From Macro to Micro:

Social Determinants of COVID-19 & Nasopharyngeal Features Associated with Disease Progression

A Study of COVID-19 in Worcester, Massachusetts

ACKNOWLEDGEMENT

This project would not have been possible without the contributions of the many people who guided, and supported me throughout my entire time at WPI. To the members of the WPI community-at-large, as well as each of my professors, you have my thanks.

This past academic year (2020-2021) was especially difficult due to the advent of the COVID-19 Pandemic, but fortunately, I was surrounded by countless individuals who helped raise me up, and pushed me every step of the way. To those of you who helped me on this journey, I would like to express my sincerest gratitude:

- Brenton Faber
- Reeta Rao
- Evan Bradley
- John Haran
- Lindsey Cincotta
- Protiva Dutta
- John Hoang Do

From Macro to Micro:

Social Determinants of COVID-19 & Nasopharyngeal Features Associated with Disease Progression

A Study of COVID-19 in Worcester, Massachusetts

A Major Qualifying Project Report submitted to the faculty of Worcester Polytechnic Institute in partial fulfillment of the requirements for the Degree of Bachelor of Science in cooperation with UMass Memorial Medical School.
Submitted May 6th 2021

Submitted By:
Christopher Tocci

Project Advisors:
Brenton Faber (WPI)
Reeta Rao (WPI)
Evan Bradley (UMass)
John Haran (UMass)

ABSTRACT

BACKGROUND

The health and wellness of every individual is determined by many aspects, including their genetics, environment, behaviors (both the individual and those around them), medical care, microbiome, and social factors affecting their life. In order to fully assess and fix and global health issue such as the COVID-19 Pandemic, analysis into each of these aspects is necessary. The root of every outbreak lies in one of these issues, and while fixing one may stop the spread of disease, it will never truly eradicate it.

METHODOLOGY

In a correlative, qualitative analysis, we obtained data on COVID-19 density in Worcester, Massachusetts from the city's Department of Public Health. We overlayed the map associated with this data with heat maps on social factors in Worcester. These social factors included different metrics for areas like housing, nutrition, income, crime, transportation, and education. We analyzed these maps based on color and opacity to determine where social factors affecting health fell in relation to COVID-19 density; we also analyzed the numbers associated with the colors/opacities to show quantitatively how areas with a high prevalence of particular social factors vs. low matched up with COVID-19 density.

Following this, a correlative machine-learning algorithm was developed to understand how members of the microbiome correlate to respiratory distress following recovery from COVID-19.

Patients with COVID-19 or those exhibiting symptoms had a nasopharyngeal swab ordered upon admission to the UMass Memorial Hospital Emergency Department. Following this, patients were prompted by research staff to enroll in the study. To map the microbiota of each participant, whole DNA/RNA extraction was carried out, followed by Illumina sequencing. Illumina reads generated were to build a model for determining differences in microbial abundance as a predictor of respiratory support. The model was used to generate median importance values for a series of clinical factors, and microbial features found in the patients. To discern how specific microbes predicted the need for respiratory support or not, the model was also used to measure differences in abundancy levels of each microbe.

RESULTS

Analysis of the social determinants of health related to COVID-19 prevalence in Worcester revealed that all metrics were at least somewhat correlated. The best correlation existed between metrics like population density, income, and education. Slightly less significant, and likely less correlated were nutrition, and crime.

Following this, data from the first model revealed that differences in microbial species' abundances may act as better health predictors than age, BMI, and CCI when assessing if a patient will need respiratory support or not following COVID-19 recovery. Data from the second model revealed that abundances of microbes was highest in those patients who did not end up needing respiratory support

DISCUSSION

Since many of the social determinants are correlated with COVID-19 density, we have determined that further analyses into each factor is necessary to identify causal factors. Additionally, this evidence provides reason for the Worcester Department of Public Health to pursue campaigns that are geared towards preventing infections and assessing issues surrounding poverty, education, nutrition, and other social factors troubling particular populations in Worcester.

In addition to this, data from the microbiology models shows us that many different microbes may actually exhibit protective attributes in the nasopharynx; the microbes in those patients who had high abundancies and no respiratory distress may interact with the immune system or fill in niches that would otherwise be filled with opportunistic pathogens. Similar to the former half of the project, more analyses must be performed to make better sense of the data.

TABLE OF CONTENTS

TABLE OF CONTENTS *(CONT.)*

LIST OF FIGURES

PROLOGUE

In light of the 2020-2021 Coronavirus Disease 2019 Pandemic, this project faced many last-minute adaptations. Initially, communications with my advisors at UMass began in September 2019. I had planned to solidify a project before the holiday break ended, and while this was the result, it would not last long. As the Pandemic continued to rage on throughout the world, it became increasingly clear that the virus would hit America in much the same way as it had elsewhere.

A number of different projects cycled through the meetings I had with my advisors during the holidays. We finally solidified one around the New Year, and the specifics unfolded throughout following months, until March 11th, 2020, when the WHO declared COVID-19 was a worldwide pandemic. All at once, project centers shut down, research grants were either suspended or cut, and the world stood at a standstill. By April 2020, I was at the end of my Junior year and still without a project. UMass had decided to divert most funding to cohorts related to the virus. Luckily, my UMass advisors, Dr. Haran and his colleague Dr. Evan Bradley were ahead of the curve and just earned a grant for a project centered on COVID-19 in March. We discussed having me join the team, and once we had a game plan for my role, I started to work.

Throughout the rest of the Spring and Summer, I assisted with patient data collection. I worked remotely within our RedCap database, and even helped call patients to enroll them near the beginning of Autumn. We planned to have me enter the lab and assist with the extraction of DNA/RNA from the oral swabs by the New Year, but UMass had prohibited all outside research members from entering; even regulars to the labs at UMass had trouble finding time, and space. I simply bided my time meeting with Dr. Bradley and Dr. Haran about once a month to discuss how the project was going and what to expect from the future. It wasn't until weeks before my final deadlines that we procured data for me to present. It was close… but in the end we got it.

All the while, I was working on the epidemiology portion of the project; this too faced a pretty significant change. At first, I planned to utilize patient data from the former research project to understand how COVID-19 affects members of the Worcester population. However, as time went on, analyses started to bear little significance due to the size of the patient pool we had. In much the same way, the project returned to the drawing board in January 2021. At this point, half the year was gone – we needed a feasible project that could be accomplished in ~15-weeks.

That is when we stumbled upon the geographical heat maps of Worcester which showed how certain social determinants of health proliferate throughout the city. Using this data, and COVID-19 density previously procured from the Worcester Department of Public Health, I was able to complete the second portion of the project.

These obstacles pale in comparison to those being faced by families all across the globe. COVID-19 has taken much away from my peers and I – our Senior year, a normal graduation, and in some ways our memories. But many have lost their lives, and so we owe to those who have suffered much greater losses to work together and endure the Pandemic. It's simple and yet insightful research shown in this project that can make the biggest difference for many. We hope that from macro- to micro- the work presented here will pave the way for future endeavors.

INTRODUCTION

Coronavirus Disease caused by SARS-CoV-2 has been confirmed in over 153 million individuals worldwide as of 4:38 PM, May 5th, 2021. These cases accounted for nearly 2% of the world's population (exactly 1.9578%). Of those confirmed cases, a total of 3.2 million individuals have died. These deaths accounted for 2.0915% of all cases, and 0.0407% of the world's population[1].

Based on this data, a population of 1000 people will see around 20 cases. And every population that sees 1000 cases will experience around 20 COVID-19-related deaths. Over this past year, COVID-19 has taken the lives of nearly 0.05% of our species and has drastically increased morbidity in peoples everywhere.

Worcester Massachusetts was no exception when it came to the number of individuals affected by COVID-19. By the time the WHO had declared the COVID-19 outbreak a global pandemic on March 11th, 2020, Worcester officials had already started cancelling large events such as the St. Patrick's Day Parade, and even began closing public schools[2,3]. On March 14th, Worcester saw its first case[4], and as of April 30th the city has seen over 23,000 cases cumulatively[5]. Current research in both epidemiology and microbiology have already pointed out the relationships that social factors and the human microbiome have with the advent of disease. However, every pathogen is different, and therefore research in these fields must constantly be conducted to reveal new relationships. In our case, we are dealing with a global pandemic unlike any other during the past century.

In order to better understand the dynamics of COVID-19 from macro-to-micro, two investigations were carried out in this project. Investigations in COVID-19 from different fields are necessary to fully understand how the virus spreads, and what populations it poses the biggest threat to.

We first sought to understand if social determinants commonly associated with health might be correlated to the prevalence of COVID-19 in Worcester, Massachusetts. Social factors such as education, income, and access to care play a large role in the health of an individual and may be just as important as virulence factors or even genes.

Using data collected from the Worcester Department of Public Health, as well as local survey agencies and the U.S. Census Bureau, we were able to overlay COVID-19 Density with different social factors, and demographics. These overlays were geographic heat maps depicting locations in the city, were (1) COVID-19 was most prevalent, and (2) where populations experiencing particular social factors lived. This data can be used in the future as evidence to support public health programs built on delivering health- and social- equity in Worcester.

The second component of this project was to investigate the oral microbiome's relationship with COVID-19. Prior literature on COVID-19 has shown that SARS-CoV-2 may attack the epithelial cells lining the upper airway, giving rise to further upper and lower respiratory infections following recovery from COVID-19. The exact mechanism as to why these infections happen is currently being investigated and was not the focus of this study. We sought to investigate whether members of the nasopharyngeal microbiome were correlated to respiratory distress in patients following recovery from COVID-19. This distress is thought to be linked with damage caused by the virus itself, and those additional infections that happen following recovery from COVID-19.

Using a machine learning algorithm trained with patient data collected through the UMass Memorial Hospital Emergency Department, a model was generated to show the predicative power of three parameters with respect to a dichotomous outcome – whether a patient will need respiratory support or not. These three parameters included the microbes present in the population, as well as clinical features (age, BMI, etc.) and demographics (age, race). In addition to this, the algorithm also modeled the relative abundance of nasopharyngeal microbes with respect to this same dichotomous outcome.

The former analysis was used to determine how accurate three parameters were at predicting the right patient outcome. The second analysis was used to show which microbes were associated with which specific outcome, and therefore only focused on one of the specific outcomes. Since it is already well known how age, and BMI can affect the body, this micro-investigation was performed to see if similar levels of importance could be determined for the microbiome.

Both the macro-analysis of social determinants and micro-analysis of the oral microbiome have provided evidence to support the correlation between COVID-19 and social determinants and with the nasopharyngeal microbiome. Further analysis into causal relationships between social factors and how microbes interact with the in the body virus may reveal ways to improve how we combat infection, and future viral outbreaks.

BACKGROUND — BIOLOGY

THE ADVENT OF SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS 2

Coronaviruses (CoVs) are enveloped viruses with a single-strand positive-sense RNA genome of approximately 26-32 kilobases in size[6] and belong to the family, Coronavirdae, and subfamily Cornavirinae. Within this subfamily, there are four genera: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus. These genera are based on the viruses' phylogenetic associations and genetic sequences. Variability between the CoVs' primary structure leads to differences in the infectivity, and lethality, as well as the host range of the virus[7].

Alphacoronaviruses and betacoronaviruses exclusively infect mammals, whereas gammacoronaviruses and deltacoronaviruses mostly infect birds. Some cases of gammacoronavirus and deltacoronavirus have been detected in mammals, including humans but these cases are usually rare[8]. All coronaviruses, regardless of genera, have been traced back to animal origins, most of which started in bats or rodents – more so in bats. Not only do these animals act as reservoirs for the virus, but they also act as vectors. Rodents and bats can infect humans directly and can also cause infection through intermediate animal hosts, usually livestock or fish. Livestock that contracts the virus is believed to play a critical role in its transmission to humans, due to their constant interactions with one another. Most of the coronaviruses which come from bats have originated in China, but have also been traced to some European, African, and Southeast Asian countries[8].

The disease and associated symptoms caused by these coronaviruses vary widely between families, subfamilies, and genera. Additionally, symptoms and progression of the virus depend heavily on the species they infect. In most cases, alphacoronaviruses and betacoronaviruses cause respiratory distress in humans, and gastrointestinal distress in all other mammals[8]. The best-case scenario for any human that contracts a coronavirus is the common cold[9]. In fact, human coronavirus (HCoV) 229E (229E), and HCoV-OC43 (OC43) are responsible for around 15-29% of all common colds.[6] However, certain strains of betacoronaviruses pose a much greater risk[6,8].

Three strains of coronaviruses have been responsible for three separate pandemics. Such adverse events were caused by Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV), Middle Eastern Respiratory Syndrome-Coronavirus (MERS-CoV), and Severe Acute Respiratory Syndrome-Coronavirus 2 (SARS-CoV 2)[9]. The SARS Pandemic started in November 2002, leading to 8,273 reported cases, and 775 deaths (CFR= 9%) between the 37 countries affected as of 2016[6]. The MERS Pandemic started in September 2012, leading to 1621 reported cases and 584 deaths (CFR= 36%) between the 26 countries affected as of 2016[6]. The major spread of these viruses during these pandemics was fast, but not long-lasting, with SARS being stopped in just over 9-months after the initial outbreak[9].

Coronaviruses such as SARS-CoV are associated with higher prevalence and mortality due to their binding affinity for the angiotensin-converting enzyme 2 (ACE2) receptor on the membranes of cells in the lungs, heart, arteries, kidney, and intestines[7]. This enhanced binding affinity for ACE2 by SARS Related-Coronaviruses (SARSr-CoV) is due to differences in these CoVs S gene and a region upstream of orf8.

The S gene encodes the spike (S) protein that contains the receptor-binding domain (RBD), while the orf8 region encodes an accessory protein[8]. Due to naturally selected mutations in the RBD SARSr-CoVs, binding locations on ACE2 normally inaccessible are turned into viral binding hotspots. SARS-CoV-2 which first appeared in China in 2019 has been associated with higher rates of mortality and infectivity than SARS-CoV due to slight differences in residues at the SARS-CoV-2 RBD-ACE2 binding interface[10]. These residue changes help stabilize binding at this interface thus greatly increasing affinity.

CLINICAL CHARACTERISTICS OF CORONAVIRUS DISEASE

Coronavirus disease is caused by SARS-CoV-2, a virus capable of entering cells with the ACE2 receptor[7]. This receptor is found in virtually every organ and body system, including the oral and nasal mucosa, nasopharynx, lung, stomach, small intestine, colon, skin, lymph nodes, thymus, bone marrow, spleen, liver, kidney, and brain. Since the respiratory system has multiple portals of entry, this system is typically the first one compromised. SARS-CoV-2 is typically spread via respiratory droplets, which can be transferred from person to person through respiratory secretions (like coughing) or by making physical contact with the virus via a surface, including the clothing and/or body of another person[7]. Viral particles have been found to remain on surfaces for up to 72 hours[11].

Once the virus enters the body and contacts the ACE2 receptor it fuses with the membrane of the cell and inserts a positive-sense, single strand of RNA. This strand of RNA is translated by the host ribosomes which allows the virus to hijack the cells to produce more viral particles. Hijacking the cell's machinery can damage and destroy the affected cells, which in turn damages (or destroys) the associated tissue. In the case of SARS-CoV-2, respiratory distress is the most cited problem[12]. Upon infection, symptoms typically appear 2-14 days after exposure to the virus[13] with a median incubation period of about 5.1 days[14]. Out of the 72,314 cases reported by the Chinese Center for Disease Control (CCDC), 81% of cases appeared mild, 14% were severe, 5% were critical, and 2.3% were fatal. Severe and critical cases were associated with hypoxia, dyspnea, shock, pneumonia, and respiratory failure[12].

Patients usually present with symptoms associated with a viral infection, including fever or chills, cough with or without sputum, shortness of breath or trouble breathing, fatigue, body aches, headache, sudden loss of taste or smell, sore throat, congestion, or runny nose, nausea or vomiting, diarrhea, malaise, and altered mental status[13]. Typically, patients reporting any of the following symptoms, especially those associated with decreased lung function will undergo a physical examination by a doctor. CT scans and chest radiography can be used to detect pneumonia and altered lung function. However, diagnosis can only be confirmed with a positive screen from a laboratory test. Physicians suspecting SARS-CoV-2 can order a real-time reverse transcription-polymerase (rRT-PCR) assay, molecular test, antibody test, or a viral culture. The rRT-PCR test is the only test that has been validated, approved, and recommended by laboratories for testing acute diagnosis of SARS-CoV-2. All other tests have either not been approved, are still being validated, or are both validated and approved, but not recommended due to constraints that are made up by the rRT-PCR test[15].

BACKGROUND — BIOLOGY *(CONT.)*

According to the World Health Organization (WHO) COVID-19 Dashboard, there have been 63,965,092 cases reported worldwide as of November 30, 2020, and 1,488,120 deaths. This number has risen steadily since the beginning of the Pandemic in December 2019[16]. According to the CDC COVID-19 Dashboard, there have been 13,822,249 cases reported in the U.S. as of November 30, 2020, and 272,525 deaths. This number has also steadily risen since December 2019[17]. The CDC has also stratified the data according to age, sex, and ethnicity by controlling for all other factors within each respective study. According to data published August 18, 2020, when compared to individuals 18-29 years old, both rates of hospitalizations and death due to infection increase every 10 years older the patient is[18]. Furthermore, when compared to white, non-Hispanic persons, cases, hospitalizations, and deaths for all other ethnic groups increases aside for cases in Asian, non-Hispanic persons[19]. The margin between how SARS-CoV-2 affects both sexes is less separated; 52.1% of all U.S. cases are female, with 47.9% male. However, only 46.2% of COVID-19 related deaths are female, with 53.8% male[20].

To date, multiple double-blind placebo-controlled studies have been performed with various drugs to control the virus and reduce symptoms while the development of a vaccination is carried out. These include compounds such as Hydroxychloroquine, Lopinavir-Ritonavir, Dexamethasone, and Tocilizumab. According to multiple research studies conducted for each drug, none showed statistically significant improvements in the health of patients over placebos. As of October 22, 2020, the only drug to have gained approval from the FDA was Remdesivir, an antiviral agent which has been previously employed to treat Ebola[15]. In a study conducted by Beigel et al, researchers followed 1062 patients previously hospitalized with COVID-19 and measured the time to recovery as their primary outcome. Patients were assigned to either receive the placebo, or Remdesivir over 10 days. Researchers found that Remdesivir significantly improved the time of recovery, from a median of 15 days in the placebo group to 10 in the treatment group. In addition to this, the team discovered multiple secondary effects of Remdesivir, including lowering of respiratory tract infections due to the virus. The team hopes to continue the studies to find ways of mitigating some of the side-effects caused by the drug[21].

Aside from Remdesivir, the virus is currently untreatable with drugs or medications. However, hospital care and therapies for reducing the pain and severity of symptoms may help alleviate some of the distress caused by the virus. These include treatments such as supplemental oxygen or nonprescription pain medications.

As of July 14th, Jackson et al have completed a phase 1 dose-escalation trial of the candidate vaccine mRNA-1273. During this trial, researchers split 45 healthy adults into three dose groups (15 individuals each) – 25ug, 100ug, or 250ug injections of the vaccine. The mRNA sequence injected encoded from the S-2P antigen, a core component of the SARS-CoV-2 spike protein. This highly unique viral particle successfully stimulated the participant's immune systems at each dose. Titers collected from each participant showed heightened levels of anti-S-2P antibodies at higher doses. The researchers followed participants and reported that in all patients with successful antibody production, no adverse health defects occurred.

These two results indicate that the group's vaccine was not only effective but also safe. The group concluded by stating that a phase 2 trial for efficacy in random participants was already underway at the time of publishing and that a phase 3 trial for population efficacy and safety would begin by the end of the summer[22].

THE GLOBAL PANDEMIC

From December 20th, 2019 to December 27th, 4 cases of pneumonia of an unknown etiology were identified in people reporting to local hospitals near Wuhan China. Before the launch of the investigation, physicians had collected four lower respiratory tract samples from the patients for research. On December 31, the CCDC launched an epidemiological and etiological investigation with Hubei provincial and Wuhan health authorities. After discovering that the four samples collected could not be identified by the local hospital staff, researchers from the CCDC ran the samples through a High Pure Viral Nucleic Acid Kit manufactured by Roche. Nucleic acids were then tested for viruses and bacteria via PCR. Results showed no specific pathogen with the detection kits, but through Illumina sequencing and nanopore sequencing, viral reads were obtained with an 85% match rate with SARS-like CoV (bat-SL-CoVZC45, MG772933.1). RT-PCR on a special region of the Beta-CoV also matched with the novel SARS-like CoV. Finally, an electron microscope was used to visualize the virus in culture, physically confirming its presence[23]. On January 8th, once the virus was isolated, CCDC officially announced a novel coronavirus as the causative agent for the outbreak[24]. After this point, the WHO temporarily named 2019-novel coronavirus (2019-nCoV) to reduce stigmatization of the Wuhan region. Once assessed by the WHO, the virus was officially named SARS-CoV-2, and the disease-associated COVID-19.

