

May 2021

The Honors College at the University of Missouri-Kansas City

Statistical Analysis of Kansas City COVID-19 Data with Respect to Race, Ethnicity, Age &
Gender

Sindhu Balakumar

Date of Submission: May 20th, 2021

Written under the direction of Professor Majid Bani-Yaghoub, Ph.D.

Department of Mathematics and Statistics

A thesis submitted in partial fulfillment of the requirements to graduate as an Honors
Scholar from the University of Missouri-Kansas City

Abstract

COVID-19 is a respiratory tract infection that has created a deadly pandemic dominating the world. This statistical analysis identifies the prevalence of confirmed COVID-19 cases in Kansas City by various demographics such as age, race, gender, and ethnicity. The main objective of this study was to determine if COVID-19 disproportionately impacted certain minority communities. We analyzed raw data of confirmed COVID-19 cases from March 2020 to October 2020, and we studied the frequency, proportion, and prevalence of COVID-19 cases in Kansas City, Missouri. Individuals between the ages of 20-29 years had the highest frequency of COVID-19 cases in the White population. Our ethnicity analysis showed that the average age of COVID positive Hispanic/Latino individuals is significantly lower than both African American and White individuals. Hispanic/Latino individuals also had the highest prevalence of COVID-19 positive cases than both White and African American individuals in Kansas City. Moreover, by using ANOVA testing, we found that females had a significantly higher prevalence in COVID positive cases and mortality than men. In conclusion, we observed that COVID-19 has impacted individuals who are Hispanic/Latino, female, or between the ages of 20-29 more than other groups.

Table of Contents

Chapter 1: Introduction.....	4
1.1 Overview.....	4
1.2 Transmission.....	5
1.3 Prevention.....	7
1.4 Symptoms.....	8
1.5 Risks.....	9
Chapter 2: Descriptive Analysis.....	11
2.1 Time Series.....	11
2.2 Race/Ethnicity Analysis.....	13
2.3 Gender Analysis.....	16
2.4 Age Analysis.....	17
Chapter 3: Inferential Analysis.....	20
3.1 Hypothesis Test.....	20
3.2 ANOVA Tests.....	21
Chapter 4: Conclusion.....	26
4.1 Potential Challenges & Future Results.....	26
4.2 Concluding Remarks.....	29

Chapter 1

1.1 Overview

COVID-19 is a respiratory tract infection that created a devastating global pandemic. This viral infection is caused by the novel SARS-CoV-2 which is a severe acute respiratory syndrome coronavirus [15]. This public health crisis has resulted in approximately 145,631,634 total cases and 3,089,271 total deaths globally as of April 23rd, 2021 [22]. The World Health Organization even declared the pandemic a Public Health Emergency of International Concern [17].

COVID-19 is a coronavirus that belongs to the subfamily Orthocoronavirinae which also includes the Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus [4]. The Alphacoronavirus and Betacoronavirus can infect mammals, but the Gammacoronavirus and Deltacoronavirus mainly infect birds [5]. COVID-19 is also known as SARS-CoV-2. The Coronavirus genome is a Betacoronavirus with a positive-sense, enveloped, single-stranded RNA and even has the largest genome of RNA viruses [4]. RNA viruses also have poor proofreading abilities and high-fidelity replication due to a specific proofreading exoribonuclease encoded that is linked to the increased fitness of CoV [24]. Increased fitness can often lead to more mutations and potentially new strains. The coronavirus family can cause respiratory, enteric, hepatic, and neurologic diseases in humans [11]. The structure of the genome plays a critical role in the way that the virus is able to infect humans.

SARS-CoV-2 has an envelope spike protein capable of recognizing the ACE-2 receptor of humans [5]. The ACE-2 receptor is a membrane protein expressed in the lungs, heart, kidney, and intestines and is often associated with cardiovascular diseases [11]. Once the spike protein of the Coronavirus binds to the ACE-2 receptor, it is cleaved and the viral and cell membranes are

fused together [5]. Since ACE-2 is expressed on a multitude of human tissues, the virus is able to spread to other organs and areas of the human body other than lung epithelial cells. After the virus enters the cell, it releases its RNA in the cytoplasm and translation and replication occur [5].

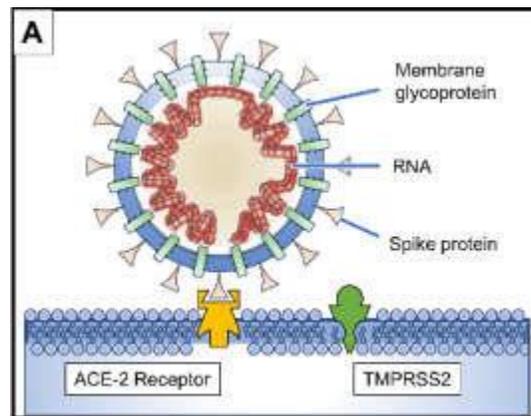


Figure 1.1: Spike Protein binds to ACE-2 Receptor [18]

1.2 Transmission

COVID-19 poses a great threat to the world because of its high fatality risk and high rate of transmission rate. This virus has a fatality risk of about 1%, which is greater than the 1957 influenza pandemic which had a fatality risk of about 0.6% [13]. In addition, individuals infected with COVID-19 are able to spread the disease to two or three other individuals on average, resulting in an exponential growth of infection [13]. This relates to the reproduction number, R , which details the average number of people each person with COVID-19 infects [23]. Epidemiologists are able to use this reproduction number to estimate the spread of infection within a population.

Researchers believe the coronavirus (COVID-19) first emerged in Wuhan, China in December 2019 before rapidly spreading throughout the world [1]. While there is no evidence to show how humans initially contracted the virus, the outbreak has been traced to a seafood market in Wuhan that sold many live animals such as bats, frogs, snakes, and birds [1].

Human transmission of this virus is supported by the fact that not all individuals who initially contracted the virus visited the seafood market. In fact, this disease is highly transmittable, especially when a susceptible individual comes in close contact with an infected individual. This virus is known to be transmitted by means of respiratory droplets or aerosols, via coughing or sneezing, that can enter the lungs via inhalation [1]. These aerosols can infect humans by depositing on surfaces or can persist in the air for minutes to hours. The *New England Journal of Medicine* studied the decay rates of coronavirus aerosols and discovered that the virus remained viable in aerosols for three hours [3]. Airborne transmission of COVID-19 is of particular concern because aerosols can decrease in size due to evaporation that better allows droplets to spread through air; moreover, such aerosols may have the ability to infect beyond six feet [2]. Maintaining a six feet distance from others is the golden rule of preventing transmission, but further research is needed to determine if airborne transmission may invalidate this edict. Aerosols are also generated by speaking and breathing, which is why masks are encouraged to prevent transmission. This virus even has a half-life of 1.1 to 1.2 hours as an aerosol and a half-life of 6.8 hours on surfaces such as plastic [8]. This trait illustrates the importance of disinfecting surfaces and maintaining proper personal hygiene. Since this virus has a greater reproductive number than the 2009 H1N1 influenza virus, it is imperative that preventative measures are taken [2]

The asymptomatic spread of this virus also contributes to the rapid transmission of the Coronavirus. COVID-19 may spread asymptotically by infected individuals who show no signs or symptoms which makes disease prevention and control more difficult. This virus has an incubation period that ranges from 2 - 14 days, and about 44% of cases spread before an

individual can even notice signs and symptoms [8]. Some patients do not show symptoms at all, making it difficult to identify and isolate individuals before they spread the virus to others.

