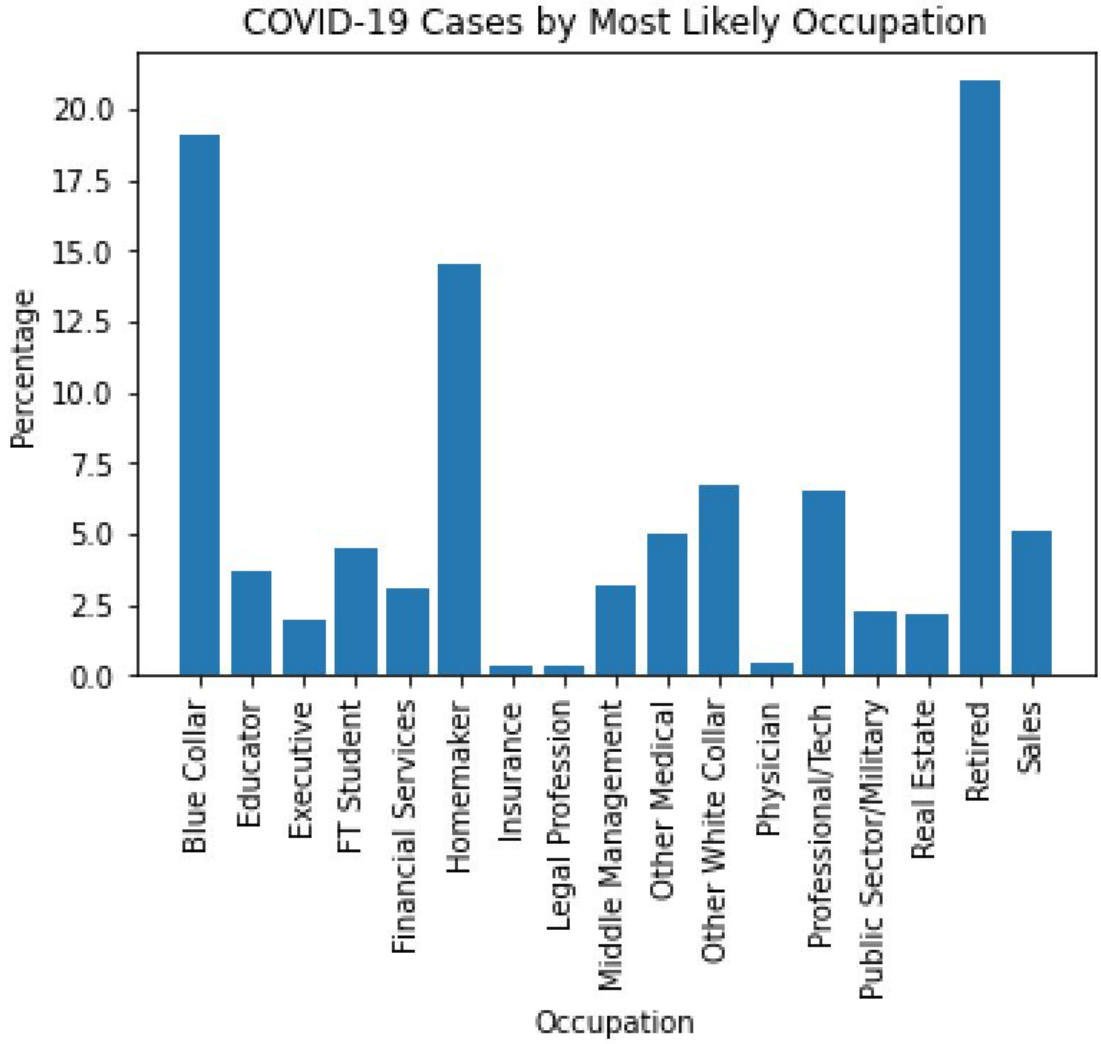
In addition to the household size, AnalyticsIQ tracks the number of children in the household via AIQ_Children_In_HH attribute. We performed the same analysis as before and found that there seems to be a slight correlation with COVID-19 diagnosis and the presence of children in the household.



3.4 Occupational Exposure

In the previous section a correlation was shown between the type of dwelling an individual resides in and COVID-19 cases. Since some job types require a lot of time indoors or exposure to many other people, we investigated the effect of occupation type on COVID-19 cases. AnalyticsIQ collects this information in the JobsIQ attribute. This attribute is one of the derived attributes (see section 1.0 for further information). There are 17 job type values available for occupational classification.

