EXCESS MORTALITY DURING THE COVID-19 PANDEMIC
BY AGE, RACE, SEX, AND CAUSE OF DEATH

A Thesis
Presented to the
Faculty of
California State Polytechnic University, Pomona

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science
In
Mathematics

By
Cameron Sowers
2022
SIGNATURE PAGE

THESIS: EXCESS MORTALITY DURING THE COVID-19 PANDEMIC BY AGE, RACE, SEX, AND CAUSE OF DEATH

AUTHOR: Cameron Sowers

DATE SUBMITTED: Spring 2022

Department of Mathematics and Statistics

Dr. Adam King
Thesis Committee Chair
Mathematics & Statistics

Dr. Michael Green
Mathematics & Statistics

Dr. Alan Krinik
Mathematics & Statistics
ACKNOWLEDGMENTS

Thank you to my family and friends for their endless support. To my lovely parents, who never once doubted that I could succeed; my mother, who especially encouraged me to pursue this master’s degree and my father, who worked hard to make sure I had the opportunity for higher education. Thank you to my brother for helping me navigate a new school during a pandemic, and to my friends for helping me deal with my stress along the way.

I would like to acknowledge the staff at Cal Poly Pomona, who provided me with such a high quality and supportive graduate school experience, even during an unpredictable two years. Specific recognition to Dr. Green and Dr. Krinik for serving on my thesis committee. And a very large thank you to Dr. King for being not only a motivational teacher, but a thoughtful advisor; without his guidance this work would not be possible.
Since March of 2020, there has been much research and discussion regarding COVID-19 specific mortality in the United States, but predominantly concerning only the first year of the pandemic. In this thesis, we attempt to demonstrate the impact of the pandemic during both 2020 and 2021, while also considering disparities between demographic groups defined by sex, race and ethnicity, and age. We examine not only deaths attributed to COVID-19, but also those not attributed to COVID, with special attention to deaths occurring as a result of typical leading causes of death in previous years.

This thesis aims to provide a more in-depth and exhaustive analysis of mortality during the coronavirus pandemic using data sourced directly from various U.S. government agencies and departments, including the National Center for Health Statistics and the Census Bureau. Much time was spent cleaning and standardizing four different data sets, and while modeling was used to create missing population projections and death estimates for 2020 and 2021, the majority of chapters will be comprised of statistical analysis of national data.

This paper found that when comparing COVID-19 mortality between race and ethnicity groups in the United States, it essential to account for age as well, since people of color tend to have a younger age distribution than the White population. Examining coronavirus mortality without regard to age leads to severe underestimation of the impact of the pandemic on Hispanic people, and to a lesser extent Black and American Indian / Alaskan Native populations. Similarly, categorizing a specific wave of the coronavirus as more deadly than the others can be an over-
simplification when considering total mortality rates rather than mortality rates within individual racial populations. During the first few months of the pandemic, higher weekly mortality rates were recorded for Black Americans than at any other time in the rest of 2020 and 2021.

Considering the coronavirus in terms of four distinguishable waves, we compared mortality during each period rather than overall and observed again that in the initial wave of the pandemic, Black populations experienced mortality rates nearly twice as high as the next population. When accounting for age as well as race, the third wave was most deadly for Hispanic Americans between 55 and 84 in particular, with substantial gaps in death rates seen compared to the other race groups. Throughout the analysis in this thesis, one common reoccurrence is that White and Asian populations tended to have lower death rates than other races. Examining data for the surges before and after the widespread availability of COVID-19 vaccines in the U.S., we found that the previous disparities in mortality between races seemed to lessen for older populations with high vaccination rates; however, while the deaths appear to have equalized among most races, Asian individuals in the fourth wave in fact experienced much lower death rates than other groups.

While there has been discussion that COVID-19 deaths were undercounted in the early weeks of the pandemic, our analysis shows that there was excess mortality in 2020 resulting specifically from diabetes, heart disease, Alzheimer’s disease, and to a lesser degree influenza and pneumonia deaths, which may be attributable to misclassification of COVID deaths. Additionally, other causes such as accidents and strokes exceeded expected deaths and while likely did not result directly from the coronavirus disease, are possible consequences of other sociological factors of the pandemic.
Contents

Signature Page ii
Acknowledgements iii
Abstract v
List of Tables viii
List of Figures x

Chapter 1  Introduction 1
   1.1 Motivation ....................................................... 1
   1.2 Typical Mortality in the U.S. .................................. 3
   1.3 Coronavirus at a Glance ............................... 4

Chapter 2  Literature Review 8
   2.1 Related Publications ......................................... 8
   2.2 Comparison to Thesis .................................... 10

Chapter 3  Data Cleaning 12
   3.1 Overview ......................................................... 12
3.2 Excess Deaths Data ........................................... 14
3.3 Annual Population Estimates ................................. 17
3.4 Vaccination Rates by Demographic .......................... 18
3.5 Leading Causes of Death Data ............................... 20

Chapter 4 COVID-19 Mortality (2020-2021) .................. 22
4.1 Deaths by Sex, Race, and Age ............................... 22
4.2 Weekly COVID-19 Deaths by Individual Demographics .... 25
4.3 Weekly COVID-19 Deaths by Age and Race Combinations ... 29
4.4 Coronavirus in Four Waves ................................... 35

Chapter 5 COVID-19 Mortality Before and After Vaccine Access 39
5.1 Vaccination Rates by Sex, Race, and Age ...................... 39
5.2 Weekly COVID-19 Mortality Before and After Vaccines .... 41
5.3 Total COVID-19 Mortality Before and After Vaccines ........ 43

Chapter 6 Non-Coronavirus Mortality ......................... 46
6.1 Weekly Non-COVID Deaths .................................. 46
6.2 Non-COVID Deaths by Sex, Race, and Age .................. 47
6.3 Leading Causes of Death (2015-2020) ....................... 52

Chapter 7 Conclusion ................................................. 57
7.1 Results ............................................................ 57
7.2 Limitations and Future Direction ............................. 59

Bibliography ............................................................. 61
List of Tables

3.1 Death Counts for Leading Causes in 2019 .................. 20

6.1 Excess Death Percentages by Cause ....................... 55
List of Figures

1.1 Weekly Deaths Before the Pandemic ......................... 3
1.2 Annual Deaths (2015-2021) ............................. 5
1.3 Overall Deaths (2015-2021) ............................. 7

3.1 Population Distributions for Each Year (2015-2019) .... 18

4.1 COVID Deaths by Sex .................................. 23
4.2 COVID Deaths by Race/Ethnicity .......................... 24
4.3 COVID Deaths by Age .................................. 25
4.4 Weekly COVID by Sex .................................. 26
4.5 Weekly COVID by Race/Ethnicity .......................... 27
4.6 Weekly COVID by Age .................................. 28
4.7 85+ Population by Race ................................ 30
4.8 Weekly COVID Deaths by Race & Age (40-64) ............ 31
4.9 Weekly COVID Deaths by Race & Age (65-84) ............ 32
4.10 Weekly COVID in Waves ................................. 35
4.11 COVID Deaths in Waves by Race & Age (40-64) ........... 36
4.12 COVID Deaths in Waves by Race & Age (65-84) ........... 37

5.1 Vaccination Rates by Individual Demographics ............ 40
5.2 Weekly 3rd and 4th Wave COVID by Race/Ethnicity & Age (40-74) 42
5.3 Total 3rd and 4th Wave COVID by Race/Ethnicity & Age (40-74) 44

6.1 Weekly Deaths Excluding COVID . . . . . . . . . . . . . . . . . . . . . . . . . 47
6.2 Non-COVID Mortality by Individual Demographics . . . . . . . . . . . . . . . 48
6.3 Non-COVID Mortality Percent Increases in 2020 . . . . . . . . . . . . . . . . . 50
6.4 Non-COVID Mortality Percent Increases in 2020 . . . . . . . . . . . . . . . . . 51
6.5 Leading Causes of Death . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53
6.6 Top Two Causes of Death . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
6.7 Other Causes of Death . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
Chapter 1

Introduction

In this first chapter, we will state motivation for our work, as well as provide introductory analysis regarding typical mortality in the United States. To give context, we will explicitly define excess mortality and give an understanding of why it is a necessary component for our purposes. Section 1.3 will specifically explain how death rates changed in 2020 and 2021 as a result of the coronavirus pandemic.