1.3 Prevention

Maintaining social distancing, wearing face masks regularly, and canceling public events and gatherings such as sporting events have all contributed to flattening the curve and slowing the spread of this virus. Flattening the curve is an approach that emphasizes the strategies to slow the spread of this virus which would spread out the peak of the epidemic, preventing hospitals from reaching max capacity [21]. However, many countries battled COVID-19 by using different responses and strategies. China took an aggressive approach in containing this virus by mandating a national lockdown and promoting temperature monitoring, wearing masks, and washing hands [8]. This clear and focused approach was successful in slowing the spread. Taiwan was proactive in containing the spread by examining all individuals entering the country and quarantining anyone infected by the virus in December [8]. South Korea was able to track and isolate infected individuals using advanced technology and high testing rates [8]. New Zealand's response was to require self-quarantine measures, contact tracing, and a national lockdown to prevent transmission.

Italy experienced a rapid increase in confirmed cases and deaths initially which can be attributed to a lack of available testing and difficulties in containing the spread [8]. Germany saw a high number of confirmed cases but relatively low fatalities due to its proactive testing program that began in January 2020 [8]. The United Kingdom reacted late but eventually restricted travel only for essential reasons and imposed fines on anyone who did not follow these regulations [8]. The differences in responses to COVID-19 have resulted in different rates of transmission and fatalities reported.

1.4 Symptoms

According to the CDC, common symptoms of COVID-19 may include fever, chills, cough, shortness of breath, fatigue, muscle/body aches, headache, loss of taste or smell, sore throat, congestion, runny nose, nausea, vomiting, or diarrhea [6]. As the name SARS-CoV-2 suggests, severe acute respiratory syndrome specifically targets lung epithelial cells and can cause respiratory illness. About 80% of individuals infected with COVID-19 experience mild to moderate illness such as pneumonia; 14% of individuals experience severe disease that may include the saturation of blood oxygen levels; and 6% of individuals experience life-threatening complications such as respiratory failure, septic shock, or organ failure [8]. While SARS-CoV-2 expresses symptoms characteristic of other Betacoronaviruses such as fever, dry cough, and difficulty breathing, this virus also shows other features that target the lower airway such as sneezing and sore throat which are also symptoms of rhinorrhea [12]. Moreover, diarrhea is a symptom of SARS-CoV-2 but is uncommon in patients infected with MERS-CoV or other SARS-CoV [12].

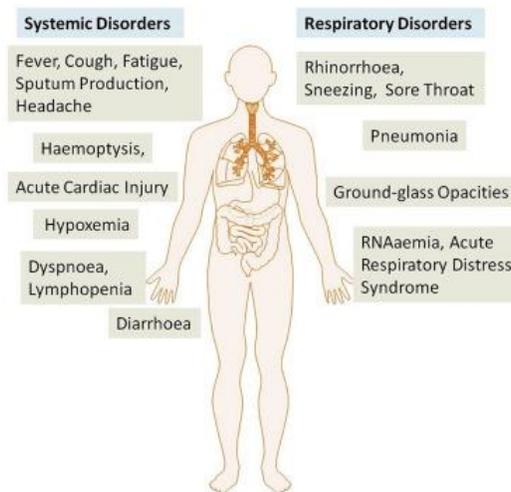


Figure 1.2: Systemic and Respiratory Disorders Relating to COVID-19 [12]

Figure 1.2 illustrates the wide range of symptoms that COVID-19 can cause. Systemic disorders may result in symptoms such as fever, cough, and diarrhea, while respiratory disorders may result in symptoms such as pneumonia, rhinorrhea, or sneezing [12]. This virus has the unique ability to impact almost every organ of the human body. It is crucial to understand how COVID-19 is able to cause such a diverse range of symptoms. Scientists have discovered that the immune system's response to viruses may lead to blood clotting, resulting in some complications such as pulmonary embolism and stroke [7]. Normally viral infections do not cause clots throughout blood vessels, but COVID-19 is unique in its ability to detrimentally impact the vascular system. Complement proteins were even found within blood vessels which activate an inflammatory response in the endothelium, or inner lining of blood vessels [7]. The endothelium also lines other organs and cavities throughout the human body which explains why this virus is able to cause various symptoms throughout the body. As mentioned before, the ACE-2 receptor is also present on many different organs, also contributing to the diversity of COVID-19 symptoms.

1.5 Risks

Individuals that are older or possess pre-existing conditions are at a greater risk of infection. Older adults with conditions such as cardiovascular diseases, cancer, hypertension, and diabetes are especially at risk [5]. It is important to note that age can impact the severity of COVID-19. Some studies have shown that ACE-2 expression in the lungs increases with age which means SARS-CoV-2 would be able to bind to more receptors and infect more host cells; however, further research must be completed to confirm this finding. In particular, individuals with the HLA-B*46:01 gene products have a lower binding capacity to the COVID-19 peptides which reduces the ability of viral antigen presentation to immune cells, increasing vulnerability

to the virus [2]. Interestingly, this specific gene variant can provide protection against leprosy due to its ability to present foreign molecules to the immune system and destroy the infected cells [25]. While this gene grew rapidly in Southeast Asia due to selective pressures, it also shows evidence of increasing susceptibility to other infectious diseases such as malaria and HIV, demonstrating a disadvantage to carriers [25]. While research is still ongoing, it is clear that this virus is able to impact each patient differently. Some experience mild symptoms, while others are hospitalized and experience severe, fatal effects. Age and the presence of pre-existing conditions can help identify which individuals are most susceptible to contracting COVID-19, but medical advances and discoveries can help in identifying the underlying factors.

Chapter 2

Globally, there are 145,631,634 total cases and 3,089,271 total deaths as of April 23rd, 2021 [22]. The following map shows the clusters of cases by global region (as of April 23rd, 2021).



Figure 2.1: COVID-19 Global Hot Spots [22]

While clusters exist globally, there is a high concentration of cases located in the North American Region, specifically the United States. In order to understand the impact of COVID-19 on the United States, it is imperative to observe the progression of positive COVID-19 cases and COVID-19 related deaths over time.

2.1 Time Series

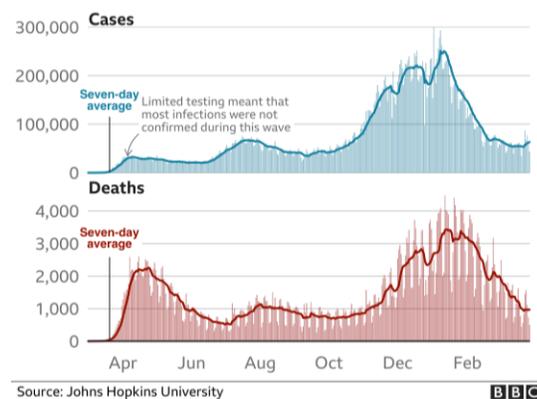


Figure 2.2: Daily Reported COVID-19 Cases and Deaths in the U.S. [22]

These graphs show the changes in daily reported COVID-19 cases and COVID-19 related deaths between the time period of April 2020 to March 2021. The frequency of daily reported cases peaked between December 2020 – January 2021. The frequency of daily reported deaths first peaked in May 2020 and then peaked again between December 2020 – January 2021. Since the frequency of new cases and deaths is different depending on the location/region, this study will primarily focus on the impact of COVID-19 with respect to Kansas City, Missouri.