Below is the distribution of COVID-19 cases by occupation.



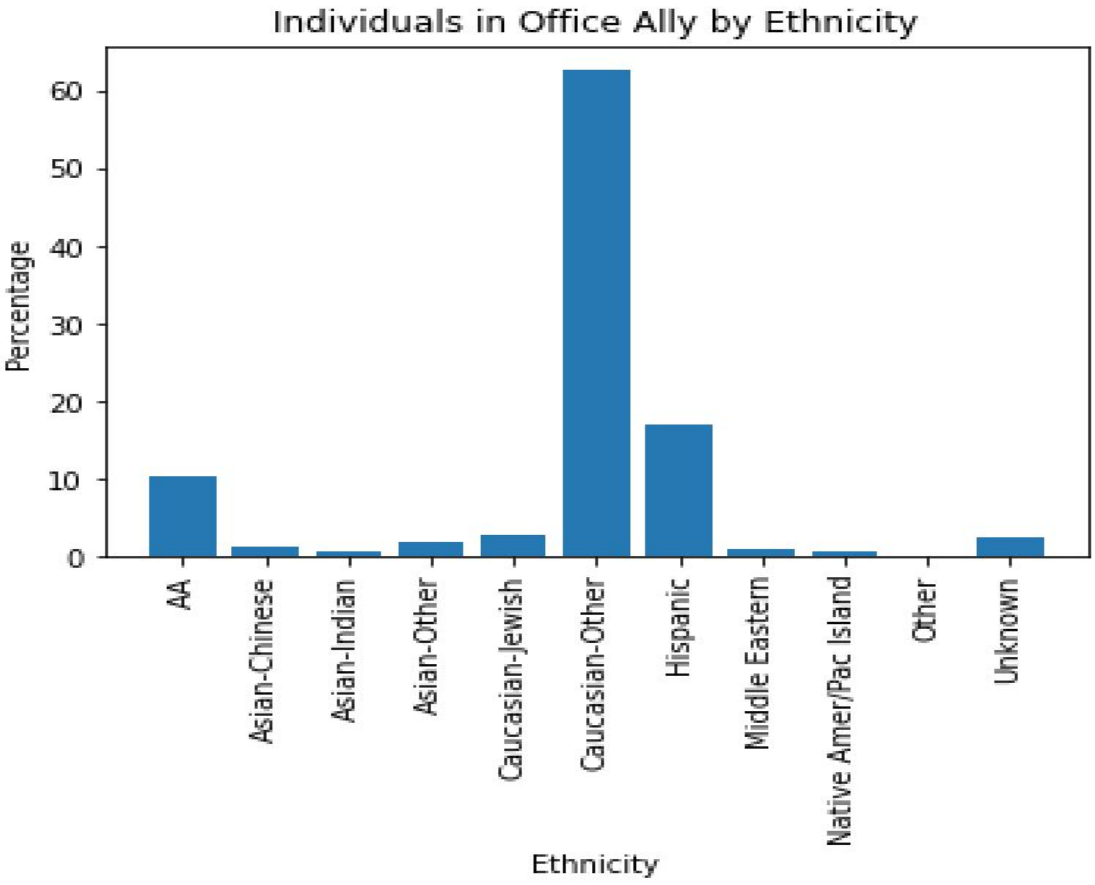
We found that compared to the general population in the AnalyticsIQ database, Blue Collar was 1% more. Also Retired increased by 3%. Notably Homemaker and Professional/Tech decrease by 2.5% and 2% respectively.

Because of the exposure of front-line workers in the medical and education fields, we take a deeper look at them next.