By January 22nd, over 425 cases of novel coronavirus had been reported in Wuhan China, with other provinces and surrounding countries starting to report cases well[24]. As of January 30th, a total of 9,976 cases were reported in over 21 separate countries, including the United States, which first reported a case on January 20th. This patient was a 35-year old man who presented to a clinic in Snohomish County, Washington, on January 19th, following a family visit to Wuhan China[25]. With thousands of patients reporting every day, the WHO declared the COVID-19 outbreak a global health emergency on January 30th. As the virus proceeded to spread exponentially throughout the world's population, the WHO officially declared COVID-19 a global pandemic[15].

UNDERSTANDING THE NASOPHARYNGEAL MICROBIOME'S ASSOCIATION WITH HUMAN HEALTH

Research surrounding the gut microbiome has extensively shown that the microbes present interact with epithelial cells, and lymphocytes and may play an extensive role in the development of the immune system[26]. Further research in understanding the microbiota's effects on infectious disease has since extended into the nasopharynx. In fact, the nasopharynx has recently become increasingly recognized for its role in respiratory health. Specifically, it's been cited to influence things such as respiratory tract development and may even act as a carrier for opportunistic respiratory pathogens[27]. In addition to this, the nasopharynx is likely one of the first regions that the SARS-CoV-2 contacts after entry into the upper respiratory tract through respiratory droplets.

BACKGROUND — BIOLOGY *(CONT.)*

In models of host disease interaction with the influenza virus, some features of the nasopharyngeal microbiome may enhance infectivity and lead to more severe disease[28] while some interactions have been shown to be protective. Murine models with healthy nasopharyngeal microbiota have less severe disease and quicker clearance of viral particles than those who did not[29]. Because of the interactivity thought to exist between the nasopharyngeal microbiome, and the immune system, changes induced by viral infection and activity may pose a threat to our health. Therefore, understanding of the microbiome may help us with treating viral infections, particularly COVID-19.

As of the 5th of May 2021, three vaccines have been approved by the FDA for administration to the general public. These include the Pfizer-BioNTech's BNT162b2, Moderna's mRNA-1273, and Johnson & Johnson's (J&J) Janssen JNJ-78436435. Vaccines in phase 3 clinical trials for the United States include those being developed by AstraZeneca, and Novavax. Vaccines produced by Pfizer-BioNTech and Moderna are mRNA vaccines and require two shots; Pfizer-BioNTech shots should be separated by about three weeks, and Moderna's should be separated by about four weeks. The vaccine produced by J&J Janssen is a viral vector (DNA) vaccine and requires a single shot. Efficacy ratings during the stage three clinical trials for these vaccines were 95.0%, 94.1%, and 66.3% for Pfizer-BioNTech, Moderna, and J&J Janssen respectively[30–33]. It is important to note that none of these vaccines' efficacies were measured against one another – therefore there is no significant difference between their effectiveness (despite what the numbers may say). All three of the vaccines rose well above the FDA's 50.0% benchmark for approval and passed with "flying colors"[34].

Vaccine Brand Name	Who Can Get this Vaccine [1]	How Many Shots You Will Need	When Are You Fully Vaccinated?
Pfizer-BioNTech	People 16 years and older	**2 shots** Given 3 weeks (21 days) apart [2]	2 weeks after your second shot
Moderna	People 18 years and older	**2 shots** Given 4 weeks (28 days) apart [2]	2 weeks after your second shot
Johnson & Johnson's Janssen	People 18 years and older	**1 shot**	2 weeks after your shot

Figure 1: COVID-19 Comparison Table: Represents the three vaccines approved by the FDA and associated characteristics.

BACKGROUND — EPIDEMIOLOGY

CRIME RATE

Crime is directly tied to health when referring to violet crimes such as homicide. Currently crime is rated among the top causes of “worldwide loss of disability-adjusted living years,”[35] and cost the U.S. health services sector “$5.6 billion a year”[35] from 2008 to 2013. On an individual level, violent crime has been linked with characteristics like untreated childhood behavioral problems, risk-seeking behaviors such as drinking and smoking, as well as low IQ and poor education achievement[35]. Many theories on criminology have tried to assess the root of criminal behaviors, and their relation to both socio-economic and personality-related factors such as the ones described above. Aside from these, issues such as poverty, discrimination, and general inequalities can drive people towards crime. In “Merton’s theory of anomie, violent crime[35]” is a form of a rebellion. This idea parallels with relative deprivation “which suggests that such discontent motivates or disinhibits individuals to engage in violent acts[35].” Based on this, it is believed that those involved with crime are acting out of anger and revenge for systems of inequality. Partaking in crime however can have drastic effects on the individuals conducting the acts and those exposed to the crime.

Crime has been shown to be detrimental to the health of the individuals involved and their communities in several indirect ways. For starters, “children and adolescents exposed to violence are at risk for poor long-term behavioral and mental health outcomes regardless of whether they are victims, direct witnesses, or hear about the crime[36].” Individuals, especially children exposed to crime may interact less socially, or stop partaking in physical activities due to psychological stress. In some cases, behavioral problems such as anxiety and depression can surface[36]. These again can lead to worsening health outcomes, and also act as predictors of future criminal behaviors.

EDUCATION

Education is defined as being “both a process and a product[37].” As a process, education attainment is an event which occurs both in formal environments such as school, and work, but also within the home and one’s community. As a product of this process, education is also the “knowledge, skills, and capacities”[37] that are acquired by the person being educated. Based on this definition, education can also be defined as a personal attribute of an individual. Not only does one’s education encompass critical thinking skills, reasoning, and problem-solving, but also emotional awareness, control, and tolerance. When any one of these attributes are missing, an individual is lacking in a vital area of education. Deficits within these areas can lead to poorer health outcomes directly, and are correlated with other factors such as income, job security, and general health awareness[37,38]. Education is therefore a “central conceptual component and essential element of health… [and] education is a contributing cause of health[37].” It therefore necessary in order to maintain the health and wellbeing of a society, to also improve upon the education individual’s within said community receive. Without basic knowledge, the ability to reason, or a capacity for emotion, a person is unhealthy[37], and likely to live a life, and make decisions that further exacerbate this position.

Education in a broad sense is directly connected to health because it can help improve agency and responsibility for taking care of one’s body.

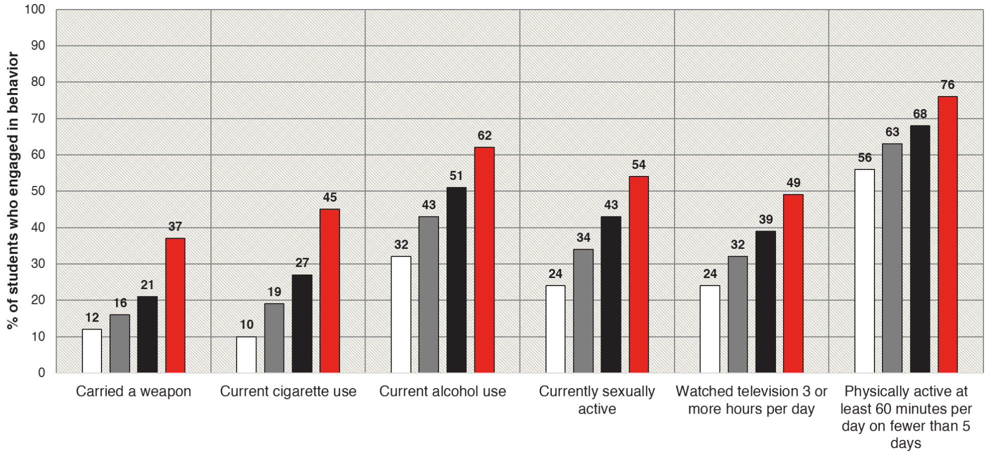


Figure 2: **Education & Student Behaviors: Represents data collected from the 2009 National Youth Risk Behavior Survey (YRBS) utilized by the CDC[37,39].**

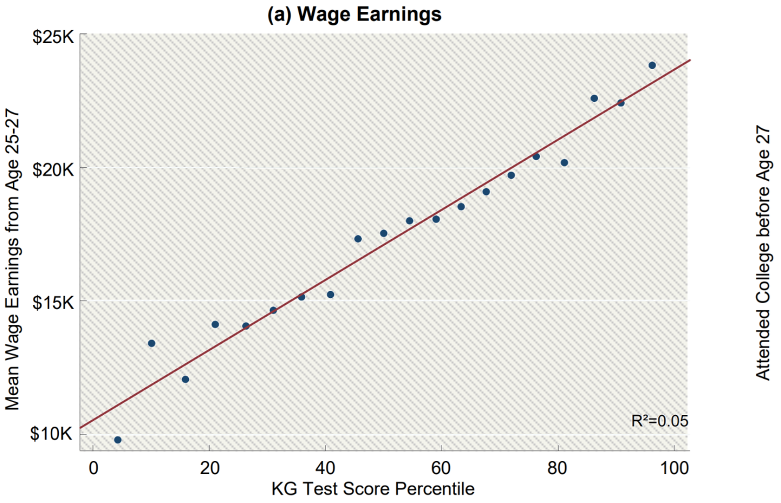
☐ Mostly A's
▨ Mostly B's
■ Mostly C's
🟥 Mostly D's/F's

When individuals are educated, that is when they are aware of what decisions will negatively impact them, they have the capacity to make a decision more beneficial to themselves. In an observational study conducted by the CDC, researchers investigated the association between grades achieved in school, and health risk behaviors. Data collected from the 2009 National Youth Risk Behavior Survey (YRBS), showed a negative correlation between grades and number of risky behavior participation (Fig. 2)[37,39].

Education is also highly associated with income, wages, access to healthcare, and access to healthy foods. In a report by Chetty et al., researchers showed numerous examples of how both test scores and grades positively correlate with mean age earnings, college attendance and even earnings after college[40]. In the figure below (Fig. 3), a linear regression of kindergarten test score percentile vs. mean wages in demonstrates this linear relationship.

An important factor to note is that education does not exist in a vacuum. Education is associated with social factors such as opportunity. While higher education based on grades can be associated with lower levels of risk-behaviors and higher earnings throughout life, it could be said that lower grades are the result of unhealthy behaviors. Furthermore, these unhealthy behaviors themselves could be caused by some other factors, such as upbringing, family social status/income, as well as relative levels of opportunity in childhood[38].

Figure 3: **Educational Outcomes Correlation to Wage Earnings: Represents data from the study conducted by Chetty et al. Researchers demonstrate a positive correlation between earnings through their mid-20s and KD test score percentile[40].**



BACKGROUND — EPIDEMIOLOGY *(CONT.)*

INCOME

Median household income is a measure often used by U.S. agencies in conjunction with family size and composition to determine what income threshold a family falls in. As defined by the U.S. Census Bureau, "if a family's total income is less than the family's threshold, then that family and every individual in it is considered in poverty[41]." Researchers have noted for decades that a positive correlation between income and health exists in the U.S. Specifically, this relationship has been shown when looking at mortality, chronic conditions prevalence, functional limitations (handicap impairments), and self-reported health status[42].

Underlying causal relationships between health and wealth are complex. This is due to the overwhelming number of factors linking health and household income. One reason often cited in literature points to the fact that those in poor health are usually unable to work or will miss multiple days greatly reducing income. Reduced income further exacerbates poor health has and individuals' abilities to take care of themselves[43]. Attending medical examinations, purchasing nutritious foods, and overall increases in stress levels are associated with poorer health outcomes, and higher incidence of disease and mortality[42].

Work-related factors such as unemployment can also act as proxies for poor health outcomes for the same reasons. Occupational injuries, and environmental hazards that otherwise diminish people's health all correlate then to decreases in income and further exacerbation of underlying conditions.

NUTRITION

Eating healthy is an important factor in maintaining one's health. Nutritious foods packed with essential vitamins, minerals, and an adequate amount of fats, sugars, and calories all contribute to an equal, balanced diet. Eating healthy lowers, the rates of chronic diseases, and can improve the immune system when fighting infectious disease. Eating unhealthy for long periods of time is highly associated with heightened prevalence of "high blood pressure, diabetes, and cancer[44]."

However, healthy eating is not always cheap or accessible. According to a review by Hilmers et al., "neighborhoods where economically disadvantaged and minority populations reside were more likely to have abundant sources of foods that promote unhealthy eating[45]." Socioeconomic, racial, and ethnic disparities within communities is associated with higher access to fast-food chains, and convenience stores. Furthermore, very limited access to grocery stores, and other outlets where healthier food-options are provided was the primary barrier for families and individuals to eat healthily[45].

In addition to the development of food deserts due to limited healthy options, access to grocery stores is also an area of concern. Low-income areas are especially in need of better transportation options to and from grocery stores; whether it's simply more frequent trips or improved bus lines. For example, bus lines that might target multiple neighborhoods at a time, thus reducing wait times. Proposed strategies for battling issues such as these include incentivizing buy healthy foods by reducing taxes associated with them, and deploying mobile markets to neighborhoods to sell fresh fruits, and vegetables[45].

TRANSPORTATION

Access to healthcare in the United States poses a problem to many, and this issue is worsened when adding in the variable of transportation. Access to transportation, and utilization of transportation is the difference between missing and making a doctor's appointment, filling a prescription, and even seeking emergency treatment. This can then lead to further exacerbation of preexisting conditions and worse health outcomes[46,47].

Having access to, or owning a vehicle is associated with the greatest utilization of health-related services, like those mentioned before. Additionally, having access to a family member or friend who owns a vehicle is correlated with higher access to care (although not as high as personally owning a vehicle). In a study conducted in 12 Appalachian counties in North Carolina, researchers found that of the 1,059 participants screened, having a driver's license, and having a relative or friend with a vehicle was highly correlated with the number of chronic and emergency visits (See Figure 4). In fact, it's reported that, "having a driver's license doubled the number of chronic care and regular care visits in a year[47]." Even when individuals were unable to access their own vehicle, being able to drive a friend or relatives, or being driven increased utilization of healthcare resources.

Within the same study, and others reported by Syed et al. use of public transportation was cited as an important tool for individuals to receive healthcare[46,47]. With that being said, utilization of a personal vehicles is likely a more reliable and convenient option for most households. When trying to make an appointment, work/life schedules, and bus/train schedules need to be taken into account.

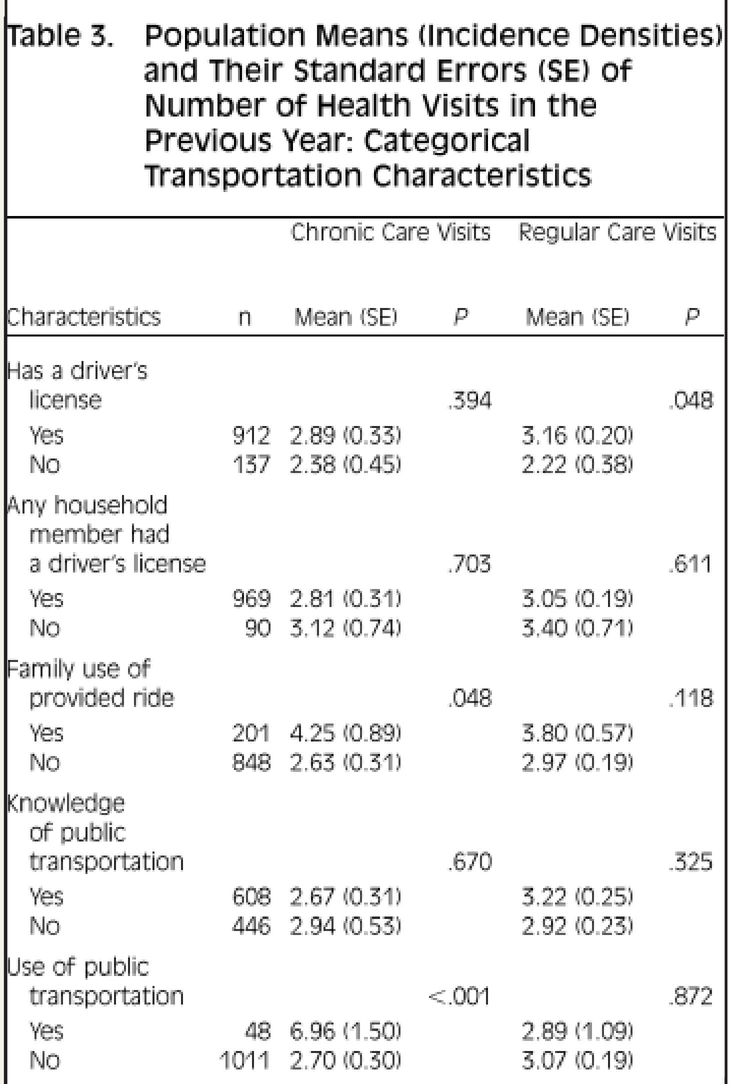


Table 3. Population Means (Incidence Densities) and Their Standard Errors (SE) of Number of Health Visits in the Previous Year: Categorical Transportation Characteristics

		Chronic Care Visits		Regular Care Visits	
Characteristics	n	Mean (SE)	P	Mean (SE)	P
Has a driver's license			.394		.048
Yes	912	2.89 (0.33)		3.16 (0.20)	
No	137	2.38 (0.45)		2.22 (0.38)	
Any household member had a driver's license			.703		.611
Yes	969	2.81 (0.31)		3.05 (0.19)	
No	90	3.12 (0.74)		3.40 (0.71)	
Family use of provided ride			.048		.118
Yes	201	4.25 (0.89)		3.80 (0.57)	
No	848	2.63 (0.31)		2.97 (0.19)	
Knowledge of public transportation			.670		.325
Yes	608	2.67 (0.31)		3.22 (0.25)	
No	446	2.94 (0.53)		2.92 (0.23)	
Use of public transportation			<.001		.872
Yes	48	6.96 (1.50)		2.89 (1.09)	
No	1011	2.70 (0.30)		3.07 (0.19)	

Figure 4: Transportation and Clinical Care Visits: Represents data collected by Arcury et al. describing the number of chronic and regular care visits with respect to transportation access characteristics[47].

In another study on transportation and health access, Rask et al. found that "walking or using public transportation to receive medical care was an independent predictor of not having a regular source of care"[46]. Public transportation is important for many individuals but doesn't provide as adequate access to medical services as would personal vehicles, relatives, and friends[48].

METHODOLOGY — BIOLOGY

PROJECT GOAL AND OBJECTIVES

The primary goals of this project were (1) to identify nasopharyngeal microbiome factors associated with COVID-19, and (2) to identify nasopharyngeal microbiome features associated with clinical outcomes in COVID-19. This team predicted that *features of the nasopharyngeal microbiome are associated with active COVID-19 infection and that nasopharyngeal microbiome features will predict progression to severe disease*. More specifically, this team's main hypothesis was that *there is an oral microbiome signature associated with the need for respiratory support*. To investigate the association between the nasopharyngeal microbiome, COVID-19, and disease progression/clinical outcomes, this project's objectives were to:

OBJECTIVE 1: Build a repertoire of at least 200 patients. (At least 100 COVID-19 positives, and 100 negative samples)

OBJECTIVE 2: Perform bulk RNA and DNA extractions on nasopharyngeal swab samples collected from patients

OBJECTIVE 3: Generate Illumina-sequencing libraries for all microbial DNA

OBJECTIVE 4: Map each individual microbial species present in samples

OBJECTIVE 5: Model how changes in the nasopharyngeal microbiome correlate to patient's clinical outcomes

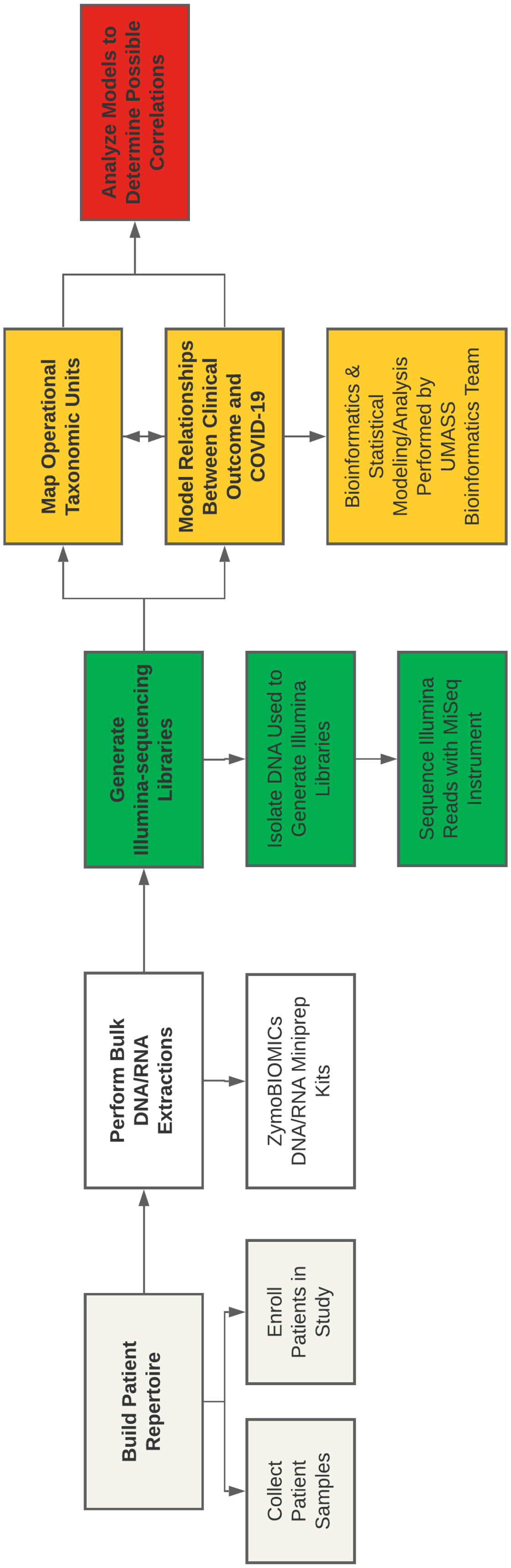


Figure 5: Microbiology Block Diagram - Represents the workflow for project goals and objectives. Colors represent distinct objectives.