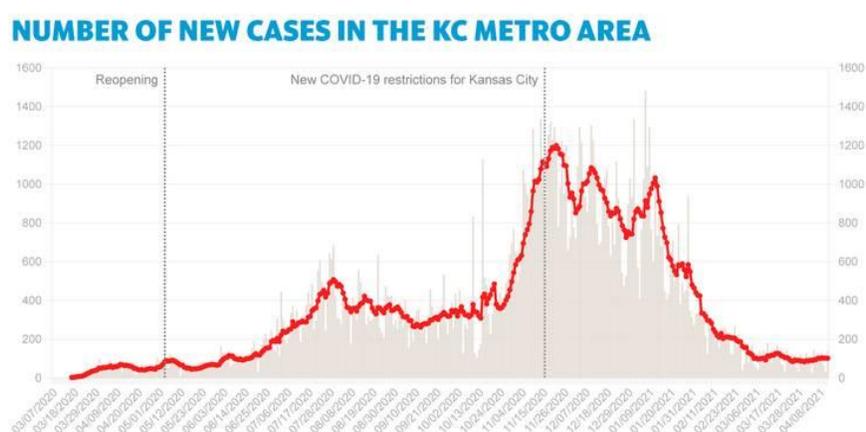


Figure 2.3: Frequency of New COVID-19 Cases in the Kansas City Metro Area [26]

This time series illustrates the frequency of new cases that occurred in the Kansas City metro area between the time period of March 2020 – April 2021. Upon the economic reopening in May 2020, we can observe an increase in COVID-19 cases until a local max in mid-July 2020. After a slight decline, cases peaked again between November – December 2020. Once new COVID-19 restrictions were enforced, positive COVID-19 declined and have remained low since. Kansas City, Missouri has experienced nearly 38,000 total COVID-19 cases and 557 deaths as of April 14th, 2021.

2.2 Race/Ethnicity Analysis:

Racial disparities in the spread of infectious diseases are important factors when studying the spread of COVID-19. In order to identify any potential racial disparities that may exist in Kansas City, Missouri, a racial analysis was completed using data listing the confirmed cases in Kansas City as of November 12, 2020. The data includes information such as the date received, epidemiology week and year, diagnosis (COVID-19), case status (confirmed), age, gender, race, ethnicity, age group, zip code, and vital status. All data was retrieved from the Kansas City Health Department and the data ranges from March 2020 – October 2020.

This statistical analysis identifies the prevalence of confirmed COVID-19 cases in Kansas City by various demographics such as race, age, and gender distributions. This research aims to identify if COVID-19 disproportionately impacts certain racial minorities or communities over others by looking at the frequency and prevalence of cases with respect to certain demographics. Analyzing COVID-19 confirmed cases by using raw data obtained from the Kansas City Public Health Department will provide the data needed to draw conclusions. Pivot tables were used to organize data and create graphical representations of cases and deaths. All graphs and tables were created using Microsoft Excel. Figure 2.4 displays the confirmed COVID-19 cases in Kansas City, Missouri with respect to race.

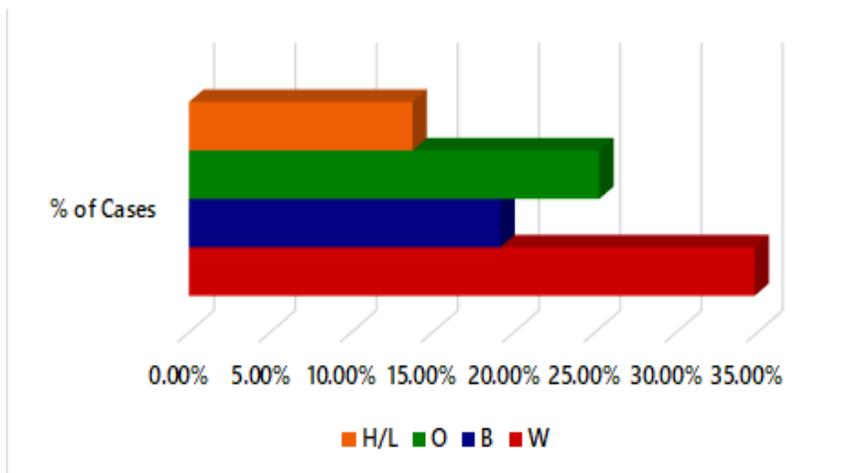


Figure 2.4: Proportion of Confirmed COVID-19 Cases in Kansas City, Missouri by Race/Ethnicity

White individuals have the greatest proportion of confirmed COVID-19 cases, while individuals who are Hispanic or Latino have the lowest proportion of cases. The “Other” category is composed of people who identify as Other, American Indian, Hawaiian/Pacific Islander, or Asian. Since White individuals make up the majority of the Kansas City population, this justifies the high percentage of COVID-19 among this racial group.

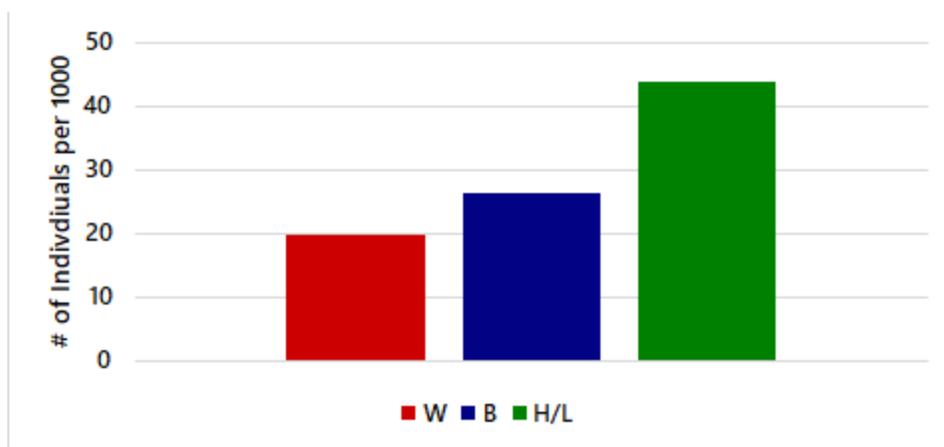


Figure 2.5: Prevalence of COVID-19 Cases by Race/Ethnicity (per 1000 individuals)

Figure 2.5 represents the prevalence of COVID-19 cases with respect to specific race/ethnicity groups: White, Black, and Hispanic/Latino. Hispanic/Latino individuals have the highest

prevalence of COVID-19 cases in Kansas City, while White individuals have the lowest prevalence. Prevalence was calculated by dividing the number of cases within a specific racial group by the total number of individuals belonging to that same racial group in Kansas City, MO and multiplying that decimal by 1000 to obtain the prevalence per 1000 individuals.

