Risk Factors Associated with COVID-19

by

Kevin Eric Fitzgerald Simmons

A Thesis submitted to the Graduate Faculty of Elizabeth City State University in partial fulfillment of the requirements for the Degree of Master of Science in Mathematics with Applied Mathematics Concentration.

Elizabeth City, North Carolina

April 2022

APPROVED BY

Mohammed Talukder, Ph.D. Committee Chairperson	Julian Allagan, Ph.D. Committee Member
Kenneth Jones, Ph.D. Committee Member	Gabriela Del Villar, Ph.D. Committee Member

ABSTRACT OF THESIS

RISK FACTORS ASSOCIATED WITH COVID-19

COVID-19 is a viral disease that began impacting the world in the latter part of 2019. In early 2020, the world would be in a state of pandemic due to the rapid spread of the disease. Statistical analysis has been an extremely useful tool in determining the impact of COVID-19. Some discussion about the COVID-19 pandemic is based on how medical and non-medical factors can influence the severity of the infection of a person. A case study is conducted regarding cases that occurred in Wake County, North Carolina. The association between demographic factors (age, sex, and race) and both ICU admittance and death are investigated using the chi-square test of association and binary logistic regression. Age and race have some significant association with severe cases of COVID-19 while sex does not.

ACKNOWLEDGMENTS

I would like to express my gratitude to all who have assisted me in my academic journey. To Mohammed H. Talukder, Ph.D., thank you not only for working with me on this project, but for the encouragement and guidance you provided. This would not be possible without your patience and dedication. To Kenneth L. Jones, Ph.D., and Julian D. Allagan, Ph.D., thank you both for all you have done for me throughout the course of my studies at Elizabeth City State University. Neither of you have ever hesitated to assist me when I have questions regarding coursework, registration for classes, or future plans. A special thanks to Gabriela del Villar, Ph.D., for taking time to serve as a committee member.

Contents

List of Figures

1 Introduction

1.1 COVID-19 Pandemic Timeline

The COVID-19 pandemic is issue in most of the world that has existed for approximately two full years. COVID-19 is caused by a strain of coronavirus and has symptoms that can include coughing, fever, loss of taste and/or smell, headache, chills, muscle aches, and shortness of breath [1]. "Severe Acute Respiratory Syndrome Coronavirus 2", or often abbreviated as SARS-CoV-2, is the virus responsible for COVID-19. This disease first arose in December of 2019 when a cluster of patients in China were suffering from pneumonia. The cause of their illness was initially unknown, but the cases could all be traced back to a market in Wuhan, China. On January 7, 2020, it would be discovered that a coronavirus was causing this string of illnesses. The first case of COVID-19 infections in the United States would be confirmed on the 20th of January. Just under two months later, a state of pandemic would officially be declared by the World Health Organization on March 11, 2020 [2]. Not long before the start of 2021, the first COVID-19 vaccine would receive approval by the U.S. Food and Drug Administration and could be administered to the public. Different variants of the disease would be encountered throughout the course of this pandemic. The COVID-19 pandemic would also play a role in the 2020 presidential election and would be one of the most discussed topics by politicians at the time [2]. COVID-19 would prove to be deadly all over the world and have a tremendous impact on the daily life of many people.

1.2 Background

There has been a fair amount of discussion on how COVID-19 has impacted people. Some research has been in regards to how being infected with COVID-19 effects people with preexisting health conditions, those from different socioeconomic backgrounds, ethnic groups, age groups, and more. Another goal of research has been to understand the impact that pandemic guidelines have had on the well-being of people and the economy. There have been studies have have concluded that individuals from racial or ethnic minority groups, along with those in lower socioeconomic class, at a greater risk of being diagnosed with COVID-19, hospitalization, admittance to intensive care, and death [3]. A solid understanding of how the virus impacts people from different backgrounds and groups is important in combating the COVID-19. This is because being infected with the virus, and harmful effects associated with isolation and pandemic guidelines, can impact everyone differently.

1.3 Purpose of this Study

The area of focus for this case study is Wake County, North Carolina. As of March 24, 2022, Wake County has had a total of 289,373 cases of COVID-19 infections with 1,040 of those resulting in death [4]. Wake County has the highest number of confirmed COVID-19 infections and the second highest number of deaths associated with the disease in comparison to all other counties located within North Carolina [5]. Throughout the county, about 73% of the population is fully vaccinated against COVID-19. Wake County has the highest population of any county in North Carolina and is home to the state's capital city, Raleigh [6]. According to the 2020 census, Wake County is home to 1,129,410 people and this number has likely grown by the time of writing. The primary goal of this project was to perform a case study of individuals from Wake County, North Carolina who have suffered from COVID-19 throughout the course of the pandemic and determine if demographic factors of age, sex, and race have any association to death and ICU admittance rates through the chi-square test of association and binary logistic regression. The primary research questions for this study are as follows:

1. Are the elderly at a higher risk of dying or being sent to the ICU?
2. Are females less likely to need intensive care or die as the result of COVID-19 infection?
3. Is race related to a higher risk of being admitted to the ICU and/or dying?
4. What are the measures of likelihood associated with the previous questions?

1.4 Limitations

There are some limitations that need to be addressed. Firstly, the data used was obtained from a larger set of data, which consisted of many incomplete items. These incomplete items were removed before conducting analysis since they do not provide enough information to be used. Secondly, variables in the data are entirely categorical and based on demographics. Data with more continuous measures for age could provide for some different possibilities regarding analysis. Also, if the data contained more information (such as income, employment status, etc.), these variables could be taken into consideration if the information is made readily available.

2 Data and Literature Review

Before conducting this study, baseline statistics and various pieces of literature were studied. This was to give an understanding of COVID-19 related statistics from a global view, research that has already been conducted and conclusions that are relevant to this study, and statistical methods of analysis that have been used and the impact of this research.

2.1 Global, National, and State Statistics

When considering the impact COVID-19 has had on the world as a whole, it is important to first take a look at the statistics regarding infections and deaths due to the disease. There have been a total of 486,761,597 total cases of COVID-19 and 6,142,735 deaths from the disease up to April 1, 2022 [7]. That is a rate of death equating to roughly 1.262% across the globe. While this rate appears to be low, COVID-19 is still causing deaths and lasting effects to those who face infection. There have also been a total of 11,054,362,790 doses of the COVID-19 vaccine administered globally. This number includes all doses that are included in a series of vaccinations. The United States has the highest number of both cases and deaths in comparison to any other country [7]. As of April 1, 2022, the United States has had a total of 79,342,899 cases of COVID-19 and 972,830 deaths attributed to the disease. That is a mortality rate of approximately 1.226% for the country, which is 0.036% lower than the global rate. COVID-19 vaccines were first available in December 2020 and have been widely distributed across the country and made available to the public in many locations. The effectiveness and safety of COVID-19 vaccinations has also been a hot topic for debate in the United States. So far, there have been 561,659,770 total vaccine doses administered in the U.S. with 217,774,183 people being considered as fully-vaccinated [26]. That is approximately 65.6% of the entire U.S. population who have completed a series of COVID-19 vaccinations. Nearly 5.08% of all administered vaccination doses have been in the United States.

Each state in the U.S. has handled the pandemic differently; some taking a more relaxed approach while others enforced strict guidelines. North Carolina's primary measures were for all citizens to wear face coverings and practice social distancing in public areas, to isolate oneself if exposed to someone who has tested positive for COVID-19 or if exhibiting symptoms of the disease, to wash hands and disinfect surfaces regularly, and recommending that individuals get fully-vaccinated if able to do so. These methods were found to be helpful in reducing the spread of the virus [8]. North Carolina has a population of 10,439,388 people as of 2020 [6]. Throughout the course of the COVID-19 pandemic, North Carolina has accumulated 2,630,506 cases and a total of 23,215 deaths as of April 3, 2022 [9]. This means that North Carolina has an overall mortality rate of approximately 0.883%, which is roughly 0.343% lower than the national rate and 0.379% lower than the global rate. While less than 1% of all COVID-19 cases in North Carolina have resulted in death, this does not mean that a considerable portion of the population has been severely impacted by complications of the disease. In regards to COVID-19 vaccinations, there have been 16,349,406 total administered doses. It was found that 6,465,525 North Carolina residents were fully vaccinated, which is 62.99% of the state's population [9]. In comparison to all other states in the U.S., North Carolina has been determined to have the 8th highest number of cases and 14th highest amount of deaths [27]. Out of the 50 states, and Washington D.C., North Carolina has sustained much higher transmission of the virus and more severe cases than most other in the country. Although mask mandates have recently been lifted for most of the state, is appears that COVID-19 may still be part of life in North Carolina for some time to come.