CLINICAL DATA COLLECTION

To investigate possible relationships between the nasopharyngeal microbiome and COVID-19, a prospective cohort study was designed at the University of Massachusetts (UMASS) Memorial Medical School. Due to the COVID-19 Pandemic, the UMASS Memorial Emergency Department (ED) faced a high influx of patients, many of whom presented with COVID-19 or COVID-19 like-symptoms. Those individuals who were suspected to have COVID-19 were flagged by staff in their electronic medical record (EMR). Research assistants (RAs) in the ED, reviewed these EMRs and identified these individuals as potential study candidates if they ordered a COVID-19 test (and had yet to result), and met other preexisting criteria; the study excluded minors, pregnant women, and prisoners. To expedite the process of data collection, individuals noted as candidates had a nasopharyngeal swab added to their routine care by nursing staff in the ED. Individuals were contacted by RAs later via telephone to obtain informed consent. All individuals who wished to participate had their swabs stored, and RAs started to document their cases. This included personal health information (PHI) such as the outcome of their COVID-19 test, past medical history, medications, microbiology labs, as well as demographics, and other related information. Patients who did not give consent or were unreachable had their swabs destroyed, and no PHI was recorded. Since the study sought to investigate correlations between COVID-19 and respiratory outcomes, data was collected from all patients regardless of the COVID-19 test result. Patients whose samples were negative were used as controls. The figure on the next page (Figure 6) was adapted from the original research grant for the study.

METHODOLOGY — BIOLOGY *(CONT.)*

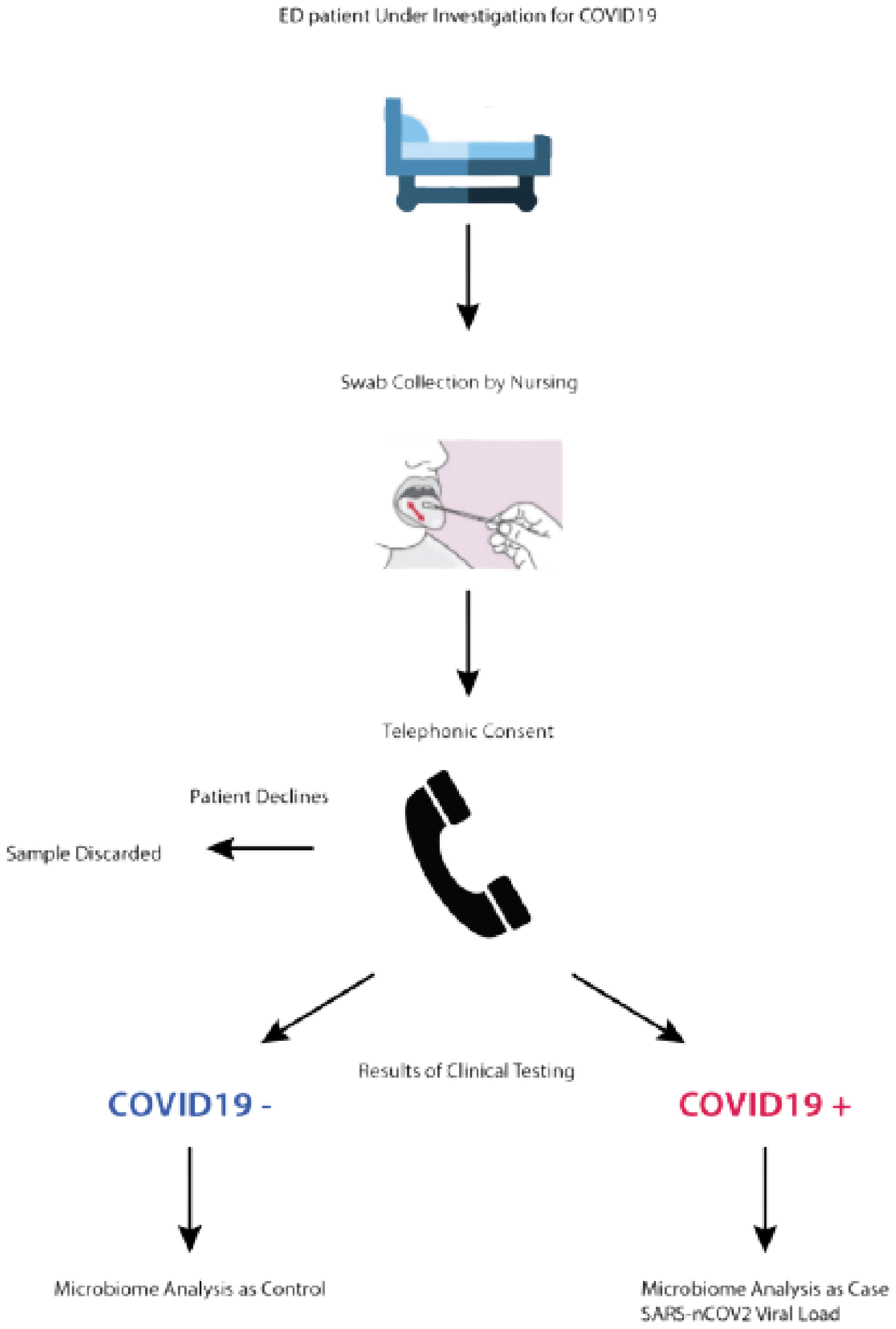


Figure 6: Patient Enrollment Flow Diagram: Represents the process of patient enrollment into the study. Adapted from grant proposal by Dr. Evan Bradley (M.D., Ph.D.) at UMASS Memorial Medical School

NUCLEIC ACID EXTRACTION

In order to collect more nuanced data covering multiple kingdoms of microbiota, bulk DNA, and RNA isolation was carried out on each of the samples, using ZymoBIOMICS™ DNA/RNA Miniprep Kit. To ensure the stability of the recovered nucleic acids, the DNA/RNA extractions were recovered separately in microfuge tubes, homogenized with RNALater and stored at -80°C. DNA samples were recovered as needed for Illumina sequencing.

ILLUMINA SEQUENCING

Metagenomic sequencing was carried out via the Illumina sequencing technique using the Illumina MiSeq instrument. Following Illumina MiSeq sample preparation, obtained PCR products were compared directly with those in GenBank by BlastN search, available from NCBI. Additionally, samples were also compared against the Ribosomal Database Project RDP II Match Facility from Michigan State University.

OTUs AND DIVERSITY SCORING

Each species of the microbiome detected in samples was defined as an operation taxonomic unit (OTUs) for ease of analysis. For example, if both E. coli, and C. elegans were present in samples, they may be defined as OTU-1 and OTU-2. Mapping of microbial species began by calculating the alpha and beta diversity the study sample.

Alpha diversity was calculated to determine the mean species diversity of a particular habitat. The relative diversity (mean diversity) generated tells how rich a specific sample being analyzed is. A higher alpha diversity indicates a more diverse microbiome population (in the context of this study). High diversity within a patient is a positive result since higher levels of diversity allow for broader comparisons across patients. This value helped accomplish the first goal, *to identify nasopharyngeal microbiome factors associated with COVID-19*. Determining the species diversity of a particular locale is essential for comparing diversity across habitats. Comparisons of species between habitats generates a value known as beta diversity.

Beta diversity was calculated to determine the dissimilarity of species between habitats (in this case between patients). True beta diversity is a ratio of regional species (the diversity of the two habitats being compared), and the local diversity for one of the habitats in the ratio. By determining the differences in nasopharyngeal microbial species between patients we began to model correlations between clinical outcomes and the microbiota. For example, if two patients with similarly diverse microbiota entered the ED, and one patient was intubated while the other received a nasal canula, modeling of beta diversity can show whether a particular microbe may be associated with that outcome. Therefore, this value helped accomplish the second goal, *to identify nasopharyngeal microbiome features associated with clinical outcomes in COVID-19*. Of course, there are many other factors associated with clinical outcomes; this is why a model was used to control for specific conditions and OTUs.

MODELING RELATIONSHIPS BETWEEN COVID-19, THE NASOPHARYNGEAL MICROBIOTA AND CLINICAL OUTCOMES

In order to perform higher level analyses and to answer questions such as the hypothesis, a machine learning algorithm was created to model the relationships between OTUs and clinical factors with respect to COVID-19.

METHODOLOGY — BIOLOGY *(CONT.)*

The algorithm was trained using half of the data from the data set, and the other half was used for analysis. The algorithm was first utilized to generate an F1 score, to understand the predicative power of the model.

Following this, the model was used to specifically test for a dichotomous outcome – whether patients would need respiratory/oxygen support or not. The model was used to generate a *median importance score* for either an OTU or a clinical feature. A graph was generated to show the relative "predictive power" of each of these variables with respect to one another. The predicative power of these variables did not tell whether the OTUs or clinical features predicted if the patient would need oxygen or not. Rather, the model showed the level of importance of the value overall. For example, in determining the severity of a fall, a doctor will often ask for a patient's age – this is because age is a great predictor of how severe the damage from a fall might be – age might be placed higher on the list of relative predictors than something like when was *the last time the patient ate*. Similarly, the OTUs and clinical features (including age) modeled using this algorithm were analyzed to see how important they were for determining if a patient would need respiratory support.

Afterwards, the model was used to show the relative abundance of microbes which scored highly in *median importance*. This model showed relative abundance against the dichotomous outcome – whether patients needed respiratory support or not. This information was used to determine the correlation that microbes have with respect to COVID-19's effects on respiratory deterioration following recovery.

METHODOLOGY — EPIDEMIOLOGY

PROJECT GOAL AND OBJECTIVES

The primary goal of this project was to determine whether social factors commonly associated with poverty were also correlated with COVID-19 case prevalence in Worcester, MA. In order to accomplish this, this project's objectives were to:

OBJECTIVE 1: Determine the most cited social factors/measures for poverty and health

OBJECTIVE 2: Establish effective strategies for measuring both, the prevalence of these social determinants on families & COVID-19 prevalence in Worcester, MA

OBJECTIVE 3: Measure the correlation between the social factors determined and COVID-19 prevalence

To determine the social factors that affected people in poverty the most, a literature review was performed. In order to gain insight into what other researchers determined as important, this literature review mostly involved peer-reviewed research articles, as well as an Interactive Qualifying Project from WPI that focused on poverty prevention. After analyzing this information, a list of ten social factors was generated. These included average family size, best schools, crime rate, food deserts, income below poverty, less than high school education, meals from fast food, median household income, transportation, and unemployment. Using data from public sources, such as the U.S. Census Bureau, correlations were drawn between the prevalence of COVID-19 and the social factors in specific areas of Worcester.

DETERMINE SOCIAL FACTORS RELATED TO POVERTY AND HEALTH

In order to determine what social factors were most related to poverty and health, I examined peer-reviewed articles published on the subjects. While reading the articles, I maintained a list of factors that came up most often. After a few weeks of reading articles, I needed to verify if I would be able to find data on these factors in the City of Worcester. To complete this, the second objective (establishing a strategy) was performed largely at the same time as identifying social factors.

ESTABLISH MEASUREMENT STRATEGIES FOR SOCIAL DETERMINANTS AND COVID-19

While collecting information on common social determinants of health, I checked if data collected by City-Data (which utilizes U.S. Census, municipal and state survey data collected from 2000 to 2020) had been used to generate a heat map of the social factor being reviewed. Maps from other sources such as Neighborhood Scout, and the U.S. Department of Agriculture, Economic Research Service Department. If these sources had a map with data for the City of Worcester, the social factor was retained for use in correlation with COVID-19 prevalence.

After collecting data on social determinants of health in Worcester, I found a map generated by the Worcester Department of Public Health that tracked COVID-19 prevalence throughout the pandemic. This map and the social determinant heat maps were used to investigate possible correlations between the prevalence of COVID-19 and common social inequities.

METHODOLOGY — EPIDEMIOLOGY *(CONT.)*

IDENTIFY CORRELATIONS BETWEEN SOCIAL DETERMINANTS AND COVID-19 PREVALENCE

In order to properly analyze the maps both qualitatively and quantitatively, a region on the COVID-19 prevalence map was drawn in the area with the highest rate of COVID-19 cases. This region, labeled as the *Worcester COVID Belt*, was then analyzed concerning the census blocks on the maps from City-Data. To observe differences in the prevalence of social inequities related to COVID-19, two maps were selected for each social determinant (when data came from City-Data). A census block within and outside the Worcester COVID Belt was selected.

Data in these blocks showed the number of individuals data was collected from, and from how many households. Additionally, it showed the number of individuals facing a particular social determinant by percentage, or count. For example, data on *income below poverty* was examined for 2 census blocks, one inside and one outside the Worcester COVID Belt. Within each of these census blocks, a particular number of families answered the survey questions, and data from these questions were used to determine the average number of people living below the poverty line. On the map, shading is used to represent differences between census blocks with respect to that social determinant. Using the exact numbers for the census blocks, and the shading, I was able to draw correlations between the social determinants examined and COVID-19 prevalence. Correlations were qualitative and drew on the fact that higher regions of COVID may align with areas where families suffer more often from particular social inequities.

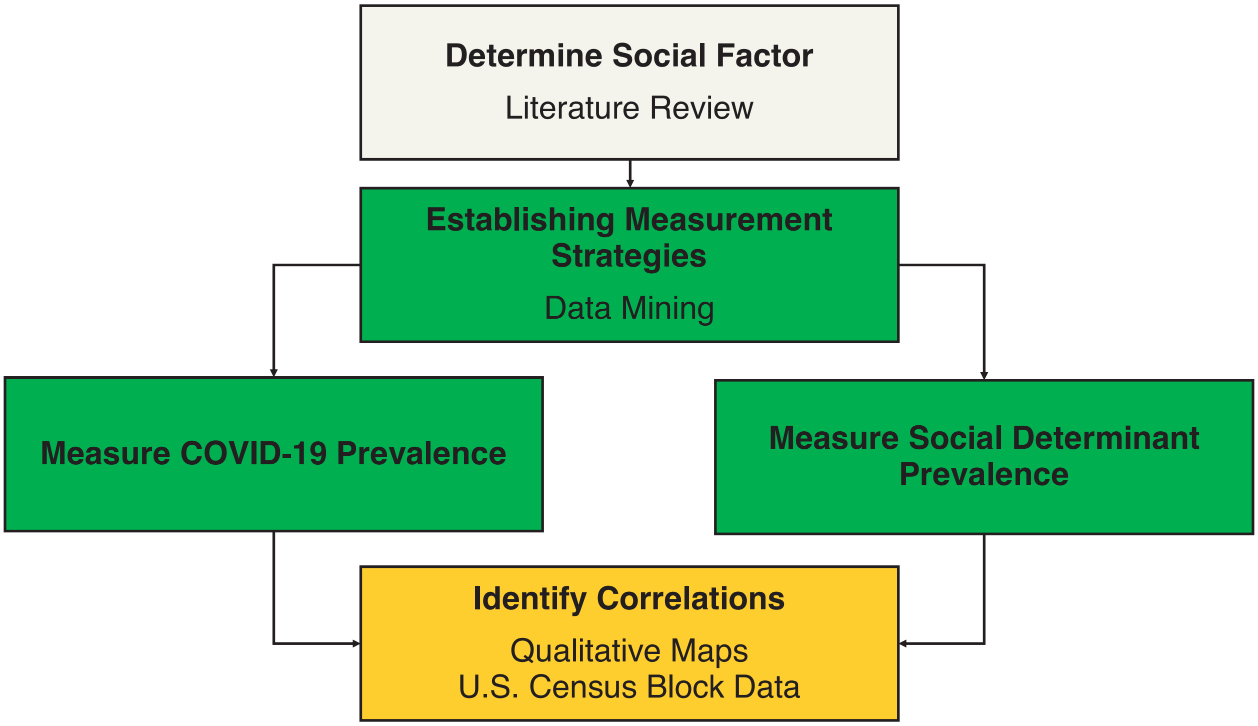


Figure 7: **Epidemiology Block Diagram - Represents the timeline for project objectives and path of action. Colors represent each distinct objective.**

RESULTS — BIOLOGY

DIFFERENCES IN MICROBIAL SPECIES' ABUNDANCES MAY ACT AS BETTER HEALTH PREDICTORS THAN AGE, BMI, AND CCI

The machine learning algorithm was used to model *predictor vs median importance* for selected microbes, clinical outcomes, and a handful of demographics related to the patient pool. Based on the algorithm's training, it generated an F1 score of 0.85, indicating that 85% of the time it would correctly assess the importance of a particular microbe, clinical outcome, or characteristic.

The model was able to assess the "weight" that each of these microbes, clinical factors, and demographics had in determining if an individual's respiratory condition would worsen following COVID-19. Specifically, the model was able to arrange values by the relative level of importance in establishing whether the patient would need oxygen or not following recovery. As was mentioned in the methodology, the use of age to assess falls is analogous here to the use of body mass index (BMI) or Charles Comorbidity Index (CCI) to assess whether respiratory support would be needed. For this study, we tried to assess whether microbes held similar weights to some of these clinical outcomes.

Outcomes are arranged top to bottom on *Fig. 8 Graphs 1-2*, by relative importance. Predictors that appear further on the right have a higher relative importance as predicted by the algorithm. *Graph 1* includes race, while *Graph 2* does not, and therefore acts as a control for this factor.

Based on the model generated from the patient data set *(Fig. 8)*, one microbe *(P. salivae)* acted as a best predictor for the dichotomous outcome – the need for oxygen or not – scoring better than age, BMI, CCI, race, and many other microbes. Additionally, as shown in Fig. 8 many different microbial species also acted as better predictors for the dichotomous outcome than CCI and race. These included C. *concisus, V. infantium, A. sp.S6 Spd3, G. elegans, A. sp. oral taxon 458, P intermedia,* and *H. sp. HMSC71H05*.