3.5 Ethnic Disparity

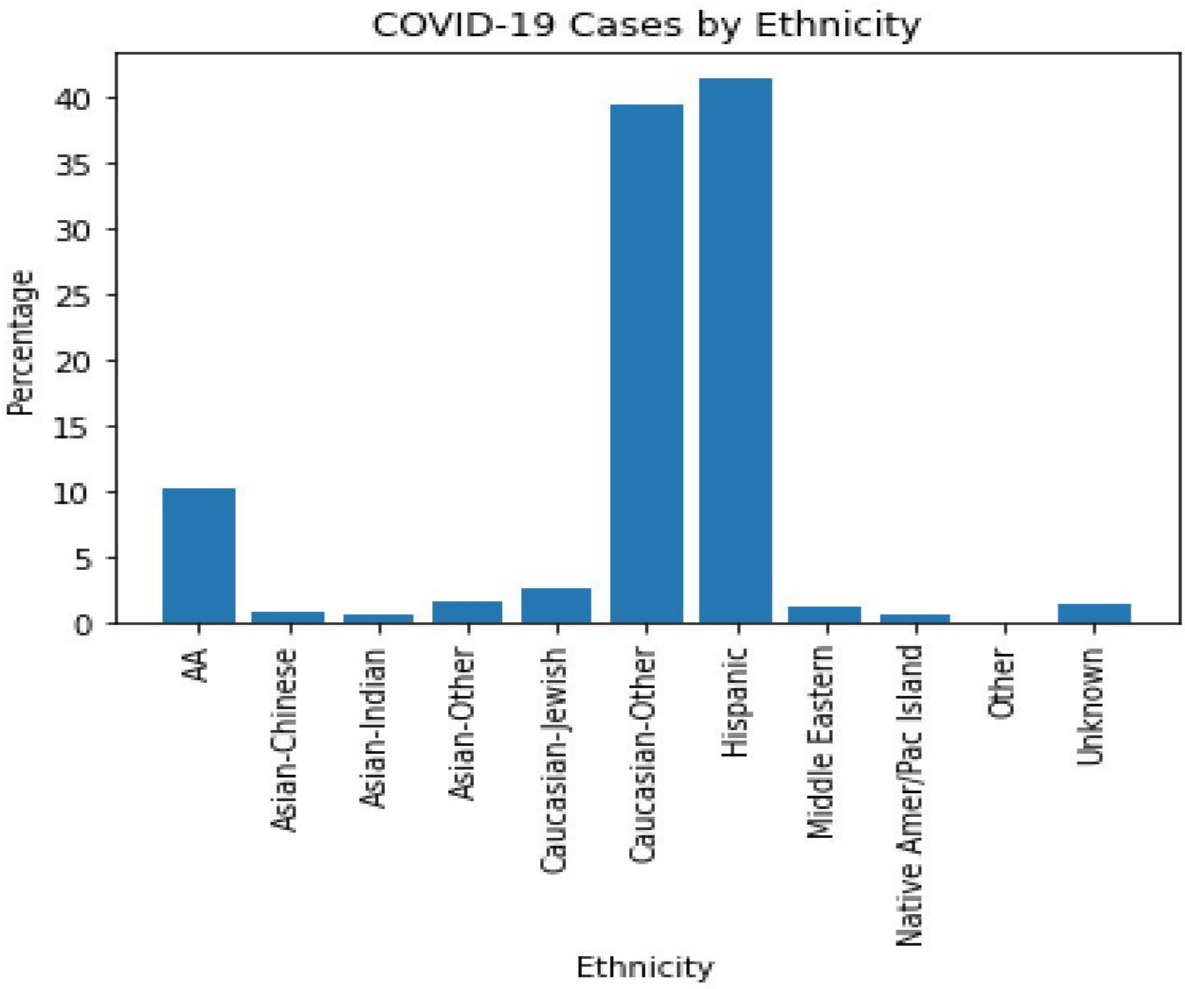
Many studies have shown that the burden of COVID-19 is not equally distributed amongst the ethnic communities in the United States.[4,5] We wanted to validate this in our data and try to see what other correlations might be impacting this disparity. AnalyticsIQ determines ethnicity via the EthnicIQ_v2 attribute. It is their proprietary Ethnicity Identification System (known and inferred). Below is the ethnicity distribution of all patients in the linked data. Specific values are 10.3% for African-American, 62.7% for Caucasian-Other, and 17.0% for Hispanic. In the

complete AnalyticsIQ database the values are 10.6%, 64.5%, and 13.2% for African-American, Caucasian-Other, and Hispanic respectively.