1.1 Motivation

In early 2020, the SARS-CoV-2 virus began spreading across the world, and on January 20th, the Center for Disease Control announced the first laboratory-confirmed case of COVID-19 in the United States. Since then, there has been much discussion over the severity and true impact of this pandemic, with additional attention on the influence demographic factors may have had on COVID-19 mortality rates. A commonly repeated notion in the early months was that COVID-19 was no worse than the common flu; however, after nearly two years of tracking and publicly shar-
ing COVID-19 data we have seen that this virus has been much deadlier than the common flu. In fact, we will show in this thesis that COVID-19 has had even more of an impact than surface-level analysis suggests, and specifically there exists an undeniable effect of age and race/ethnicity when examining variation in coronavirus deaths.

*Excess mortality* is defined as the number of deaths during a crisis that exceed normal expectations and may have any cause. In terms of COVID-19, we want to examine this concept specifically to give “a more comprehensive measure of the total impact of the pandemic” than we could from the virus specific death count alone (Ritchie et al., 2020). In this thesis, we will be examining both total and COVID-19 specific death counts during 2020 and 2021 according to the time of year and various demographic factors, with the goal of determining the true impact of the COVID-19 pandemic on people of specific race, sex, and age backgrounds, but also on our country as a whole. This paper focuses primarily on the deaths that occurred in the United States between January 1, 2020 and December 31, 2021, with additional consideration given to historic death counts between 2015-2019 for comparison. Throughout this thesis we will explore in-depth breakdowns of mortality among different demographic groups, accounting for vaccination rates among these various communities, separated into distinct coronavirus waves, and with regard to causes other than COVID-19. First though, we think it is necessary to establish a general understanding of typical death totals in recent years and examine how this statistic has changed as a result of the pandemic.
1.2 Typical Mortality in the U.S.

If we want to understand the impact of the pandemic and measure excess mortality during the past two years, it is first necessary to set a baseline using data from previous years. We examined data detailing weekly deaths in America from 2015-2019 with specification for race/ethnicity, age group, and sex. In Chapter 3, we will give a more in-depth explanation of these demographic breakdowns as defined by the data retrieved from the CDC, but for now the aim is to communicate introductory analysis regarding typical mortality rates among these groups in the five years prior to the rapid spread of COVID-19.

Looking at Figure 1.1, we see that deaths per week tend to be higher towards both the beginning and end of each year, likely related to “seasonal changes in behavior and the human body, as well as increased exposure to respiratory diseases” (Gasparrini et al., 2015). In addition, we found that the yearly death totals have

![Weekly Death Counts 2015–2019](image)

Figure 1.1: Weekly Deaths Before the Pandemic
been slightly increasing each year from 2015 to 2019 (Table 1.1), which is not surprising since the U.S. population has also grown annually by between 0.6% and 0.7% in that same period. In 2019, there were a little over 2.83 million deaths nationally, so we would have expected the death counts in the following years to be slightly larger than that amount. The United States Census Bureau (Sabo & Johnson, 2022) states that the average annual increase in deaths was 1.6% during the last decade, so we will use that as an approximation for what the expected mortality increase in 2020 and 2021 should have been without the pandemic.

1.3 Coronavirus at a Glance

During approximately the first two years of the pandemic in the U.S., a total of 836,206 people were reported to have died of COVID-19; this is broken down into 389,794 deaths in the year 2020, and 446,412 in 2021. When examining the total number of deaths every year, we expect that number to slightly increase in conjunction with population growth—between 2015 and 2019, the average increase in annual deaths was 1.4%. However, Figure 1.2 shows that in 2020 there was a 20.6% increase in deaths compared to the previous year and in 2021 where we might expect a roughly 3% change from 2019, there was instead an 18.7% increase.

We can also look at the non-COVID related death counts, in order to understand a more complete picture of excess mortality in the last few years. It appears that there was a larger than expected increase in deaths from 2019 to 2020, even when we exclude the ones attributed to COVID-19; there was a 6.8% increase, which is higher than our previous average yearly rise (Figure 1.2). There are a few causes we might attribute these excess deaths to, the most plausible of which is that the
true number of COVID-19 deaths was undercounted in the early months of the pandemic as a result of a scarcity in testing. According to a study by Weinberger et.al. (2020), since early pandemic responses were managed at the state level, the excess in weekly deaths observed began in several states “before increases in the availability of COVID-19 diagnostic tests and were not counted in official COVID-19 death records”.

Another possible contribution could be non-covid related deaths arising from a lack of routine healthcare due to either overwhelmed hospital systems or patient hesitancy to visit a highly contagious environment. The CDC released a report that stated by June 2020, an estimated 41% of Americans “delayed or avoided medical care”, 12% of which being urgent or emergency care (Czeisler, 2020). Later in this thesis, the typical leading causes of death will be examined to more fully understand
this unusual increase in deaths observed during 2020, even after removing those resulting from the coronavirus disease.

In contrast, when looking at 2021 there was only a 2.9% increase in deaths not related to COVID-19 compared to the total deaths two years prior, which is not unreasonable when again considering the average yearly increase in deaths was 1.6% over the last decade (Figure 1.2). When excluding deaths due to coronavirus in the second year of the pandemic, deaths dropped below the non-COVID death total in 2020. This indicates that in the second year of the pandemic there was little undercounting of deaths due to the coronavirus, most likely since testing was more widely available and uniformly required as compared to the previous year. Unfortunately, data on the leading causes of death for 2021 has not yet been published by the CDC, and so while we note the presence of this decrease in non-COVID deaths as compared to 2020, we will not explore it any further than the simple speculation given above.

While looking at the total mortality across recent years is important, more specific insights arise when examining the data recorded as deaths per week. Figure 1.3 shows the weekly deaths for previous years and pandemic years, including additional information for 2020 and 2021 to specify death counts excluding those attributed to the coronavirus. We see that for the pandemic weeks beginning after March of 2020, death counts are consistently higher than for the pre-pandemic years. In the past five years the average weekly death count in the United States was between 51,469 and 54,463, and it is clear that 2020 and 2021 surpass the 2019 average by over 50% during the pandemic’s most deadly weeks. Again, it is worthwhile to consider the death counts excluding COVID-19 deaths, and Figure 1.3 demonstrates support for our earlier statement that COVID-19 related deaths were potentially
undercounted in the beginning of the pandemic. There is a noteworthy spike in non-COVID deaths for 2020 between weeks ten and twenty (March – May), which mirrors the first wave of increased deaths in the figure; most likely, this smaller peak is reflecting undiagnosed COVID-19 deaths in the first two months of the pandemic.
Chapter 2

Literature Review

Before diving into an explanation of the data sets and the analysis we have compiled for this thesis, it is essential to acknowledge and understand other recent publications researching the effect of the coronavirus pandemic on excess deaths in the United States. This chapter serves to discuss related literature and give consideration to possible similarities and differences with the work in the rest of this paper.

2.1 Related Publications

The goal here is to understand major results found in other studies of shorter periods of COVID-19 and see in later chapters how this thesis detailing a full two years of excess mortality compares. The first piece of literature we will discuss is an article from the Morbidity and Mortality Weekly Report published in October of 2020. This piece specifically focuses on the differences in excess deaths across race and age groups; the report is examining excess weekly mortality overall, without distinction between deaths attributed to COVID-19 and those not.
The researchers modeled expected weekly deaths for the desired period using “overdispersed Poisson regression models with splines”; they note that different methods of expected death counts could lead to different excess death results, and may vary from study to study (Rossen et al., 2020). According to the article, there were higher than expected deaths recorded every week between March 2020 and October 2020, totaling an estimated 299,028 excess deaths. In addition, the authors found that the highest percentage increase was seen in adults between the age of 25 and 44; within race and ethnicity groups the largest increase was among the Hispanic population. It is important to note that age and race are investigated separately in this report, which is a topic we will discuss in Chapter 4.