2.2 Relevant Studies

A topic for many discussions regarding COVID-19 is how preexisting factors can impact the outcome of COVID-19 infection. These factors may be purely medical, based on demographics, or based on socioeconomic standing. Obesity, a challenge that many people across the world face, comes with numerous issues. People suffering from obesity are more likely to be diagnosed with cardiovascular diseases and diabetes, and there is also a relationship between reduced respiratory function and weaker immune systems than individuals within a health weight range [10]. When obese individuals are infected with COVID-19, the likelihood of it having a severely negative impact rises. Studies have presented evidence that there is a high frequency of obese patients who require intensive care and are also obese. This implies that these two conditions when combined can be extremely harmful and that obesity is a risk factor for hospital admittance and death [10]. It has also been shown that many health conditions associated with obesity have contributed to more extreme cases of COVID-19 [11].

There is evidence to suggest that there is a link between obesity and poverty in the United States, possibly through limited access to proper food and physical health resources [12]. This is important as there has been research into the relationship between COVID-19 severity and socioeconomic class. Some evidence suggests that people with poor and/or crowded living conditions, high stress levels, low education, and those who are unemployed are at higher risk correlate with a more severe impact of the virus than people with more favorable conditions [11]. Poverty is a crisis of it's own in the United States that effects many people. The U.S. census bureau reports that roughly 37.2 million people lived in poverty in the United States in 2020 [13]. That is a rather large group of people in the country that are at greater risk of suffering from a severe case of COVID-19.

Race is a topic that often accompanies discussion of poverty. In 2020, Black Americans had a poverty rate of 19.5% which is the highest rate in comparison to Whites (8.2%), Hispanics (17.0%), and Asians (8.1%) [13]. Black Americans are not only more likely to live in impoverished conditions, but it is also known that a higher percentage suffer from diabetes, hypertension, asthma, and obesity in comparison to White Americans [14]. These conditions have a relationship to the severity of COVID-19. Taking all of these factors into consideration, it has been shown that a larger rate of Black Americans require medical care or suffer death than their White counterparts after COVID-19 infection [14]. Poverty rates, along with a higher predisposition to medical ailments, could also have an impact on minority races and ethnic groups.

Gender has been mentioned in the discussion of COVID-19 severity as well. The susceptibility of infection between the two sexes shows a small difference when comparing infection totals for the pandemic. It has been asked before if there are fundamental differences in how many diseases, not just COVID-19, effect men and women. Quite a few studies show that men and women are infected with COVID-19 at approximately the same rate [15].

While the infection rates are rather close between men and women, this does not mean that the rate of severe cases of the virus is the same. When analyzing hospitalization and death rates, it is prevalent that men are more likely have severe cases of COVID-19. Research has shown that both internationally and in the United States, men see higher rates of hospitalization and death [11][15]. There are some underlying social and medical trends that differ between men and women that may aid in explaining this. In the United States, a higher percentage of men smoke than women, men are less likely to seek medical help for ailments, and men are shown to suffer from hypertension more often then women [15]. Smoking has been shown to be harmful to the human body in numerous ways, especially in regards to the respiratory system. Hypertension is a condition that is mentioned frequently when discussing conditions that may lead to higher COVID-19 hospitalizations and deaths of males in select communities. Another potential explanation as to why men are more severely impacted by COVID-19 infections may lie biological differences with women, as opposed to environmental. A 2016 study suggests that due to hormonal and genetic differences, women often have a more effective immune response to viral diseases than men [16]. This means that overall, an average women may be more able to overcome COVID-19 without the need of extreme medical care than men.

The last factor that is known to have an association with COVID-19 severity that will be discussed is age. COVID-19 infections are possible for individuals of any age. It is certain that people coming from all age groups been infected by the virus throughout the pandemic. People from all of these groups have been hospitalized, admitted to ICU, and/or died as result of their infection. While it is important to understand that these things can happen to anyone, those of a younger age have been shown to be less likely of suffering from extreme cases of the disease. Despite children having immune systems that are not fully developed, it is believed by some researchers that the true number of COVID-19 infections in children is much higher than the reported numbers but due to mild symptoms or asymptomatic cases that simply go unreported [17]. Overall, COVID-19 has not been a disease with a tremendous negative impact on younger individuals who get infected. The same can not be said for older people, however. According to data posted by the CDC, an extremely large percentage of COVID-19 deaths in the United States are attributed to people who are 50 or older at the time of death (about 93.3%) [18]. The percentages of deaths for each of CDC's defined age groups are as follows: 18.2% for ages 50 - 64, 22.6% for ages 65 - 74, 25.9% for ages 75 - 84, and 26.7% for ages 85 and up. Infections for each age group do not follow the same trend as the death per age group. The 50 years and older group only makes up approximately 29.9% of the total infections in the United States (18.3% ages 50 - 64, 6.7% ages 65 - 74, 3.3% ages 75 - 84, and 1.6% 85 years and older). As people age, it is common for their lungs to become weaker and for their immune response to less effective at fighting off disease, leading to a higher risk of hospitalization and death in the elderly once suffering from COVID-19 [11].

2.3 Statistical Analysis

Statistical analysis has played a major part in how investigating the impact of COVID-19, the pandemic, and how the response has impacted human life. Many people have debated the effectiveness of preventative methods, laboratory tests, vaccines, and government intervention regarding slowing the infection rates and lowering deaths related to COVID-19. An analysis of 169 countries provided evidence that the mortality of COVID-19 is closely related to less testing and poor government response (and some other factors), meaning that increased testing and government effectiveness could possibly reduce deaths [19]. The pandemic has has a tremendous impact on the economy in most countries. Statistical studies have not only determined the true impact that the virus has had, but models of the potential effects of COVID-19 on employment and large businesses have been constructed in an attempt to provide an idea of what to expect in times of uncertainty [20]. As previously discussed, poverty is one of the conditions that is related to an increased risk of COVID-19 death. Increased unemployment rates and damage to businesses can lead to an increase in poverty, meaning that the mortality rate of COVID-19 could continue impacting those in this situation for quite some time. These are just a couple examples of the significance of statistical studies of the COVID-19 pandemic.

2.4 Data

The data for this project was compiled and made available for use by the Center for Disease Control. The page containing the data is updated on a monthly basis (as of March, 2022). The CDC has maintained a detailed table containing non-identifying, public-use data of individuals who have been infected by COVID-19 since January 2020 in the United States. All data has been provided by health departments and hospitals from across the country. The data consists of many factors for each case, but for the purpose of this project only some were taken into consideration. The important factors for this project are resident state, resident county, age group, sex, race, hospitalization status, ICU status, and death status. The location for cases was filtered to only include those from North Carolina, and then only ones from Wake County. The reason for choosing Wake County is that it is the most populated county in the state of North Carolina, it is the county that contains the state capital, and it provided an ideal amount of data for the purposes of this project. Mostly incomplete data items (elements that did not include information for selected factors) were removed from this set, leaving a sample size of 30,811 individuals who were infected with COVID-19 throughout the course of the pandemic. This was done to remove potential issues caused by incomplete data when performing the analysis and to have a more workable sample size. All non-hospitalized people with an "NA" label on ICU status were assumed to not be admitted to ICU since they never admitted to hospital to begin with. It is important to note a few things regarding this data. Not every single case was declared as a COVID-19 infection through lab results, but some were declared as such by a medical professional due to being symptomatic following exposure to someone who was confirmed to be COVID positive. The set of data used for this project does not consist of all COVID-19 cases in Wake County, North Carolina. These cases may not be made up entirely of different people. It is possible, but unknown due to all identifying information being excluded, that the cases present in the data me be attributed to the same person who was infected with COVID-19 on more than one occasion.