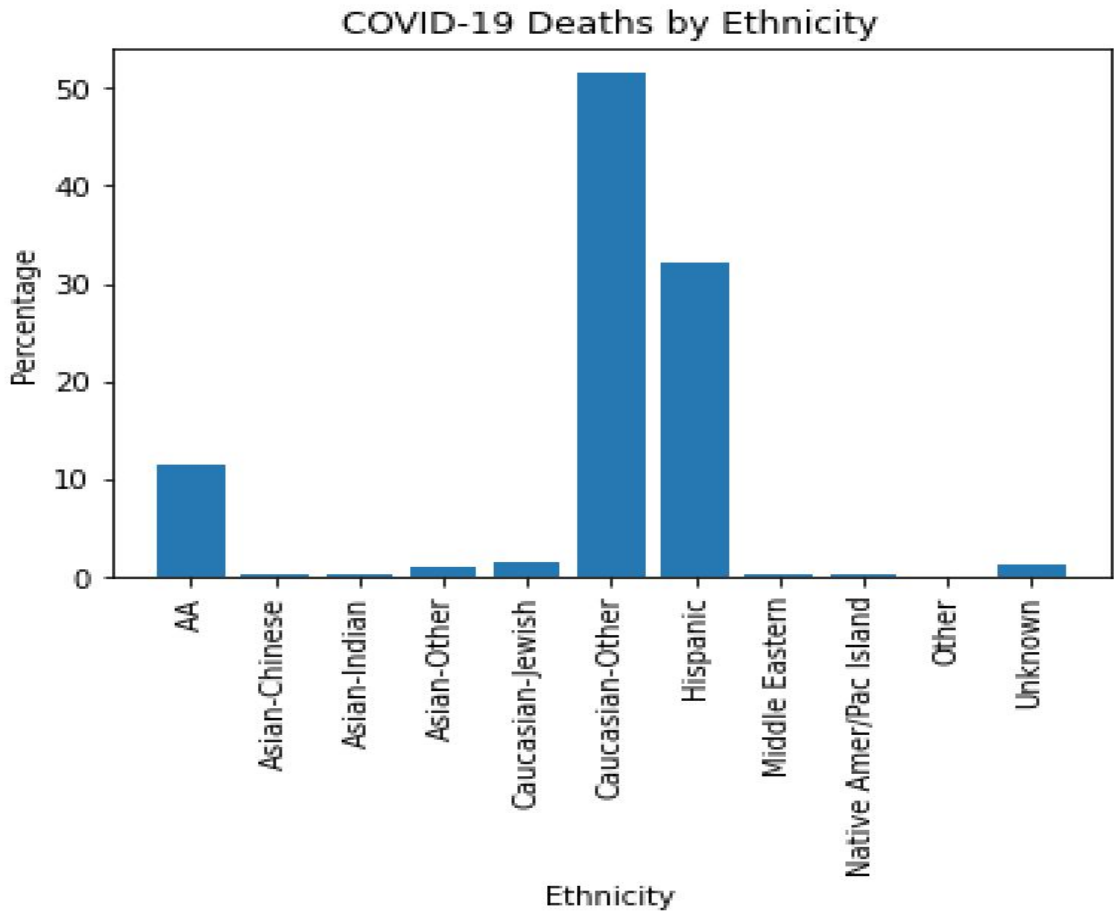
COVID-19 Cases

Below is the distribution of COVID-19 cases by ethnicity. Compared to the ethnic distribution of the general patient population we see a substantial increase amongst Hispanics and a substantial decrease amongst Caucasians.