RESULTS — BIOLOGY *(CONT.)*

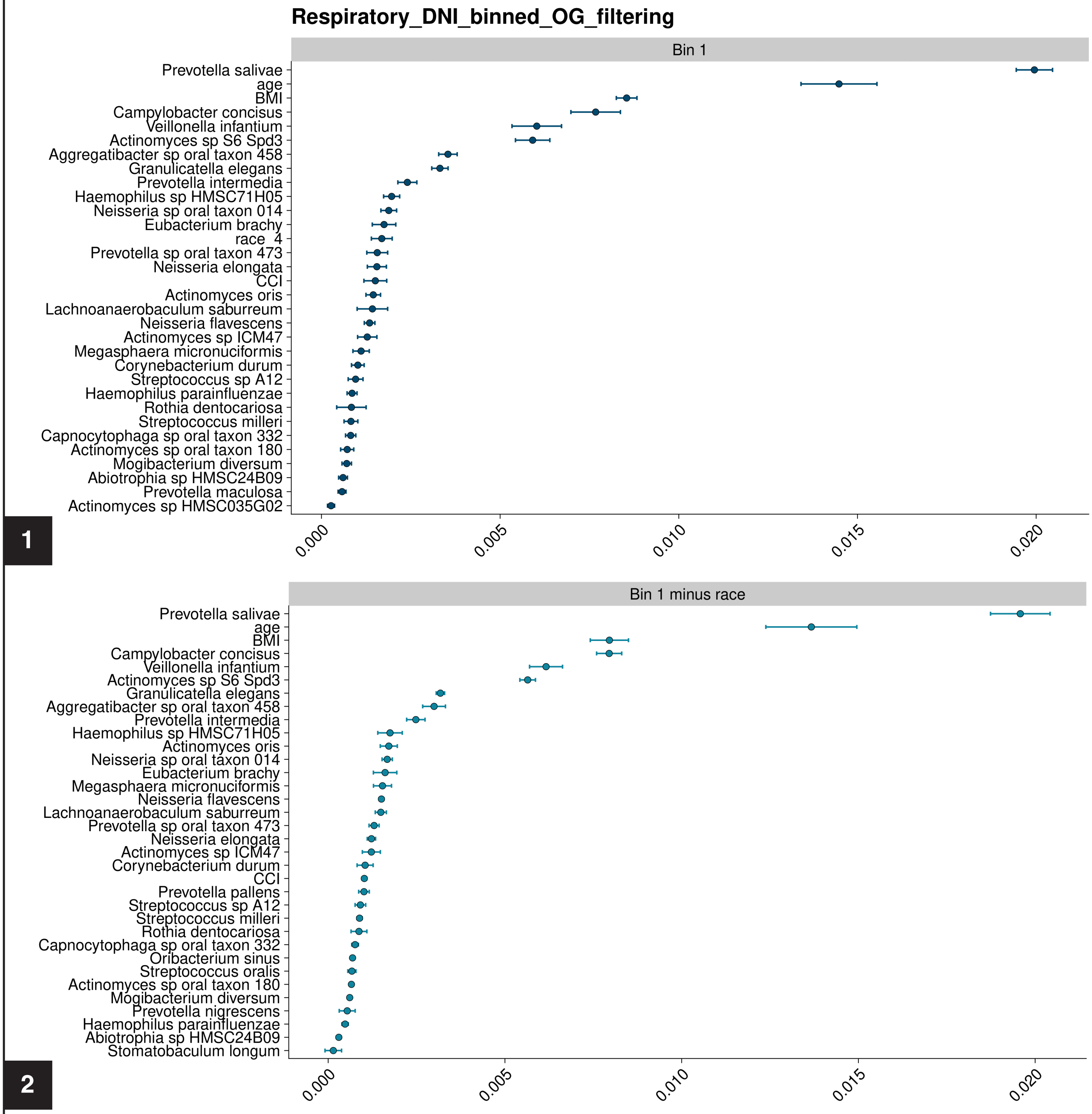


Figure 8: Predictor (Y) vs. Median Importance Graphs (X): Represents the relative importance of each health-related feature with respect to a dichotomous outcome: Need for respiratory/oxygen support or not. Graph 1 includes race in the series of predictors; graph 2 does not include race. *Prevotella salivae* was the best predictor of the need for respiratory support or not in both bins. Following *P. salivae* were age, BMI, and several other microbes. In Graph 1, when race was accounted for, it acted as a better predictor than more than half of the microbial features. In both Graphs 1-2, the Charles Comorbidity Index (CCI) was ranked below several microbial features, age, and BMI. In Graph 2, a higher number of microbial features acted as better predictors of the need for respiratory support or not over CCI than when race was accounted for.

RESULTS — BIOLOGY *(CONT.)*

CERTAIN MICROBIAL SPECIES MAY EXHIBIT PROTECTIVE ATTRIBUTES WITHIN THE NASOPHARYNX

The machine learning algorithm was also used to model *relative microbial abundance vs. dichotomous outcome*. As shown in *Fig. 9* the model was able to access the level of abundance of each microbe within patients with respect to the outcome they eventually had. The model uses box and whisker plots to display the data. Each dot represents a single patient; dots that are higher indicate higher level of that microbe in that patient. 0 is representative of those patients who did not go on to need respiratory/oxygen support. 1 is representative of those patients who did eventually need respiratory/oxygen support.

It's important to keep in mind that there is no significant difference between 0s and 1s; we are not comparing differences in abundances between the groups. We are determining if differences between the groups predicts a particular outcome. Additionally, due to speed at which data was processed, each microbe is on a different scale.

The top eight microbes were chosen to show the model's ability to measure their abundance with respect to the patient outcome. This includes the microbe *P. salivae*, the best predictor of patient respiratory distress. For each of the eight microbes *(Graphs 1-8)* shown in *Fig. 9* their relative abundance is high for patients who did not experience a need for respiratory support. This shows us that higher level of these particular microbes is correlated to patient's lack of a need for respiratory support.

RESULTS — BIOLOGY *(CONT.)*

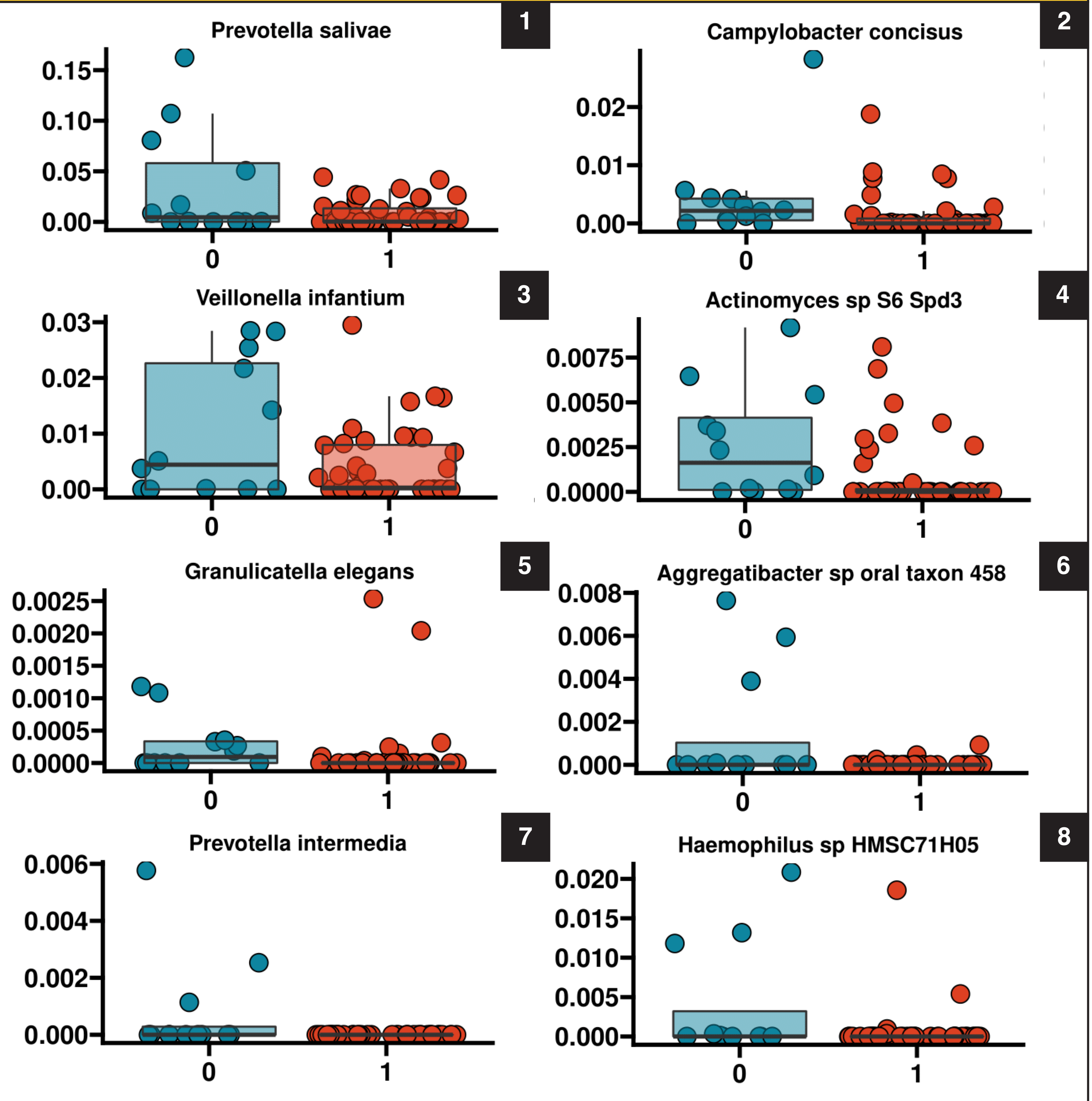


Figure 9: **Microbial Feature Abundance (Y) vs. Dichotomous Outcome (X): Need for respiratory support or not – Represents the relative abundance of particular microbial features associated with either no need for respiratory support (0, blue), or the need for respiratory support (1, red). Box-and-whisker plots should not be compared within the same graphs, as there is no statistical significance in their differences. Additionally, each graph is associated with its own relative scale. Rather, each relative abundance and outcome should be viewed independently of one another. Graph 1 *(P. salivae)* shows that a higher abundance of microbe in patients is associated with an absence of the need for respiratory support. Similar findings are also seen for the microbial species in Graphs 2-6, 8. In Graph 7 differences in abundances of bacteria were poorly associated with both the need or the absence of the need for respiratory support.**

RESULTS — EPIDEMIOLOGY

WORCESTER COVID BELT

Worcester is a city in central Massachusetts within Worcester County. According to data from the U.S. Census Bureau Population Estimates Program data, as of July 1st, 2019, the city of Worcester had a population of 185,428. Trends suggest that this has increased over the past 2 years[49]. Based on data collected from the Worcester Department of Public Health, officials were able to generate a heat map depicting COVID-19 case density within the city as shown in *Figure 10*. Darker regions in the map are indicative of higher prevalence of COVID-19. Regions with no dots within the city show regions where there are no residential areas, or where data was not significant. The red circle highlights the region with the darkest regions of a map, and thus represents the *Worcester COVID Belt*. This region stems from the southcentral region of the city to the northeast before ending at the city limit[50]. This region of high COVID-19 density lines up with a high level of accuracy with other map showing different social determinants of health. Determinants with two maps include data aggregated from the U.S. Census Bureau surveys delivered between 2000-2020.

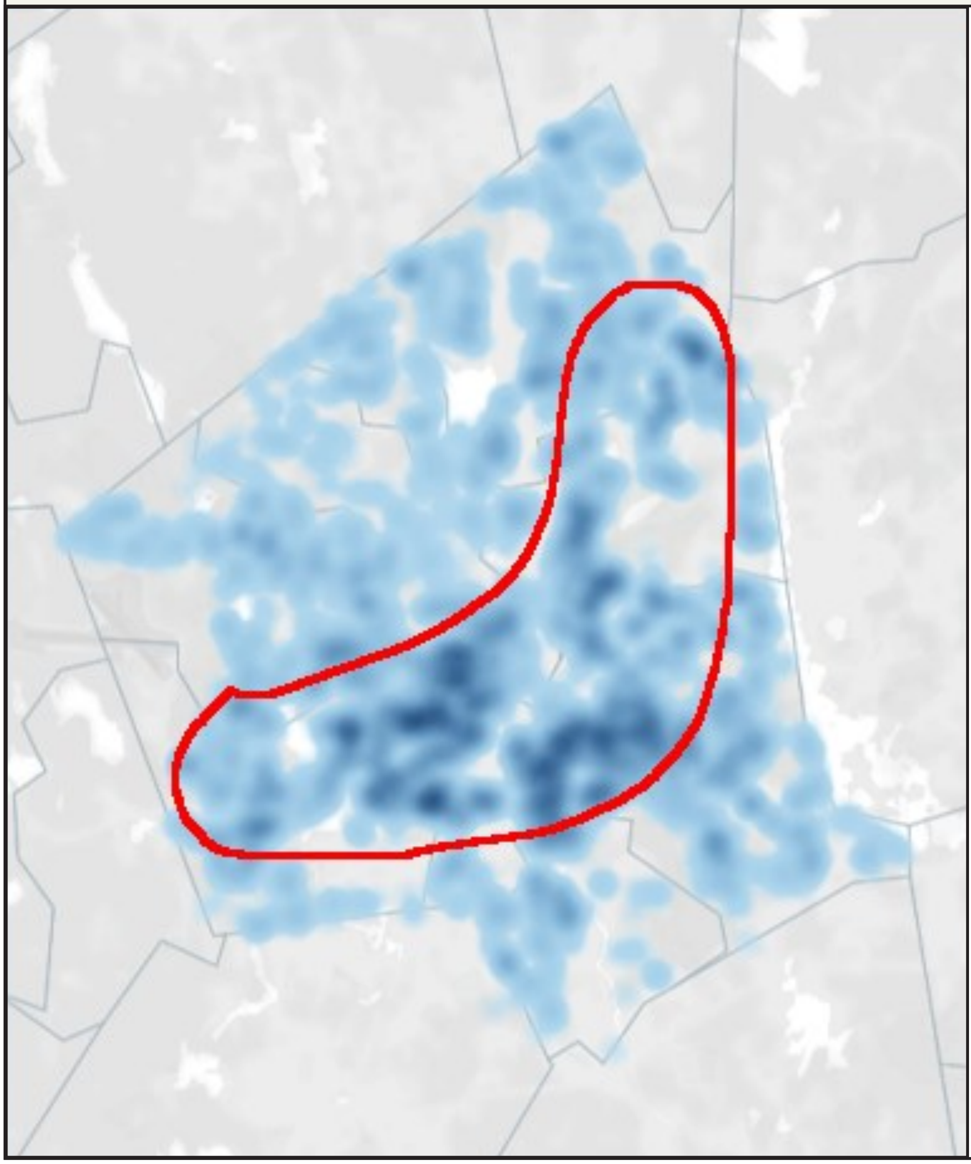
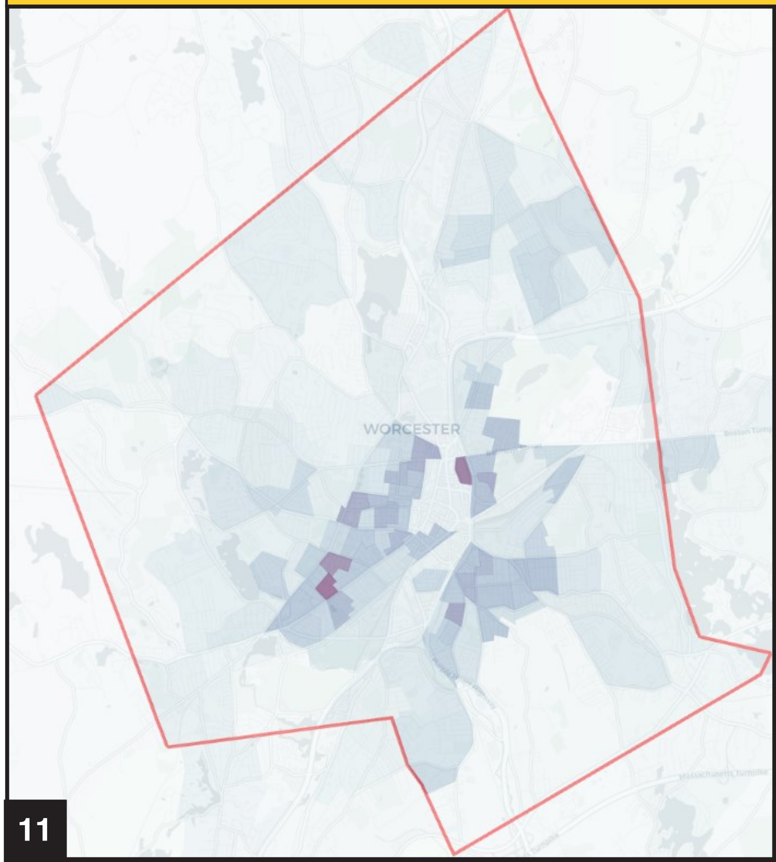
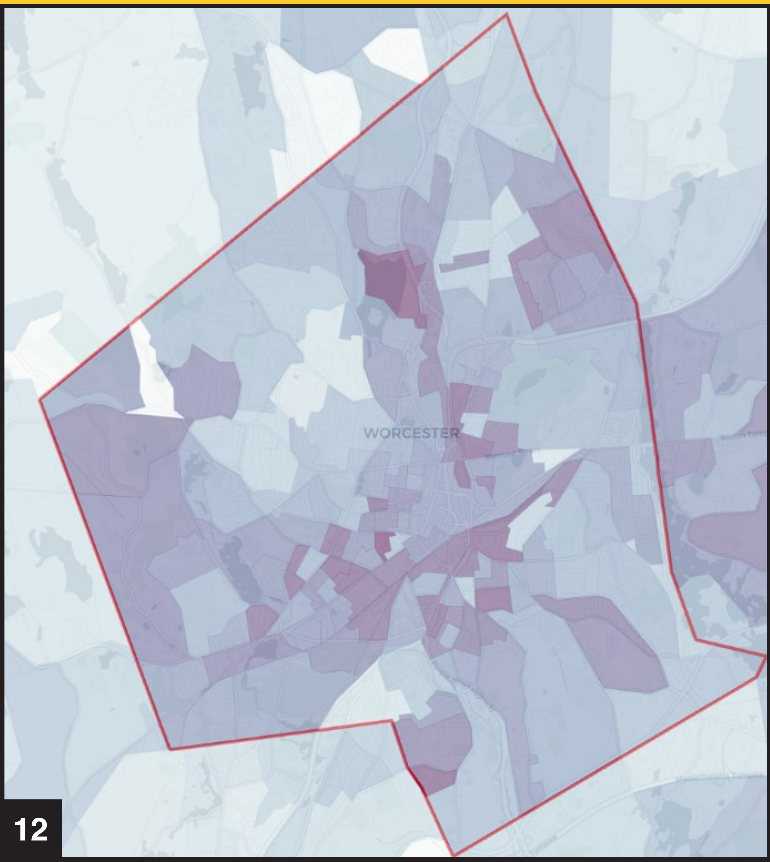


Figure 10: Worcester COVID Belt: Represents COVID-19 cases by geographic dispersion. Darker regions indicate higher prevalence of COVID-19. Redline surrounds the areas with the highest prevalence of COVID-19 cases, dubbed the "Worcester Covid Belt."