Steven Woolf et al. (2021) discuss excess mortality staring in March of 2020 through the end of the year, which encompasses a larger time period than the last publication, but still only examines the first year of COVID-19. In the months examined, this study found that nearly 23% more total deaths were reported than expected, with a higher death rate among Black populations than for both White and Hispanic groups; the authors state this “[reflects] the racial disparities in COVID-19 mortality”, which accounted for 72.4% of excess deaths in the U.S. (Woolf et al., 2021). As well as examining overall excess mortality, there is also some attention given to increases in deaths not directly credited to COVID-19. At different times during 2020, heart disease and Alzheimer deaths each had mortality rates above the expected amount (Woolf et al., 2021). While this paper gives less consideration to weekly excess mortality as opposed to overall deaths, we find the inclusion of excess mortality as a result non-COVID causes to be very valuable and has the potential to be expanded upon.

There is less research examining the most recent year of the pandemic, yet
“COVID-19 Cases and Deaths by Race/Ethnicity: Current Data and Changes Over Time” was published by the Kaiser Family Foundation in February of 2022. The authors detail analysis of COVID-19 in regard to cases, hospitalizations, and deaths among various race and ethnicity groups, for total population and age adjusted populations. Noting that “people of color are generally younger than White people”, the data was adjusted for age and thus showed larger death disparities in Hispanic, Black, and American Indian and Alaska Native populations (Hill & Artiga, 2022). When considering coronavirus weekly mortality during the pandemic, the disparities between race and ethnicity groups lessened at times, but widened during surges of new COVID-19 variants in which “people of color [were] disproportionately impacted” (Hill & Artiga, 2022). This publication gave a brief insight to more recent pandemic mortality, with emphasis to racial disparities, but limited to only COVID-19 related deaths.

### 2.2 Comparison to Thesis

As mentioned before, there has been much consideration given to the effect of coronavirus in the year 2020, and less investigation more recently. The major benefit of our work is to show a more complete history of the pandemic and how mortality has changed throughout 2020 and 2021. In addition, we wanted to give an in-depth exploration of racial, age-based, and sex disparities for not only COVID-19 related deaths, but those not attributed to the coronavirus as well. Many reports either regard the demographics individually, or with essential consideration to possible intersections; we will attempt to explore both these analytical approaches, with the purpose of shedding light on their drastic differences and emphasizing the impor-
tance of adjusting for age when considering race. Through many different avenues, we intend to lay out a thorough understanding of mortality during the coronavirus pandemic, combining the various research topics we have seen explored in these other pieces of literature.
Chapter 3

Data Cleaning

Here, we will explain the data that was used for this paper and summarize any cleaning done to ensure it was ready for analysis. Section 3.1 will provide a basic description of each data set and where it was obtained, while the following four sections will outline the individual modifications that needed to be made.

3.1 Overview

In order to obtain a complete understanding of excess mortality between January of 2020 and December of 2021, we have compiled four data sets which give various insights into the effects of this pandemic. The four data sets included detailed death counts by various demographic groups (before and during the coronavirus pandemic), population estimates for the same time period, vaccination totals broken down by demographic, and historic leading causes of death as compared to the first year of the pandemic. The AH Excess Deaths by Sex, Age, and Race and Hispanic Origin data set (2022) was sourced from data.cdc.gov and provided by the National Center for Health Statistics; it includes weekly death counts in the U.S. broken
down by race/ethnicity, sex, and age group for 2015-2021. Starting in 2020, there are additional counts specifically of weekly COVID-19 attributed deaths. This is the main data that was used for analysis, especially in any discussion of COVID-19 specific mortality among various subsets of the population. However, to completely understand the impact on certain demographic groups and compare the results meaningfully, the population estimates of each group were needed to normalize the death counts.

The second data set is the *Annual Estimates of the Resident Population by Sex, Age, Race Alone or in Combination, and Hispanic Origin for the United States: April 1, 2010 to July 1, 2019* (2020). It was sourced directly from census.gov and includes detailed data regarding sex, age, and race/ethnicity population groups according to the 2010 census and projections for the subsequent nine years. This information was used to extrapolate the desired population data for each combination of demographic factors for the pandemic years of 2020 and 2021, which allowed us to have a more accurate understanding of the mortality rates of various U.S. population groups as stated before. Next, the *COVID-19 Vaccination Demographics in the United States* (2022) data set was also retrieved from data.cdc.gov and provided by the CDC from the Immunization Information Systems databases; this information was helpful in giving background on vaccine protection in the U.S. It gives single dose, full vaccination, and booster counts in the U.S. grouped by a single demographic factor. New counts were updated on a daily basis, with the numbers for previous dates remaining in the data set as a record.

Lastly, we included the data on the leading causes of death from 2015-2020 to examine the full picture of excess mortality due to causes other than the coronavirus. This data reports information regarding the top causes of death in the U.S. and their
yearly counts. We had to manually collect and source the information from annual
reports produced by the National Center for Health Statistics, since a collective
data set for the desired years could not be located. The purpose of including this
information is to analyze the yearly change in death counts for various causes,
focusing specifically on if there was a significant difference in 2020, possibly related
to any socioeconomic or healthcare related effects after over nine months of the
pandemic.

The main task in preparing these four data sets for statistical analysis was to
standardize the various demographic groups so that the data could be compared
and used in conjunction with each other. It should be explicitly stated what lev-
els will be used in this thesis for each demographic group, as determined by the
death information collected and population groups defined on the U.S. census. For
detailing race/ethnicity the groups are defined as follows: White, Black, American
Indian or Alaska Native (AIAN), Asian, Native Hawaiian or Other Pacific Islander
(NHPI), and Hispanic. Note that all five race groups besides Hispanic indicate
non-Hispanic descent. Female and male are the two categories used to define sex.
Lastly, the age groups we will use are 0-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44,
45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, and 85+.

3.2 Excess Deaths Data

The \textit{AH Excess Deaths} data set has values for total deaths and COVID-19 spe-
cific deaths separately, and in addition includes “(predicted) weighted provisional
counts”. The weighting is predicted by the NIH using previous years to account for
underreported deaths in recent weeks of the data. We will only be using the raw
(unweighted) death counts in this paper and will instead be cutting off the most recent data in order to handle the potential underreporting issue. The data also includes other values such as average death counts for time of year and the percent change by week in 2020 and 2021. Since we will mainly be focusing on demographic disparities rather than historic weekly death trends, this portion is removed from the data as well.

The relevant factors for the purpose of this thesis include multiple demographic categories that give a detailed account of the death counts in various groups, as well as week number and year of the death counts. In this data set, the “Race/Ethnicity” variable is broken into seven groups: the six defined previously and an extra “Other” identifier. The “Sex” factor has only two levels: male and female. The “Age Group” categorical variable is separated by five-year age gaps, with the exception of the first group “0-14 Years” and the last age group “85+” as noted before; however, there is an additional “Not Stated” level. The “Week” variable will be used to examine COVID-19 specific death trends during the pandemic years; weeks consist of 1-53, in which the 53rd week is only used for the year 2020.

Another thing to note is that there were many duplicate death counts recorded in this data set, since some values were listed under either “All Race/Ethnicity Groups”, “All Sexes”, or “All Ages” in combination with specific other groups. For example, there was information regarding deaths among Hispanic 0-14 year olds, but without regard to sex—coded as “All Sexes”. These general values were included to have quick access to larger demographic subsets of the population; however, for this thesis we were only interested in examining the most stratified data available and grouping them together as needed later. Thus, these repeats were removed from the data set without loss of information, which we verified by
summing the number of total deaths in the altered data set for each year and comparing them to official yearly death counts from the CDC.