Descriptive frequencies of the data were computed to provide some insight into how many deaths and cases that required the individual be admitted to the ICU can be contributed to each demographic. This was done by constructing cross-tabulations with death and ICU being compared to age group, sex, and race. Then, the values obtained from these tables were used to generate both pie and bar graphs to assist in visualizing the statistics.

The age group containing the most overall infections was 18 to 49 years (16,650 infections, 54.0%), followed by 0 to 17 years (6,220 infections, 20.2%), then 50 to 64 years (4,875 infections, 15.8%), with the lowest being the 65+ years age group (3,066 infections, 10.0%). There was a slight difference in the total number of females (16,748 infections, 54.4%) infected in comparison to males (14,063 infections, 45.6%). The most prevalent race in the data set was white (20,018 infections, 65.0%), followed by Black (9,776 infections, 31.7%), with Asian (997, 3.2%) and American Indian/Alaska Native (20, 0.1%) making up the rest. Of all of the cases considered for this project, 924 of them (3.0%) resulted in hospitalizations and 121 (0.4%) resulted in admittance to ICU. Lastly, 221 (0.7%) of the cases resulted in the infected individual dying due to COVID-19 complications. Figure 1 provides an overview of the number of infections attributed to each age group. The pie chart in figure 2 depicts the infections for males and females included in the data. Lastly, figure 3 shows the number of infections for each of the races included in the data set.

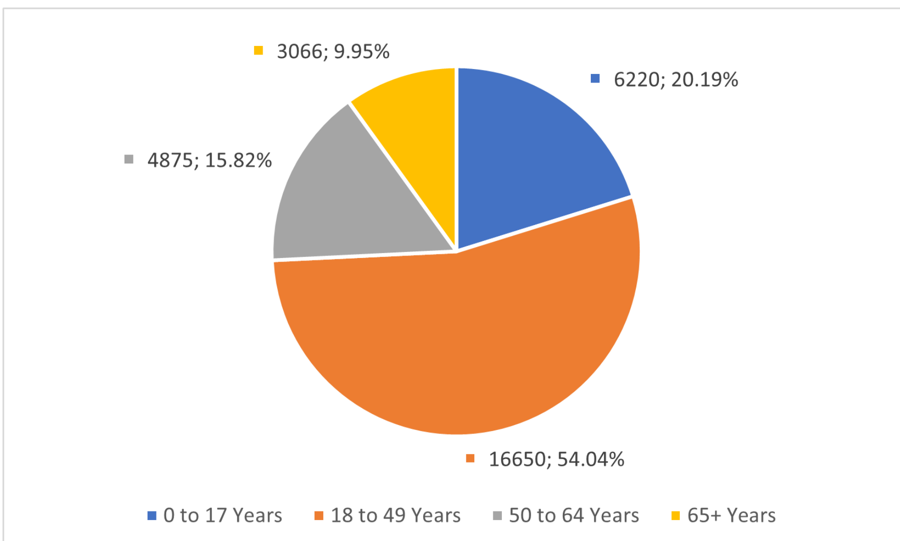

Figure 1: Infections for Each Age Group

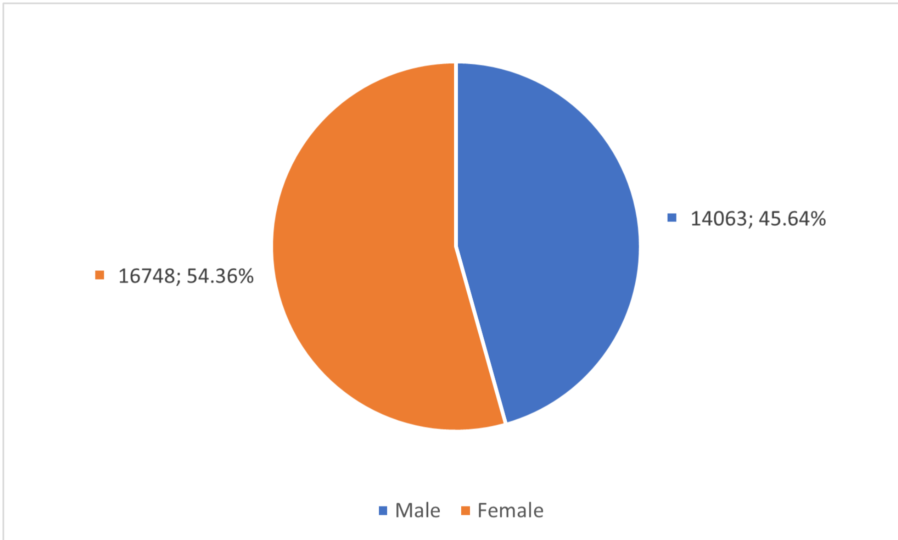

Figure 2: Total Cases for Each Sex

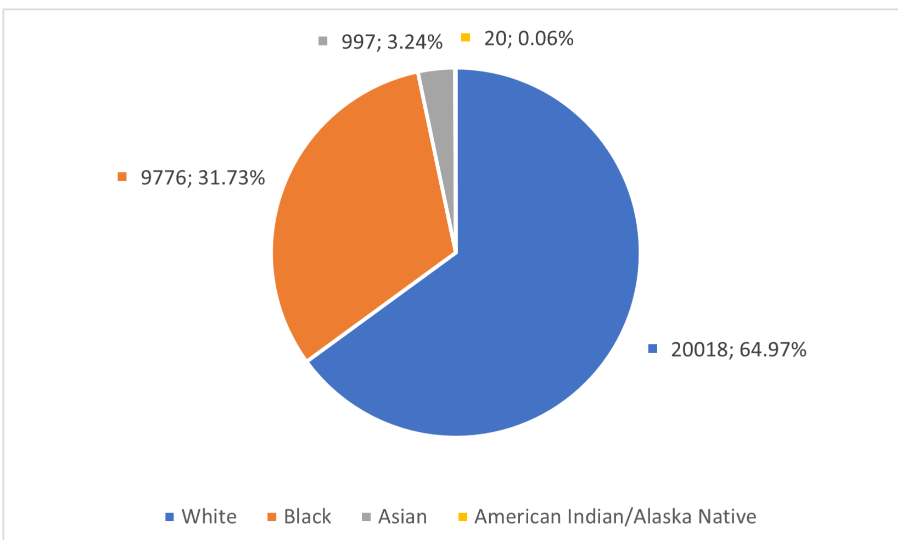

Figure 3: Total Cases for Each Race

When observing the number of severe cases in each of the four age groups included in the data, it appeared that the number of hospitalizations, ICU admittances, and death rose with age. For those 0 through 17 years old, only 30 cases required hospitalization, 4 saw intensive care, and there were no deaths. For the 18 through 49 years group, there were 199 hospitalizations, 18 ICU entrances, and 0 deaths. The 50 through 64 years group saw 226 hospitalizations, 40 ICU admittances, and still no deaths. Lastly, the 65 years and older group had 469 hospitalizations, 59 instances of ICU entry, and 221 deaths. It is important to note that all deaths were found to be attributed to people who were over the age of 65 at the time of death. The bar graph in figure 4 summarizes these totals with respect to each age group.