An ethnicity analysis is based on an individual identifying as a particular ethnicity. A person who identifies as Hispanic, Latino, or Spanish may be of any race. According to the U.S. Census Bureau, this identification may relate to “heritage, nationality, lineage, or country of birth” [28]. In addition, 2.5 million Hispanic-Americans changed their race identification from “some other race” to “White” [27]. The following graph shows the distribution of individuals who identify as non-Hispanic or Hispanic by race.

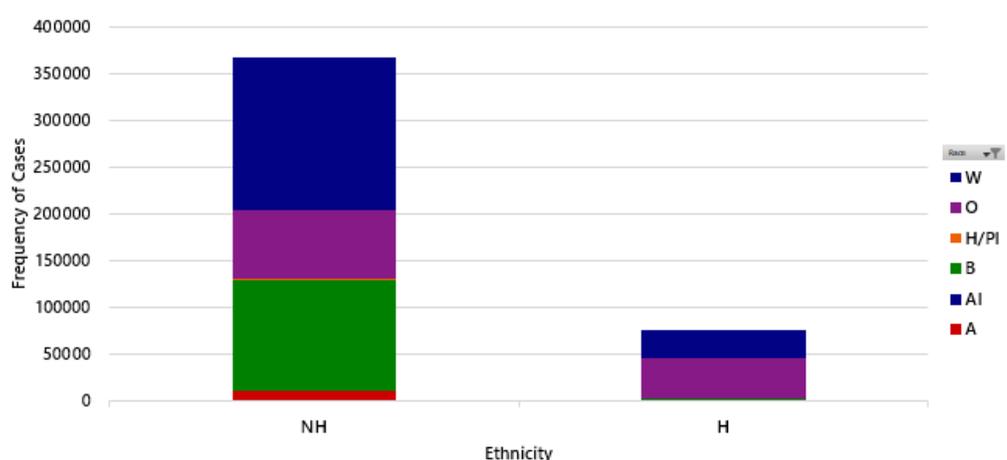


Figure 2.6: Distribution of Race and Ethnicity of COVID-19 Cases

Non-Hispanic individuals have a greater number of COVID-19 cases than Hispanic individuals. Also, people who identify as Hispanic are either White or belong to the “Other” category. It is beneficial to look at the diversity of race within the Hispanic/Latino demographic. This graph also illustrates the intersectionality of race and ethnicity with respect to COVID-19 cases.

2.3 Gender Analysis

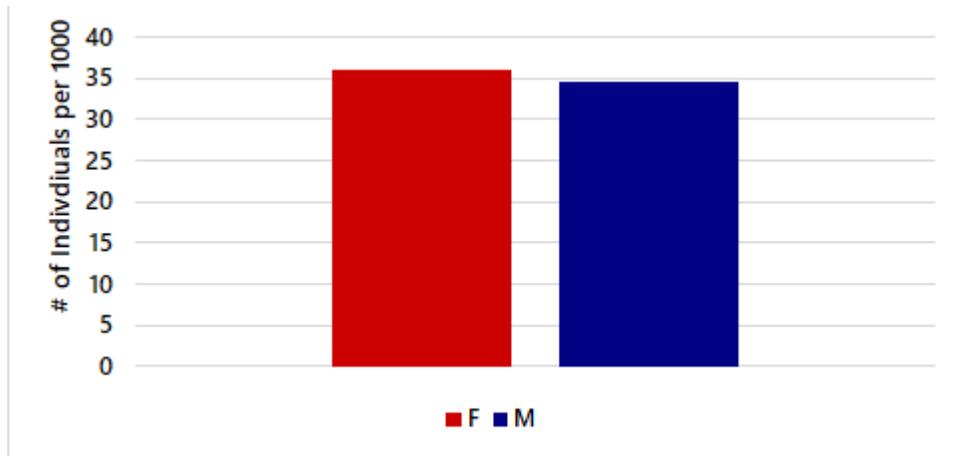


Figure 2.7: Prevalence of COVID-19 Cases by Gender (per 1000 individuals)

Figure 2.7 represents the prevalence of COVID-19 cases with respect to gender. Females have a slightly higher prevalence than males. Prevalence was calculated by dividing the number of cases within a specific gender by the total number of individuals identifying as that gender in Kansas City, Missouri and multiplying that decimal by 1000 to obtain the prevalence per 1000 individuals.

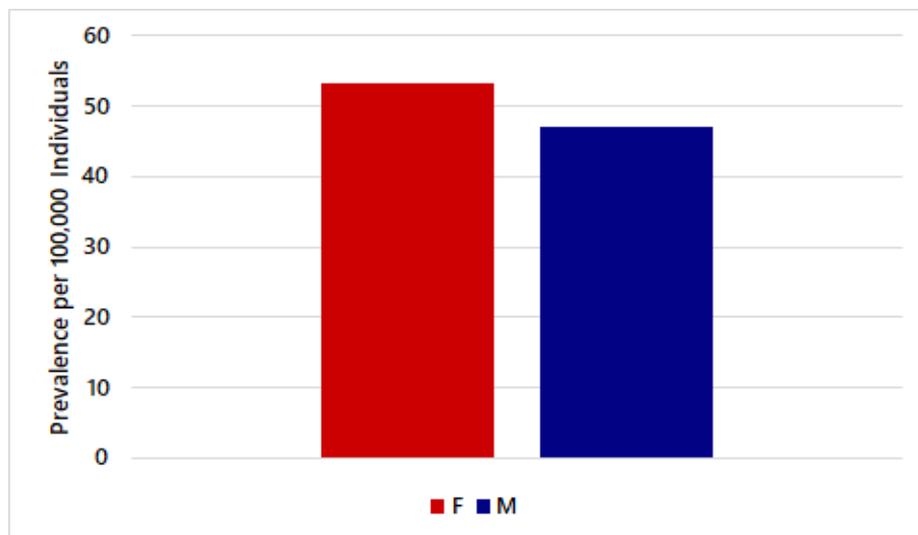


Figure 2.8: Prevalence of Mortality by Gender (per 1,000,000 Individuals)

Figure 2.8 represents the prevalence of mortality with respect to gender. Females have a higher prevalence in both proportion of cases and mortality within this data set. Prevalence was calculated by dividing the number of deaths within a specific gender by the total number of individuals belonging to that same gender in Kansas City, Missouri and multiplying that decimal by 1,000,000 to obtain the prevalence per 1,000,000 individuals.