To account for the discrepancies in demographic definitions, certain values were removed completely from the data set, since redistributing such a small proportion of the deaths was determined to not be necessary. Although this death data includes an “Other” category for race and ethnicity, the identifier was ultimately dropped since the census data we will be using for normalizing does not include this particular category. The number of deaths classified as “Other” was relatively small compared to the total death counts—only 0.9% of total deaths and 1.1% of COVID-19 deaths. Since these instances attributed to such a small percentage of both the COVID-19 and the overall deaths, redistributing these values would have mostly resulted in small decimal values for each granular demographic combination and week per year. Regarding the levels in the “Age Group” variable, there was the concerning “Not Stated” category included. This value was in fact found to mainly be used in in combination with “All Race/Ethnicity” or “All Sexes”, and thus was nearly completely removed from the data in the initial cleaning step. The few instances that remained were only used in combination with the “Other” category for race or contributed to a very small percentage of the total deaths, and thus was also removed from the data resulting in negligible difference. Therefore, it was decided that all the aforementioned values could be removed from the data set without major consequence, and we could interpret the results for excess deaths with uniform demographic labels.
3.3 Annual Population Estimates

In this data, the race groups were categorized similarly to those in the previous data set; however, it did not have overall totals for Hispanic populations. Census race data is instead grouped into White, Black, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander. Then within each of these groups, the population counts are further separated by Hispanic and Non-Hispanic origin. In order to gather accurate population estimates for Hispanic residents in the U.S. during 2015 to 2019, we combined the Hispanic totals within the various race groups for each sex and age group subset. However, after adding in the Hispanic population counts and examining all of the population data together, we saw an overcounting of approximately 2-3% in total yearly population. This is presumed to have been a result of the option on the 2010 census to select more than one race; a 2010 census brief stated that nine million people (2.9% of the total population), reported multiple races (United States Census Bureau, 2012). To adjust for this discrepancy, each population group was decreased by a factor of the overcounted percentage for that specific year. It is important to note that this was done based on the assumption that each group had equal amounts of multiracial reports, which is most likely inaccurate but is a negligible difference for the purpose of this analysis.

The next step needed to prepare this data set for further analysis was to determine population predictions for the years 2020 and 2021, specifically predictions for each existing demographic subset. To extrapolate the necessary data, we broke the population counts into every combination of race/ethnicity, sex, and age group, which totaled 192 subgroups. Each subgroup of the data was plotted across the years 2015-2019 to determine how to fit the regression model (Figure 3.1). The trends seemed to be relatively similar; thus, they were all fitted with both a linear
regression model and a quadratic regression model, which were then used to predict the populations for the subsequent two years. The average of these population values was calculated and added into the existed population data set, to be used moving forward.

### 3.4 Vaccination Rates by Demographic

Since the COVID-19 Vaccination Demographics data was documented daily, we first needed to extract only the subset of the data necessary for the purpose of this thesis. We chose the vaccination data recorded as of 6/1/2021, because we will later use it when comparing two COVID-19 waves: one before and one after vaccinations were available for most Americans. The data also needed some cleaning in regard to standardizing the demographic categories to be consistent with the
previous two data sets. The race/ethnicity groups were nearly the same as before, separating Hispanic individuals from the other five races; however, one difference in this data is the existence of a “Non-Hispanic Multiple Races / Other” category. These counts for each vaccination level were redistributed among the non-Hispanic groups (White, Black, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander) based on their population proportions extrapolated for the year 2021, as determined from previous work with the population data set. Note that since the multiple/other race category specifies non-Hispanic ethnicity, the proportions were calculated to be out of the population total for only five of the six groups.

In addition, there was a certain percentage of missing demographic information in the data set. There were two additional counts included besides the defined levels for race, sex, and age group which detailed known and unknown subtotals for each group (i.e. the number of vaccinations where sex is known and the number of vaccinations where sex was not recorded). In the provided data, there was a minor error in the unknown race totals for the fully vaccinated counts, which we manually corrected based on the known demographic totals given. For the sex and age group demographics, the missing information was not a large concern since at least 99.1% and 99.98% of vaccination records included their known sex and age identifiers respectively. Unfortunately, for race and ethnicity as low as 74.1% of the data included this information. To account for this shortage in reporting, we decided to redistribute the unknown counts for each vaccination level similarly as before but this time based on the population proportion of all six race/ethnicity groups. After dealing with the unknown demographic information and the multiple/other race category, the vaccination data was ready to be analyzed appropriately.
3.5 Leading Causes of Death Data

The causes of death data detailed the top ten leading causes of death in the U.S. and their counts for 2015 through 2019. These ten causes remained the same during this five-year period and are: heart disease, cancer, accidents, stroke, chronic lower respiratory diseases, Alzheimer’s disease, diabetes, influenza and pneumonia, kidney disease, and suicide (Table 3.1). In 2020, the global spread of coronavirus made COVID-19 the third leading cause of death in this country, and so it was included in that year’s data as well. In the interest of analyzing excess mortality and to determine how the pandemic may have indirectly affected the yearly death rates, we needed to model what the death counts would have approximately been.

<table>
<thead>
<tr>
<th>Cause of Death</th>
<th>Death Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Disease</td>
<td>659,041</td>
</tr>
<tr>
<td>Cancer</td>
<td>599,601</td>
</tr>
<tr>
<td>Accidents</td>
<td>173,040</td>
</tr>
<tr>
<td>Respiratory Disease</td>
<td>156,979</td>
</tr>
<tr>
<td>Stroke</td>
<td>150,005</td>
</tr>
<tr>
<td>Alzheimer’s Disease</td>
<td>121,499</td>
</tr>
<tr>
<td>Diabetes</td>
<td>87,647</td>
</tr>
<tr>
<td>Kidney Disease</td>
<td>51,565</td>
</tr>
<tr>
<td>Influenza &amp; pneumonia</td>
<td>49,783</td>
</tr>
<tr>
<td>Suicide</td>
<td>47,511</td>
</tr>
</tbody>
</table>

Table 3.1: Death Counts for Leading Causes in 2019
under non-pandemic circumstances. To do this, we defined a linear regression model for each cause of death using only the 2015-2019 data to extrapolate the expected death counts by cause in 2020. With these projections, the difference in deaths for each cause in 2020 was ready to be considered.
Chapter 4

COVID-19 Mortality (2020-2021)

This chapter will focus specifically on deaths attributed to the coronavirus disease. The first two sections will provide analysis of mortality by individual demographic factors, while the later two give further consideration to race when also accounting for age. Section 4.3 examines weekly death rates and Section 4.4 divides 2020 and 2021 into four distinct COVID-19 waves, so we can analyze death rates across those groupings.

4.1 Deaths by Sex, Race, and Age

The first analysis we did when examining coronavirus specific deaths during 2020 and 2021 was to understand how deaths were distributed among the various demographic groups overall. In order to compare amongst the different levels in each group, the mortality rates shown are normalized by the populations of each specific subgroup. Note as well that the rates are scaled to represent deaths per 100,000 people. The first demographic identifier we investigated was sex.

In Figure 4.1, it is clear that COVID-19 deaths are slightly higher for men than
Figure 4.1: COVID Deaths by Sex

women during both years of the pandemic, however the difference is not anything too concerning. Although, one observation of interest is that this gap does increase in 2021 as compared to 2020, possibly due to lower vaccination rates among men compared to women in the U.S., which will be further discussed in Chapter 5.

Next, we looked at differences in COVID-19 deaths among various race and ethnicity groups. When considering mortality throughout the entire pandemic, deaths by population proportion were highest amongst Black Americans, followed by White, Hispanic, and American Indian / Alaska Native. On the lower end for mortality rates were Native Hawaiian / Pacific Islander and Asian individuals. This seems slightly contradictory to many conclusions stated in Chapter 2 regarding the effect of race on coronavirus deaths; however, it is important to remember the mortality rates shown in Figure 4.2 are only taking race and ethnicity totals into consideration, without regard to any other factors like age. Additionally, if we pay closer attention to the difference between the two years, we see that mortality for people who are White in the U.S. surpassed that of people who are Black in 2021. In
fact, mortality rates among Black, Hispanic, and Asian racial groups either slightly decreased or stayed the same. Alternately, mortality rates for the AIAN, NHPI, and White groups increased in the second year of the pandemic, most drastically for people who are Native Hawaiian / Pacific Islander and White. Considering again that vaccines became widely available in the Spring of 2021, this difference very likely could be related to vaccine hesitancy within certain racial groups. We will explore this possibility more in the next chapter.