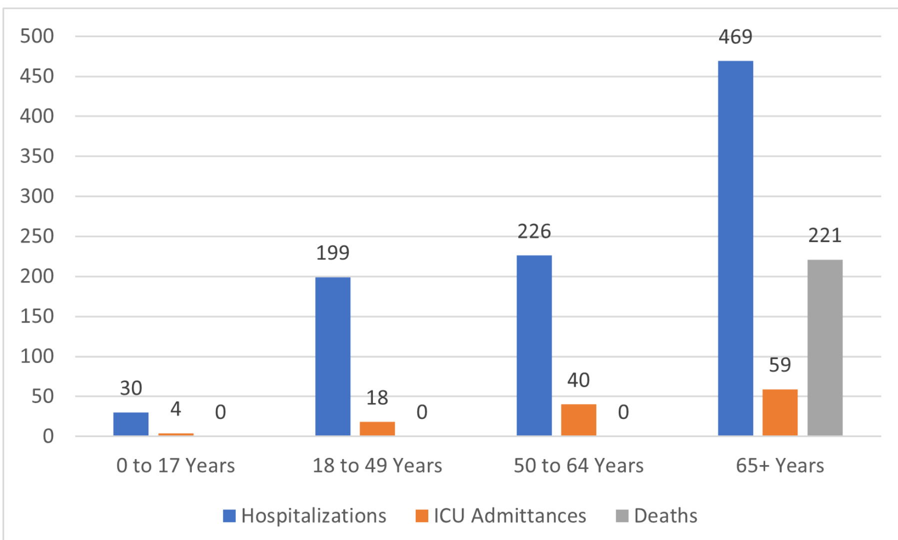

Figure 4: Severe Cases for Each Age Group

There some differences found in the amount of severe cases with regards to sex. A total of 408 hospitalizations, 50 ICU admittances, and 89 deaths were attributed to males. For females, there 516 cases requiring hospitalization, 71 entries to the ICU, and 132 deaths. There are more severe cases among women, but there are also more overall cases attributed to women as well. Figure 5 presents a bar graph depicting the amount hospitalizations, ICU entries, and deaths for each sex.

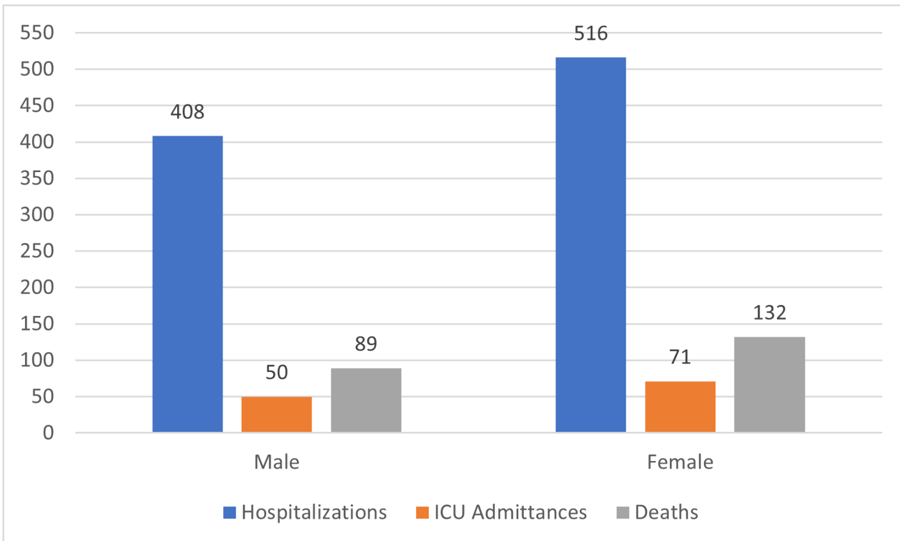

Figure 5: Severe Cases for Each Sex

The cases portrayed in this data could be attributed to either White, Black, Asian, or American Indian/Alaska Native people. There were a total of 502 hospitalizations, 59 ICU admittances, and 713 deaths from White people. The next largest group with respect to total number of infections was Black people. This group saw 406 hospitalizations, 61 cases requiring entry to the ICU, and 48 deaths. The Asian and American Indian/Alaska Native groups saw both significantly smaller numbers of both infections and severe cases. Only 15 hospitalizations and 1 ICU entry are attributed to

Asians. This group had no deaths. The American Indian/Alaska Native group only sustained a single hospitalization with no cases requiring ICU admittance or resulting in death. The bar graph in figure 6 presents all of these amounts.

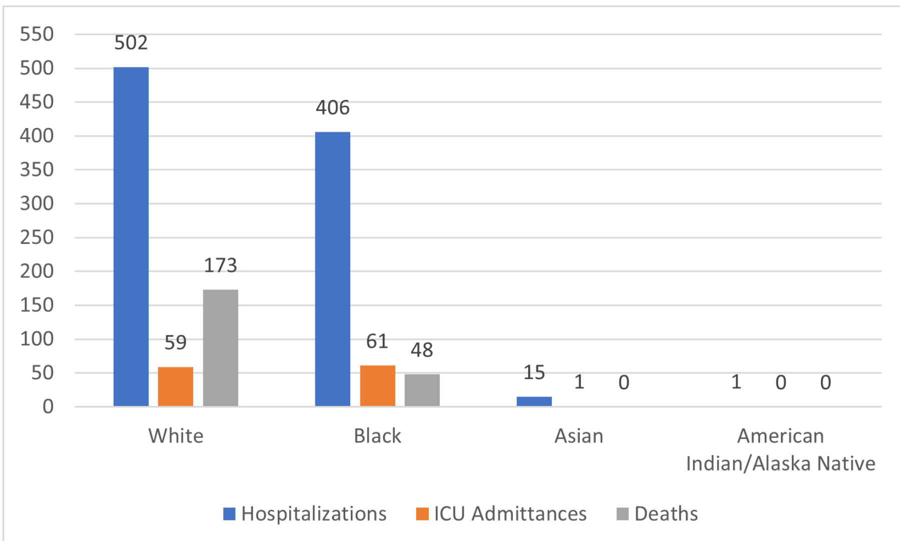

Figure 6: Severe Cases for Each Race

These graphs are important as they assist in visualizing how COVID-19 has effected each of the groups that have been discussed. With over 30,000 cases being considered for this analysis, only providing numbers and percentages may not be sufficient for understanding how these numbers are distributed for each of the demographic subgroups.

Comparing the total of all COVID-19 cases represented in the data to the frequency of hospitalization, ICU admittance, and death provided great insight as to how COVID-19 infections have impacted people from Wake County throughout the course of the pandemic. The 18 to 49 years old age group saw the highest number of infections than any other group based around age. Based on some information from Johns Hopkins, this may due to a high frequency of people in their 20s and 30s who have jobs that put them at a high risk of COVID-19 exposure, attend colleges that are subject to outbreaks of the disease, and some many may simply not follow proper methods of prevention due to a lower likelihood of severe cases [24]. Despite having the lowest number of infections, the 65 years and older age group had the highest frequency of hospitalizations, ICU admittances, and deaths. The reasoning behind this may be due to reasons discussed earlier that put older people at a higher risk of having a severe case of COVID-19.

The total number of males and females who were infected with COVID-19 did not differ too much. There were slightly more females infected than males. This trend continued on with hospitalizations, ICU admittances, and deaths. Slightly more females were both infected and saw severe cases of COVID-19 in comparison to males.

The frequencies with regards to race displayed some unexpected trends. There were just over twice as many COVID-19 cases attributed to Whites as opposed to Black people, with a rather small portion being from Asians and American Indians/Alaska Natives. White people had the highest number of infections that resulted in hospitalization or death, but the frequency of cases resulting in ICU admittance was very close between Whites and Blacks. The reasoning behind this trend is unknown.

3 Methods

The analysis of the data for this project is completed using SPSS Statistics by IBM. This software is chosen do to its efficiency of statistical computations and implementation of models. First, the chi-square test of association is utilized to observe if there is some statistical association between the demographic factors of age, sex, and race with regards to ICU entry and death. This is done to provide some insight into whether or not someone from a different group may be more subject to the selected outcomes. Then, binary logistic regression analysis of the data is performed to determine the risk associated with the sex, race, and age in regards to dying and ICU admittance due to COVID-19. Regression analysis should provides some insight into the statistical associations that were found through the chi-square test. SPSS was used in obtaining the statistical frequencies and in all of the statistical computations for this project.

3.1 Chi-Square Test of Association

The chi-square test of association (also commonly referred to as simply the chi-square test) is a statistical method of testing the association shared between categorical variables and hypothesis testing [28]. It is commonly utilized in situations where the association between two binary variables are being investigated. Often, a contingency table or cross-tabulation will be implemented to observe frequencies with regards to the variables. Figure 7 provides an example of one of these tables.