2.4 Age Analysis

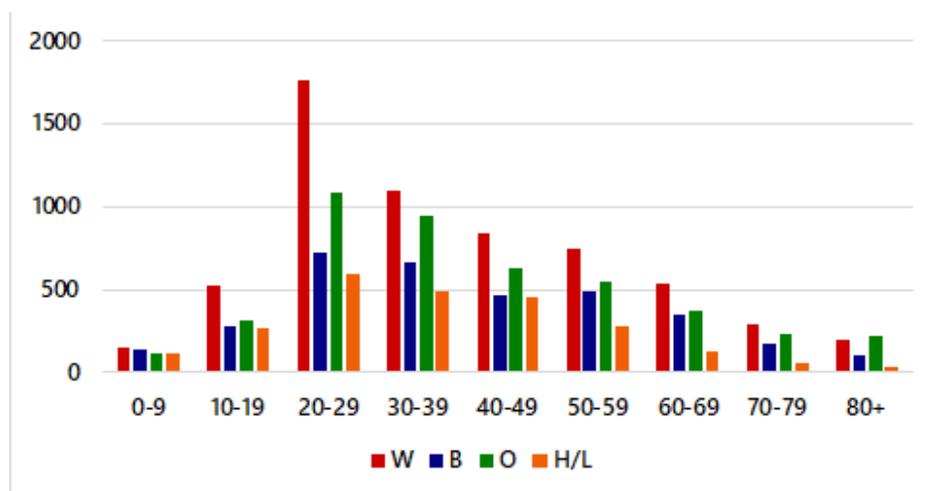


Figure 2.9: Crude COVID-19 Cases by Age & Race/Ethnicity

Analyzing this graph shows us that individuals between the ages of 20-29 have the highest number of confirmed cases. This graph only shows data for individuals who are White (W), Black (B), Other (O), and Hispanic/Latino (H/L). Moreover, White individuals who are between 20-29 years old have the highest number of confirmed cases. In order to accurately analyze the prevalence of COVID-19 cases we would need to also take into account the specific demographics of Kansas City, Missouri.

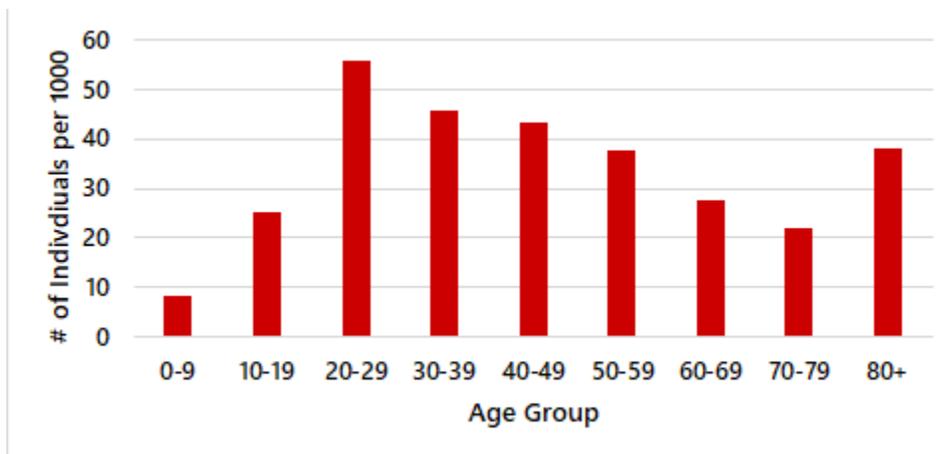


Figure 2.10: Prevalence of COVID-19 Cases by Age Group (per 1000 individuals)

Figure 2.10 shows the prevalence of COVID-19 cases with respect to age group. Individuals between the ages of 20-29 years have the highest prevalence. Individuals in the 80+ age group have the second highest prevalence in cases. Prevalence was calculated by dividing the number of cases within a specific age group by the total number of individuals belonging to that age group in Kansas City, Missouri and multiplying that decimal by 1000 to obtain the prevalence per 1000 individuals.

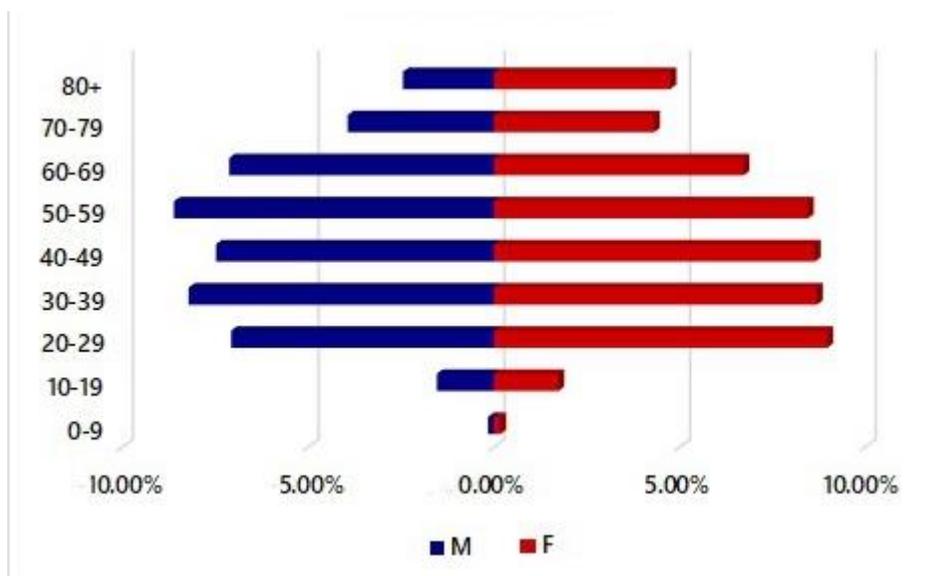


Figure 2.11: Proportion of COVID-19 Cases by Gender & Age Group

Figure 2.11 is a population pyramid illustrating the proportion of COVID-19 cases with respect to gender and age group. This figure shows that most of the confirmed COVID-19 cases impact individuals between the ages of 20-60 years. According to the U.S. Census Bureau, the median age of Kansas Citizens is 37.7 years old and individuals between the ages of 18-64 make up 61% of the population (U.S Department of Commerce). We can observe how COVID-19 has impacted age groups by comparing the percentage of individuals diagnosed with COVID-19 vs. the total population of Kansas Citizens.

The following graph shows the average age of COVID-19 patients with respect to Race/Ethnicity.

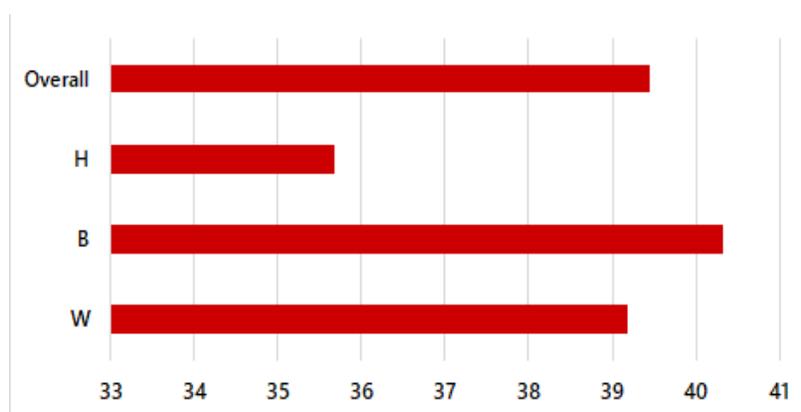


Figure 2.12: Average Age of COVID-19 Patients by Race/Ethnicity

This graph shows that the average age for Hispanic individuals diagnosed with COVID-19 is 35.7 years which is lower than the overall average age of 39.4 years. The average age of White individuals testing positive for COVID-19 is 39.2 years and the average age of Black individuals is 40.2 years which is similar to the overall average age of COVID-19 patients in our data.