The last and likely most drastic contribution to differing mortality rates resulting from the coronavirus is age. As expected, the death rates were extremely high for individuals aged 85 and older, but in Figure 4.3 we clearly illustrate exactly how drastic the disparity was. In 2020, mortality among people in the highest age category was over twice as high as the next closest age group, 80-84 year old individuals. It is again interesting to compare between the first and second year, and we see in
this case that the top two age groups saw a decrease in COVID-19 mortality; while this could be attributed in part to vaccine availability, for the 85 and over group it is also possible that the actual population total was much lower than the projection we calculated for 2021, as a result of high excess mortality for that age range in the year prior. Since COVID-19 deaths seem to have a very strong dependence on age, in Section 4.3 when examining mortality for specific combinations of race/ethnicity and age group, we will focus specifically on those aged 40 and older.

4.2 Weekly COVID-19 Deaths by Individual Demographics

While still considering the different demographic categories independently, it is worthwhile to analyze the coronavirus death rates over time, and thus, in this
section we will examine weekly mortality over the course of 2020 and 2021. First to consider again is the effect of sex on weekly COVID-19 death rates, and we observe similar results as in the previous section. Figure 4.4 shows that men see slightly higher weekly mortality rates than women during the pandemic, which becomes most apparent during COVID-19 surges when the death rates are highest. However, it is interesting to note that after the most recent wave of coronavirus, the weekly mortality rates for men retain a slightly larger gap above women than for previous waves. This would again indicate the possible effect of vaccines, which became widely available in the U.S. between the third and fourth COVID-19 waves.

Since we already have an idea of the difference in coronavirus mortality rates between various race and ethnicity groups, looking at these rates on a weekly basis will produce a more detailed understanding of when exactly such variations were most prominent. Looking at Figure 4.5 overall, the death rates seem to be relatively similar in the third and fourth COVID-19 waves, yet the same cannot be said for the beginning of the pandemic. The most obvious and concerning consideration
to be made in Figure 4.5 is that the first wave of coronavirus deaths was much more deadly for Black Americans than other groups. This might be a result of a few realities, such as increased likelihood of living in densely populated areas and decreased likelihood of “equitable healthcare access” (Ray, 2020). In addition, New York City has a very large Black population and this area was hit harder than any other U.S. city in the initial wave. At its peak, death rates during the first few months of the pandemic were over twice as high for Black individuals than for Hispanic, White, and Asian people, and over four times as high as AIAN and NHPI populations. In the following wave of summer 2020, weekly mortality rates among Hispanic Americans slightly overcame the rates of Black Americans, but both these groups retained consistently higher weekly death rates than the other races.

Another noticeable element is the second half of 2021, the weekly death rates of Asian Americans was consistently less than half that of any other race/ethnicity group. As restrictions and mask mandates have been lifted gradually, COVID-19
protocols have become less of a requirement and more a suggestion left up to individual decision. Since wearing masks is common and even culturally encouraged in East Asia, a possible contribution to why people in the U.S. who are Asian were experiencing lower recent death rates is continued mask wearing and social distancing even after the requirement by local governments. Looking at a figure which indicates mortality rates are relatively similar between all races and ethnicities during much of the pandemic is surprising, and contradictory to other literature. There likely exists many more conclusions to be made about COVID-19 mortality when considering race, but which require more detailed stratification of the demographic groups and will be explored in section 4.3.

Finally, age group as an indicator of coronavirus mortality is extremely important. Similar to sex disparities, the age group variation in weekly death is most prevalent near the peak of each wave. Figure 4.6 demonstrates just exactly how deadly the coronavirus was for older Americans; at the height of the pandemic dur-

![COVID Death Rates by Age Group](image)

Figure 4.6: Weekly COVID by Age
ing the December 2020 through February 2021 surge, over 1 in 1,000 people aged 85 and older were dying every week. While the next most affected group was those aged 80-84, they experienced mortality rates less than half as great as the previous group; this is not to be discarded as insignificant however, as during the same period over 1 in 2,000 people in this population were dying every week. Figure 4.6 indicates that age is by far the strongest indicator of COVID-19 mortality rates and has a staggering impact on the number of deaths observed.

4.3 Weekly COVID-19 Deaths by Age and Race Combinations

From Section 4.2, it is obvious that age is an extremely important factor to consider when discussing the impact of COVID-19 on mortality rates. Now in this section, we will demonstrate why age is not just important, but absolutely necessary for mortality analysis during the pandemic, even when considering other demographic indicators. The previous examination of individual demographics gave some insight into race and ethnicity inequality when looking at weekly coronavirus deaths; however, for many weeks these rates were not extremely different, which contradicted the findings of other studies mentioned in Chapter 2. To investigate further, we decided to stratify the data not only by race/ethnicity over time, but also with additional consideration to age group; since we have seen that older age strongly increases the mortality of a population, holding this variable constant will allow for more clear results of any racial disproportion in deaths.

Note that each set of mortality rates is normalized by the population of each subset of race/ethnicity and age group; since the population sizes for American
Indian / Alaska Native and Native Hawaiian / Pacific Islander are much smaller relative to the other groups, stratifying even further for age and lags in reporting created difficult to interpret death curves. As a result, they were removed from this portion of the analysis. In addition, populations aged 85 and older will not be given much consideration here, since as seen in Figure 4.7, White populations have an extremely elevated percentage of older and high-risk individuals. Given the information in the previous section that this age category had much higher COVID-19 mortality than any other group, this particular subset of the population is likely what was skewing the overall mortality rates before and allowing for misleading conclusions to be made.

Looking at Figures 4.8 and 4.9 on the next pages, we see a different picture of COVID-19 mortality rates than before. For people between the ages of 40-79, at almost every week during 2020 and 2021 the mortality rate for Hispanic and Black individuals was noticeably higher than that of White and Asian individuals. In
COVID Death Rates in Ages 40–44

COVID Death Rates in Ages 45–49

COVID Death Rates in Ages 50–54

COVID Death Rates in Ages 55–59

COVID Death Rates in Ages 60–64

(a) COVID in Ages 40–44
(b) COVID in Ages 44–49
(c) COVID in Ages 50–54
(d) COVID in Ages 54–59
(e) COVID in Ages 60–64

Figure 4.8: Weekly COVID Deaths by Race & Age (40–64)
Figure 4.9: Weekly COVID Deaths by Race & Age (65-84)

the previous section, race did not seem to affect total mortality rates very much after the initial COVID-19 wave, but when separating the deaths further by age it is abundantly clear that coronavirus mortality was not consistent among racial groups. During coronavirus death peaks, Black and Hispanic mortality for certain age groups reached 2-3 times higher than weekly deaths for White and Asian Americans.

While for nearly every race, the weekly coronavirus mortality rate reached its
deadliest during the winter between 2020 and 2021, this is in actuality an inaccurate statement for Black Americans. Based on the figures shown, for Black populations aged 55 and older, the first wave consistently reached the highest death rates per week. In fact, in every age group examined here the weekly mortality rate in the first wave was higher for Black individuals than any of the other three races and ethnicities. This tells us that the initial few months of the pandemic in the U.S.—which consisted of a slow federal response and lack of available testing—was disproportionately affecting Black Americans and resulted in a deadlier start to the pandemic for this group as compared to other populations.