		Category 2	
		1	2
Category 1	1	x_{11}	x_{12}
	2	x_{21}	x_{22}

Figure 7: Example of a Contingency Table

Consider a set of data consisting of n people. This table separates everyone bases on two classifications, "Category 1" and "Category 2". These categories are broken down further into two sections, "1" and "2". The variables within the table provide the number of people that fit into that group. It is important to not that the sum of these values will result in the number of people in the sample. Mathematically speaking,

$$x_{11} + x_{12} + x_{21} + x_{22} = n.$$

Two variables that needs to be discussed are "observed value" and "expected value." Observed value is simply the number of individuals belonging to one of the four groups based purely on the data. This value should already be in the contingency table. The expected value is an estimation of how many people belong to one of the groups based on the total values for the respective rows and columns obtained from adding observed values. The following equation can be used to determine expected value:

$$E = \frac{(RT)(CT)}{n}$$

where E = Expected Value, RT = Row Total, and CT = Column Total. The totals are calculated by added the values found in each cell in the row or column respective to the specific grouping that is being investigated.

Now that observed and expected values have been discussed, the chi-squared value (or statistic) can be found. This value will be denoted as χ^2. This value is derived from a chi-square distribution and is used when comparing observed and expected values [28]. It is also rather important when discussing the testing of a null hypothesis. In many studies, the null-hypothesis may be defined as there being no relation between the two variables and the alternative-hypothesis being that there is some significant association [28]. The following equation defines χ^2:

$$\chi^2 = \sum \frac{(O-E)^2}{E}.$$

χ^2 is needed to determine whether or not the null-hypothesis should be rejected, however, some more information is needed. The corresponding p-value is used to measure the significance of χ^2 and the association between the variables. To find the p-value, one must find the area under the graph (to the right of χ^2) of a chi-square distribution with the same degrees of freedom as the data being investigated. The equation for finding degrees of freedom is:

$$df = (R-1)(C-1)$$

where R = the number of rows and C = the number of columns found in the contingency table. Once the χ^2 and the degrees of freedom are found, one can use a variety of tools (such as a table p-values associated with chi-Square distributions or computational software) to find the needed p-value. The p-value (which is short for probability value) is a the measure of making an error when rejecting the null-hypothesis. So, a low p-value implies that there is a lower probability of making an error if accepting the alternative hypothesis. This value will be compared to a predetermined significance level (often something like $\alpha = 0.05$ or $\alpha = 0.01$). In this case, if the p-value is greater than α, then the null-hypothesis is not rejected, meaning there is no association between the variables. If the p-value is less than α, then the null-hypothesis will be rejected and the alternative-hypothesis accepted, implying that there is some association between the variables.

3.2 Binary Logistic Regression

Binary logistic regression is an extremely useful tool when performing analysis in studies where there are results that are split into to two categorical values (for example: yes or no, lived or died, employed or unemployed, etc.) [21]. When utilizing this method, one or more characteristics (independent variables) are chosen and the goal is to determine their influence on some outcome (dependent variable). It is used to predict group membership based on the information that is given for some data set and measures how some independent variable(s) can impact this outcome. An example of this in many medical based studies where certain living conditions of patients may influence the severity of a condition, such dog owners surviving a heart attack [21]. Independent variables may be discrete or continuous while the dependent variable must be dichotomous. In literature that covers binary logistic regression, these dichotomous variables, or ones that can only take one of two different values, are often referred to as "dummy variables". A detailed explanation of the work used in binary logistic regression is necessary in understanding results.

An important question to ask is "why should binary logistic regression be used, rather than some other method of analysis?" Linear regression is able to be used in various situations, but there are instances in which is not a viable method. If the selected dependent variable is dichotomous, then it may be necessary to use binary logistic regression instead of a variation of linear regression. When conducting research on a set of data and investigating the relationship between 1 categorical variable and some dependent variable, it may not be necessary to conduct binary logistic regression. This method excels in cases where a predictor variable is continuous or there are multiple predictor variables that are influencing the outcome variable [25].

The first component needed in understanding binary logistic regression is the computation of odds. Conceptually, odds are similar to probability but there are some fundamental differences that will be discussed. Probability is the likelihood of some outcome occurring with respect to all possible outcomes or over an entire population while odds is the direct comparison between the occurrences of two different outcomes. Consider the following example to reinforce the difference. Suppose there is a sample of 100 people where 25 have the same disease. The probability of some individual being sick is $25/100 = 0.25$. The odds of an individual being sick can be computed by dividing the number of sick people in the population by the number of non sick people. So, the odds of being sick are $25/(100 - 25) = 25/75 \approx 0.333$. The odds of not being sick can easily be calculated by taking the reciprocal of the odds of being sick (which would be $75/25 = 3$).First, probability is known to have a range of 0 to 1 (or 0% to 100%). There is a 25% chance that a person selected from the group will be sick. Odds have a range of 0 to $+\infty$. A value less than one means the focused outcome is less likely to happen while a value greater than one means that it is more likely to happen. If the odds values is equal to one, that means that both outcomes are equally likely to happen. It is observed that the odds of the person being sick, as opposed to well, is approximately 0.333. This means that a randomly chosen person is one-third times less likely to be sick than well. The odds for a person being well are 3, which implies that it is 3 times more likely that a randomly selected individual is not ill.

The next step in the process is understanding odds ratios. These are a little different than simple odds. An odds ratio is a direct comparison between the odds of two different events. Suppose the odds of someone being a smoker is 0.125 and the odds of someone being infected with illness are 0.500. The odds ratio (with being ill as the reference group) can be simply computed as $0.125/0.500 = 4$. This value would conclude that the odds of being sick are 4 times more likely for someone who is a smoker (for the sake of this example).

Odds ratios have some interesting qualities. They are typically not normally distributed or related to scores that may be given by quantitative predictor variables (the values in the previous examples are purely quantitative values) [21]. Due to these conditions, the odds ratio does not provide much use for other applications, particularly regression analysis, but it can be transformed using the natural logarithm to give it a more applicable form.

When the natural logarithm is taken of an odds ratio, it gives a new outcome, or dependent, variable. This variable is referred to as the "logit" and will be denoted as L_i. The logit variable has some benefits that make it more effective than the odds ratio for the dependent variable for a regression model [21]. L_i does not have any fixed upper or lower limits, it is often normally distributed, and it can often be related to scores provided by quantitative predictor variables. The main disadvantage of the logit variable is that it can not be interpreted without some alteration, whereas probability values are easier to interpret. The equation

$$L_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$$

can be used to determine the value of L_i using a linear function of scores that are determined for different predictor variables. Binary logistic regression generates values for L_i that can be used to estimate the predicted probability of the target outcome occurring for an element with specific characteristics. When all independent variables are considered for the previous equation, this is often called the full model. The equation

$$L_i = \beta_0$$

does not represent any indicator variables, and it is referred to as the null model. The equation will return the log odds of the selected outcome occurring when considering no conditions. In binary logistic regression, independent predictor variables are implemented to the equation to enhance the odds of predicting the target outcome for an individual. If the variable X_1 is binary, it will take the value 0 or 1 in the equation. The logit variable is related to the idea of predicted probability by the following equation:

$$L_i = ln(\frac{p_i}{1-p_i}).$$

This shows that L_i, which can be found using a linear function, can be expressed in terms of probability. This also means that once the β_i constants are calculated, the computed value for L_i can be used to determine the probability of the selected outcome occurring for an individual based off of their qualities that correspond to the predictors [25]. Consider that there are 100 cars on a lot, and at the end of the day, 21 of them have been purchased. Without any consideration of variables that may contribute to certain cars being purchased, the probability of a car on that lot being purchased is 0.21. Certain independent predictor variables (such as car color, engine size, miles per gallon rating, etc.) can be implemented to compute odds ratios to allow the B_i values to be found. Then, L_i equation is used to find the logit value, or log odds, based on the values that accompany these predictor. Recall that the logit variable was related to predicted probability through the previous equation. Once L_i is found for an individual, predicted probability can be estimated as

$$p_i = \frac{e^{L_i}}{1+e^{L_i}}.$$

The predicted probability of one belonging to the selected outcome category can be found using this relation [25].