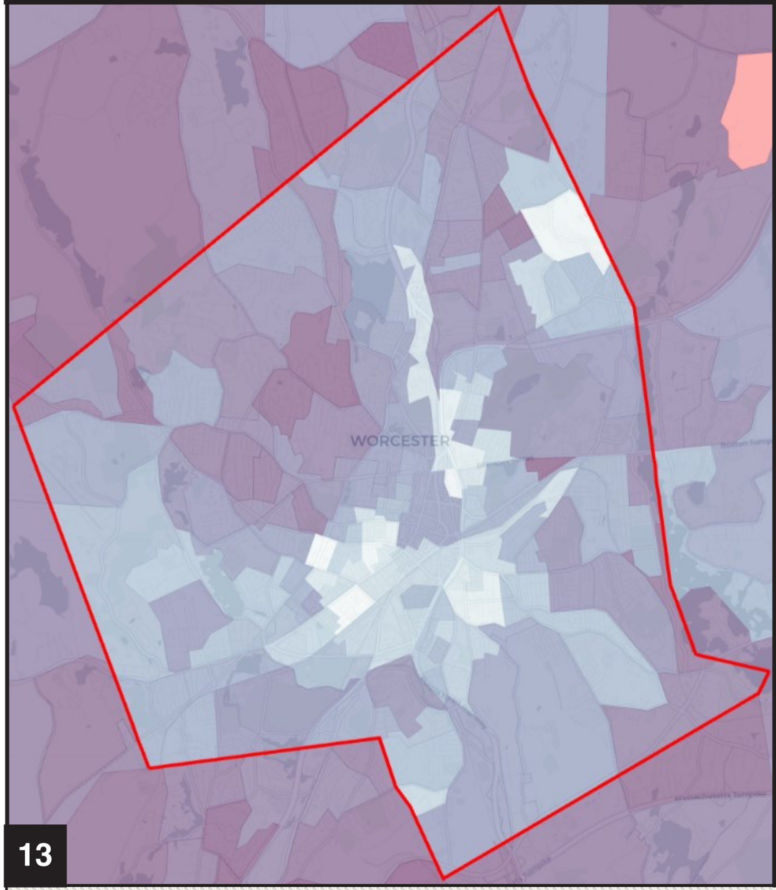
RESULTS — EPIDEMIOLOGY *(CONT.)*



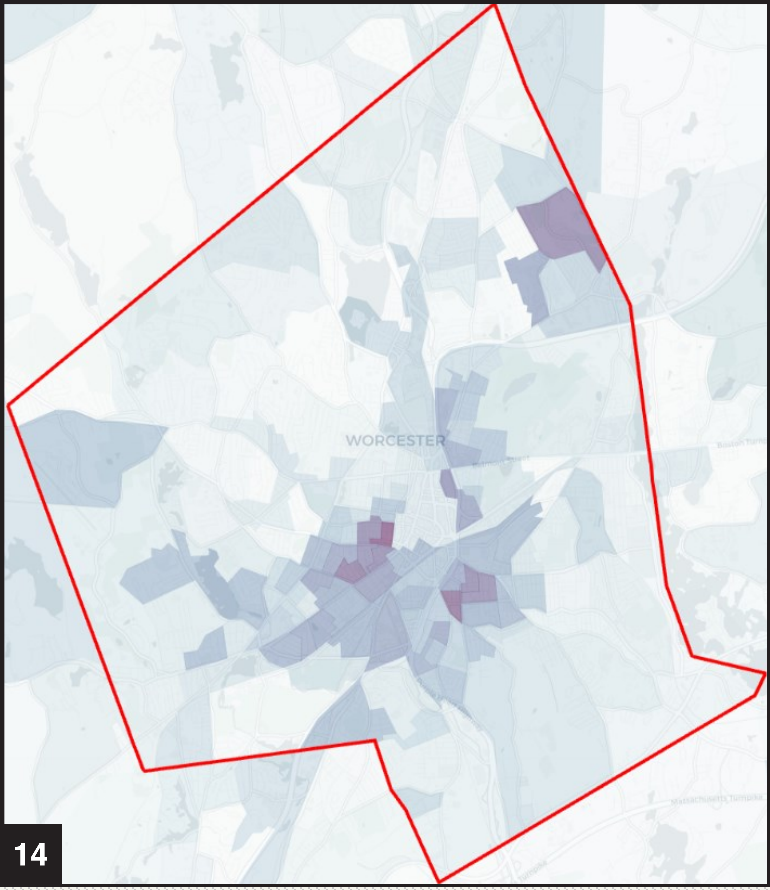
11



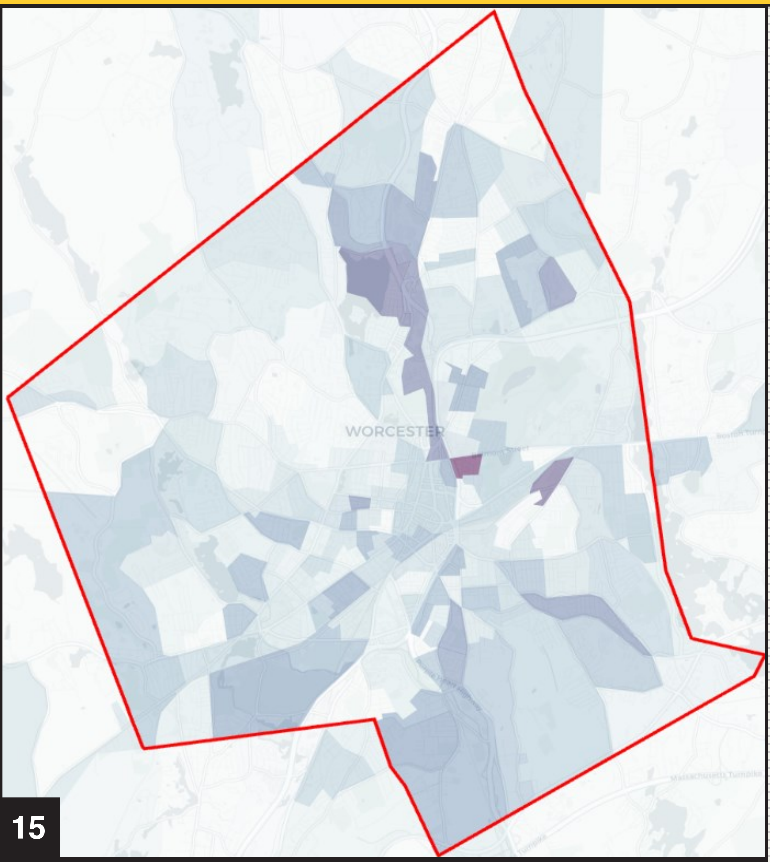
12



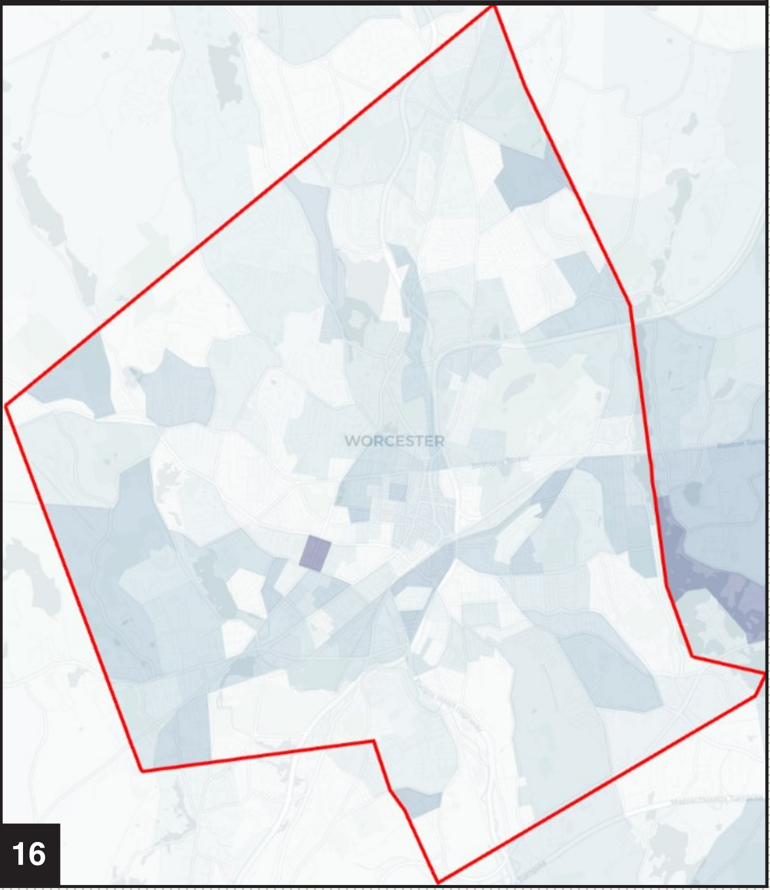
13



14



15



16

DEMOGRAPHICS

Based on data presented in *Figures 11-16*, the highest population density *(Fig. 11)*, and most racially diverse region *(Fig. 12)* lie in the center of the *Worcester COVID Belt*. Blocks within the center of the belt have the fewest percentage of white individuals, and the highest percentage of Hispanic individuals. Areas closer to the center also have a relatively higher percentage of black individuals, though not nearly as much as the Hispanic population. Finally, blocks with the highest percentage of Asian individuals appear at the southern end of the *Worcester COVID Belt*, and along the western city limit.

Figure 11: Demographics 1: Represents population density of Worcester. Darker regions indicate areas with a higher number of residents.

Figure 12: Demographics 2: Represents total racial diversity. Darker regions indicate a higher level of racial diversity.

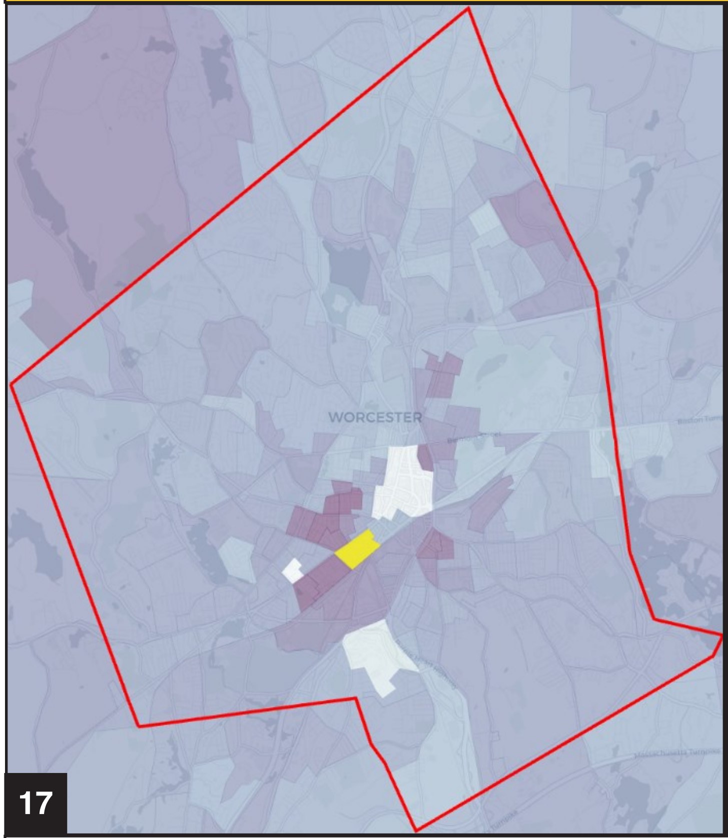
Figure 13: Demographics 3: Represents the percentage of white individuals/households in the census block. Darker regions indicate a higher percentage of white individuals.

Figure 14: Demographics 4: Represents the percentage of Hispanic individuals/households in the census block. Darker regions indicate a higher percentage of black Hispanic individuals.

Figure 15: Demographics 5: Represents the percentage of black individuals/households in the census block. Darker regions indicate a higher percentage of black individuals.

Figure 16: Demographics 6: Represents the percentage of Asian individuals/households in the census block. Darker regions indicate a higher percentage of Asian individuals.

RESULTS — EPIDEMIOLOGY *(CONT.)*



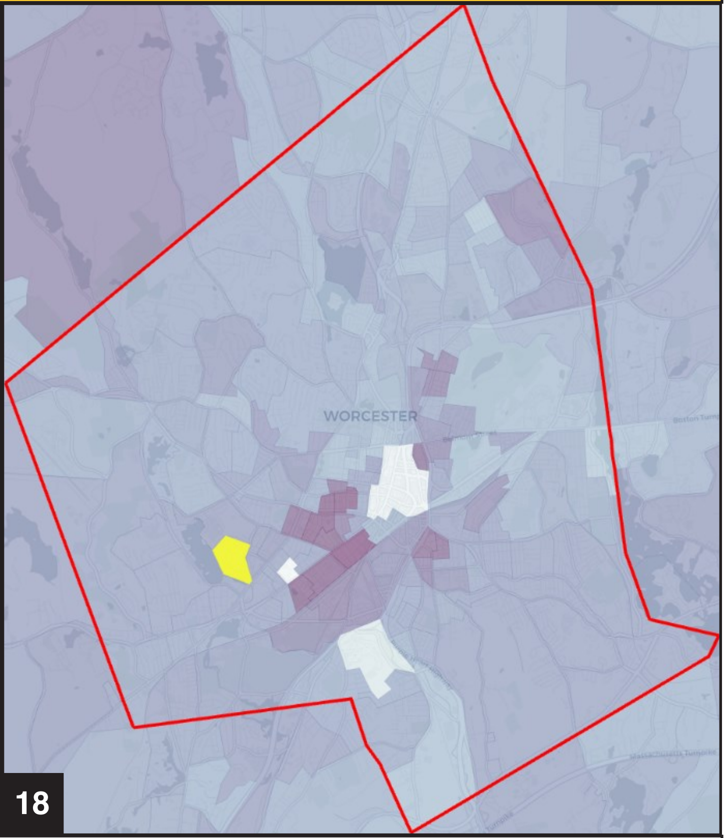


Figure 17: **Average Family Size 1: Represents census block 25027.731300.3 (yellow). Darker regions indicate larger average family size.**

Figure 18: **Average Family Size 2: Represents census block 25027.731101.4 (yellow). Darker regions indicate larger average family size.**

AVERAGE FAMILY SIZE

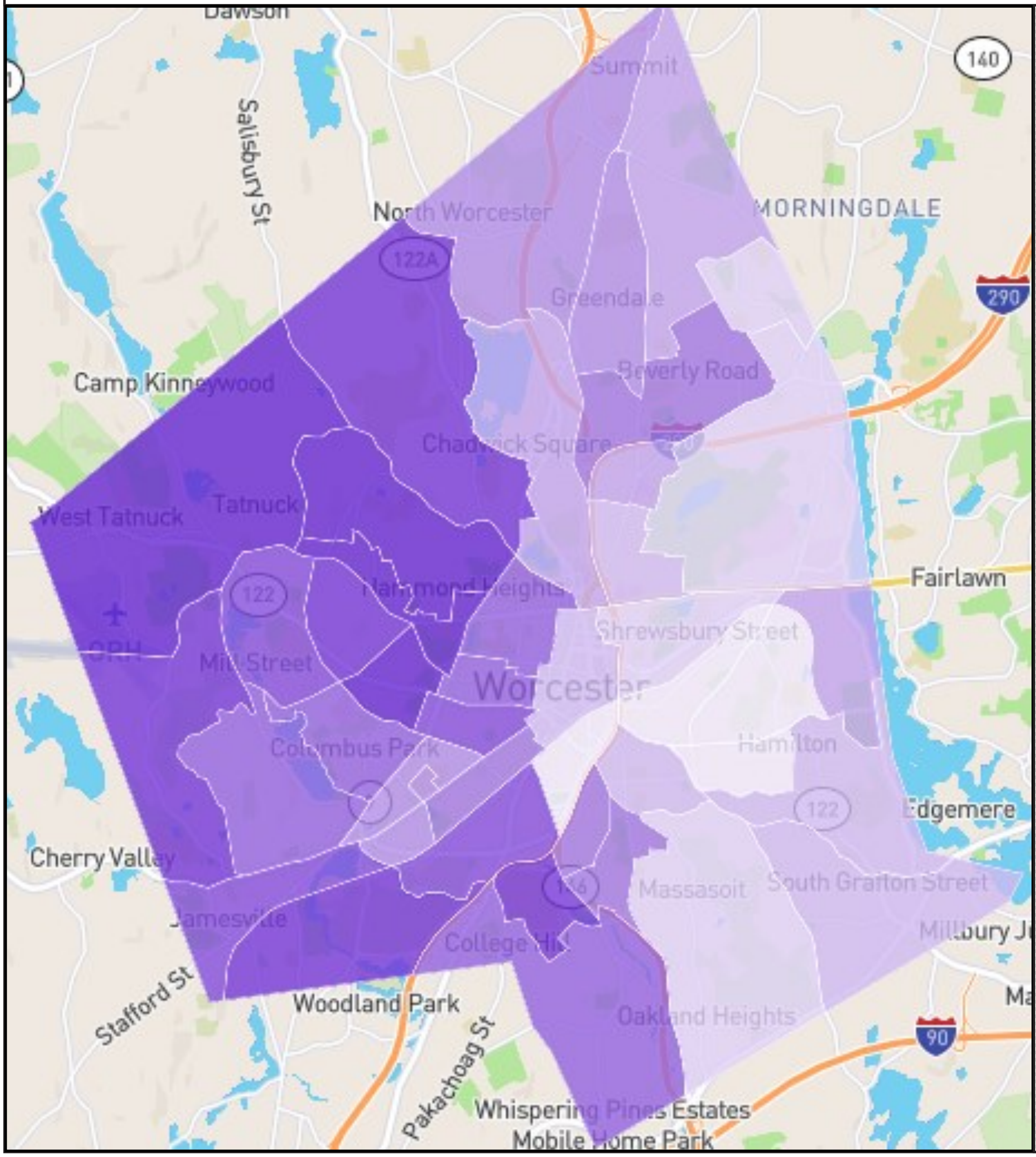
Average family size (# members per household) *(Figs. 17,18)* positively correlates with COVID-19 density shown in *Figure 10*. Census block 25027.731300.3 *(Fig. 17)* is located in the central area of the *Worcester COVID Belt*, and includes data from 576 households. Average family size within this block is 3.87. Contrasting this is census block 25027.731101.4 *(Fig. 18)*, located just northwest of the belt. Data collected from this block came from 643 households and yields an average family size of 2.78.

RESULTS — EPIDEMIOLOGY *(CONT.)*

BEST SCHOOLS

Data presented in *Figure 19* qualitatively shows areas in higher proximity to better schools based on test scores. Shading on the map positively correlates with COVID-19 density *(Figure 10)*. Darker regions (where schools where students have higher test scores) mostly appear in regions surrounding the *Worcester COVID Belt*. Central regions of the belt are much lighter indicating less proximity to schools with students who have higher test scores.

Figure 19: Best Schools: Represents best schools by test scores map of Worcester. Darker regions indicate areas with closer proximity to better schools.



RESULTS — EPIDEMIOLOGY *(CONT.)*

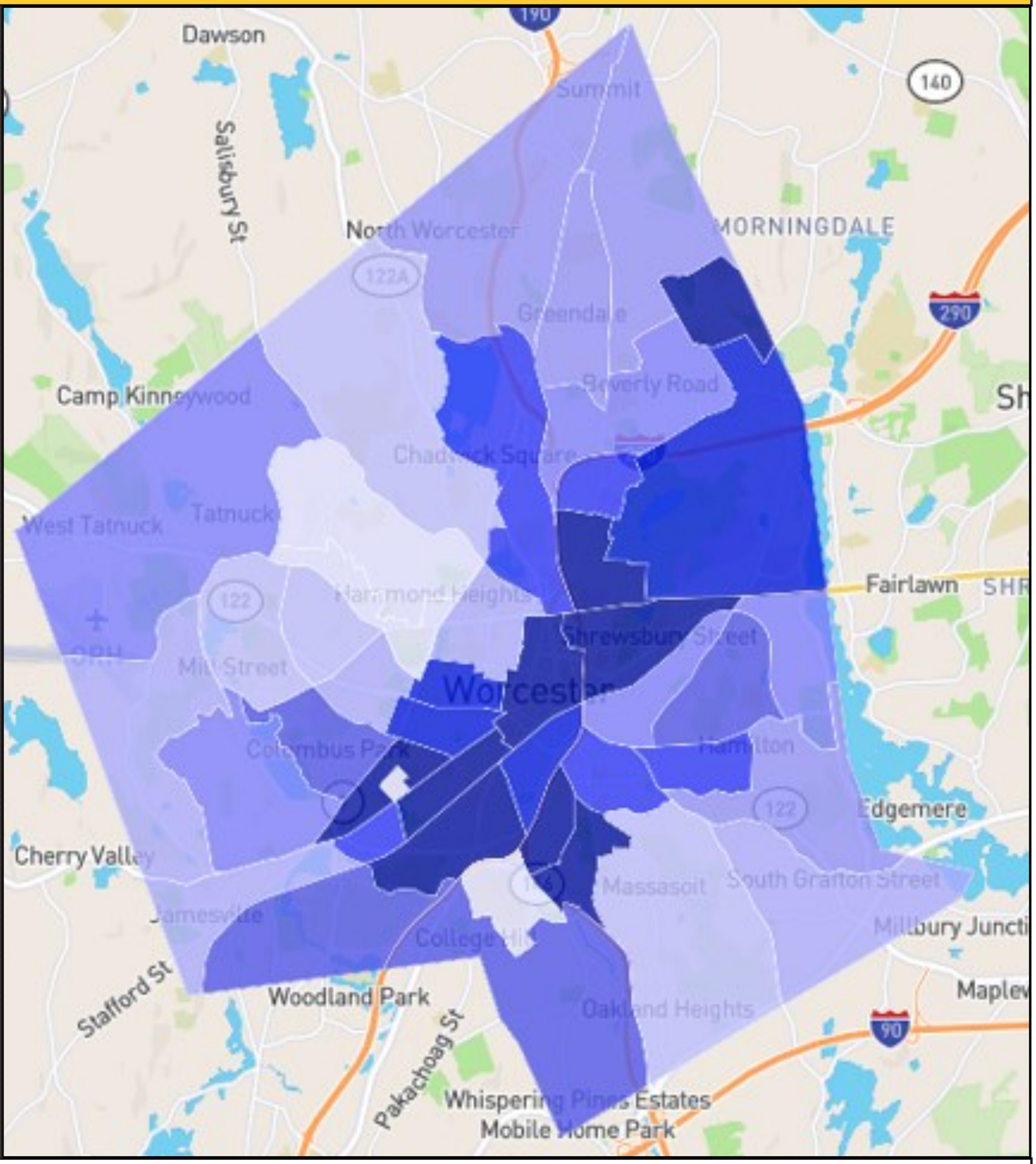


Figure 20: **Crime Rate: Represents crime rate map of Worcester. Darker regions indicate areas with higher prevalence of all crime**