We must also consider the impact that controlling for age has on the data for other pandemic weeks. Specifically for individuals of Hispanic origin, the mortality compared to other races when not accounting for age was severely underestimated and overlooked in most weeks. Between July of 2020 and July of 2021, mortality rates for Hispanic Americans were higher than every other race group shown in Figures 4.8 and 4.9—with exception to a few weeks during the spring of 2021. The most noticeable differences occurred at the peaks of the two waves during this time; examining Figure 4.8(e) of mortality for the age range 60-64, the death rates per week for Hispanic individuals were nearly double that of the next highest group. Interestingly, we see these racial gaps begin to close in the most recent wave of the pandemic, with weekly mortality rates being very similar among White, Hispanic, and Black groups. We can see Hispanic mortality rates no longer tended to be noticeably higher, and instead Black mortality is shown to be on par with or slightly higher than that for Hispanic groups in the fourth wave. This is again another indication that examining vaccination rates will give some insight and possible explanation of the change in racial mortality at the end of the second year
of the COVID-19 pandemic.

The lack of high Hispanic mortality represented when analyzing race and ethnicity on its own can be linked to the difference in age distribution among the various populations. According to William Frey (2020) at Brookings Metropolitan Policy Program, in 2019 the median age of White Americans was 43.7, but 29.8 and 34.6 for Hispanic and Black Americans respectively. Additionally, the population of White Americans has actually begun to gradually decline since 2017, while Hispanic populations grew by a rate of 20% in the last decade; Frey cites natural growth of the population as the primary contributor, with only one-fourth of the increase attributed to immigration (Frey, 2020). This means that the White population in the U.S. is older, while the Hispanic populations tend to be younger than other race and ethnicity groups.

In the context of COVID-19 mortality, White Americans are actually expected to be more heavily affected than other groups. This is likely the reason that when examining race without age consideration, the White population seemed to have death rates on par with Black and Hispanic populations, when in reality, for many age groups this was untrue. After understanding that Hispanic populations—and to a lesser degree Black populations—tend to be younger, in combination with the results in Figures 4.8 and 4.9, we can confidently state that analysis on the impact and mortality of COVID-19 without consideration to age is incomplete and will likely be extremely misleading.
Based on the analysis of weekly death rates, it is clear that coronavirus mortality during the past two years differed within the various demographic groups over time. Interested in investigating this further, we decided to compare deaths separated into distinct COVID-19 waves so that we can also consider American Indian / Alaska Native and Native Hawaiian / Pacific Islander populations as well. Looking at Figure 4.10, we see that the deaths can be separated into four individual waves, and a fifth wave beginning at the end of 2021. For the purposes of this analysis, we will only consider the four complete coronavirus waves; the dates of each wave are as follows: first wave (12/29/2019 – 6/27/2020), second wave (6/28/2020 – 10/3/2020), third wave (10/4/2020 – 7/3/2021), and fourth wave (7/4/2021 – 11/13/2021).

Figures 4.11 and 4.12 visualize COVID-19 death rates by wave and race, with each individual plot specifying a particular age group. As stated in Section 4.3, we
Figure 4.11: COVID Deaths in Waves by Race & Age (40-64)
can see the trend of high Black mortality within the first wave for all ages, and high Hispanic mortality demonstrated within the second and third surges. By including the two additional racial groups we can now observe how they compare with rest of the population, but to discuss this, we will make separate inferences for ages 40-64 and 65-74. Looking at Figure 4.11 for the younger populations shown, AIAN and NHPI populations tended have death rates closer to Hispanic and Black groups, rather than the lower impacted White and Asian groups. The death rates among
these two groups was often very similar to or even higher than that of the two previously noted most affected races. In fact, for 40-49 year-olds, AIAN mortality exceeded that of Hispanic people in the third wave, and during the fourth wave, death rates for NHPI individuals well surpassed that of other races between the ages of 45-54. It is important to note, that death rates among the aforementioned age groups are generally much lower than for the older populations, but analysis within these subsets is still useful for internal comparison.

The data visualizations for ages 65-74 in Figure 4.12 give slightly different conclusions. While AIAN groups retained mortality nearer to the population highs, NHPI groups often had rates similar to that White and Asian populations, and for ages 70-84 even had the consistently lowest mortality of any race and ethnicity group during the largest COVID-19 wave. This had a huge impact on lessening overall mortality for the Native Hawaiian and Pacific Islander population, considering how deaths predominantly existed among older populations and in the third wave.

The analysis in this section revealed even further than in Section 4.3 how the overall mortality of each race group is distributed by age. We saw that AIAN individuals experienced relatively high death rates throughout the second and third waves of the pandemic, and for the younger ages (40-49) they even had the largest rates. While the younger NHPI populations also saw high mortality on par with Hispanic, Black, and AIAN groups, for populations 65-84 they experienced death rates more similar to that of White and Asian people, at times even retaining the lowest mortality of any race or ethnicity group. Stratifying by age, race, and COVID-19 wave gave significant insights that could not have been realized by simply looking at the demographics separately.
Chapter 5

COVID-19 Mortality Before and After Vaccine Access

Similar to Chapter 4, we will be again looking at COVID-19 specific deaths; however, in this chapter, we aim to determine how mortality differed among various demographic groups when considering the availability of vaccinations in the United States. The first section will summarize the vaccination rates within certain demographic populations to give context for the following two sections, which provide weekly and surge mortality information.

5.1 Vaccination Rates by Sex, Race, and Age

As stated before, public vaccinations for the coronavirus began in early 2021, with open access nationally by mid Spring; we will be using vaccination data as of June 1, 2021 with the assumption that by July of 2021 all individuals included will be considered fully vaccinated—completed the required series of shots and sufficient time has passed since receiving it. In the next section we will explore the change
in mortality before July 4, 2021 and after, but for now we will simply give an understanding of vaccination rates among the individual population groups.

In Figure 5.1(a), we can see that women indeed have higher vaccination rates than men, which gives a possible explanation for why deaths varied more by sex in the most recent wave of the pandemic than before. For the purpose of the analysis

Figure 5.1: Vaccination Rates by Individual Demographics
in the next section, we will be focusing specifically on the vaccination rates among different ages, as well as for the various race and ethnicity groups. While it would be beneficial to have more stratified information detailing combinations of these two demographics, that information has not been released by the CDC and so we will proceed with what is shown in Figure 5.1.

The highest percentage of vaccinations are for Asian and White groups with approximately 48% of their respective populations receiving vaccinations by June of 2021, while the other groups ranged between 34% and 39%. With this in mind, we will explore in the next section whether the difference in vaccinations by race and ethnicity is reflected in the mortality rates. Looking at age in Figure 5.1(c), age groups 65-74 and 75+ have very high vaccination rates at around 75%, which is not surprising considering they were at very high-risk of COVID-19 and thus some of the first populations allowed to receive a vaccine. Note that vaccination rates drop quite a bit for younger ages: 65% for the 50-64 age group and even lower with 53% for individuals ages 40-49. This indicates that we expect the older populations specifically to experience lower mortality rates in the fourth wave of COVID-19 than seen in the third.

5.2 Weekly COVID-19 Mortality Before and After Vaccines

After examining vaccination rates for various demographics groups, we will now consider that data in the context of weekly COVID-19 deaths for the December 2020 through February 2021 surge, as compared to the following wave beginning in July of 2021, which resulted from the COVID-19 Delta variant. Note that we
will again be excluding the oldest population group, as well as AIAN and NHPI races, as we did in the analysis for Chapter 4. The most prominent decreases in mortality are seen among Americans aged 65-74, as pictured in Figure 5.2(c). Each race group saw drastic decreases in weekly deaths, but especially within Asian and Hispanic populations; Black populations did not see as large of a decrease as the aforementioned groups, and actually had slightly higher mortality rates than Hispanic populations, which we know from Section 4.3 has not occurred in most age
categories since the initial wave of the coronavirus. Looking at Figure 5.2(a) and 5.2(b) shows that this observance holds true for ages 40-49 and 50-64 as well, with weekly mortality among Black individuals overtaking even the maximum death rates of any race in the third wave for the younger age group. This may be in part attributed to vaccine hesitancy amongst Black populations, who have the second lowest vaccination rate, and which is 5% lower than the Hispanic vaccination rate.