This method of analysis may be tedious to complete by hand, especially when working with large data sets with many variables, but SPSS allows for a user to utilize binary logistic regression in a manner that does not require much, if any, hand computation. One can upload the data set to the program. Then, if not already done so, the dependent/outcome variable will be used need to be changed to a binary format. If this alteration cannot be done without compromising the data, then another method of analysis should be used. One can now select the option to perform binary logistic regression, set the dependent and predictor variables, any parameters the user wants set, and then run the analysis. The results will be computed by the program, but it is the responsibility of the user to properly interpret the output. While the odds ratios are a key component of the outcome, a researcher must also consider the statistical significance of variables in the model and how accurately the computed model fits the data.

The first section of the output includes the "Case Processing Summary" and "Dependent Variable Encoding." These sections don't contain much information, but they inform the user of how many cases were included in the analysis, how many were missing, and how the outcome variable was coded

(0 or 1). The "Categorical Variables Codings" sections presents a table that provides the frequencies associated with each predictor variable. The next section, labeled "Block 0" computes the null model (excluding all predictor variables) for binary logistic regression. The main purpose of this section is be used standard measure to be compared to the full model after predictor variables are implemented. The final section of the output is labeled "Block 1." It is within this section that most goodness-of-fit and significance measures are found, along with the final odds ratios along with their confidence intervals. The main focus of the output here are the sections β, e^{β}, the 95% confidence interval for e^{β}, and the p-values. Recall that β are the log-odds used to determine the value of the logit variable, e^{β} is the odds of the select outcome occurring for the particular group, and the p-value is used for significance testing.

3.3 Re-categorization of Data

Binary logistic regression is a chosen method for this project due to the categorical structure of the data. The data consists primarily of classifications between two main groups for outcome variables. These are whether an individual was admitted to ICU or not, and if the infected person died or lived. As discussed in the previous subsection, it is necessary to have the outcome variable in a binary (or dichotomous) format to conduct binary logistic regression analysis.

While it is not necessary to have indicator variables be binary, age groups and race were re-categorized to form two distinct groups for potentially clearer results. Due to the classification of race being primarily split between black and white, with a much lower number of infections occurring in Asian and American Indian/Alaska Native people (approximately 3.3% between the two) the deciding variable was set to infections in Black versus non-Black people. The means that for the purposes of analysis, there are 9,776 and 21,035 COVID-19 cases attributed to Black and not Black people, respectively. This distinction was chosen as Black people make up a large portion of infections in the data set and they have the highest poverty rate of any race in the United States according to national data [13] and there has been evidence to suggest that Black people have a higher likelihood of dying from COVID-19 [14]. The distinction could have been made between Whites and individuals who are not white, but the data includes Hispanic/Latino people in the White race group. Hispanics and Whites are often viewed as separate races. According the the United States Census Bureau, Hispanic or Latino could be considered part of different races [23].

The grouping of age was based off of larger groups of 0 to 17, 18 to 49, 50 to 64, and 65 years or older as opposed to each patients exact age being given. This was also modified into a binary structure. By separating the number of patients along one of the years defining the age groups, it is easy to define two distinct groups. 65 years old was selected as the boundary age for the age grouping.

4 Results

The chi-square test of association is performed multiple times for the data. The association between the following variables is tested: age and ICU, age and death, sex and ICU, sex and death, race and ICU, and then finally race and death. Binary logistic regression is then performed a few times on the data. This is because there are two different outcome variables (ICU admittance and death) that must be analyzed separately. The predictor variables for the analyses are the binary age group with the border age of 65 years, the infected person's sex, and whether the individual was Black. The names of these variables in output are are follows: "Under65" for people who are under 65 years old or those who are 65 years and older, "Sex" for the male or female categorization, and "Black" for the individuals whose race is either Black or not Black. Note that $\alpha = 0.05$ is the value used in all significance testing.

4.1 Results of Chi-Square Test

All uses of the chi-square test are completed in SPSS, which provides an output containing all χ^2 values and p-values. The degrees of freedom for each test are 1.

First, the chi-square test of association is used to investigate any associated between age and ICU entry due to COVID-19. Figure 8 provides the contingency table for this test.

		ICU?		
		No	**Yes**	**Total**
Under 65?	**No**	3007	59	3066
	Yes	27683	62	27745
	Total	30690	121	30811

Figure 8: Table Comparing Age to ICU Entries

The test yields the results of $\chi^2 = 204.184$ with a p-value that is less than 0.001. Clearly, $0.001 < 0.05$ and therefore the null-hypothesis of no association is rejected. This means that there is some level of association between being 65 or older and requiring intensive care due to a severe case of COVID-19.

For the second use of the chi-square test, the association between age and death due to COVID-19 is observed. The table in figure 9 summarizes the frequencies used here.

		Died?		
		No	**Yes**	**Total**
Under 65?	**No**	2845	221	3066
	Yes	27745	0	27745
	Total	30590	221	30811

Figure 9: Table Comparing Age to Deaths

The results here provide the values $\chi^2 = 2014.333$ and a p-value that is sufficiently close to 0. It can be concluded here that the null-hypothesis should be rejected and that there is some statistically significant association between age and death present in the data.

The third and fourth use of the chi-square test are with regards to sex and ICU admittance followed by sex and death. Figures 10 and 11 provide cross-tabulations for sex and ICU admittance, and sex and death.

		ICU?		
		No	**Yes**	**Total**
Sex?	**Female**	16677	71	16748
	Male	14013	50	14063
	Total	30690	121	30811

Figure 10: Table Comparing Sex to ICU Entries

		Died?		
		No	**Yes**	**Total**
Sex?	**Female**	16616	132	16748
	Male	13974	89	14063
	Total	30590	221	30811

Figure 11: Table Comparing Sex to Deaths

For the examination of sex and ICU entry, small value of $\chi^2 = 0.914$ was found and the accompanying p-value was equal to 0.339. For the test between sex and death, it was found that $\chi^2 = 2.588$ with a p-value of 0.108. It is obvious that both $0.339 > 0.05$ and $0.108 > 0.05$. This means that for both test, the null-hypothesis is accepted. In other words, there is no statistical association between sex and ICU admittance or between sex and death due to COVID-19 based on the data.

The fifth use of the chi-square test is to analyze the potential association between race and ICU entry. This contingency table is depicted in figure 12.

		ICU?		
		No	Yes	Total
Race?	Not Black	20975	60	21305
	Black	9715	61	9776
	Total	30690	121	30811

Figure 12: Table Comparing Race to ICU Entries

It was found here that $\chi^2 = 19.577$ and its corresponding p-value is smaller than 0.001. It is known that $0.001 < 0.05$ and therefore the null-hypothesis of no association is rejected. It can be concluded that the race of an individual is associated with the likelihood of them requiring intensive care due to a COVID-19 infection.

Finally, the chi-square test is conducted one last time with regards to race and death due to COVID-19. This contingency table is depicted in figure 13.

		Died?		
		No	Yes	Total
Race?	Not Black	20862	173	21305
	Black	9728	48	9776
	Total	30590	221	30811

Figure 13: Table Comparing Age to ICU Entries

It was determined here that $\chi^2 = 10.296$ with a p-value of 0.001. It is observed that $0.001 < 0.05$, and the null-hypothesis shall be rejected. This implies that there is some significant association between race and dying from COVID-19.

4.2 Results of Binary Logistic Regression

Binary logistic regression is carried out a few times for this study and all results are compiled for analysis. As mentioned in chapter 3, the predictor (or independent) variables used in the test were if the as follows: the individual is 65 years of age or older, the individual is Black, and the individual is either male or female. For the odds and predicted probabilities, it is important to note the values that are associated with each classification for the predictor variables. For the "Under65" variable, the values are assigned as 0 = under 65, and 1 = older than 65. Regarding the "Sex" variable, a value of 0 = male, and 1 = female. Lastly, for the "Black" predictor variable, 0 = the person is Black, and 1 = the person is some other race. The observed outcome (or dependent) variables are ICU admittance and death. The results obtained for logistic regression each will be reviewed in detail. For significance testing, it is assumed that $\alpha = 0.05$.