Chapter 3

3.1 Hypothesis Test

Hypothesis testing can be used to complete a statistical comparison of relative proportions of confirmed positive COVID-19 cases by gender in Kansas City. Hypothesis testing can determine if data results are valid by observing the chance that results have happened by chance. The data consists of the proportions of males and females in Kansas City, Missouri, who tested positive for COVID-19.

	Females	Males
COVID-19 Cases	9251	8233
Total Population	250430	235974

\hat{P}_1 is the proportion of females with confirmed COVID-19 cases while \hat{P}_2 is the proportion of males with confirmed COVID-19 cases.

$$\hat{P}_1 = \frac{9251}{250430} = 0.0369$$

$$\hat{P}_2 = \frac{8233}{235974} = 0.0349$$

Next, the estimated standard error was calculated:

$$SE(\hat{P}_1 - \hat{P}_2) = \sqrt{\frac{0.0369(1-0.0369)}{250430} + \frac{0.0349(1-0.0349)}{235974}} = 5.336 \times 10^{-4}$$

Using this estimated standard error, the 95% confidence interval for $\hat{P}_1 - \hat{P}_2$ was calculated with the equation: $\hat{P}_1 - \hat{P}_2 \pm z^* SE(\hat{P}_1 - \hat{P}_2)$.

$$= (0.0369 - 0.0349) \pm (1.96 * 5.336 * 10^{-4})$$

$$= 0.0021 \pm 0.0010$$

$$= (0.001, 0.003)$$

Thus, we are 95% confident that the true difference in population proportions lies between 0.001 and 0.003. In order to determine statistical significance, a null hypothesis and an alternative hypothesis were identified:

$$H_0: \hat{P}_1 = \hat{P}_2$$

$$H_a: \hat{P}_1 \neq \hat{P}_2$$

Next, the pooled sample proportion was calculated:

$$\hat{P} = \frac{9251+8233}{250430+235974} = 0.0359$$

The estimated $SE_0(\hat{P}_1 - \hat{P}_2)$ was found using the pooled sample proportion:

$$SE_0(\hat{P}_1 - \hat{P}_2) = \sqrt{0.0359(1 - 0.0359)\left(\frac{1}{250430} + \frac{1}{235974}\right)} = 5.341 \cdot 10^{-4}$$

The Z score can be found using this $SE_0(\hat{P}_1 - \hat{P}_2)$ and the initial $\hat{P}_1 - \hat{P}_2$.

$$Z = \frac{0.369 - 0.0349}{5.341 \cdot 10^{-4}} = 3.840$$

Based on this Z score, the p-value is 0.000123. Since $p < 0.05$, there is strong evidence of a difference in the population proportions. Thus, the difference between female and male COVID-19 cases is statistically significant and we can reject the null hypothesis ($H_0: \hat{P}_1 = \hat{P}_2$).

3.2 ANOVA Tests

In this statistical analysis, a one-way analysis of variance (ANOVA) was used to determine if there exist any statistically significant differences between the means of three or more independent groups. The mean ages of the following races/ethnicities were compared: White, Black, and Hispanic. The following plot shows the distribution of age by race/ethnicity.

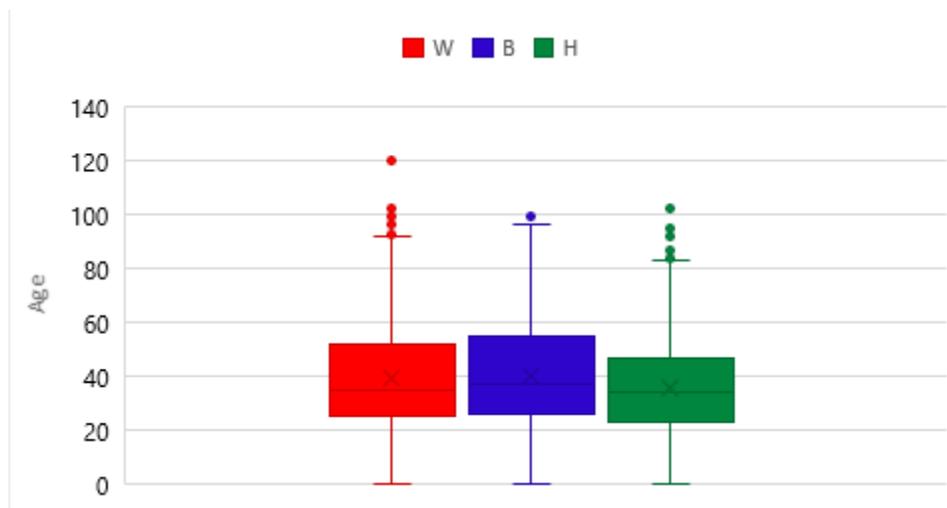


Figure 3.1: Age Distribution by Race/Ethnicity

This plot shows that the average age of White individuals testing positive for COVID-19 is 39.1 years, the average age of black individuals is 40.3 years, and the average age of Hispanic individuals is 35.7 years. This means that on average, Hispanic individuals are impacted by COVID-19 at a younger age than White and black individuals.

It is also important to note the null and alternative hypotheses:

H_0 : the mean values are the same with respect to each race and ethnicity

H_a : there is at least one mean value which is different than the others

The following chart shows the results of a one-way ANOVA test comparing these three groups: White, Black, and Hispanic.

ANOVA: Single Factor with respect to Race/Ethnicity and Age

Groups	Count	Average	Variance
White	6142	39.18772	365.1475
Black	3384	40.31797	384.2199
Hispanic	2429	35.67559	293.8412

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	32432.19	2	16216.1	45.54312	0	2.996483
Within Groups	4255633	11952	356.0603			
Total	4288065	11954				

This ANOVA test determined the P-value which is 0. Since this value is less than 0.05, this indicates strong evidence against the null hypothesis which states the mean values are the same with respect to each race and ethnicity since there is less than a 5% probability that the null is correct. Therefore, the null hypothesis can be rejected and there is evidence to accept the alternative hypothesis which states that there is at least one mean value which is different from the others. Next, a post-hoc test was used to determine if there exists a statistically significant difference between each group. A two-sample T-test assuming equal variances compared each relationship among the three groups. The two-tail p-value of each comparison was calculated and compared to the Bonferroni correction value of 0.0167.

Post-Hoc Test:

	W vs. B	B vs. H	H vs. W
P(T<=t) two-tail	0.00637	1.07657E-20	3.62E-15
Bonferroni correction	0.016667	0.016667	0.016667
P < 0.01267	True	True	True

From this table, each p-value was less than the Bonferroni correction value of 0.0167. This means that the average age of Hispanic individuals is significantly lower than both Black and White individuals.

A one-way ANOVA test was used to observe the differences between race by time.

ANOVA: Single Factor with Respect to Time and Race/Ethnicity:

Groups	Count	Sum	Average	Variance
W	35	6146	175.6	23041.07
B	35	3384	96.68571	3601.692
H	35	2429	69.4	1447.894

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	212921.5	2	106460.8	11.3697	3.49E-05	3.085465
Within Groups	955082.3	102	9363.552			
Total	1168004	104				

From this test, the observed P-value is 3.49E-05. Since this value is less than 0.05, this indicates strong evidence against the null hypothesis which states the mean values are the same with respect to EPI week and Race/Ethnicity.