It is important to note that decreases in weekly mortality for Hispanic individuals are not consistent among all the age groups shown in Figure 5.2, as well that mortality actually increased for White groups aged 40-64 despite this race having the second highest vaccination rate. This indicates there is, as suspected, differing levels of vaccination rates by age within each individual race population. As mentioned previously in this thesis, the age distribution of races and ethnicities are a necessary consideration, and without this data it is irresponsible to make definitive statements about the effect of vaccines on mortality for specific subsets of Americans. Thus, we have simply communicated the findings in this section, with the intent of advocating for further exploration into these novel conclusions when more vaccination data is available.

5.3 Total COVID-19 Mortality Before and After Vaccines

Since we could not consider American Indian and Alaska Native or Native Hawaiian and Pacific Islander populations in the weekly data visualized in the previous section due to small population sizes, Figure 5.3 visualizes overall mortality during the third and fourth coronavirus waves for each race and ethnicity group. We see that for
ages 40-49, there is not very significant variation between the two waves and the death rates were all relatively low, so we will focus on the two older populations. The most change is reflected in ages 65-74; while the fourth wave of COVID-19 was less deadly in general than the previous one, interestingly we see that mortality seems to have actually equalized among most races and ethnicities. In the third wave, there was a large discrepancy between Hispanic mortality and the other races, with Black and AIAN having the next highest, and the lowest death rates among
NHPI, White, and Asian individuals. Yet, after the introduction of vaccines, these gaps closed drastically. This observation is also seen for ages 50-64, although the disparities were less extreme.

Another important remark is that Asian populations are shown in Figure 5.3 to have experienced extremely low mortality in the most recent COVID-19 wave. As stated before, the six races and ethnicities appear to endure similar death rates in the fourth wave, with the exception of Asian individuals whose were at least half as high as the next closest population. This may be the combined result of already being one of the groups with lower mortality in general and the fact that Asian individuals have the highest rate of vaccinations among Americas as of 6/1/2021. The results in this section likely had other contributing influences besides vaccination in the U.S., but it is unlikely that any had a stronger impact.
Chapter 6

Non-Coronavirus Mortality

To get a full understanding of excess mortality, it is also necessary to examine deaths which were not attributed to COVID-19. In the following pages, we will compare non-COVID death rates in 2020 and 2021 to previous years, taking into account sex, race, and age. To investigate further, we will examine various causes of death in 2020 and determine how the counts for each differ from what was expected.

6.1 Weekly Non-COVID Deaths

We know from our introductory analysis that regarding non-COVID mortality, 2020 was the primary year which experienced an elevated increase in annual deaths—6.8% increase from 2019 as compared to the average 1.6% over the last decade. Looking at Figure 6.1, we see that there is an unusual spike in the spring of 2020, which as stated before is likely a result of undiagnosed COVID-19 deaths at the beginning of the pandemic. Now that we have established in the previous chapters that the first wave of coronavirus was especially deadly for Black Americans, and to a lesser degree Hispanic Americans, it is worthwhile to see if this observation is also
reflected in non-COVID related deaths. This will be explored in the next section.

Another noticeable trend for both the year 2020 and 2021, is the lack of a decrease in death rates during the middle of the year. As stated in Section 1.2, deaths typically tend to be lower in non-winter months, yet Figure 6.1 shows that mortality in this period was slightly higher than expected, even when accounting for annual increase. During the late spring and early summer time, state and local governments in the United States began to lessen COVID-19 restrictions as deaths decreased compared to the prior winter surge, which could possibly have some relation to increased non-coronavirus deaths; more investigation is required and will be addressed in section 6.3.

### 6.2 Non-COVID Deaths by Sex, Race, and Age

In this section, the goal is to more deeply understand the effect that the coronavirus pandemic may have had on deaths not necessarily attributed to COVID-19 and
explore if there is a difference among the various demographic groups. First, we will examine the non-COVID mortality rates during the pandemic as compared to two pre-pandemic years. In Figure 6.2, it is clear that there were noticeably larger increases, among all demographic factors, in mortality rates for deaths not directly

(a) Non-COVID Deaths by Sex

(b) Non-COVID Deaths by Race/Ethnicity

(c) Non-COVID Deaths by Age

Figure 6.2: Non-COVID Mortality by Individual Demographics
related to COVID-19 in both 2020 and 2021 as compared to the slight increase between 2018 to 2019. It seems that 2020 specifically had a large impact—especially considering mortality in 2021 should have been higher of the two years as a result of the previously discussed annual mortality growth. However, to be thorough we will still investigate this change in mortality for both years. These graphs do not give a very clear understanding of exactly how drastic the increase in non-coronavirus mortality was during the pandemic, and thus further exploration was deemed necessary.

Instead, we will examine mortality in a slightly different way than in Section 4.1 and as above; instead of looking at the yearly death rates for each demographic group, the following figures will show the percent change in deaths each year as compared to the previous year. We will compare the percent increase in mortality during 2020 to the average annual increase between 2015 and 2019. Since non-COVID mortality rates seem to have been also affected by pandemic influences in 2020, we calculated the two-year percent death increase between 2019 and 2021 instead. This will then be compared against the expected two year increase as again indicated by previous average annual increase. Unlike in Chapter 1 where an overall annual death increase was given, we will instead be comparing mortality increases against 2015-2019 averages calculated for each demographic group separately. Note also that these mortality increases are calculated as a percentage of each individual demographic population, not the overall population.

Looking at Figure 6.3, it is clear that in 2020, nearly every demographic saw notable increase in deaths not related to COVID-19. Regarding sex, the increases for both male and female populations seem relatively similar. In contrast, there was slightly variation across racial groups. The largest increase was seen for Black and
Figure 6.3: Non-COVID Mortality Percent Increases in 2020

American Indian and Alaska Native individuals, with a yearly mortality increase that was higher than the previous average by 12.3% and 11.9%; Asian and Hispanic populations followed closely behind. On other other end, the Native Hawaiian and Pacific Islander group saw a much lower excess in non-coronavirus deaths than other races during the first year of the pandemic. Interestingly, when looking at
non-COVID mortality by age, the most prominent increases were seen in individuals ages 15-44. A few age groups that were seeing average annual mortality decrease over the last five years even had mortality rates increase in 2020.

To compare, we can look at the 2021 two-year mortality changes side-by-side with the expected non-pandemic increase, as visualized in Figure 6.4. Overall, 2021

![Figure 6.4: Non-COVID Mortality Percent Increases in 2020](image-url)
appears to have had lower excess deaths than seen in 2020, although the distribution of those excess deaths did slightly change across demographics. While during the first year, excess mortality appeared similar for both sexes, in the second pandemic year non-coronavirus deaths rates among females appear to be very consistent with expected increase over a two-year period; thus, the above average increase in deaths was seen in primarily the male population. Concerning race and ethnicity, the AIAN population once again experienced a much larger than anticipated increased death rate, while every other group had rates closer to the anticipated death rate over two years; notice that even though the excess deaths were less drastic than in 2020, every group except for NHPI had some amount of elevated change in mortality rates. Age group mortality increase in 2021 looks very similar to the distribution for 2020, with the exception of smaller gaps between the observed and projected mortality increase for older ages (55-79).

6.3 Leading Causes of Death (2015-2020)

The final analysis to be conducted regarding non-coronavirus mortality during the pandemic is on the leading causes of death in the United States. As stated before, since this data has not been released yet by the CDC for 2021, we will only examine changes in typical deaths for 2020, which was observed to have more severe excess mortality anyways. To review, the leading causes of death between 2015-2019 were: heart disease, cancer, accidents, stroke, chronic lower respiratory diseases, Alzheimer’s disease, diabetes, influenza and pneumonia, kidney disease, and suicide; in 2020, COVID-19 became the third leading cause of death in the U.S (Figure 6.5). Since we have established in the previous two sections that there was definitely
Figure 6.5: Leading Causes of Death

larger than expected increase in mortality even when excluding deaths attributed to the coronavirus, it is worthwhile to investigate death counts for other causes. As mentioned in Chapter 3, we used linear regression on five years of data to project what the death totals should have been without the influence of a pandemic in 2020. The projections for each individual cause are shown via dashed lines in Figures 6.6 and 6.7 on the next page.