Binary logistic regression is first carried out with age, sex, and race being predictors and the outcome being admission to the ICU. The table in Figure 14 provides the information obtained through binary logistic regression.

	β	e^{β}	95% Confidence Interval	p-value
Under65(1)	2.297	9.947	[6.914, 14.309]	$<$ 0.001
Sex(1)	0.016	1.106	[0.705, 1.465]	0.931
Black(1)	−1.006	0.366	[0.254, 0.546]	$<$ 0.001
Constant	−5.554			

Figure 14: Age, Sex, and Race as Predictors for ICU Admittance

For Under65(1) (which if for people 65 years and older), the odds ratio is 9.947 (with 95% confidence interval of [6.914, 14.309]) and a p-value of $<$ 0.001. Therefore, the odds here are statistically significant and it can be said that someone who is 65 years of age or older is 9.947 times more likely to require intensive care than someone who isn't. The confidence interval being greater than 1 implies that there is a significant connection between older age and needing ICU entry. Now observe the values accompanying the Sex group. The odds ratio here is 1.016, which is extremely close to 1 and a 95% confidence interval of [0.705, 1.465]. This could imply that Females are 1.016 times more likely to need ICU admittance than their male counterparts, however, the p-value is 0.931. Since $0.931 > 0.05$, and the odds are close to 1, and the confidence interval crosses over 1, it can be understood that sex is not a statistically significant variable. Recall that an odds value of 1 means that the both outcomes are equally likely to happen. In other words, being a female (or male) does not increase nor decrease the likelihood of being admitted to the ICU. Finally, the Black group has an odds ratio of 0.366 with p-value of ¡ 0.001 and 95% confidence interval of [0.254, 0.546]. This is statistically significant and it can be said that someone who is not Black is 0.366 times as likely to need intensive care than someone who is Black. If we take the inverse of this ratio (i.e. $[e^(\beta)]^{-1} = \frac{1}{e^(\beta)}$) to find the odds ratio for the opposing classification. So, $\frac{1}{0.366} \approx 2.732$. This means that a Black individual is about 2.73 times more likely to be admitted to ICU than some someone of another race.

Binary logistic regression is then carried out again with the same predictor variables but with the dependent variable being death. Recall back to chapter 2 where the number of deaths attributed to each age group are discussed. All of the COVID-19 related deaths present in the data set were from individuals who were 65 years or older. With no deaths occurring in the group of people under 65 years old, the logistic regression model is severely impacted. The results are presented in Figure 15.

	β	e^{β}	95% Confidence Interval	p-value
Under65(1)	18.646	125269277.54	$[0, 3.078 \times 10^{213}]$	0.938
Sex(1)	-0.026	0.974	$[0.737, 1.289]$	0.856
Black(1)	0.033	1.034	$[0.742, 1.441]$	0.844
Constant	-21.212			

Figure 15: Age, Sex, and Race as Predictors for Death

Notice that the odds ratio present for the age group variable is extremely high (125,269,277.54) along with a high p-value (0.938). This means that the age group predictor is not statistically significant with regards to death. The sex and race groups are also found to be insignificant here. These results are not useful due to these issues.

Binary logistic regression is perform again but this time ignoring age. The predictor variables here are sex and race, and the outcome variable is death. The decision to ignore age here was on the premise that it was severely skewed towards the older of the two age groups and that this may have impacted the other variables. The exclusion of age, since all deaths were found to be in a single category, allowed for a more preferable model to be created that contained some result that are able to be used. Figure 16 provides a table depicting the results for this model.

	β	e^{β}	95% Confidence Interval	p-value
Sex(1)	0.243	1.275	$[0.973, 1.670]$	0.078
Black(1)	0.532	1.702	$[1.235, 2.346]$	0.001
Constant	-5.459			

Figure 16: Sex and Race as Predictors for Death

Notice that notice that the significance values are much lower than in the previous table. Regarding the sex of an individual, it was found that odds of dying are approximately 1.275 time higher of a female dying as opposed to a male with a 95% confidence interval of [0.973, 1.670] and a p-value of 0.078. Since $0.078 > 0.05$, this information is not statistically significant. The confidence interval is observed to cross over the value of one, which coincides with sex being a statistically insignificant predictor for death from COVID-19. Considering race as the predictor, it was found that a person who is not Black was about 1.702 times more likely to die of COVID-19 than someone who is Black with a 95% confidence interval of [1.235, 2.346]. The p-value here is < 0.001, which is smaller than our test value, α. Therefore, this information is statically significant.

5 Discussion

Almost all instances of the chi-square test for association result in rejection of the null-hypothesis. Recall that the null-hypothesis was that there is no association between the selected demographic factor and either death or ICU admittance as a result of COVID-19. The alternative-hypothesis is that there is an association between the two said variables. Age and race were found to have some association with both death and entry to the ICU. The only demographic factor found to not have some statistically significant association with either death or ICU admittance was sex. This means that there may be no substantial difference in the rate in which men and women suffer from severe COVID-19 cases based on the data.

While not all of the obtained results from the binary logistic regression are statistically significant, there are still some notable findings. With regards to ICU admittance, it was found that the odds of someone 65 or older are approximately 9.947 times more likely to require this level of medical care as opposed to someone younger than 65 based on the data obtained from Wake County, North Carolina. Based on statistics published by the CDC on March 24, 2022, approximately 11.6% of all COVID-19 cases were attributable to people 65 years and older while an astonishing 75.2% of all deaths were from this group [18]. While this is not necessarily ICU admission data, it would make sense that the group with the highest frequency of deaths could also require intense medical care more often as well. Due to all of the deaths depicted in the set of data used in this project belonging to the 65 years and older group, binary logistic regression could not be used to provide statistically significant odds of death predicted by a persons age. As discussed earlier, some sources indicate that that weaker lungs and immune response is observed in the elderly [11], which may be the reasoning behind this trend. Due to the extreme increase in odds of ICU admittance and all deaths being attributed to the older age group, elderly people (especially those who are older than 65) should exercise caution if they find themselves infected with COVID-19.

The analysis of the data with regards to race also presents some interesting results. Prior research indicated that it was likely that Black people were at a higher risk of severe cases of COVID-19 rather than other races due to higher poverty rates [13] and a greater presence of conditions known to contribute to severity of the disease [11][14] in comparison to other races (especially Whites). Based on the sample data from Wake County, North Carolina, it was found that the odds of a Black person being sent to the ICU were approximately 2.732 times higher than that of someone of any other race (particularly White, Asia, or American Indian/Alaska Native) based on the data. Similar results were expected to be found when considered death as the selected outcome, but the opposite would occur. It was found that people who are not Black are about 1.702 times more likely to die as a result of COVID-19 infection than someone who is Black. Both of these results were found to be statistically significant. The question remains though: what is causes the differing results for the group of Black people in Wake County with regards to death and ICU entry?

The sex of those who have been infected with COVID-19 is used as a predictor variable in the analyses, but the results show there is a statistically insignificant difference in odds between males and females dying or being sent to the ICU based on the severity of their case. Women are found to be 1.016 times more likely to be admitted to intensive care and 1.275 times more likely to die from COVID-19. However, these findings were found to be statistically insignificant, meaning that these findings are not of importance. It can be understood that males and females are share approximately the same likelihood of requiring intensive care and dying as a result of COVID-19.

The results of binary logistic regression on the data overall are consistent with the findings of the chi-square test of association. It is determined that age and race could be significant predictors of a severe case of COVID-19 and it is found that individuals older than 65 and Blacks are at a higher risk of ICU entry while people who are not Black are at a higher risk of death due to COVID-19. Sex is found to have no statistically significant association with death or ICU entry through the chi-square test, and similar results were observed with binary logistic regression. While age is determined to have a relationship with dying from COVID-19, it could not be used properly as a predictor variable for binary logistic regression due to all deaths occurring in the 65 years or older group.