CRIME RATE

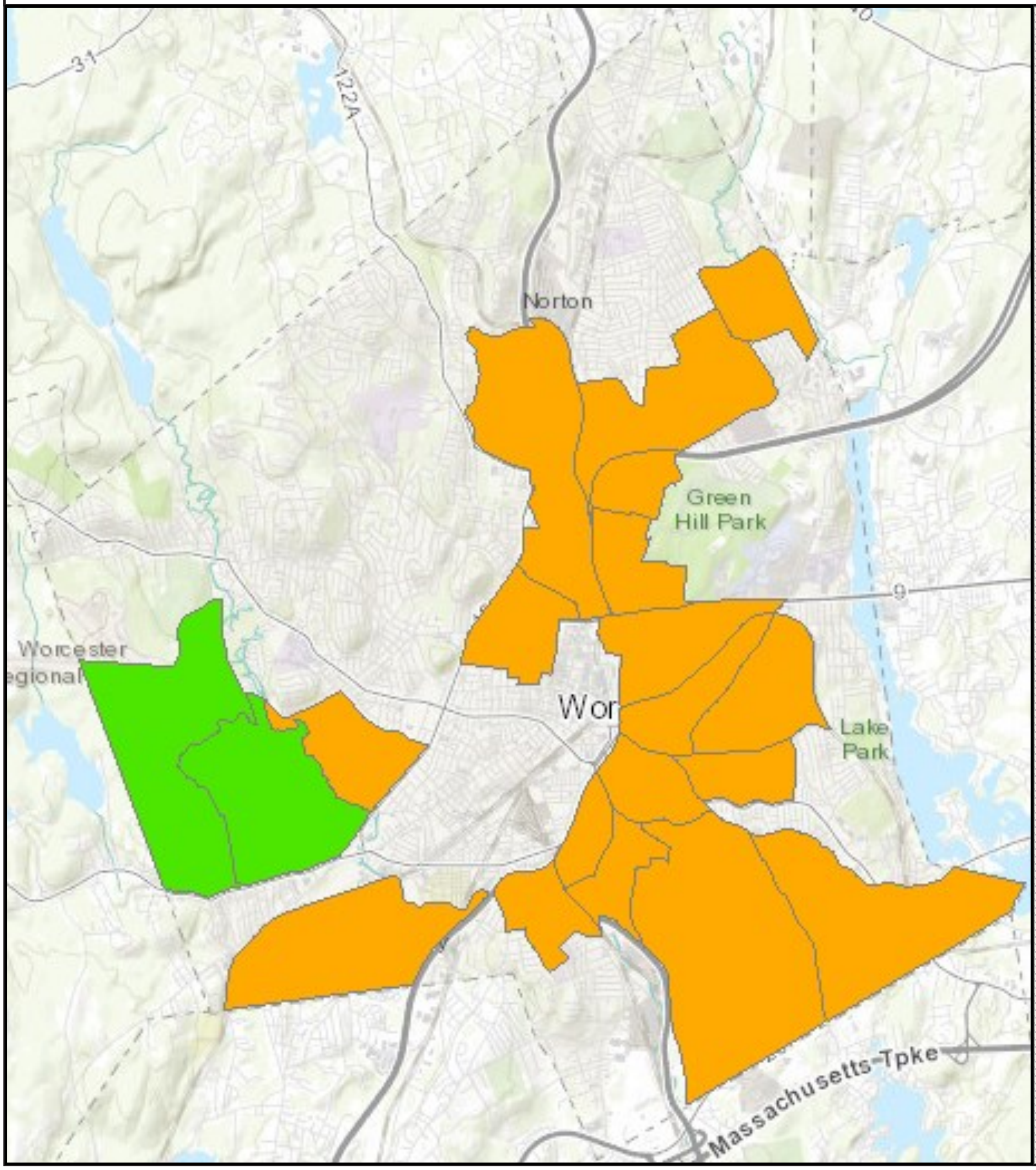
Data presented in *Figure 20* qualitatively shows areas with higher total crime rate. Shading on the map positively correlates with COVID-19 density *(Figure 20)*. Darker regions (where crime rate is highest) mostly appear in the central region of the Worcester Covid Belt. Regions surrounding the belt are much lighter indicating lower crime rates.

RESULTS — EPIDEMIOLOGY *(CONT.)*

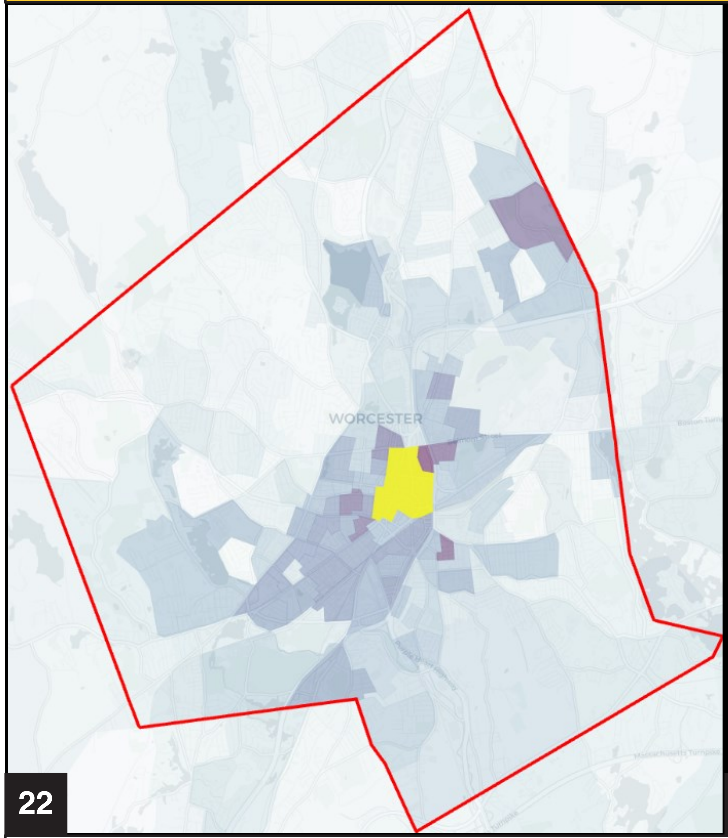
FOOD DESERTS

Data presented in *Figure 21* qualitatively shows areas where food deserts exist. Shading on the map positively correlates with COVID-19 density *(Figure 10)*. Green represents regions where most residents are 1 mile (urban) or 10 miles (rural) from a supermarket. Orange represents regions where most residents are 1/2 mile (urban) or 10 miles (rural) from a supermarket.

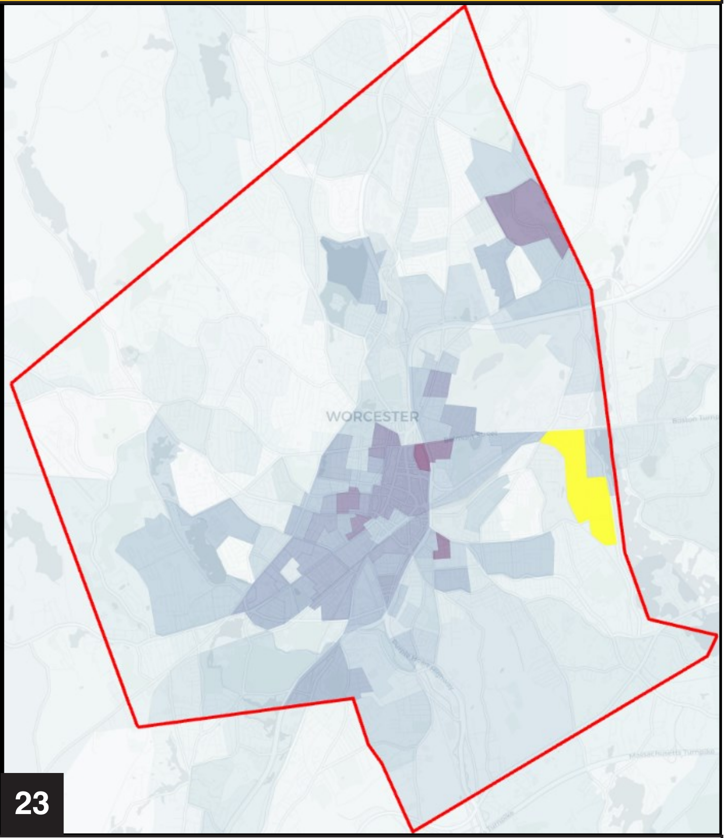
Figure 21: **Food Deserts: Represents food desert map of Worcester. Green regions represent regions where most residents are 1 mile (urban) or 10 miles (rural) from a supermarket. Orange represents regions where most residents are 1/2 mile (urban) or 10 miles (rural) from a supermarket.**



RESULTS — EPIDEMIOLOGY *(CONT.)*



22



23

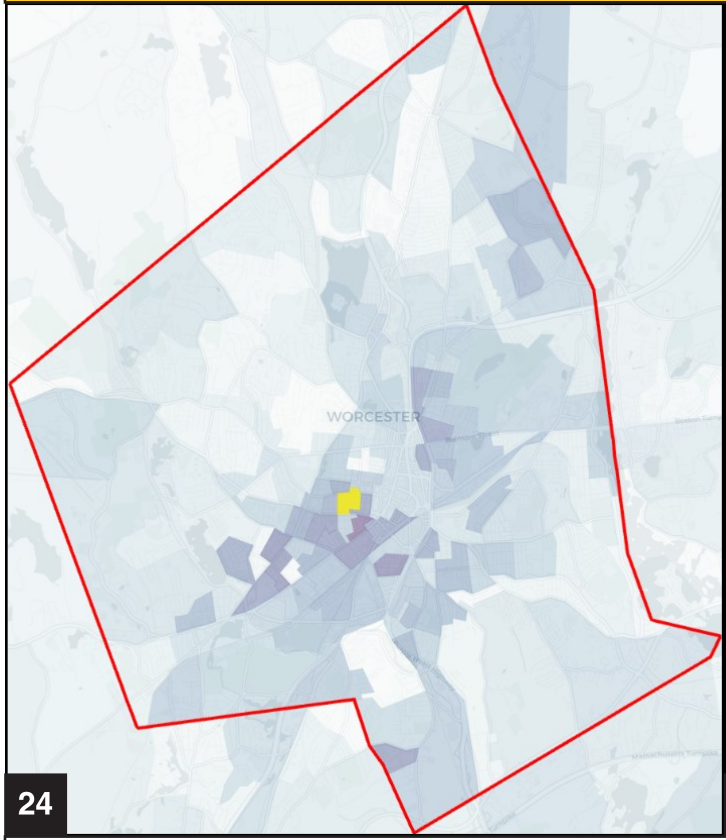
Figure 22: Income Below Poverty 1: Represents census block 25027.731700.1 (yellow). Darker regions indicate higher percentage below poverty.

Figure 23: Income Below Poverty 2: Represents census block 25027.732201.2 (yellow). Darker regions indicate higher percentage below poverty.

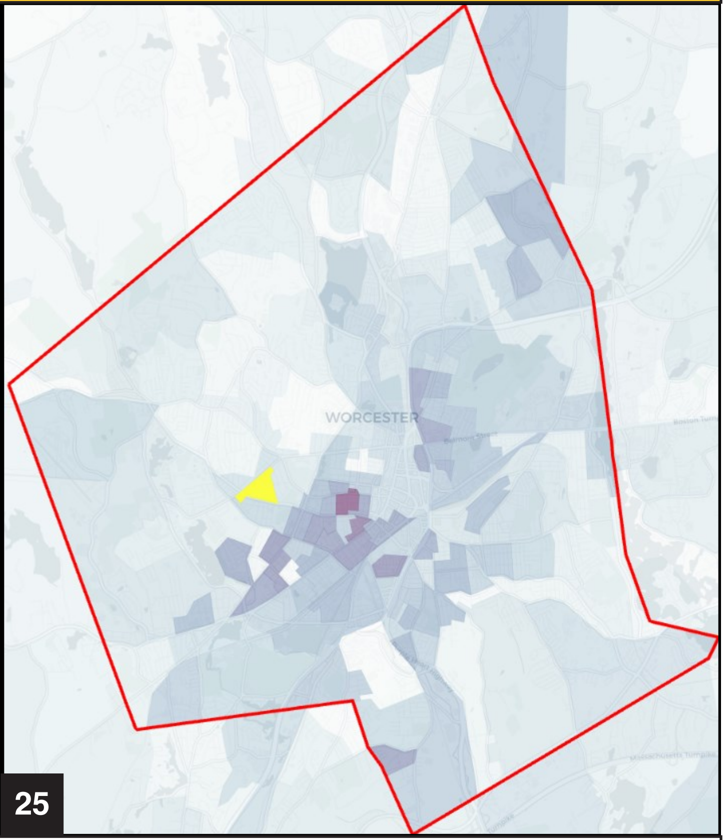
INCOME BELOW POVERTY

Income below poverty (% of individuals) *(Figs. 22,23)* positively correlate with COVID-19 density shown in *Figure 10*. Census block 25027.731700.1 *(Fig. 22)* is located in the central area of the *Worcester COVID Belt*, and includes data from 815 households. Percentage of families whose income is below the poverty line within this block is 45.0%. Contrasting this is census block 25027.732201.2 *(Fig. 23)*, located just east of the belt. Data collected from this block came from 1,009 households and yields an average percentage of families living below the poverty line of 2.24%.

RESULTS — EPIDEMIOLOGY *(CONT.)*



24



25

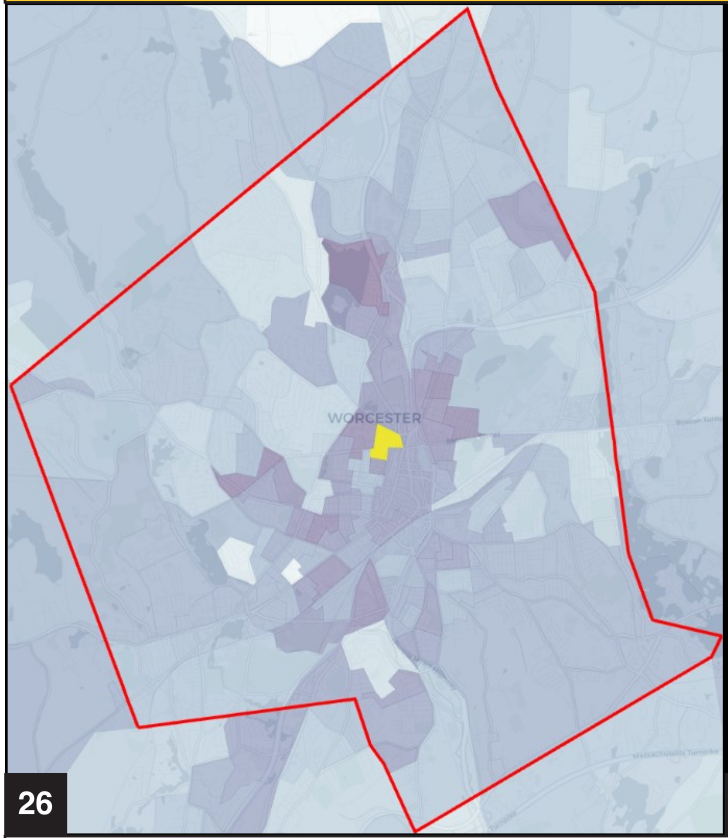
Figure 24: **Less than High School Education 1: Represents census block 25027.731500.3 (yellow). Darker regions indicate higher percentage of individuals who did not finish high school.**

Figure 25: **Less than High School Education 2: Represents census block 25027.731102.3 (yellow). Darker regions indicate higher percentage of individuals who did not finish high school.**

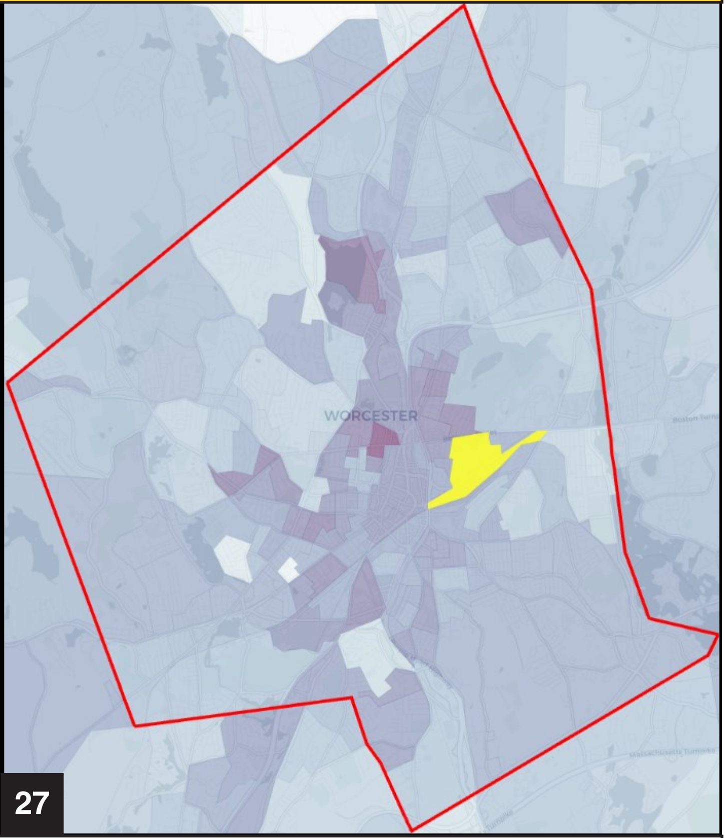
LESS THAN HIGH SCHOOL EDUCATION

Less than high school education (% of individuals) *(Figs. 24,25)* positively correlates with COVID-19 density shown in *Figure 10*. Census block 25027.731500.3 *(Fig. 24)* is located in the central area of the *Worcester COVID Belt*, and includes data from 447 households. Less than high school education within this block is 53.4%. Contrasting this is census block 25027.731102.3 *(Fig. 25)*, located just northwest of the belt. Data collected from this block came from 372 households and yields a less than high school education rate of 3.81%.

RESULTS — EPIDEMIOLOGY *(CONT.)*



26



27

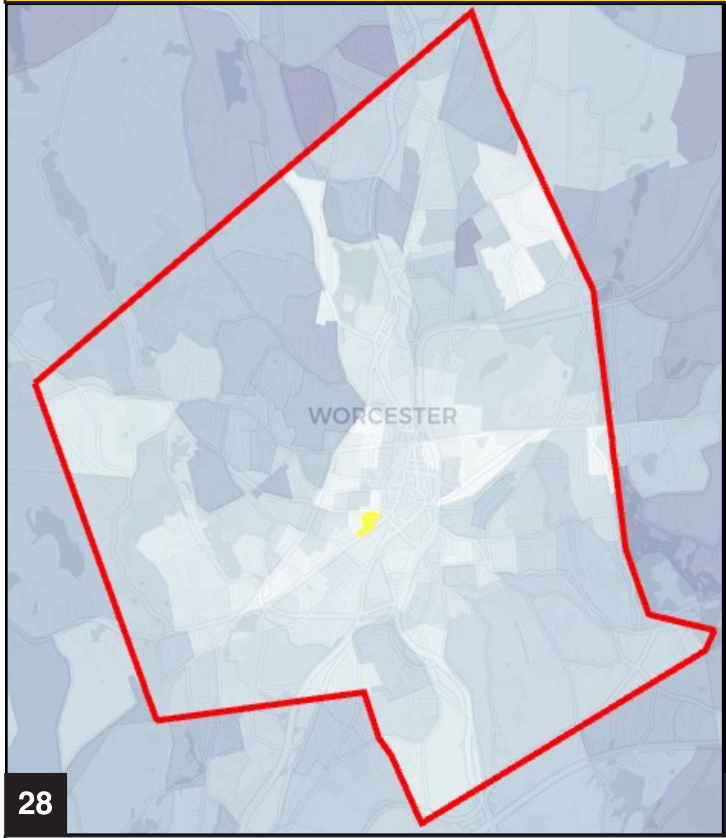
Figure 26: **Meals from Fast Food 1: Represents census block 25027.731600.1 (yellow). Darker regions indicate higher number of fast-food meals eaten per week.**

Figure 27: **Meals from Fast Food 2: Represents census block 25027.731800.6 (yellow). Darker regions indicate higher number of fast-food meals eaten per week.**

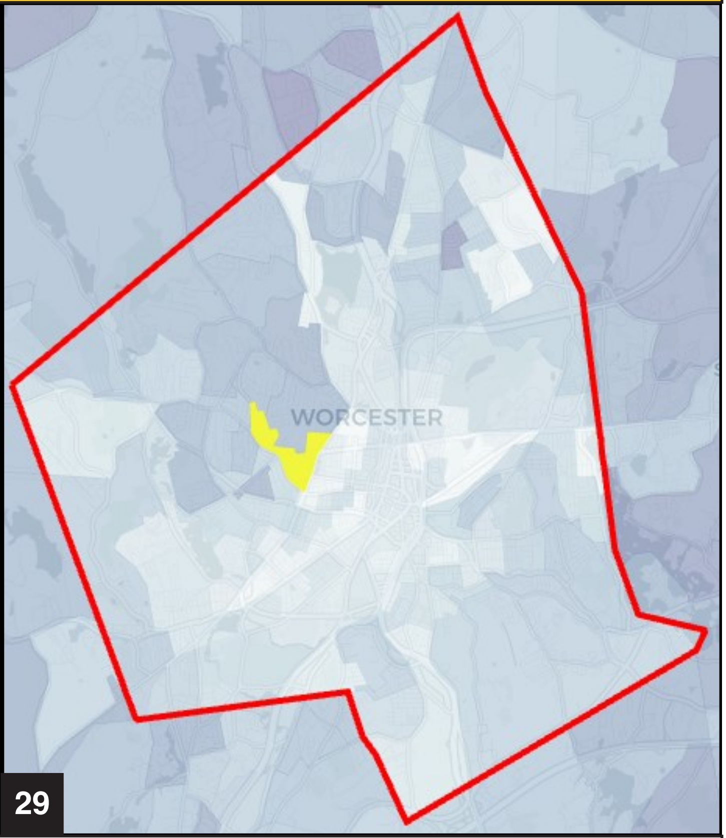
MEALS FROM FAST FOOD

Meals from fast food (#/week) *(Figs. 26,27)* positively correlates with COVID-19 density shown in *Figure 10*. Census block 25027.731600.1 *(Fig. 26)* is located within the central area of the *Worcester COVID Belt*, and includes data from 545 households. The average number of fast-food meals eaten per week within this block is 2.48. Contrasting this is census block 25027.731800.6 *(Fig. 27)*, located just southeast of the belt. Data collected from this block came from 816 households and yields and average of 1.47 meals from fast-food eaten per week.