COVID-19 Deaths

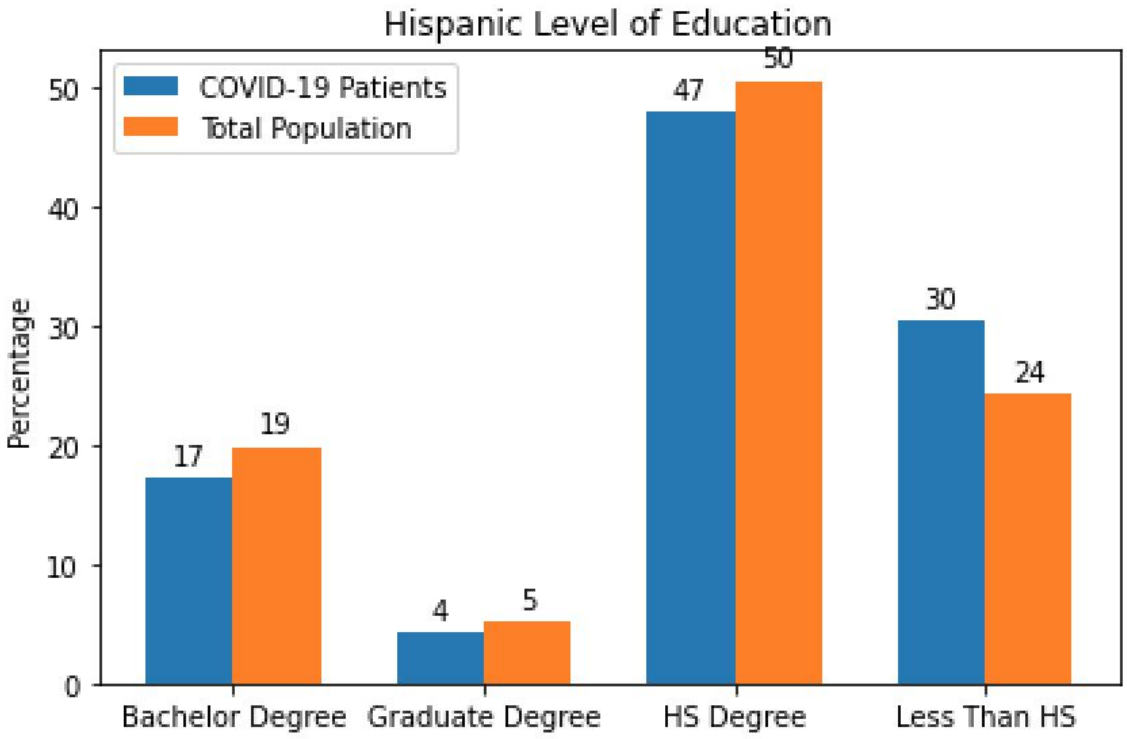
We also analyzed the ethnic distribution of COVID-19 deaths. Again, we found a substantial increase in the percentage for Hispanics and decrease for Caucasians. There is also a very slight (about 1.2%) increase for African-Americans.



In the next two sections we examine even more attributes related to the Hispanic COVID-19 population.