6 Conclusion

6.1 Summary

The COVID-19 pandemic has had an extreme impact on life for people all over the world. COVID-19 is a viral disease with symptoms that primarily impact the respiratory function of an infected individual. Due to the widespread effects of this disease, there has been research into how COVID-19 impacts people with specific preexisting conditions. These conditions may be specifically medical (such as diabetes or hypertension) or non-medical (age, race, or sex). The case study of COVID-19 data for Wake County, North Carolina was conducted to investigate the relationship between the aforementioned non-medical factors and ICU admittance and death due to COVID-19 complications. A set of data containing non-identifying information for 30,811 COVID-19 cases from Wake County was used in this project. Descriptive statistics were obtained to give an idea of how many people from each defined group compiled the number of ICU entries and deaths. The chi-square test of associations is utilized to provide insight into how the demographic factors of age, sex, and race may be associated to ICU admittance and death. Finally, binary logistic regression is utilized to provide and idea of the odds associated with requiring intensive and death with regards to age group, sex, and race. Logistic regression analysis also provided some model equations that can be utilized to estimate the predicted probabilities of ICU admittance and death of a person given some of their age, sex, and race.

6.2 Primary Findings of this Study

Based on the data for Wake County, North Carolina, it was determined through the chi-square test of association that age and race share some association with ICU admittance and death due to COVID-19. Sex was determined through the chi-square test to not have a statistical association with death or ICU entry.

Binary logistic regression shows that people 65 years and older are 9.947 times more likely to be admitted to the ICU than anyone younger and Black people are about 2.732 times more likely to need intensive care than people of other races. It is also determined that people who are not Black are around 1.702 times more likely to die from COVID-19 than someone who is black.

All of the 4 research questions that were asked at the start of this study have now been answered.

6.3 Ways to Improve Future Studies

It would potentially be beneficial to perform this analysis on a larger set of data, potentially for complete individual data for the entire state of North Carolina or some other territory that has been impacted by the COVID-19 pandemic. This may provide more substantial and accurate results than this analysis was able to do. Another way to improve on this research would be to include more potential predictor variables. If more accurate information regarding age, financial standing, living conditions, preexisting medical conditions, ethnicity, or more happens to be readily available, it would be interesting to see if those factors have any relationship with hospitalization, ICU admittance, and death due to COVID-19. Following these suggestions may provide more substantial and statistically significant results that can be used to predict the likelihood of someone facing intensive care and/or death as a result of COVID-19.

References

[1] Symptoms of COVID-19. (1 March 2022). Center for Disease Control. https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

[2] CDC Museum COVID-19 Timeline. (5 January 2022). Center for Disease Control. https://www.cdc.gov/museum/timeline/covid19.html

[3] Khanijahani, A., Iezadi, S., Gholipour, K., Azami-Aghdash, S., Naghibi, D. (2021). A systematic review of racial/ethnic and socioeconomic disparities in COVID-19. International Journal for Equity in Health, 20(1), 1–30. https://doi-org.ecsu1891.idm.oclc.org/10.1186/s12939-021-01582-4

[4] COVID-19 Cases in Wake County. (24 March 2022). Wake County COVID-19 Information. https://www.wakegov.com/covid-19-information

[5] COVID-19 North Carolina Dashboard. (24 March 2022). NC Department of Human and Health Services. https://covid19.ncdhhs.gov/dashboard

[6] 2020 Census, North Carolina - Total Popluation by County. (14 August 2021). North Carolina General Assembly. https://www.ncleg.gov/Files/GIS/Base_Data/2021/Reports/PL94_171_2020_CountyPop.pdf

[7] WHO Coronavirus (COVID-19) Dashboard. (3 April 2022). World Health Organization. https://covid19.who.int/

[8] Lio, C.F., Cheong, H.H., Lei, C.I., Lo, I.L., Yao, L., Lam, C., Leong, I.H. (29 April 2021). Effectiveness of personal protective health behaviour against COVID-19. BMC Public Health. https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-021-10680-5

[9] North Carolina - COVID-19 Overview. (3 April 2022). Johns Hopkins University Medicine. https://coronavirus.jhu.edu/region/us/north-carolina

[10] Caci, G., Albini, A., Malerba, M., Noonan, D.M., Pochetti, P., Polosa, R. (4 August 2020). COVID-19 and Obesity: Dangerous Liaisons. Journal of Clinical Medicine. https://www.mdpi.com/2077-0383/9/8/2511

[11] Rashedi, J. Poor, B.M., Asgharzdeh, V., Pourostadi, M., Kafil, H.S., Vegari, A., Tayebi-khosroshahi, H., Asgharzadeh, M. (December 2020). Risk Factors for COVID-19. Le Infezioni in Medicina. https://www.infezmed.it/media/journal/Vol_28_4_2020_2.pdf

[12] Levine, James A. (1 November 2011) Poverty and Obesity in the U.S. American Diabetes Association. https://diabetesjournals.org/diabetes/article/60/11/2667/33452/Poverty-and-Obesity-in-the-U-S

[13] Shrider, E.A., Kollar, M., Chen, F., Semega, J. (14 September 2021). Income and Poverty in the United States: 2020. United States Census Bureau. https://www.census.gov/library/publications/2021/demo/p60-273.html

[14] Golestaneh, L., Neugarten, J., Fisher, M., Billett, H.H., Gil, M.R., Johns, T., Yunes, M., Mokrzycki, M.H., Coco, M., Norris, K.C., Perez, H.R., Scott, S., Kim, R.S., Bellin, E. (August 2020). The association of race and COVID-19 mortality. eClinicalMedicine. https://www.sciencedirect.com/science/article/pii/S2589537020301991

[15] Mukherjee, S., Pahan, P. (6 January 2021). Is COVID-19 Gender-sensitive?. Journal of Neuroimmune Pharmacology. https://link.springer.com/article/10.1007/s11481-020-09974-z

[16] Ruggieri, A., Anticoli, S., D'Ambrosion, A., Giordani, L., Viora, M. (16 March 2016). The influence of sex and gender on immunity, infection and vaccination. https://annali.iss.it/index.php/anna/article/view/423/251

[17] Mustafa, N.M., Selim, L.A. (July 2020) Characterization of COVID-19 Pandemic in Pediactic Age Group: A Systematic Review and Meta-Analysis. Journal of Clinical Virology. https://www.sciencedirect.com/science/article/pii/S1386653220301372

[18] Demographic Trends of COVID-19 cases and deaths in the US reported to CDC. (16 March 2022). Center for Disease Control. https://covid.cdc.gov/covid-data-tracker/#demographics

[19] Liang, L., Tseng, C., Ho, H.J., Wu, C. (24 July 2020). Covid-19 mortality is negatively associated with test number and government effectiveness. Scientific Reports. https://www.nature.com/articles/s41598-020-68862-x

[20] Hyman, M., Mark, C., Imteaj, A., Ghiaie, H., Rezapour, S., Sadri, A.M., Amini, M.H. (13 August 2021) Data analytics to evaluate the impact of infectious disease on economy: Case study of COVID-19 pandemic. Patterns. https://www.sciencedirect.com/science/article/pii/S2666389921001586

[21] Warner, Rebecca M. (2008). Applied Statistics: From Bivariate Through Multivariate Techniques. Sage Publications Inc. (pp. 931 - 974).

[22] Latif, Khawaja Fawad. Binary Logistic Regression using SPSS. https://researchwithfawad.com/index.php/lp-courses/data-analysis-using-spss/binary-logistic-regression-analysis-in-spss/

[23] About the Topic of Race. (1 March 2022). United States Census Bureau. https://www.census.gov/topics/population/race/about.html

[24] Maragakis, L.L. (2 December 2020). Coronavirus and COVID-19: Younger Adults Are at Risk, Too. Johns Hopkins Medicine. https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/coronavirus-and-covid-19-younger-adults-are-at-risk-too

[25] Berger, Dale. (2016). Introduction to Binary Logistic Regression. Claremont Graduate University. http://wise.cgu.edu/wp-content/uploads/2016/07/Introduction-to-Logistic-Regression.pdf

[26] COVID-19 Vaccinations in the United States. (3 April 2022). Center for Disease Control. https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-total-admin-rate-total

[27] US COVID-19 Cases and Deaths by State. (3 April 2022). USAFacts. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/

[28] Ugoni, A. Walker, B.F. (November 1995). The Chi Square Test: An Introduction. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2050386/