RESULTS — EPIDEMIOLOGY *(CONT.)*



28



29

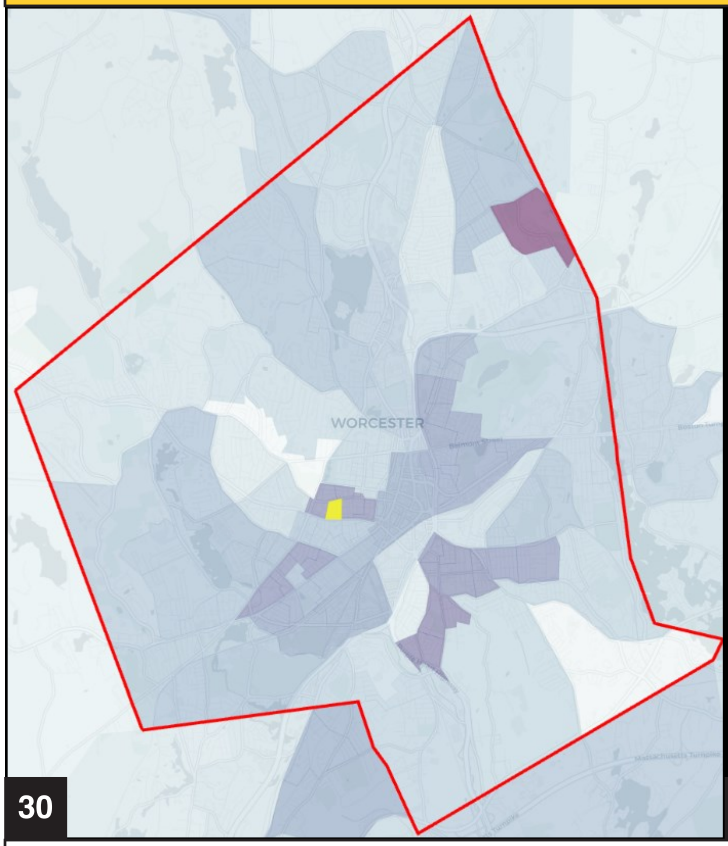
Figure 28: **Median Household Income 1: Represents census block 205027.731400.1 (yellow). Darker regions indicate higher median household income.**

Figure 29: **Median Household Income 2: Represents census block 25027.730802.2 (yellow). Darker regions indicate higher median household income.**

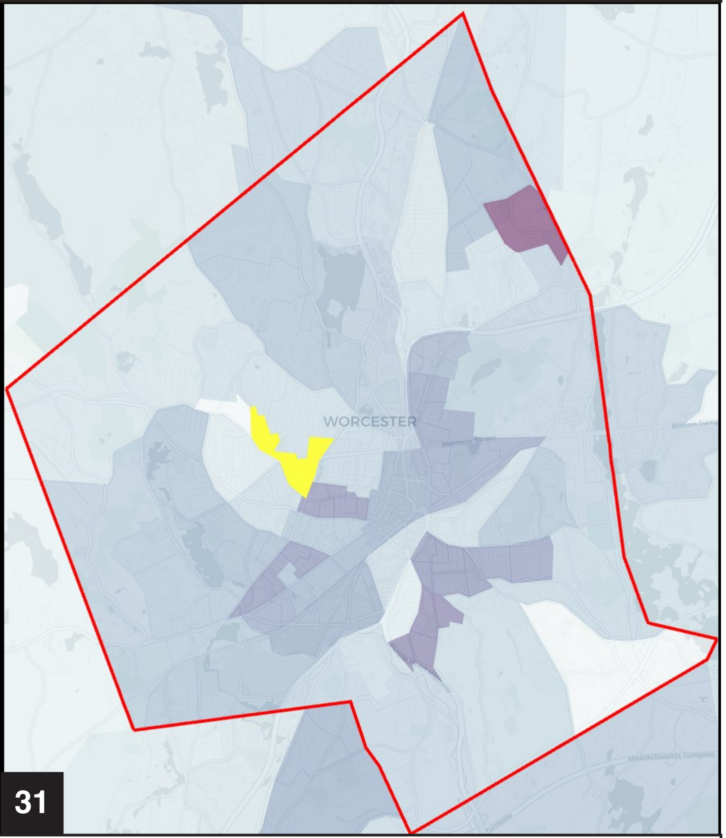
MEDIAN HOUSEHOLD INCOME

Median household income ($ per family) *(Figs. 28,29)* positively correlates with COVID-19 density shown in *Figure 10*. Census block 25027.731400.1 (Fig. 28) is located within the central area of the COVID-19 belt and includes data from 477 households. Median household income within this block is $15,270. Contrasting this is census block 25027.730802.2 *(Fig. 29)*, located just northwest of the central COVID-19 belt region. Data collected from 526 households in this region yields a median household income of $79,730.

RESULTS — EPIDEMIOLOGY *(CONT.)*



30



31

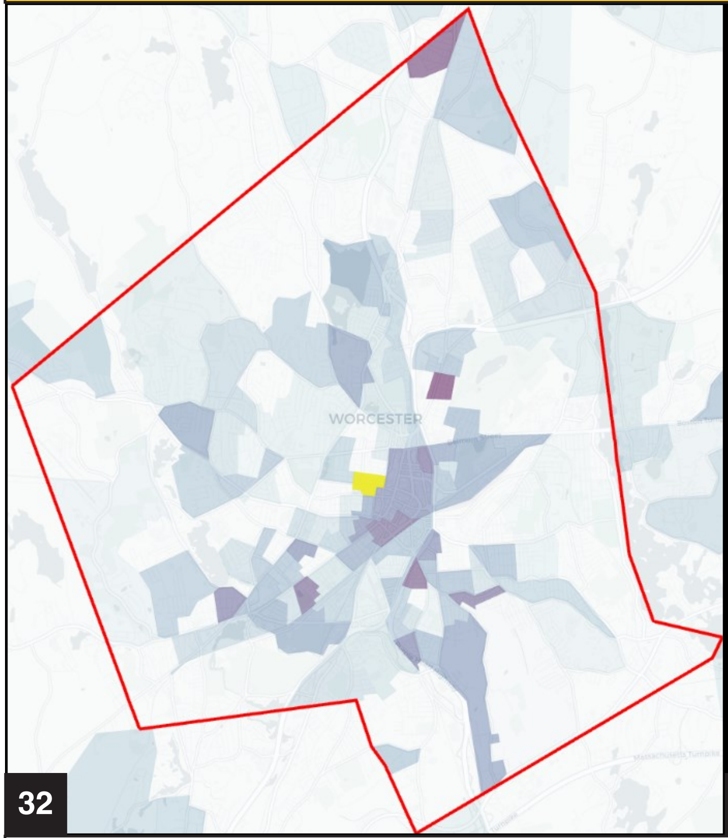
Figure 30: **Unemployment 1: Represents census block 25027.731500.4. (yellow). Darker regions indicate higher levels of unemployment.**

Figure 31: **Unemployment 2: Represents census block 25027.730802.2 (yellow). Darker regions indicate higher median household income.**

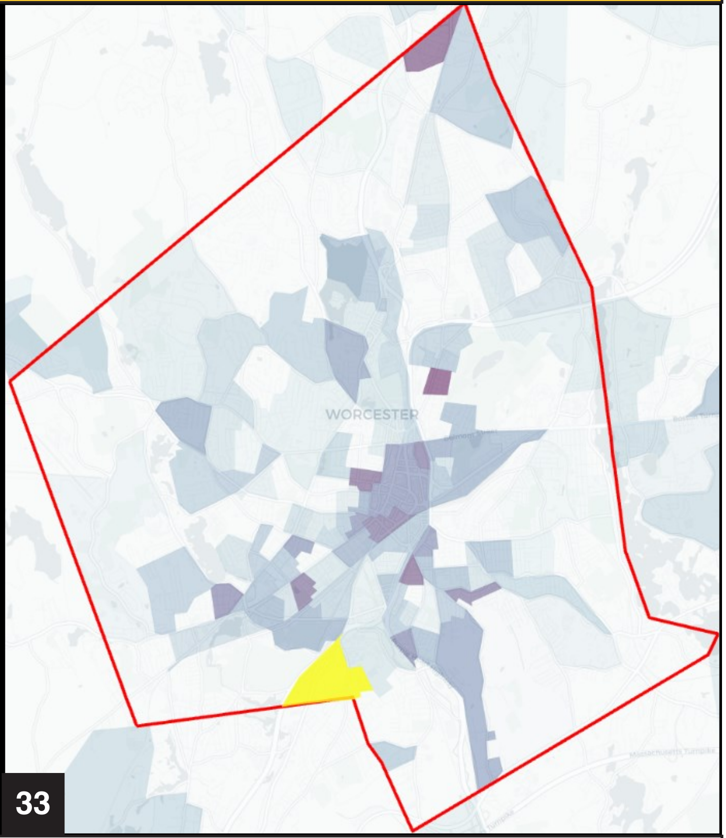
UNEMPLOYMENT

Total Unemployment (% of individuals) *(Figs. 30,31)* positively correlates with COVID-19 density shown in *Figure 10*. Census block 25027.731500.4. *(Fig. 30)* is located within the central area of the *Worcester COVID Belt*, and includes data from 196 households. Total unemployment (%) within this block is 12.0%. Contrasting this is census block 25027.730802.2 *(Fig. 31)*, located just northwest of the belt. Data collected from this block came from 526 households and yields an unemployment rate of 1.10%.

RESULTS — EPIDEMIOLOGY *(CONT.)*



32



33

Figure 32: **Usage of Public Transportation 1: Represents census block 25027.731600.3 (yellow). Darker regions indicate higher use of public transportation to travel to work.**

Figure 33: **Usage of Public Transportation 2: Represents census block 25027.732901.4 (yellow). Darker regions indicate higher use of public transportation to travel to work.**

USAGE OF PUBLIC TRANSPORTATION

Usage of public transportation as a means to get to work (% of individuals) *(Figs. 32,33)* positively correlates with COVID-19 density shown in *Figure 10*. Census block 25027.731600.3 *(Fig. 32)* is located within the central area of the COVID-19 belt and includes data from 787 households. Usage of public transportation within this block is 16.2%. Contrasting this is census block 25027.732901.4 *(Fig. 33)*, located just southwest of the central COVID-19 belt region and includes data collected from 444 households. Usage of public transportation within this block is 3.72%.

DISCUSSION — BIOLOGY

The overall goal of this portion of the project was to determine if a correlation existed between the nasopharyngeal microbiome and a clinical outcome related to COVID-19; the need for respiratory/oxygen support or not was chosen. This is because COVID-19 has been shown to have a dramatic affect on the lungs in prior studies and reports. When developing a model to understand the complex relationships of the microbiome and a clinical outcome, all other factors like age, race, etc. need to be controlled for. This is the principal reason for why a machine-learning based algorithm was used; this algorithm scored particular clinical and demographic factors as well as the microbiota to determine (1) what factors best predicted the need for oxygen or not, and (2) how specific microbial species abundances correlate to one of these two outcomes.

DIFFERENCES IN MICROBIAL SPECIES' ABUNDANCES MAY ACT AS BETTER HEALTH PREDICTORS THAN AGE, BMI, AND CCI

Overall, we see that many different microbes have higher median importance values assigned to them by the algorithm, meaning that they are more likely to indicate the correct dichotomous outcome a patient eventually experienced. The fact that *P. salivae* and many of the other microbes scored so highly has two major implications. First, it means that microbes like *P. salivae* may have some correlation with COVID-19 and its affects on the respiratory system. Second, it means that physicians may want to consider an individual's microbiota alongside other clinical factors when assessing how their COVID-19 may progress.

CERTAIN MICROBIAL SPECIES MAY EXHIBIT PROTECTIVE ATTRIBUTES WITHIN THE NASOPHARYNX

Interestingly, this data showed us that a high abundance of many nasopharyngeal species was correlated with "not needing respiratory support," contrary to the original predictions of this research. Therefore, the microbes found in this population cannot be used as clinical proxies for determining if an individual will need to be kept in the hospital for further examination. However, this finding does open the doors to many interesting biological questions.

FUTURE IMPLICATIONS

Given the fact that the presence of many nasopharyngeal species is associated with the absence of a negative outcome, what does this tell us about our microbes? Are they somehow protecting the body from the harmful aftereffects of COVID-19? And if so, how? Immunology research conducted over the last decade has presented the idea that many commensal microbes which are largely present in the population, may interact with the immune system to prevent infection, and protect the body while infiltrated[29,51]. In fact, one of the studies showed that direct "manipulation of commensal bacteria, expression of antiviral defense genes and interferon responsive pathways were altered in the steady state[29]." In the case of a viral infection, researchers pointed out that such an interference may leave just enough time for viruses to infect the body, putting both humoral and cellular immunity at a great disadvantage. This increases the time it takes for the body to recover. Another research team concluded that for gut bacteria, their functions go, "well beyond the digestive system[51]." While these studies don't draw attention explicitly to the nasopharynx, they lend reason to believe that perhaps microbes in the throat play a similar role.

DISCUSSION — BIOLOGY *(CONT.)*

Despite all of the work done in the past, current research, and even this project, it's clear that the microbiome still remains an elusive part of our bodies, and our species. Further basic research in cross sectional fields like this one, are needed to understand the microbiome's role within our bodies, and how we can use or manage it to enact change in our lives.

DISCUSSION — EPIDEMIOLOGY

The health and wellbeing of a particular population relies on several factors directly related to the healthcare system. Access to care, the means to afford treatment, and knowledge of what treatment options exist are critical for individuals suffering from disease or disability. Additionally, both social and structural elements of one's life play a large role in their health. In this study, examinations of particular social and structural determinants of health were carried out to identify possible relationships with the advent of COVID-19. In particular, crime, education, income, nutrition, and transportation, and related factors were examined with respect to demographics, and COVID-19 density in Worcester Massachusetts.

CRIME

Crime within one's environment not only poses a physical threat to safety but can also wear down on individuals emotionally and mentally. Areas where crime rate is highest generally fall within regions with high population densities, and where citizens have lower socioeconomic status (SES). In this regard, crime can also exacerbate stress caused by these preexisting conditions. High crime rate can also be the root of low SES, preventing people from working, or by pushing them towards bad decision-making habits.

In this analysis, crime included data solely on the total crime within the City of Worcester as shown in *Fig. 20*. Regions in Worcester where COVID-19 density was the highest *(Fig.10)* fell within the same regions that had the highest rates of total crime.

With regard to COVID-19, high crime rate may directly interfere with one's ability to receive treatment. It may also indirectly affect how the body respond's due to a number of reasons again related to stress and mental health. These factors may then contribute to further spread of the virus.

EDUCATION

Educational attainment derived from schooling provides students with skills in critical thinking and reasoning. In turn, these skills help students maneuver through life, providing them with increased income, job security, and health awareness. Individuals who lack in education are therefore more likely to suffer as a result. Possible connections between disease and lack of education have been shown in literature surrounding poverty for decades. This literature, "demonstrates a positive correlation between educational attainment, and health status indicators such as life expectancy, obesity, morbidity from acute and chronic diseases, health behaviors, and more[52]." Education attainment also has a generational effect, where the education of the parents, particularly the mother correlates to better health and lower disease rates in children.

DISCUSSION — EPIDEMIOLOGY *(CONT.)*

In this analysis, education included data on best schools, and less than high school education, as shown in *Figs. 19*, and *24-25*. Regions in Worcester where COVID-19 density was the highest *(Fig. 10)* fell within the same regions that had the lowest proximity to good schools, and also in regions where there was a high percentage of residents with less than a high school education.

With this in mind, we can see that there may be a correlation then between the density of COVID-19 within this region and education. Perhaps higher rates of COVID-19 stem from individual's lack of access to proper information (which stems from a lack of knowledge/critical thinking skills).

INCOME

Income and employment affect health, "by buffering individuals against the financial threat of large medical bills while also facilitating access to health-promoting resources such as access to healthy neighborhoods, homes, land uses, and parks[52]." Like education, income is related to outcomes such as mortality, and morbidity due to chronic conditions.

In this analysis, income included data on average family size, income below poverty, median household income, and unemployment as shown in *Figs. 17-18, 22-23, 28-29,* and *30-31*. For average family size, blocks with households that had larger family size were generally found within the central region of the *Worcester COVID Belt*. In one block, families had an average of about 3.87 people/household, compared to smaller households which had about 2.78. Blocks where households had smaller family sizes were mostly found outside of the belt. This determinant alone does not show much, but when it's paired with other income-based determinants provides an important lens to view poverty within the *Worcester COVID Belt*. Blocks in which more households had earnings below the poverty level also fell within the belt. This contrasted many blocks situated around the belt where households earned above the poverty level *(Figs. 22-23)*. Furthermore, blocks that contained households with lower median family incomes *(Figs. 28-29),* and higher unemployment rates *(Fig. 30-31)* also fell within the *Worcester COVID Belt*. Directly surrounding this region, median family incomes increase, and unemployment decreases. This means that blocks with households that had the lowest median income, and highest unemployment, generally had the highest family sizes. Naturally, it's expected that with increasing family size, more income would be coming in to afford an extra body, or perhaps that extra individual is also employed to bring more income in. However, this is not the case; this is the lens mentioned earlier. Larger family sizes may further exacerbate income instability within these blocks. Households may experience increased COVID-19 prevalence.

With this in mind, we can see that there may be a correlation between income and job security and the prevalence of COVID-19. Individuals from households living below the poverty line, or where a large number of people are unemployed might be related to a lack of education. These determinants might also be directly related to a lack of access to care; it might be that individuals who earn less are less likely (or less able) to utilize the healthcare system. Therefore, their COVID-19 may never get diagnosed or it's diagnosed late. In either case, this extra time may lead to further spread. Income is also related to each of the other major determinants analyzed (crime, education, nutrition, and transportation).

Lack of stable income can make it much more difficult to afford reliable transportation, purchase healthy foods, get a proper education, and can increase problem-behaviors (like stealing food) which lead to jail-time. The reverse is also true here, in that lack of a proper education, transportation, etc. may also make it difficult to earn a reliable income.

NUTRITION

Nutrition is directly tied to one's health, including how the body responds to treatments, and medications. Improvements to the immune system from eating a well-balanced diet can bolster the body's ability to fend of foreign pathogens such as viruses. Furthermore, unhealthy eating habits may exacerbate preexisting conditions and ailments; this gives way for more problems to arise in the future like more infections.

In this analysis, nutrition includes data on food deserts, and the number of meals eaten from fast food chains per week as shown in *Figs. 21* and *26-27*. Areas within Worcester where COVID-19 is the most prevalent *(Fig. 10)* fall within the same areas where grocery stores are furthest away (farther distances) and where households eat the highest number of fast-food meals per week.

With respect to COVID-19, accessibility concerns with nutritional foods may push people to eat at fast food more. It might also be a problem with income, or that local grocery stores are in areas where crime rate is highest. Irrespective of the reason, lack of nutritional foods will dampen the innate and adaptive immune responses to pathogens. Less fighting ability and longer inoculation of the virus may lead to spreading it more, increasing prevalence.

TRANSPORTATION

Telehealth, and in-home nurse/patient visits may not be accessible to everyone, might not serve to cure a viral infection. Many still rely on going into a hospital or visiting a doctor/clinic for care. If transportation became inaccessible to a family or an individual, that might offset one's ability to receive proper, timely care.

In this analysis, transportation included data solely from usage of public transportation, as shown in *Figs. 32-33*. Areas within Worcester where COVID-19 is the most prevalent *(Fig. 10)* fall within the same areas where public transportation is the used most often.

In regard to COVID-19, lack of stable personal transportation may lead to an inability to receive timely care. It might also lead to difficulties with receiving an education, an income, and getting a proper nutrition. When compared with a personal vehicle, public transportation is far less reliable. On top of all this, more time spent on public transit also places people in closer proximity to one another for longer durations of time. This might be a vector for the spread of disease, like Sars-CoV-2[53].