Post-Hoc Test:

	W vs. B	B vs. H	H vs. W
P(T<=t) two-tail	0.005618	0.026276258	0.000151
Bonferroni	0.016667	0.016667	0.016667

correction			
P < 0.01667	True	False	True

This shows that the frequency of cases among White individuals is statistically significant. The difference between individuals who are Black and Hispanic is not statistically significant.

Chapter 4

4.1 Potential Challenges & Future Works

The novelty of COVID-19 has perplexed researchers and scientists due to its relatively recent appearance in history. While researchers have made progress in understanding this virus, one potential challenge is the uncertainty and lack of knowledge regarding COVID-19. While technological advances and scientific breakthroughs are occurring, and medical research and knowledge are increasing, there is still much we have yet to discover about this virus. Although the CDC and WHO are continuing to gather more information about this pandemic, it is difficult to reach definite conclusions due to mutations and variants of the virus. Our society becomes better informed as more knowledge becomes available. Nonetheless, there is still a lack of knowledge regarding immunity, transmission, vaccination, and potential treatments. Despite this new and unique pandemic, history may provide us with the best source of data from which we can draw meaningful conclusions. This can be done by researching the epidemiology of historical infections in order to gain a better understanding of COVID-19.

Studying medieval medicine and applying this to modern day viruses could be a source of knowledge. Researchers at the University of Warwick have studied the uses of medieval remedies and practices for antibacterial properties. Bacteria can live as individual cells or as a multicellular biofilm which can provide bacteria with protection and defense against antibiotics [9]. This study found that a “Bald’s eyesalve remedy” using ingredients such as onion, garlic, and wine displayed antibacterial activity against bacteria that infect diabetic foot ulcers [9]. This is an innovative example of how medical advancements can stem from the application of natural substances from the past.

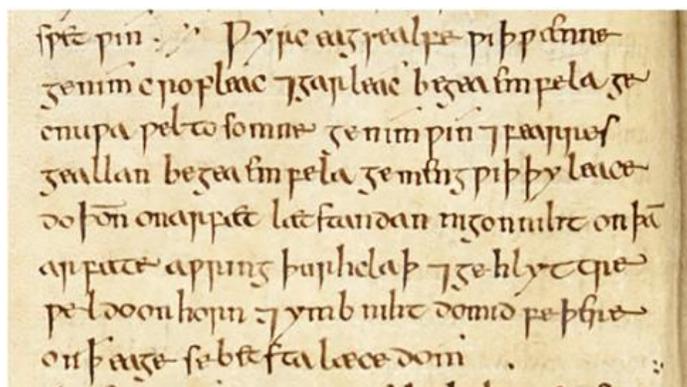


Figure 4.1: Bald's Eyesalve Remedy Recipe [16]

The Bald's eyesalve remedy was used during the Anglo-Saxon period to treat a “wen” or lump in the eye which is caused by a bacterial infection [16]. While this remedy utilizes simple, household ingredients such as onion, garlic, and wine, the salve exhibits antimicrobial properties. Garlic and onions contain ajoene and allicin which are able to prevent the multicellular biofilm from forming; onions contain the antimicrobial peptide Ace-AMP1; bile is able to limit bacterial growth in the small intestine; wine contributes antimicrobial molecules; and copper salts prevent bacterial growth and contribute to immune defense [16]. It is important to note that the “wen” that this salve targets is assumed to be a sty which is typically caused by the *Staphylococcus aureus*, a gram-positive bacterium [16]. The methicillin-resistant *Staphylococcus aureus* (MRSA) is a mutated strain that raises questions regarding antibiotic resistance. Since MRSA is resistant to current antibiotics, researchers saw potential in utilizing medieval remedies and natural substances instead.

One study at the University of Nottingham found that the Bald's eyesalve is successfully able to kill planktonic and biofilm cultures of the *Staphylococcus aureus* when the recipe is followed carefully [16]. Despite the remedy being 1,000 years old, the salve killed MRSA in chronically infected animal tissue. This particular study has excellent implications for the world of antibiotic resistance because plants and other natural substances may be used to treat

conditions that current medical advances are unable to combat. The Bald's eyesalve yielded positive results when the various ingredients worked together in unison to kill the gram-positive bacteria. This suggests that perhaps a combination of compounds working together could be the solution to battling the spread of COVID-19. Moreover, this study illustrates the great potential and promise in utilizing natural substances to treat ongoing, global challenges when existing medicines and technology are unsuccessful.

Obtaining reliable data and research could be difficult since this requires the need to transcribe historic texts. While there are countless books detailing ancient treatments and remedies used, some of these texts are written in English, Middle English, Latin, Old English, or other vernaculars. Finding texts that have been transcribed or edited to a modern edition is crucial. These texts also include archaic language that is unfamiliar or require the reader to transcribe scripts into Middle English and then finally Modern English. This may require a computational script reader, allowing for greater efficiency. Processing many texts and information can be time consuming, but Natural Language Processing tools can help analyze texts. Natural Language Processing (NLP) is a type of artificial intelligence that allows for machines to process the human language and perform data analysis [20]. Monkey Learn is an example of NLP that uses pre-trained models to perform text analysis, allowing the user to build customized models that identify insights from the data [20]. Thus, we can use technology to read and understand the human language in an efficient, time-conserving manner.

Given the nature of our research, it is imperative to obtain accurate recipes because imprecise measurements of ingredients could alter our data. For example, the Bald's Eyesalve mentioned earlier was successful due to the combined efforts of specific ingredients working in

unison. It may also be difficult to find recipes that are successful which requires trial and error. However, medieval medicine is a field that holds great potential.

This virus is concerning because it often causes many common symptoms that are also associated with other conditions such as the cold or flu. Examining other conditions that cause symptoms associated with COVID-19 may help us identify possible treatment options by studying drug repurposing. Drug repurposing is an excellent strategy because drugs that were successful in treating a particular condition may also be used to treat a new condition. This is of particular interest with COVID-19 given the urgency in finding a method to combat this virus. Utilizing prior drugs may be a viable option because this can identify possible treatments while saving the time needed to develop a new drug, undergo human clinical trials, and get it approved. This approach would involve applying new screening techniques to identify antivirals that have already been tested for safety, reducing the time it would take to discover a treatment [13]. It is also cost effective since the average cost of developing new drugs is over one billion dollars [14]. Researchers are currently working on repurposing drugs used to treat cancer, neurodegenerative diseases, viral infections, asthma, allergies, kidney disease, and cardiovascular disease [10]. If successful, this will be an excellent method of identifying potential treatments for COVID-19.