Regarding the top two leading causes, it is noticeable that deaths due to heart disease were definitely higher than expected (Figure 6.6). Additionally, Figure 6.7 demonstrates noticeable excess deaths caused by accidents, strokes, Alzheimer’s, and diabetes. According to the CDC, heart disease, diabetes, and Alzheimer’s disease are common comorbidities and risk factors for severe COVID-19; thus, it is highly possible that these excess “non-coronavirus” deaths are in fact misdiagnosed COVID-19 deaths from the initial wave of the pandemic as discussed in Section 6.1.
Figure 6.6: Top Two Causes of Death

Figure 6.7: Other Causes of Death
Table 6.1, presents that there was also almost 2% higher flu deaths in 2020 than expected, which may also be un-classified coronavirus deaths early in the year.

Alternatively, the elevated deaths for accidents and strokes are likely to have been a result of less straightforward pandemic factors. For example, a study conducted in Alabama by Adanu, et.al (2021) found that “although traffic volume and vehicle miles traveled had significantly dropped during the lockdown, there was an increase in the total number of crashes and major injury crashes compared to the period prior to the lockdown order”. The increase in deaths resulting from strokes is a possible consequence of delayed or avoided medical care (Czeisler, 2020), as stated in Section 1.3. This exploration allows us to pinpoint exactly what the excess deaths discussed earlier in the chapter may have been caused by, and gives the

<table>
<thead>
<tr>
<th>Cause of Death</th>
<th>Percent Difference from Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Disease</td>
<td>4.44%</td>
</tr>
<tr>
<td>Cancer</td>
<td>0.23%</td>
</tr>
<tr>
<td>Accidents</td>
<td>10.98%</td>
</tr>
<tr>
<td>Stroke</td>
<td>4.86%</td>
</tr>
<tr>
<td>Respiratory Disease</td>
<td>-4.52%</td>
</tr>
<tr>
<td>Alzheimer’s Disease</td>
<td>5.99%</td>
</tr>
<tr>
<td>Diabetes</td>
<td>14.20%</td>
</tr>
<tr>
<td>Influenza &amp; pneumonia</td>
<td>1.91%</td>
</tr>
<tr>
<td>Kidney Disease</td>
<td>0.89%</td>
</tr>
<tr>
<td>Suicide</td>
<td>-7.00%</td>
</tr>
</tbody>
</table>

Table 6.1: Excess Death Percentages by Cause
opportunity to commentate on the true mortality impact of the pandemic instead of simply regarding deaths resulting from the illness directly.
Chapter 7

Conclusion

In this final chapter, we will restate the methods used and summarize the principal results of the thesis. Additionally, Section 7.2 discusses possible limitations of the research, and specifies topics for further expansion of this work.

7.1 Results

This thesis consisted of thorough data management that allowed for worthwhile statistical inference. We were able to examine data directly from the CDC detailing weekly death counts by race and ethnicity, sex, and age group, for 2015-2021, as well as Census population estimates, vaccination rates, and typical causes of death in previous years. Using regression models, we created projections for missing 2020 and 2021 population data, as well as approximating what the total deaths for each of the ten leading causes would have been in 2020 without the rise of COVID-19.

After gathering, cleaning, and analyzing U.S. mortality data on deaths preceding and during the coronavirus pandemic, various insightful conclusions were made. The most important statement we will make as a result of the research in
this thesis is that age is an essential indicator of COVID-19 mortality, and analysis comparing death rates between race and ethnicity groups cannot be complete without accounting for age as well. Age is extremely necessary because of two primary factors: older age heavily influences the deadliness of COVID-19 and age distributions are not equivalent across all race and ethnicity groups in this country, with White people tending to be much older than people of color. The population most overlooked when age is not considered is Hispanic people; when also stratifying by age we observed that Hispanic groups had mortality rates consistently higher than any other race during the second and third wave of the pandemic. Not accounting for age also neglected relatively high death rates among Black and American Indian / Alaska Native groups compared to other races, but not to the same degree as it did for Hispanic Americans. COVID-19 mortality breakdowns differed within each of the four waves examined, with more deadly weeks for Black populations in the initial wave of coronavirus than any other time during 2020 and 2021.

With special examination of the COVID-19 wave following vaccine introduction, we observed that previous disparities in mortality between the six race groups was nearly eliminated among older populations who experienced very high vaccination rates. The exception to this stabilization of mortality is Asian individuals, who maintained extremely low mortality rates between 7/4/2021 and 11/13/2021, when they had previously experienced death rates on par with White individuals.

Regarding deaths not directly attributed to the coronavirus, the majority of excess deaths in the past two years occurred in 2020. We observed percentage increases in mortality for individual sex, age, and race populations, and found that most prominent above average non-COVID mortality was seen in individuals ages 15-44, while men had higher percent increases in mortality than women as well.
Pertaining to race, well above average mortality was seen for Black and American Indian and Alaska Native individuals, followed by Asian and Hispanic populations.

By analyzing data detailing the typical leading causes of death prior to the pandemic, we even further illustrated the impact of COVID-19. The year 2020 saw higher than anticipated deaths resulting from diabetes, heart disease, Alzheimer’s disease, and to a lesser extend influenza and pneumonia, all of which are common COVID-19 risk factors and possible sources of undiagnosed coronavirus deaths. Other causes such as accidents and strokes exceeded their expected death counts as well and have possible relation to less obvious pandemic influences. With the results noted in this thesis, we have demonstrated a more complete picture of the destructive mortality experienced during the coronavirus pandemic.

7.2 Limitations and Future Direction

It is important to note that the findings of this study have been presented in light of some limitations. The first major constraint we found was a lack of more detailed data, specifically to explore further our inferences in Chapter 5 regarding the effect of vaccines on mortality among demographic groups. The vaccine information provided by the CDC was limited to only one identifying marker, whether that be age, race, or sex. As we have continuously stated in this thesis, COVID-19 mortality is immensely influenced by age; thus, it was difficult to make conclusions on the effect of vaccines among race and ethnicity or sex populations, when we had no knowledge of how vaccination within those groups was distributed for different ages. In addition, less than 75% of vaccination data had a known race or ethnicity associated, which also comprised a large amount of unverified data.
The second limitation followed as a result of a lack in precise data, which was our method of redistributing unspecified deaths, population estimates, and vaccination counts. As noted in Chapter 3, we needed to standardized demographic groups among the three data sets, and thus chose to either reallocate or remove data which was stated as unknown, or for race specifically was categorized as “Other” or “Multiple Races”. While our method was to redistribute the data proportional to the populations of the desired groups, this has pitfalls. It is likely that the multiracial population does not distribute evenly among all races, and the same for unreported vaccinations, and we acknowledge that error likely exists. Even though the amount redistributed was not particularly large compared to the rest of the data, it is still a possible source of error. In addition, population estimates for 2020 and 2021 were calculated using previous five years data and without consideration to the pandemic; this causes concern particularly related to older populations in 2021 who experienced very high mortality rates in the previous year. It is likely that population estimates for 2021 are slightly higher than the actual counts and could be somewhat skewing the data to underestimate death rates for older individuals.

The analysis conducted in this thesis has many avenues for further exploration. With the demonstrated impact of age on race and ethnicity mortality, it would be interesting to compare if COVID-19 inferences change when also considering the age distribution between males and females. As more detailed data is collected, the results of vaccination effect within individual age and race subsets of the population could be heavily expanded upon. Additionally, mortality information for causes not related to the coronavirus in 2021 will be crucial in checking if the death increases in heart disease and other probable COVID-19 misclassifications were also seen for a second year, when testing was more available.
Bibliography


census-data-shows-the-nation-is-diversifying-even-faster-than-predicted/


https://data.cdc.gov/NCHS/AH-Excess-Deaths-by-Sex-Age-and-Race-and-Hispanic-/m74n-4hbs


[19] United States Census Bureau (2020). Annual Estimates of the Resident Population by Sex, Age, Race Alone or in Combination, and Hispanic Origin for the United States: April 1, 2010 to July 1, 2019 [Data file]. Re-