SOCIAL FACTORS ASSOCIATED WITH COVID-19 ARE CONNECTED AND COMPOUNDING

Individuals who often affected by these social factors may be affected by more than one. In many cases this is because the social factors are connected. When income starts to fall for example, this can lead to difficulties affording nutritional food, or reliable transportation, etc.

DISCUSSION — EPIDEMIOLOGY *(CONT.)*

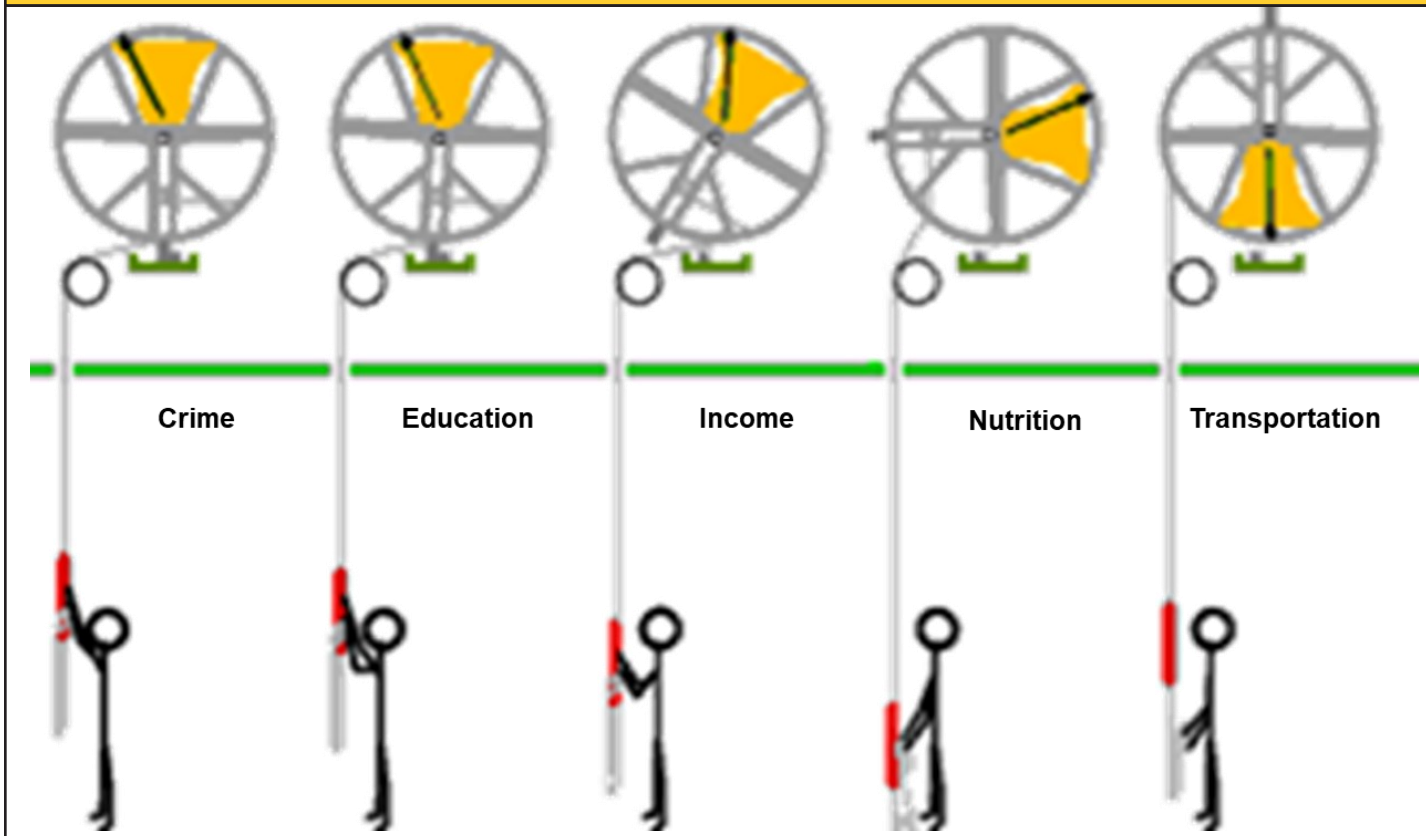


Figure 34: **Bell Ringer's Model: Represents the interconnectedness and compounding nature of social factors, and their effects on an individual's health.**

In addition, the issues are compounding, meaning that because they are connected, multiple social issues can apply stress to an individual's health all at once; in this sense, they almost act like a positive feedback look.

One way to understand this relationship is through a model called the Bell Ringer's Model *(Fig. 34)*. Members of the bellringer group work together to ring bells and play a song. Each bellringer rings the bell by pulling a rope at a particular time and set distance. If an individual pulls a rope too much or too little, or they time the pull incorrectly, the bell falls out of sync with the others. This mistake is usually simple to correct and requires patience and time for the bellringer to get back into the melody of the bells. However, as more bellringers fall out of sync, it becomes increasingly more difficult for the other bellringers to maintain their position, and if they have already fallen out - to get back into sync. Social factors as determinants for an individual's health work in a very similar capacity.

When an individual is faced with a problem in one of these social determinants, it is likely that the individual can still maintain their health. If the individual is faced with problems in two or more social determinants, it becomes increasingly more difficult to fix the issues (again due to the connectedness of some of these social factors). In some cases, a single social factor might be detrimental to the individual's health, but what the model proposes, is that as more bells (social factors) fall out of sync, it becomes increasingly more difficult to fix.

NOT THE GREAT EQUALIZER

Corona Virus Disease 2019 caused by the spread of SARS-CoV-2 has been dubbed "The Great Equalizer" by many mainstream media outlets, celebrities, and even political figures[53]. While the title fits in some aspects of the virus, it is often misinterpreted. It is true that SARS-CoV-2 does not discriminate who it infects; this is because it is a virus and has no innate capacity to do so. Ironically, this same characteristic is found within all microbes, yet none have ever been called "The Great Equalizer" before. Certainly, Influenza caused by the Influenza Virus is a much more prominent equalizer, seeing as it infects people regardless of their outward appearance year-after-year.

The fact of the matter is that COVID-19, like most infectious diseases can and will infect anyone regardless of "who" they are. But how people are affected by the virus is vastly out of the hands of the person experiencing the ailment. To some degree, genes, and geographic distribution can affect susceptibility to disease[54] but this data does not exist in a vacuum. Biological differences due to changes in genes, "have a small effect on disease risk[54]." These genes are not directly linked to individual susceptibility, but again the affects the disease has on them.

In addition to genetics, behaviors, environment, medical care, and social factors all affect an individual's ability fight disease[55]. Data utilized here from Worcester's Department of Public Health, as well as the U.S. Census Bureau correlate highly with COVID -19 density. Again, it's critical that none of these factors are viewed by themselves. Each one of the issues presented are connected and compounding.

Those who face excessive social burden, especially minority populations are not being targeted by the virus, nor is there any evidence to suggest that particular behaviors are to blame for COVID-19 prevalence. Social burden caused by a multitude of factors, such as discrimination, and systemic prejudice may be the root for social issues like lack of nutritional foods, high crime rate, and low educational attainment. It is these social factors that place individuals at a disproportionately high level of risk for contracting COVID-19. These same individuals also face increased stress throughout the infection. Together these higher rates of infection and stress levels may lead to more infections, and longer lasting symptoms.

From this perspective, we now see that it is not just genes, or behaviors that cause increased COVID-19 prevalence, but also a system which places unfavorable outcomes on particular groups of people[56].

In order to improve the health and wellness of all people, social factors need to be addressed as well. Clinicians see patients, but are only trained to examine the evidence presented to them in order to create a diagnosis. Similarly, public health workers always view the larger picture, and don't spend much time individually assessing every single individual's problems. While this strategy still remains adequate for assessing health problems in the U.S., a more collaborative approach to healthcare is needed. Physicians need to assess and make sense of each of their patient's lives, and surroundings, and public health workers need to ensure that the data they use to create health/social programs addresses needs at a more individual level.

BIBLIOGRAPHY

1. World Health Organization. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data [Internet]. 2021 [cited 2021 May 5];Available from: https://covid19.who.int/

2. Augusus E. Worcester St. Patrick's Parade Canceled for March 15 | City of Worcester, MA [Internet]. 2021 [cited 2021 Mar 21];Available from: http://www.worcesterma.gov/announcements/worcester-st.-patrick-s-parade-canceled-for-march-15#id8_CioR6KeJeP1l27R4iu6g

3. Augusus E. COVID-19 Update, Worceser Public Schools Closed March 13 | City of Worcester, MA [Internet]. 2021 [cited 2021 Mar 21];Available from: http://www.worcesterma.gov/announcements/covid-19-update-worceser-public-schools-closed-march-13#id2P_Tkcv5qEAAFA7nI_IKLg

4. Augusus E. COVID-19 Update for March 14; First Positive Case | City of Worcester, MA [Internet]. 2021 [cited 2021 Mar 21];Available from: http://www.worcesterma.gov/announcements/covid-19-update-for-march-14-first-positive-case#idHt_PLhlrbJSDO_1S79nAnA

5. Augustus E. COVID-19 Update for April 30; 23,302 Cases | City of Worcester, MA [Internet]. City Worcester. 2021 [cited 2021 May 5];Available from: http://www.worcesterma.gov/announcements/covid-19-update-for-april-30-23-302-cases#idXBtcJGNT_kS9ke5SD6PFOQ

6. Su S, Wong G, Shi W, et al. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses [Internet]. Trends Microbiol. 2016 [cited 2021 Jan 30];24(6):490–502. Available from: https://pubmed.ncbi.nlm.nih.gov/27012512/

7. Hamming I, Timens W, Bulthuis MLC, Lely AT, Navis GJ, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J Pathol [Internet] 2004 [cited 2021 Jan 30];203(2):631–7. Available from: https://pubmed.ncbi.nlm.nih.gov/15141377/

8. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses [Internet]. Nat. Rev. Microbiol. 2019 [cited 2021 Jan 30];17(3):181–92. Available from: www.nature.com/nrmicro

9. Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis [Internet]. J. Med. Virol. 2020 [cited 2021 Jan 30];92(4):418–23. Available from: /pmc/articles/PMC7167049/

10. Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by SARS-CoV-2. Nature [Internet] 2020 [cited 2021 Jan 30];581(7807):221–4. Available from: https://doi.org/10.1038/s41586-020-2179-y

11. van Doremalen N, Bushmaker T, Morris DH, et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. N Engl J Med [Internet] 2020 [cited 2021 Jan 30];382(16):1564–7. Available from: http://www.nejm.org/doi/10.1056/NEJMc2004973

12. Wu Z, McGoogan JM. Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases from the Chinese Center for Disease Control and Prevention [Internet]. JAMA - J. Am. Med. Assoc. 2020 [cited 2021 Jan 30];323(13):1239–42. Available from: https://jamanetwork.com/

13. Centers for Disease Control and Prevention. Symptoms of Coronavirus | CDC [Internet]. [cited 2021 Jan 30];Available from: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

14. Lauer SA, Grantz KH, Bi Q, et al. The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application. Ann Intern Med [Internet] 2020 [cited 2021 Jan 30];172(9):577–82. Available from: /pmc/articles/PMC7081172/

15. Cennimo D. Coronavirus Disease 2019 (COVID-19) Workup: Approach Considerations, Laboratory Studies, CT Scanning [Internet]. Medscape. 2021 [cited 2021 Jan 30];Available from: https://emedicine.medscape.com/article/2500114-workup#c2

16. World Health Organization. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus Disease (COVID-19) Dashboard [Internet]. [cited 2021 Jan 30];Available from: https://covid19.who.int/

17. Centers for Disease Control and Prevention. CDC COVID Data Tracker [Internet]. [cited 2021 Jan 30];Available from: https://covid.cdc.gov/covid-data-tracker/#datatracker-home

18. Centers for Disease Control and Prevention. Risk for COVID-19 Infection, Hospitalization, and Death By Age Group | CDC [Internet]. [cited 2021 Jan 30];Available from: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-age.html

19. Centers for Disease Control and Prevention. Risk for COVID-19 Infection, Hospitalization, and Death By Race/Ethnicity | CDC [Internet]. [cited 2021 Jan 30];Available from: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html

20. Centers for Disease Control and Prevention. CDC COVID Data Tracker [Internet]. [cited 2021 Jan 30];Available from: https://covid.cdc.gov/covid-data-tracker/#demographics

21. Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the Treatment of Covid-19 — Final Report. N Engl J Med [Internet] 2020 [cited 2021 Jan 30];383(19):1813–26. Available from: http://www.nejm.org/doi/10.1056/NEJMoa2007764

22. Jackson LA, Anderson EJ, Rouphael NG, et al. An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. N Engl J Med [Internet] 2020 [cited 2021 Jan 30];383(20):1920–31. Available from: http://www.nejm.org/doi/10.1056/NEJMoa2022483

23. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med [Internet] 2020 [cited 2021 Jan 30];382(8):727–33. Available from: http://www.nejm.org/doi/10.1056/NEJMoa2001017

24. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N Engl J Med [Internet] 2020 [cited 2021 Jan 30];382(13):1199–207. Available from: http://www.nejm.org/doi/10.1056/NEJMoa2001316

25. Holshue ML, DeBolt C, Lindquist S, et al. First Case of 2019 Novel Coronavirus in the United States. N Engl J Med [Internet] 2020 [cited 2021 Jan 30];382(10):929–36. Available from: http://www.nejm.org/doi/10.1056/NEJMoa2001191

26. Harper A, Vijayakumar V, Ouwehand AC, et al. Viral Infections, the Microbiome, and Probiotics [Internet]. Front. Cell. Infect. Microbiol. 2021 [cited 2021 May 5];10:925. Available from: www.frontiersin.org

27. Man WH, De Steenhuijsen Piters WAA, Bogaert D. The microbiota of the respiratory tract: Gatekeeper to respiratory health [Internet]. Nat. Rev. Microbiol. 2017 [cited 2021 May 5];15(5):259–70. Available from: www.nature.com/nrmicro

28. Scheblauer H, Reinacher M, Tashiro M, Rott R. Interactions between Bacteria and Influenza A Virus in the Development of Influenza Pneumonia [Internet]. 1992 [cited 2021 May 6]. Available from: https://about.jstor.org/terms

29. Abt MC, Osborne LC, Monticelli LA, et al. Commensal Bacteria Calibrate the Activation Threshold of Innate Antiviral Immunity. Immunity [Internet] 2012 [cited 2021 May 4];37(1):158–70. Available from: /pmc/articles/PMC3679670/

BIBLIOGRAPHY *(CONT.)*

30. Centers for Disease Control and Prevention. Different COVID-19 Vaccines | CDC [Internet]. Centers Dis. Control Prev. 2021 [cited 2021 May 6];Available from: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines.html

31. Centers for Disease Control and Prevention. Pfizer-BioNTech COVID-19 Vaccine Overview and Safety | CDC [Internet]. Centers Dis. Control Prev. 2021 [cited 2021 May 6];Available from: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines/Pfizer-BioNTech.html

32. Centers for Disease Control and Prevention. Moderna COVID-19 Vaccine Overview and Safety | CDC [Internet]. Centers Dis. Control Prev. 2021 [cited 2021 May 6];Available from: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines/Moderna.html

33. Centers for Disease Control and Prevention. Johnson & Johnson's Janssen COVID-19 Vaccine Overview and Safety | CDC [Internet]. Centers Dis. Control Prev. 2021 [cited 2021 May 6];Available from: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines/janssen.html

34. Berg S. What doctors wish patients knew about the Johnson & Johnson vaccine | American Medical Association [Internet]. Am. Med. Assoc. 2021 [cited 2021 May 6];Available from: https://www.ama-assn.org/delivering-care/public-health/what-doctors-wish-patients-knew-about-johnson-johnson-vaccine

35. Wolf A, Gray R, Fazel S. Violence as a public health problem: An ecological study of 169 countries. Soc Sci Med [Internet] 2014 [cited 2021 Apr 4];104(100):220–7. Available from: /pmc/articles/PMC3969091/

36. U.S. Department of Health and Human Services. Crime and Violence - Healthy People 2030 | health.gov [Internet]. 2021 [cited 2021 Mar 30];Available from: https://health.gov/healthypeople/objectives-and-data/social-determinants-health/literature-summaries/crime-and-violence

37. Hahn RA, Truman BI. Education improves public health and promotes health equity. Int J Heal Serv [Internet] 2015 [cited 2021 Mar 23];45(4):657–78. Available from: /pmc/articles/PMC4691207/

38. Center on Society and Health. Why Education Matters to Health: Exploring the Causes [Internet]. [cited 2021 Apr 1];Available from: www.societyhealth.vcu.edu.

39. Centers for Disease Control and Prevention. Health-Risk Behaviors and Academic Achievement [Internet]. Available from: www.cdc.gov/HealthyYouth/health_and_academics

40. Chetty R, Friedman JN, Hilger N, Saez E, Schanzenbach DW, Yagan D. How Does Your Kindergarten Classroom Affect Your Earnings? Evidence From Project STAR. 2010.

41. United States Census Bureau. How the Census Bureau Measures Poverty [Internet]. [cited 2021 Mar 31];Available from: https://www.census.gov/topics/income-poverty/poverty/guidance/poverty-measures.html

42. Larrimore J. Does a higher income have positive health effects? Using the earned income tax credit to explore the income-health gradient. Milbank Q [Internet] 2011 [cited 2021 Feb 16];89(4):694–727. Available from: /pmc/articles/PMC3250638/

43. Dupas P, Miguel E. Impacts and Determinants of Health Levels in Low-Income Countries [Internet]. 2017 [cited 2020 Dec 19]. Available from: http://dx.doi.org/10.1016/bs.hefe.2016.09.003

44. Healthy People 2020. Access to Foods that Support Healthy Eating Patterns | Healthy People 2020 [Internet]. [cited 2021 Apr 8];Available from: https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-health/interventions-resources/access-to-foods-that-support-healthy-eating-patterns

45. Hilmers A, Hilmers DC, Dave J. Neighborhood disparities in access to healthy foods and their effects on environmental justice. Am J Public Health [Internet] 2012 [cited 2021 Apr 8];102(9):1644–54. Available from: /pmc/articles/PMC3482049/

46. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: Transportation barriers to health care access. J Community Health [Internet] 2013 [cited 2021 Mar 22];38(5):976–93. Available from: /pmc/articles/PMC4265215/

47. Arcury TA, Preisser JS, Gesler WM, Powers JM. Access to Transportation and Health Care Utilization in a Rural Region. J Rural Heal 2005;21 (1):31–8.

48. Rask KJ, Williams M V., Parker RM, Mcnagny SE. Obstacles Predicting Lack of a Regular Provider and Delays in Seeking Care for Patients at an Urban Public Hospital. JAMA J Am Med Assoc [Internet] 1994 [cited 2021 Apr 1];271(24):1931–3. Available from: https://jamanetwork.com/journals/jama/fullarticle/375572

49. United States Census Bureau. U.S. Census Bureau QuickFacts: Worcester city, Massachusetts [Internet]. [cited 2021 Feb 14];Available from: https://www.census.gov/quickfacts/fact/table/worcestercitymassachusetts/PST045219

50. City of Worcester - Innovation Office. Worcester COVID Equity Dashboard - City of Worcester, MA | Tableau Public [Internet]. [cited 2021 Mar 31];Available from: https://public.tableau.com/profile/cityofworcester-ma.innovationoffice#!/vizhome/WorcesterCOVIDEquityDashboard/EquityDashboard

51. Ichinohe T, Pang IK, Kumamoto Y, et al. Microbiota regulates immune defense against respiratory tract influenza a virus infection. Proc Natl Acad Sci U S A [Internet] 2011 [cited 2021 May 4];108(13):5354–9. Available from: doi/10.1073/pnas.1019378108

52. Weinstein JN, Geller A, Negussie Y, Baciu A. The National Academies Press Communities in Action: Pathways to Health Equity [Internet]. 2017 [cited 2021 Apr 10]. Available from: https://lccn.loc.gov/2017005055

53. Mein SA. COVID-19 and Health Disparities: the Reality of “the Great Equalizer.” J Gen Intern Med [Internet] 2020 [cited 2021 Apr 25];35(8):2439–40. Available from: /pmc/articles/PMC7224347/

54. Medline Plus. What does it mean to have a genetic predisposition to a disease?: MedlinePlus Genetics [Internet]. [cited 2021 May 3];Available from: https://medlineplus.gov/genetics/understanding/mutationsanddisorders/predisposition/

55. World Health Organization. Social determinants of health [Internet]. [cited 2021 May 3];Available from: https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1

56. Chowkwanyun M, Reed AL. Racial Health Disparities and Covid-19 — Caution and Context. N Engl J Med [Internet] 2020 [cited 2021 Apr 21];383(3):201–3. Available from: http://www.nejm.org/doi/10.1056/NEJMp2012910