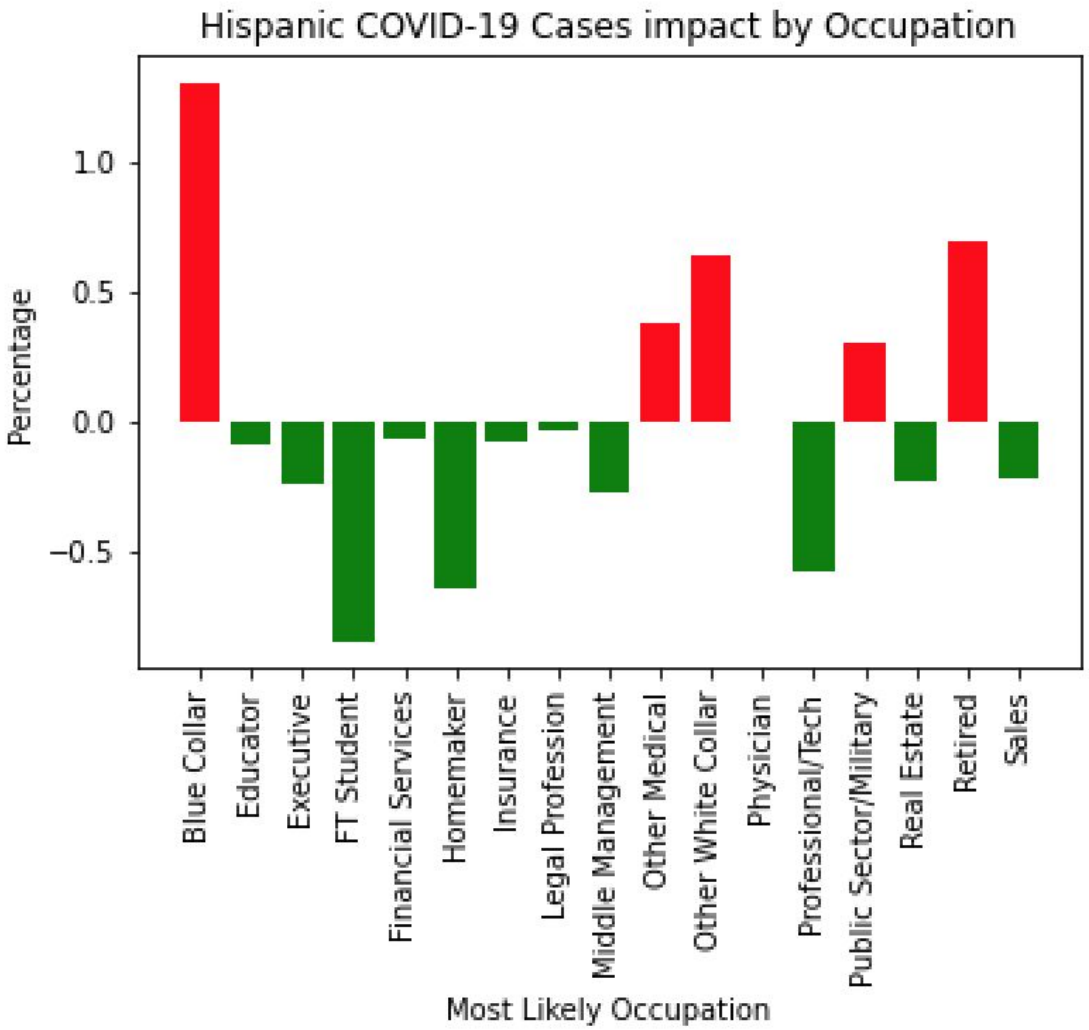
Hispanic Level of Education

To determine the correlation between COVID-19 diagnosis and the level of education we examined the AIQ_Education_v2 attribute. But this time we only selected the Hispanic population. Based on this attainment of education, our results show a correlation with an increase for those less than a High School diploma and a decrease for all others.



Hispanic Occupation

To determine the correlation between COVID-19 diagnosis and the type of work we examined the JobsIQ attribute. But this time we only selected the Hispanic population. Based on this (most likely) occupation our results show a strong correlation with increases for Blue Collar, Retired, Other Medical, Other White Collar, and Public Sector/Military groups. Presumably these occupations represent many of the “essential workers” occupations.



4.0 Summary and Future Work

We have shown many correlations between social determinants and COVID-19 diagnosis including income level, education attainment, housing situation, occupation, and ethnicity. Further we looked into finer grained determinants such as the education levels and occupations of the Hispanic populations.

As a novel virus, everyone is susceptible to COVID-19 infection if exposed to it. As such, social determinants of health play a very critical role for COVID-19 infection (COVID exposure risk). Many social determinants of health—including poverty, physical environment (eg, air quality, homelessness), and ethnicity—can have a considerable effect on COVID-19 outcomes.[7] Our study supported some of the factors but not all. Interestingly, in our data, we found that the Hispanic population has a relatively high infection rate as well as COVID-19 related deaths, but

not in the African-American population as other studies indicated.[8] Our data is still a limited population. A larger data set will provide a more complete representation of the total population.More risk factors are going in to be evaluated in our next steps.

In the future, these social determinants should be evaluated for their predictive accuracy for severe outcomes (hospitalization, ICU, etc.) and can provide individual risk scores. Many of these social determinants may predict future trends in the COVID-19 diagnosed population. For example tracking the correlations within the Educator and FT Student occupations may be fruitful. We intend to conduct further analysis which includes even more demographic data such as age, gender, location, and others. We also want to further refine our methodology for assigning COVID-19 related deaths by examining the date for the last claim record and the date of death for a patient.

We plan to incorporate de-identified personal level EHR data within the COVID-19 Research DB consortium. In addition, there is a wealth of public data sources on clinical, social and economical factors, as well as jurisdiction level policies with regards to COVID-19. Non COVID-spefiic resources such as CDC Social Vulnerability Index are valuable too.

We will use a combination of statistical methods including multiple regression, machine learning and causal inference to further analyze the SDoH on COVID-19 illness burden.

5.0 References

1. The data, technology, and services used in the generation of these research findings were generously supplied pro bono by the COVID-19 Research Database partners, who are acknowledged at https://covid19researchdatabase.org/
2. https://datavant.com/how-we-do-it/
3. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-how-is-covid-19-transmitted
4. https://www.medrxiv.org/content/10.1101/2020.05.07.20094250v1
5. https://www.npr.org/sections/health-shots/2020/05/30/865413079/what-do-coronavirus-racial-disparities-look-like-state-by-state
6. https://www.sciencedaily.com/releases/2020/04/200430191258.htm
7. Abramsa, E. and Szeflerc, S.COVID-19 and the impact of social determinants of health.Lancet Respir Med. 2020 Jul; 8(7): 659–661.
8. Hooper, M., Nápoles, A., Pérez-Stable, E.,,COVID-19 and Racial/Ethnic Disparities,*JAMA*. 2020;323(24):2466-2467. doi:10.1001/jama.2020.8598