4.2 Concluding Remarks

COVID-19 has detrimentally impacted the United States. The frequency of new COVID-19 cases and deaths in the United States peaked in the fall-winter months of 2020-2021. COVID-19 cases peaked in Kansas City between November – December 2020. By analyzing raw data, we observed that White individuals have a higher frequency of cases than other racial groups. While more White individuals have been diagnosed with COVID-19, it is important to note that this is also the largest demographic in Kansas City. Prevalence testing was essential in

determining the overall burden of disease within specific groups of Kansas Citians. The data shows that Hispanic individuals had the highest prevalence in cases among all racial/ethnic groups. Hispanic individuals also had the lowest average age among all COVID-19 positive cases. Females had a higher prevalence in both cases and mortality than men. Moreover, individuals between the ages of 20-29 had the highest prevalence in cases, followed by the 80+ age group. Moreover, hypothesis testing showed that there is a statistically significant difference between the proportions of males and females with COVID-19. When comparing individuals who are White, Black, and Hispanic/Latino, ANOVA testing showed that there was one mean age value was different from the others. A post-hoc test showed that the average age of Hispanic individuals was significantly lower than both Black and White individuals. Also, ANOVA testing showed that the frequency of cases among White individuals was also statistically significant. The significance of the main findings of this research is that it provides results specifically for Kansas City, Missouri. When searching the World Health Organization's database for global literature on COVID-19, there are only fifteen papers published about Kansas City, Missouri. Thus, this research serves to contribute to the advancement of COVID-19 literature and knowledge.

References

1. Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of advanced research*, 24, 91–98. <https://doi.org/10.1016/j.jare.2020.03.005>
2. Godri Pollitt, K. J., Peccia, J., Ko, A. I., Kaminski, N., Dela Cruz, C. S., Nebert, D. W., Reichardt, J., Thompson, D. -C., & Vasiliou, V. (2020). COVID-19 vulnerability: the potential impact of genetic susceptibility and airborne transmission. *Human genomics*, 14(1), 17. <https://doi.org/10.1186/s40246-020-00267-3>
3. van Doremalen N, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *New England Journal of Medicine*. 2020. <https://doi.org/10.1056/NEJMc2004973>.
4. Li, H., Liu, S. M., Yu, X. H., Tang, S. L., & Tang, C. K. (2020). Coronavirus disease 2019 (COVID-19): current status and future perspectives. *International journal of antimicrobial agents*, 55(5), 105951. <https://doi.org/10.1016/j.ijantimicag.2020.105951>
5. Vellas, C., Delobel, P., de Souto Barreto, P., & Izopet, J. (2020). COVID-19, Virology and Geroscience: A Perspective. *The journal of nutrition, health & aging*, 24(7), 685–691. <https://doi.org/10.1007/s12603-020-1416-2>
6. Symptoms of Coronavirus. (n.d.). Retrieved September 20, 2020, from <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>
7. Kwon, D. (2020, May 18). From Headaches to 'COVID Toes,' Coronavirus Symptoms Are a Bizarre Mix. Retrieved September 22, 2020, from <https://www.scientificamerican.com/article/from-headaches-to-covid-toes-coronavirus-symptoms-are-a-bizarre-mix1/>

8. Triggler, C. R., Bansal, D., Farag, E., Ding, H., & Sultan, A. A. (2020). COVID-19: Learning from Lessons To Guide Treatment and Prevention Interventions. *mSphere*, 5(3), e00317-20. <https://doi.org/10.1128/mSphere.00317-20>
9. University of Warwick. (2020, July 28). Medieval medicine remedy could provide new treatment for modern day infections. *ScienceDaily*. Retrieved October 10, 2020 from www.sciencedaily.com/releases/2020/07/200728113545.htm
10. Benavides-Cordoba V. (2020). Drug Repositioning for COVID-19. *Colombia medica (Cali, Colombia)*, 51(2), e4279. <https://doi.org/10.25100/cm.v51i2.4279>
11. Jin, Y., Yang, H., Ji, W., Wu, W., Chen, S., Zhang, W., & Duan, G. (2020). Virology, Epidemiology, Pathogenesis, and Control of COVID-19. *Viruses*, 12(4), 372. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/v12040372>.
12. Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of autoimmunity*, 102433.
13. Gates, B. (2020). Responding to Covid-19—a once-in-a-century pandemic?. *New England Journal of Medicine*, 382(18), 1677-1679.
14. Senanayake, S. L. (2020). Drug repurposing strategies for COVID-19.
15. Pawar, A. Y. (2020). Combating devastating COVID-19 by drug repurposing. *International Journal of Antimicrobial Agents*.
16. Harrison, F., Roberts, A. E., Gabriliska, R., Rumbaugh, K. P., Lee, C., & Diggle, S. P. (2015). A 1,000-year-old antimicrobial remedy with antistaphylococcal activity. *MBio*, 6(4).
17. Timeline: WHO's COVID-19 response. (2020). Retrieved September 29, 2020, from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline/>

18. Rabi, Firas & Al-Zoubi, Mazhar & Kasasbeh, Ghena & Salameh, Dunia & Al-Nasser, Amjad. (2020). SARS-CoV-2 and Coronavirus Disease 2019: What We Know So Far. *Pathogens*. 9. 10.3390/pathogens9030231.
19. United States. (n.d.). Retrieved October 20, 2020, from <https://www.worldometers.info/coronavirus/country/us/>
20. 10 Natural Language Processing Tools – SaaS & Open-Source [Internet]. MonkeyLearn Blog. 2020 [cited 2021Apr26]. Available from: <https://monkeylearn.com/blog/natural-language-processing-tools/>
21. Kenyon C. Flattening-the-curve associated with reduced COVID-19 case fatality rates- an ecological analysis of 65 countries [Internet]. *The Journal of infection*. The British Infection Association. Published by Elsevier Ltd.; 2020 [cited 2021Apr26]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7162747/>
22. Home [Internet]. Johns Hopkins Coronavirus Resource Center. [cited 2021Apr26]. Available from: <https://coronavirus.jhu.edu/data/new-cases-50-states/missouri>
23. Adam D. A guide to R - the pandemic's misunderstood metric [Internet]. *Nature News*. Nature Publishing Group; 2020 [cited 2021Apr26]. Available from: <https://www.nature.com/articles/d41586-020-02009-w>
24. Graepel KW;Lu X;Case JB;Sexton NR;Smith EC;Denison MR; Proofreading-Deficient Coronaviruses Adapt for Increased Fitness over Long-Term Passage without Reversion of Exoribonuclease-Inactivating Mutations [Internet]. *mBio*. U.S. National Library of Medicine; [cited 2021Apr26]. Available from: <https://pubmed.ncbi.nlm.nih.gov/29114026/>

25. An immunity gene evolved in Southeast Asia to protect against leprosy [Internet]. ScienceDaily. ScienceDaily; 2017 [cited 2021Apr26]. Available from: <https://www.sciencedaily.com/releases/2017/05/170516121714.htm>
26. Moore K. Kansas City metro adds more than 100 new COVID-19 cases ... [Internet]. Kansas City metro adds more than 100 new COVID-19 cases, no new deaths on Wednesday. Kansas City Star; 2021 [cited 2021May10]. Available from: <https://www.kansascity.com/news/coronavirus/article251186244.html>
27. Cohn N. More Hispanics Declaring Themselves White [Internet]. The New York Times. The New York Times; 2014 [cited 2021May10]. Available from: <https://www.nytimes.com/2014/05/22/upshot/more-hispanics-declaring-themselves-White.html>
28. Hispanic Origin [Internet]. The United States Census Bureau. U.S. Department of Commerce; 2020 [cited 2021May10]. Available from: <https://www.census.gov/topics/population/hispanic-origin.html>